Technical Report 1373

Assessing Character in U.S. Army Initial Entry Training

Tatiana H. Toumbeva
Aptima, Inc.

Frederick J. Diedrich,
Consulting Principal Scientist

Scott M. Flanagan
Sophia Speira

Andrew Naber, Kristy Reynolds, Jessica Shenberger-Trujillo, Patrick Cummings, Krista L. Ratwani
Aptima, Inc.

Giovanna Ubillus, Charles Nocker, Corrine M. Gerard
Guidehouse LLP

Elizabeth R. Uhl, Jennifer S. Tucker
U.S. Army Research Institute

August 2019

United States Army Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

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

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

MICHELLE ZBYLUT, Ph.D.
Director

Research accomplished under contract
for the Department of the Army by

Aptima, Inc.

Technical reviews by

Michele A. Calton, U.S. Army Research Institute
Jonathan Kochert, U.S. Army Research Institute

NOTICES

DISTRIBUTION: This Technical Report has been submitted to the Defense Information Technical Center (DTIC). Address correspondence concerning ARI reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 6000 6th Street Building 1464 / Mail Stop: 5610), Fort Belvoir, VA 22060-5610.

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

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
07-05-19	Final	Sep 2016 – Sep 2018

4. TITLE AND SUBTITLE
Assessing Character in U.S. Army Initial Entry Training

5a. CONTRACT NUMBER
W911NF-16-F-0010

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
633007

6. AUTHOR(S)
Tatiana H. Toumbeva (Aptima, Inc.); Frederick J. Diedrich (Consulting Principal Scientist); Scott M. Flanagan (Sophia Speira); Andrew Naber, Kristy Reynolds, Jessica Shenberger-Trujillo, Patrick Cummings, Krista L. Ratwani (Aptima, Inc.); Giovanna Ubillus, Charles Nocker, Corrine M. Gerard (Guidehouse LLP), Elizabeth R. Uhl, Jennifer S. Tucker (U.S. Army Research Institute)

5d. PROJECT NUMBER
A792

5e. TASK NUMBER
433

5f. WORK UNIT NUMBER

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

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

8. PERFORMING ORGANIZATION REPORT

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

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

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

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Technical Report 1373

12. DISTRIBUTION/AVAILABILITY STATEMENT:

Distribution Statement A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter POC: Jennifer S. Tucker, Ph.D.

14. ABSTRACT (200 words max)

The U.S. Army's commitment to the development of adaptive Soldiers inculcated in the Army Values is a critical element of Initial Entry Training (IET) a sub-set of Initial Military Training. In IET, trainees not only learn and are assessed on technical and tactical skills, but also on character – how they live and uphold the Army Values in everyday activities. This research showcases a strategy and provides tools for assessing and tracking character in Basic Combat Training (BCT), a component of IET. Trainee responses on peer evaluations and ethical decision-making questions were digitally captured using audience response clicker technology. An Excel-based tracking tool was developed to automatically store and analyze the assessment data. The character assessment tools were developed in Part 1 and piloted in Part 2, revealing that instructional techniques and technology interacted to differentially affect learning outcomes. In Part 3, a longitudinal evaluation was conducted, providing initial empirical reliability and validity evidence for the assessment tools. This research seeks to meet the key challenge of assessing character in a manner that facilitates Soldier development while also providing Army Leaders with data analytics to inform programmatic decisions. Copies of the character assessment tools and user guides are available from the Defense Technical Information Center.

15. SUBJECT TERMS
Character, training, assessment, ethical decision-making, clicker technology, Army Values, locus of control, clustering

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Jennifer Tucker
Unclassified	Unclassified	Unclassified	Unlimited Unclassified	68	19b. TELEPHONE NUMBER 703-545-2490

Technical Report 1373

Assessing Character in U.S. Army Initial Entry Training

Tatiana H. Toumbeva
Aptima, Inc.

Frederick J. Diedrich,
Consulting Principal Scientist

Scott M. Flanagan
Sophia Speira

Andrew Naber, Kristy Reynolds, Jessica Shenberger-Trujillo, Patrick Cummings, Krista L. Ratwani
Aptima, Inc.

Giovanna Ubillus, Charles Nocker, Corrine M. Gerard
Guidehouse LLP

Elizabeth R. Uhl, Jennifer S. Tucker
U.S. Army Research Institute

Ft. Benning Research Unit
Jennifer S. Tucker, Chief

August 2019

Approved for public release; distribution is unlimited.

ACKNOWLEDGMENTS

We would like to thank the 194th Brigade 1-46th Infantry Battalion Command Team and Drill Sergeants, as well as Chaplain (LTC) Anthony Randall, COL Matthew Scalia, Col (R) Scott King, COL Mark Huhtanen, and LTC Sheldon Morris who served as proponents of this work and provided valuable time, input, and feedback.

ASSESING CHARACTER IN U.S. ARMY INITIAL ENTRY TRAINING

EXECUTIVE SUMMARY

Research Requirement:

The U.S. Army must be prepared to respond rapidly to ambiguous and asymmetric threats especially when laws, rules, and regulations do not provide a clear, right course of action. Soldiers are trained to internalize and demonstrate a shared commitment to the Army Values, and act in accordance with the legal and moral foundation of the Army Profession (ADRP-1; U.S. Department of the Army, 2015a). The Army's strong commitment to the development of agile and adaptive Soldiers inculcated in the Army Values is a major focus during Basic Combat Training (BCT). In BCT, trainees not only learn and are assessed on technical and tactical skills, but also on character – how they live and uphold the Army Values in everyday activity. A key training and measurement challenge is assessing character within the time constraints of BCT. Assessments should support Soldier learning while enhancing leader's understanding of training effectiveness, which ultimately informs the Army's character development initiative.

Procedure:

The research presented here showcases a strategy developed and tested for assessing character in the BCT environment. The approach includes tools to enable consistent peer feedback regarding Army Values-based behaviors as well as tools to assess Soldiers' reactions to situations involving ethical decision-making. A three-part study was conducted. Character assessment tools were developed in Part 1, piloted in a quasi-experimental study in Part 2, and evaluated longitudinally in Part 3. The assessment tools were tested and refined at each stage of the research effort to maximize utility and sustainability in the target environment

Findings:

Assessment tools were developed to measure character over the course of basic training, facilitate discussions about Army Values, expedite the peer evaluation process within basic training, digitally capture assessment data, and enable management of character assessment data. Initial reliability and validity evidence of the ethical decision-making questions was provided relative to peer evaluation results. Utility and usability feedback from trainees and instructors about the tools was positive – particularly given that this process introduced efficiencies that enabled a quick and easy capture of longitudinal data in a manner that simultaneously boosted learner engagement.

A notable outcome of this research is the finding that technology alone did not promote learning. Results indicated that while the assessment tools encouraged thought and introspection in learners, reaping the benefits of these tools ultimately depended on the instructor. Thus, the implementation of such technologies must be supplemented with instructional strategies that align with the primary learning objectives and instructor training on how to appropriately harness the tools to support learning. Overall, the findings suggest that events in basic training bolstered locus of control and successfully instilled the Army Values in trainees.

Utilization and Dissemination of Findings:

Collectively, the character assessment tools were designed to provide programmatic feedback to unit leadership as different instructional approaches to character assessment and development are pioneered in the future. Although the mere act of taking character assessments as part of the training curriculum may have some developmental benefit for the trainee, the full potential of the character tools lies in individualized feedback opportunities for the trainees. In the current research, the tools were used to produce discussion and assess character primarily at the group-level; however, the assessment data could be used to provide customized, targeted formative feedback to the individual learner. As such, these tools also enable the provision of developmental feedback to individual Soldiers that may be used to maximize not only collective, but also individual learning processes and outcomes. Whereas the explicit focus of the current research was on the Army BCT environment, similar assessment tools could be developed and employed to support learning in any instructional context.

A final training was conducted with Drill Sergeants on how to implement the assessment tools in BCT. User guides that outline how to use and manage the character tracking tool, video-based ethical decision-making questions, and clicker software and hardware were developed and provided as a supplement to the Drill Sergeant training. Literature-backed best practices were also provided to the Drill Sergeants on writing effective ethical decision-making questions and using clicker technology to facilitate discussion. The character assessment tools and user guides are included as a DVD.

ASSESSING CHARACTER IN U.S. ARMY INITIAL ENTRY TRAINING

CONTENTS

Page

INTRODUCTION 1
Current Research 2

PART 1 OVERVIEW: DEVELOPMENT OF A CHARACTER ASSESSMENT SYSTEM AND APPROACH 2
Development of Ethical Decision-Making Questions 3
Development of an Army Values Rubric for Peer Evaluations 7
Development of a Database and Tracking Tool 8

PART 2 METHOD: INITIAL TESTING OF THE ASSESSMENT SYSTEM AND APPROACH 11
Participants and Procedure 11
Materials 12
Data Analysis 15

PART 2 RESULTS 15
SJT Performance 15
Locus of Control 15
Trainee Feedback 15
Drill Sergeant Feedback 16

PART 2 DISCUSSION 16

PART 3 METHOD: LONGITUDINAL EVALUATION OF THE ASSESSMENT SYSTEM AND APPROACH 17
Participants and Procedure 18
Materials 18
Data Analysis 20

PART 3 RESULTS 20
Relationships between Focal Variables 20
Group Differences 22
Closer Examination of Ethical Decision-Making Performance 22
Results across Training Phases 24
Drill Sergeant and Trainee Feedback 26
Exploratory Analyses 26

PART 3 DISCUSSION 32

GENERAL DISCUSSION 33

REFERENCES 35

APPENDIX A. ARMY VALUES RUBRIC FOR PEER EVALUATIONS A-1

APPENDIX B. OVERVIEW OF THE CHARACTER TRACKING TOOL TABS B-1

APPENDIX C. GUIDELINES FOR DEVELOPING SCENARIO-BASED QUESTIONS ON ETHICAL DECISION-MAKING C-1

APPENDIX D. VIDEO-BASED ETHICAL DECISION-MAKING QUESTIONS AND SCORING KEY D-1

APPENDIX E. GUIDANCE FROM ACADEMIC LITERATURE ON USING CLICKER TECHNOLOGY E-1

LIST OF TABLES

TABLE 1. KOHLBERG'S STAGES OF MORAL DEVELOPMENT 4

TABLE 2. DESCRIPTIVE STATISTICS, CORRELATIONS, AND RELIABILITY ESTIMATES FOR FOCAL VARIABLES BASED ON RED, WHITE, AND BLUE PHASE ASSESSMENT DATA 21

TABLE 3. MULTIPLE REGRESSION ANALYSIS RESULTS OF VIDEO-BASED SJT PERFORMANCE PREDICTING PEER EVALUATIONS 23

TABLE 4. DESCRIPTIVE STATISTICS FOR FOCAL VARIABLES ACROSS TRAINING PHASES 24

TABLE 5. SUMMARY DESCRIPTIVE STATISTICS FOR COMPLETE DATASET USED IN CLUSTER ANALYSES 29

TABLE 6. SUMMARY OF EACH CLUSTER BASED ON DISTINGISHING CHARACTERISTICS 30

LIST OF FIGURES

FIGURE 1. ASSESSMENT SYSTEM AND APPROACH 3

FIGURE 2. CHARACTER TRACKING TOOL INDIVIDUAL DASHBOARD EXAMPLE 10

FIGURE 3. CHARACTER TRACKING TOOL VIDEO DASHBOARD EXAMPLE 10

FIGURE 4. CHARACTER TRACKING TOOL PEER COMMENTS SUMMARY TAB EXAMPLE 11

FIGURE 5. MEAN PEER RATINGS ACROSS PHASES FOR TOP AND BOTTOM 10 PERCENT FROM RED PHASE 27

Assessing Character in U.S. Army Initial Entry Training

The U.S. Army must be prepared to respond rapidly to a variety of missions many of which will involve ambiguous and/or asymmetric threats or situations. In all cases, Army professionals are expected to adhere to applicable laws, regulations, or rules (e.g., Laws of Armed Conflict), particularly in combat or in any application of lethal force. However, situations of uncertainty occur where the rules do not provide a clear, correct course of action. In these cases, Army professionals must base their decisions and actions on the moral principles of the Army Ethic, ensuring the protection of the inalienable rights of all people. In this way, Army professionals live by and uphold the moral foundation of the Army Ethic and reinforce the Army culture of trust among fellow Army professionals and with the American people.

The Army profession is built on the consistent demonstration of character, strengthening the Army culture of trust, and contributing to character development through leadership, teaching, coaching, counseling, and mentoring (ADRP-1; U.S. Department of the Army, 2015a). Individuals demonstrate character when they act in accordance with the Army Ethic. Internalizing the Army Ethic (i.e., the guiding laws, values, and beliefs of the Army Profession) as a personal belief is essential for development as a member of the Army. Over time and across all activities, leaders must embody the seven Army Values (Loyalty, Duty, Respect, Selfless Service, Honor, Integrity, and Personal Courage), which are fundamental to the Army Ethic. Leaders must also purposefully set conditions for character development opportunities in day-to-day operations, model these principles themselves, and communicate expectations that others also embody the Army Values. When Soldiers fall short, counsel must be provided on the gaps between their personal values and those that the Army espouses. Reinforcing ethical standards increases the likelihood of ethical decisions and actions, and promotes an ethical climate (FM 6-22; U.S. Department of the Army, 2015b). As an integral part of the Army Leader Development Strategy (U.S. Department of the Army, 2013), character development is a continuous process within the Army as an institution and between leaders and subordinates. When integrated within sequential and progressive education and training, this process strengthens the resolve of trusted Army professionals to live by and uphold the Army Ethic (see Center for the Army Profession and Ethic, 2017).

Character is forged early in Soldiers. The development of character in accordance with the Army's expectations begins as early as enlistment, and critically, in Initial Entry Training (IET), a sub-set of the U.S. Army Training and Doctrine Command (TRADOC) Core Function of Initial Military Training (IMT). Drill Sergeants, as the Soldiers' first leaders, have the opportunity to demonstrate and shape character development in profound ways during Basic Combat Training (BCT). BCT is a component of IET/IMT and the focus of this research.[1] The culture of trust in the Army is built upon a foundation of learning to live the Army Values. Therefore, character is just as critical an objective as Soldiers learning to operate their weapons, navigate, and attain physical fitness. Trainees are assessed not only on technical and tactical skills, but also on character in terms of how they uphold the Army Values in everyday activities. A key training challenge is assessing character within the time constraints of BCT. Assessments

[1] Trainees undergoing BCT are trained to become non-combat arms Soldiers, and eventually fill a wide range of job positions within the Army (e.g., logisticians, medics).

must also support Soldier learning while enhancing leader situational understanding of training effectiveness to ultimately inform the Army's character development initiative.

Current Research

The research presented here showcases an assessment strategy and tools developed and tested for capturing and tracking character in the BCT environment. The emphasis was on creating formative assessment tools that could be used to provide feedback and guide learning (Bransford, Brown, & Cocking, 2000). The focus was not on creating tools that could be used to grade a Soldier for the purpose of assigning a final score or to provide a determinant for those lacking sufficient character to be considered fit for becoming a Soldier. Rather, the starting assumption was that all Soldiers enter the Army from a variety of backgrounds with differing conceptions of what character means. One role of BCT is to start to move trainees from their starting point to a place where they internalize and live in accordance with how the Army sees character, primarily through the embodiment of the Army Values.

Given this emphasis, the goal of the current research was to develop and evaluate tools to enable consistent peer-based feedback regarding values-based behaviors as well as tools to assess responses to situations involving ethical decision-making. Together, these tools were designed to provide developmental feedback to individual trainees and programmatic feedback to unit leadership (i.e., battalion and company command team) as different instructional and assessment approaches to character development are pioneered in the future. Furthermore, although not an explicit goal initially, given the data collected and interest demonstrated by BCT leadership, examining the different developmental paths trainees take as they progress through BCT became an additional focus of this research.

This research comprised three parts. In order to optimize the tools for the assessment context, multiple rounds of iterative tool review and refinement followed each part. Both qualitative and quantitative information from end users provided insights into how these tools were likely to operate in the target training environment. In Part 1, character assessment tools were developed for use in the BCT environment. In Part 2, the assessment tools were piloted using a quasi-experimental research design. In Part 3, a rigorous examination of the assessment tools was conducted by gathering data longitudinally over the course of a cycle with an entire company undergoing BCT. A cycle of BCT consists of Red (beginning), White (middle), and Blue (end) phases, each spanning approximately three weeks. Each part of this research is described in detail in the upcoming sections, along with the implications and limits of the assessment approach, and suggestions for how to harness the tools to support learning.

Part 1 Overview: Development of a Character Assessment System and Approach

The primary components of the assessment system are illustrated in Figure 1. In broad terms, from left to right, the system hinges on two sources of data, responses to ethical decision-making questions and peer evaluations. Soldier responses are fed into a database and automated tracking tool.

Soldier responses to the ethical decision-making questions and peer evaluations were captured via clickers, which are off-the-shelf polling tools that can be used in a classroom

setting. The clicker technology allows answering of multiple choice and text-based responses to various questions while enabling Drill Sergeants to view responses in real time. The system was built on existing and approved tools for use with multiple individuals at a single time given high instructional tempos.



Figure 1. Assessment system and approach.

Development of Ethical Decision-Making Questions

The ethical decision-making questions were created based on Army Values videos from the Center for the Army Profession and Ethic (CAPE; see http://cape.army.mil/case-studies/). Drill Sergeants frequently used these videos with new BCT trainees. The ethical decision-making questions were designed in the form of situational judgment test-style items (SJT). SJTs are low-fidelity simulations comprising a scenario and set of responses that are intended to elicit similar judgment processes that would occur in real world contexts (Motowidlo, Hanson, & Crafts, 1997). After trainees view the CAPE videos, SJTs that pose short hypothetical moral scenarios are presented with the clicker system. The SJT prompts the trainees to think critically about the scenarios and select a response using the clickers.

Army doctrine notes that "[c]haracter forms over time through education, training, and experience in a continuous, iterative process" (FM 6-22; U.S. Department of the Army, 2015b, p. 5-1). This notion is consistent with theories that suggest moral development moves from lower levels emphasizing immediate concerns about one's self, up to higher levels emphasizing individuals as independent moral agents who live by their own ethical principles (e.g., life, liberty, justice) (Kohlberg, 1971; Rest, Narvaez, Bebeau, & Thoma, 1999). These principles may interact with, conflict with, or take precedence over society's broader view of morality (Kohlberg, 1971). These stages are shown in Table 1.

The basis of ethical decision-making and individual differences in moral reasoning among Soldiers certainly vary. Nevertheless, a Soldier's character is revealed in the decisions and behaviors that conform (or not) to the moral principles of the Army Ethic. Army professionals are expected to decide what is right (ethical, effective, and efficient), and demonstrate the character, competence, and commitment to act accordingly (U.S. Department of the Army, 2015a).

Table 1

Kohlberg's Stages of Moral Development

	Level	Orientation	Definition	Example
Increasing Levels of Moral Development	**1**	*Pre-Conventional:* Punishment and obedience	Reasoning determined by individual receiving rewards and punishments.	"I should not steal because I will get in trouble."
	2	*Pre-Conventional:* Instrumental relativist	Reasoning determined by exchanges between others regarding rewards and punishments.	"I should not steal because they may steal back from me – but if they steal from me first, then I would steal from them."
	3	*Conventional:* Interpersonal concordance	Reasoning based on obedience to authority.	"I should not steal from someone, even if they steal from me, because others will think less of me."
	4	*Conventional:* Law and order	Reasoning based on importance of doing "one's duty."	"I should not steal from someone, because it is their property."
	5	*Post-Conventional:* Social-contract	Reasoning based on good intentions, and obedience to authority, laws, and rules.	"I should not steal from someone, because we all rely on respecting the property of others."
	6	*Post-Conventional:* Universal ethical-principle	Reasoning through individual principles of conscience; including likely views of everyone affected by a moral decision.	"I should not steal, but sometimes, it is necessary to steal in order to help someone in a difficult situation."

Situational judgment test scenario development. There are various approaches used to develop SJTs across a variety of contexts, jobs, and competencies (cf., Motowidlo et al., 1997). In contrast to traditional multiple-choice items where responses are keyed as correct or incorrect, an SJT item includes multiple responses that may be considered degrees of correct, and as such, offer a continuum of poor to best responses among the alternatives. Several requirements guided the development of the SJTs, aligning with existing practices and available technology. The SJT items were (a) built on video case studies from CAPE, (b) developed for use with clicker technology, (c) focused on providing targeted assessment and feedback based on Army Values, and (d) suitable for generating discussion on the Army Values. As such, this assessment method is a hybrid design built on moral reasoning development, SJT item development strategy, and discussion-based training.

Producing the SJTs followed an iterative process of development, review by DS, and refinement. The CAPE videos were selected based on feedback from Drill Sergeants on the types of videos typically used in BCT and their potential to trigger meaningful discussion.

Specifically, it was required that the described situation in the video was sufficiently relevant, detailed, and engaging, and the dilemma was at the appropriate level of complexity for trainees in BCT (e.g., not overly technical). Based on the selected CAPE videos and accompanying facilitator materials, critical ethical challenge(s) depicted in the video were identified. Most videos focused on one ethical challenge, but contained details about other potential ethical challenges in the narrative. These ethical challenges formed the stem of the SJT item. The next step was to develop potential responses to that scenario-based question that ranged in degree of correctness along with a scoring key.

"Correct" answers on a moral reasoning test are easy to describe, but challenging to develop. In writing responses, the researchers considered the levels of moral development identified by Kohlberg (1971, see Table 1). One consideration for researchers was understanding that individuals at different levels of moral development may notice and interpret a moral dilemma differently. The total SJT item comprised of a synopsis of the video, a brief recounting of the relevant details of the ethical challenge, a question, and four behavioral responses to that question. The four response options reflected different levels of moral development, though the levels of moral development reflected varied for each item. Developing items to map onto each of the six levels of moral development would have been difficult given the nature of the scenarios and it was unclear that level of detail would be useful for enabling formative feedback. Moreover, six response options were initially developed as examples for a subset of the items; however, upon review, Drill Sergeants indicated that having six response options would be too complex and time-consuming, and indicated that four would be more useful and manageable given course constraints. Therefore, four items were loosely mapped to the levels of moral development moving from personal consequences to more abstract principles. An example question and response options highlighting the Army Value of respect were: "PVT Gress describes a Soldier in her unit who isn't a team player and is disrespectful to others. What should you do in this situation? I should... (a) keep an open mind about this Soldier and give her opportunities to show that she can rise to the occasion as a team player; (b) never help PVT "My Way" if she ever needs it (incorrect/least correct response); (c) approach this Soldier in private and tell her that she is failing to earn the unit's respect (most correct response); (d) tell other Soldiers about PVT "My Way's" behavior."

Based on guidance from Drill Sergeants and unit leadership, a set of CAPE videos were selected that aligned with the learning objectives at each BCT phase. The selection of videos was guided by the Army Values that were considered to be the most salient in each phase of training. For example, based on Drill Sergeants and unit leadership feedback, loyalty takes longer than the other values to develop as part of BCT, and as such, should be assessed last. Therefore, in Red Phase, three basic training videos focusing on Respect, Duty, and Personal Courage were used. In White Phase, one basic training video focusing on Integrity, and two operational videos focusing on a more complex moral/ethical dilemma in a combat setting were used. In Blue Phase, one basic training video focusing on Loyalty, and two different operational videos were used. Basic training videos focus on one Army Value (e.g., Integrity) and depict a situation involving a moral dilemma encountered in the context of basic training (e.g., whether or not to cheat in a Basic Rifle Marksmanship competition). The operational videos, on the other hand, depict situations involving moral dilemmas that Soldiers have faced in combat and/or while deployed. As such, the operational videos encompass multiple Army Values and focus on

evaluating trainee ability to apply what they know about the Army Values to complex, challenging situations.

Multiple items were developed for each selected video. Although a particular Army Value may be more relevant to a particular scenario, the individual items were not meant to reflect any particular value in isolation. The dilemmas that Soldiers face in practice, and the scenarios used in this project, require Soldiers to apply Army Values in totality. That is, Soldiers bring a set of values that may be more or less relevant to a particular ethical situation. Furthermore, each Soldier may uniquely weigh the relevance of particular Army Values within complicated ethical dilemmas. Given this complexity, a Soldier's character allows him/her to appropriately weigh the unique values that correspond to the particular situation. Indeed, a single SJT item may assess a number of constructs simultaneously: knowledge of rules and regulations, the ability to distinguish between plausible acceptable responses, and deciding between these responses. For these same reasons, SJTs offer a uniquely valuable medium for assessing and training character in an Army context.

Scoring key development. The research team (five research psychologists and one military subject matter expert [SME]) developed a scoring key for the CAPE video SJTs to enable performance tracking and trending as part of the assessment system. Responses ranged in degrees of correctness. Each of the four responses to a scenario was scored on a scale of 0, 25, 50, 75, or 100. Although all scenario items included a single *best* response scored as 100, each scenario item could have any mix or distribution of scorings for alternate responses. Responses that were considered definitively incorrect – responses that overtly violated regulations or would directly and immediately harm another Soldier – were scored as 0. Responses varying in appropriateness (correctness) were scored 25, 50, or 75, depending on the extremity of the responses.

Scoring key validation. The scoring key developed by the research team was validated with SMEs and further refined based on qualitative and quantitative data obtained as part of a focus group. Three senior Drill Sergeants and one Drill Sergeant participated in the focus group. Focus group participants were on average 32.75 years old, reported 17.50 months in their current role, 12.38 years of military service, and all had completed at least a Bachelor's degree. Drill Sergeants are considered SMEs for validating the scoring key due to their extensive experience in the Army in addition to their experience working with basic trainees. The participants watched each CAPE video, read the corresponding SJTs, and responded to a paper-based survey. SMEs rated each response option using a sliding scale ranging from *very inappropriate* to *very appropriate* response. The participants were instructed to consider each response option independently relative to the depicted scenario, and place an 'x' anywhere on the provided line to indicate the degree to which they believed each response option was appropriate. For example, the 'x' was placed closer to the right end of the line if they believed the response was fairly appropriate given the situation. Participants also provided verbal feedback to help refine the question and response option content.

Following the focus group, the survey data were analyzed by assigning a numerical value for each response option based on the placement of the 'x' along the line. Although not initially included in the survey, five-equally spaced tick marks were added to each line after the data collection to more accurately quantify the placement of the 'x' along the line. The tick marks

represented 0, 25, 50, 75 and 100 points. If the x was placed close to the middle of the line (i.e., the third tick mark), then the response option was assigned a score of 50. All participant scores were averaged and compared to the scores provided by the research team. The Drill Sergeant scores for 74 of the 88 response options (22 total SJTs with four response options each) aligned with those initially assigned by the research team. Thus, the scores for the 74 response options were unchanged following the Drill Sergeant focus group. For the remaining 14 response options, the previously assigned score changed by one 25-unit increment (e.g., from 0 to 25 points, 50 to 75 points, 75 to 100, or vice versa in each case). The research team reached consensus on scoring responses through discussion with Army SMEs, and ranking each response in the scenario relative to the others, before assigning the final 25-unit score.

Notably, different combinations of points were obtained to properly account for nuances in the provided response options. As such, although responses were scored using a 0-100-unit scale with 25-unit increments, duplicate scores were allowed. For instance, in a few cases, one response option was marked as *very appropriate* (i.e., 100) and all other options as *very inappropriate* (i.e., 0 points). In the majority of cases, however, the scoring distribution across scenario response options was more diverse, where the *least* and *most appropriate* responses were identified (i.e., 0 and 100 points, respectively), and the remaining options fell somewhere in the middle depending on the perceived degree of response appropriateness (i.e., some combination of 25, 50, and/or 75 points).

Overall, this scoring approach provided sufficient variability to represent degrees of correct through the continuum of moral reasoning. This fine-grained consideration of correctness also enabled more precise performance tracking and trending over time as part of the tracking tool. Furthermore, the approach provided the necessary flexibility for development of a wider variety of realistic response options to complex dilemmas Soldiers may face during training and in combat. In each CAPE video, different dilemmas were highlighted, details presented, and unique characters described. Rather than force a particular distribution or moral reasoning stage, the research team developed scenarios and responses that represented the dilemmas, details, and individual Soldiers as shown in the videos. That is, scoring was in service of numerically representing the most plausible and morally-challenging responses that could be written rather than writing responses in order to match a scoring rubric.

Development of an Army Values Rubric for Peer Evaluations

As part of the assessment process, BCT trainees used clickers for peer evaluations that were previously completed on paper using a battalion-approved peer evaluation form. For one component of the peer evaluation, trainees rated each member of their squad on each of the seven Army Values. Only short definitions of each Army Value accompanied the battalion peer evaluation form. To help create a shared mental model about how the values can manifest behaviorally in BCT, a rubric was developed (see Appendix A). This rubric supplemented the peer evaluation process in order to improve the usability and reliability of the peer evaluation results. That is, the Trainees were asked to refer to the rubric when making their ratings and writing comments about their peers. Thus, instead of writing general comments about their peers' performance, they were asked to use the rubric to shape their comments. The rubric was created in the form of a behaviorally anchored rating scale (BARS). BARS comprise brief, specific, observable behaviors (i.e., behavioral anchors) at different proficiency levels. BARS

enhance assessment accuracy and consistency by reducing ambiguity and helping raters make an evaluation based on relevant factors (Guion, 2011; Smith & Kendall, 1963). Furthermore, behavior-based assessment supports formative feedback, self-awareness, learning, and growth (FM 6-22; U.S. Department of the Army, 2015b).

In the current study, the BARS were designed to prompt the rater to think about the extent to which trainees act in accordance with each Army Value. When used for peer evaluations, the BARS can provide trainees with a common frame of reference for what the Army Values mean in the context of BCT. This is critical since trainees come from different backgrounds, and as such, may interpret and evaluate each value differently. In the context of peer evaluations, the Army Values rubric can also help trainees generate actionable formative feedback as part of the required comments that accompany ratings. An iterative development and refinement process was followed to develop the anchors for the Army Values BARS. Extant seminal work on BARS development also served to guide this process (Bernardin & Beatty, 1984; Smith & Kendall, 1963), namely identifying critical performance dimensions and associated critical incidents relevant for the target context. This approach has been used in previous research conducted in a BCT setting; specifically, Muraca-Grabowski (2015) developed BARS on the Army Values from examples provided by BCT trainees (rather than Drill Sergeants as in the current study). Based on guidance from DS, the BARS focused only on the Army Values, as they are fundamental in character development and lie at the core of BCT.

A series of focus groups were conducted with Drill Sergeants to identify (a) how each Army Value is defined and operationalized within the context of BCT, (b) when each value is the most likely to manifest, (c) how each value manifests, and (d) how malleable and measurable the values are in the context of BCT. Next, observable behavioral indicators for each Army Value were developed across a three-point proficiency scale (*poor/below standard, average/at standard, excellent/above standard*). Several cycles of anchor development, review by SMEs, and revision were completed over the course of several months to ensure the content and format befit the intended workflow and use within BCT.

The resulting rubric consisted of an average of five behavioral examples at each proficiency level for a given Army Value. For instance, a behavioral example of a Soldier who is Above Standard on Integrity included "Pauses and thinks through options and makes the right ethical decisions even under extreme pressure and regardless of context (e.g., conflicts between own beliefs and mission/Army values)." A behavioral example of a Soldier who is At Standard on Integrity included "Generally makes the correct ethical decision but may be affected by conditions (e.g., the presence of Drill Sergeant or may need prompting)." A behavioral example of a Soldier who is Below Standard on Integrity included "Lies, steals, cheats (e.g., takes food out of chow hall, brings contraband in bay)."

Development of a Database and Tracking Tool

An Excel-based tracking tool was developed to automatically store, manage, and analyze data gathered as part of the assessment system, and ultimately, help unit leadership (e.g., battalion and company commanders) and Drill Sergeants understand programmatic trends, identify struggling Soldiers, and optimize training. The tracking tool was iteratively tested and refined based on data and insights gathered during each phase. The tracking tool contains a

simple, user-friendly interface to minimize burden on the end-users. Given that it is Excel-based, the tool can be used on a secure Army network and adapted as needed in the future. Along with basic analytical capabilities, the tool presents multiple tabs of assessment data in dynamic tables and graphs tailored to the desired level of analysis. Peer and CAPE video assessment results can be viewed in individual, squad, platoon, cycle, and company dashboards and tracked over time (within and across cycles).

A systematic process was followed to gather requirements for the development of a sustainable and usable tracking tool. Cycles of iterative development, review, and revision were conducted to ensure the tool contained the most useful type of data and displayed optimally at each level and within each tab for the intended audience (e.g., Drill Sergeants vs. unit leadership). Among the requirements gathered were unit-level composition and maximum number of the following: (a) trainees in squad; (b) cycles per year; (c) Drill Sergeants per squad; (d) peer evaluations sessions per cycle; (e) peer rating questions per evaluation session; and (f) CAPE video SJTs across phases of BCT. The current and future peer and video assessment objectives and needs were discussed with Drill Sergeants and unit leadership to prioritize the type of data to be gathered, stored, and managed as part of the tool. A standardized process for using these assessment tools was also recommended and discussed with end-users.

An important capability of the tracking tool is integration with the clicker software output. Peer and video assessment data can be exported from the clicker software in Excel, and then imported into the tracking tool. Because of the built-in features of the tool, minimal end-user effort is required. During the import process, the clicker software output is automatically transformed into a data entry error free format that is immediately imported into the tracking tool. Standardization of the peer and video assessment processes across squads, cycles, and companies is critical to the long-term sustainability and utility of the tool. However, the tool is flexible enough to accommodate variations across squads, platoons, cycles, and companies, as well as trainee transfers into or out of a unit. Furthermore, the tool can be expanded to include other types of results in the future (e.g., marksmanship, physical fitness test scores).

The current tabs include dynamic dashboards, peer evaluation summary results (including compiled peer comments), and video question performance at different levels (individual, squad, platoon, cycle) across BCT phases and across cycles (see Appendix B for a description of each tab in the tool). These tabs are formatted for easy printing. Below are screenshots of example tracking tool tabs. Figure 2 shows the individual dashboard, which is an interactive tab that can be used to view compiled individual-level peer and video results for a selected phase and Soldier. The dynamic charts may be used to view result details or higher-level trends and comparisons (e.g., Soldier performance relative to his/her squad or platoon). Similar tabs were developed for squad, platoon, and cycle assessment results that allow comparisons, trending, and tracking at different levels.

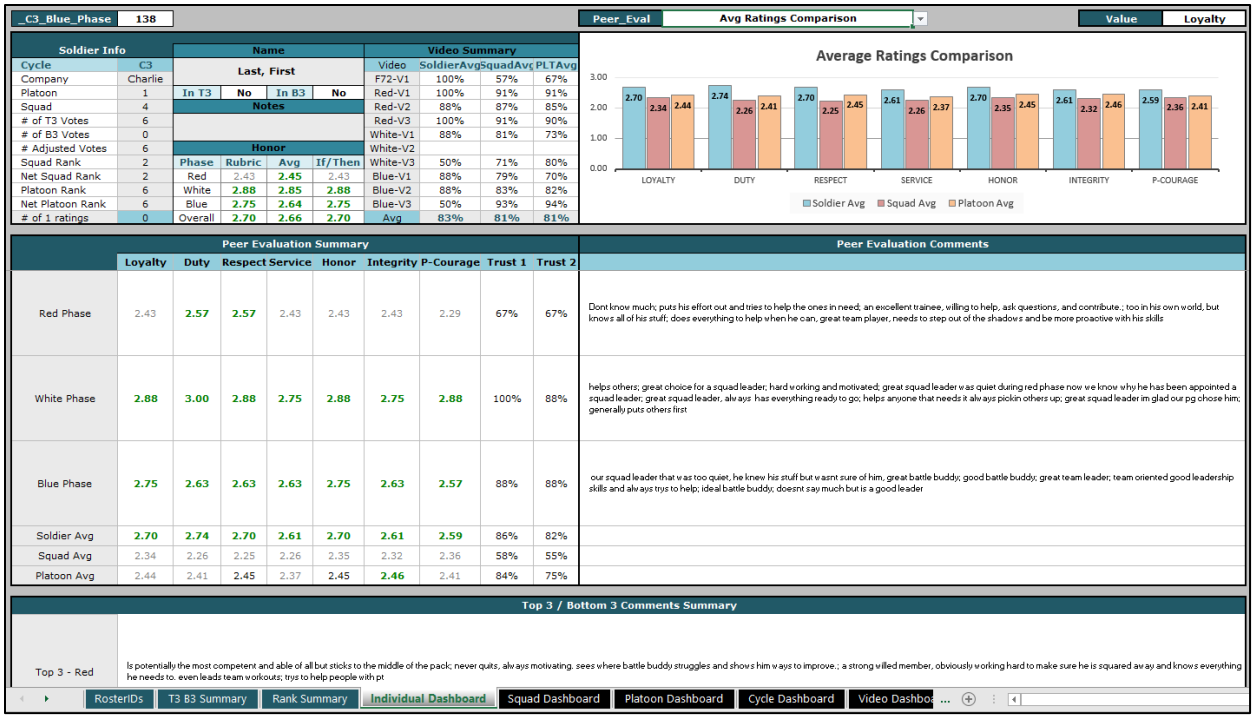


Figure 2. Character tracking tool individual dashboard example.

Figure 3 shows an example video dashboard, which is an interactive, dynamic tab that allows end users to track performance on the CAPE video SJTs over time. Response breakdowns are presented by platoon, squad, and cycle relative to the SJT scoring key developed as part of this phase.

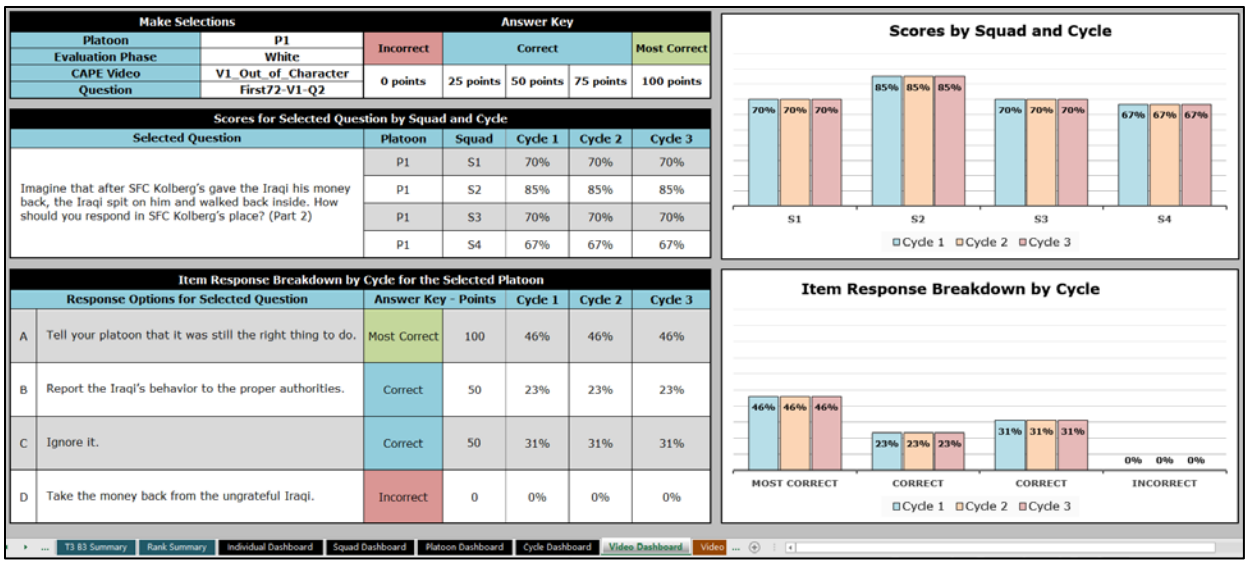


Figure 3. Character tracking tool video dashboard example.

Figure 4 shows an example comments summary tab that displays aggregated peer evaluation comments for each Soldier by phase. Data are organized by company, cycle, platoon,

and squad. The tab also displays whether each Soldier was voted in the Top 3 or Bottom 3 in the platoon based on the Army Values for each assessment period.

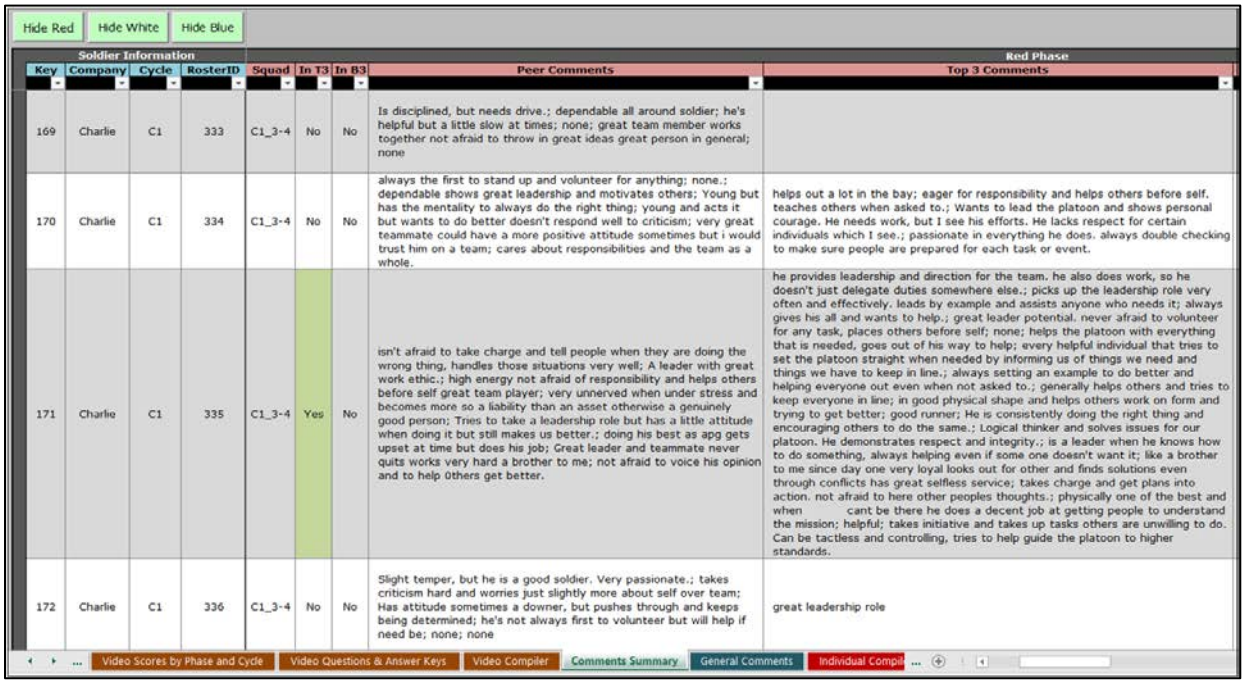


Figure 4. Character tracking tool peer comments summary tab example.

Part 2 Method: Initial Testing of the Assessment System and Approach

A pilot study was conducted to gather initial utility and usability feedback on the assessment tools (i.e., CAPE videos, SJTs, Army Values rubric) and the method for capturing peer and video question responses via clickers. The study also examined how conducive the tools were for facilitating values-based discussion and Soldier learning in a platoon classroom setting. The specific focus was on knowledge retention, critical thinking, perceptions of discussion quality, and engagement across platoons.

Participants and Procedure

A total of 143 trainees across three platoons from the same company in BCT participated in the pilot study. Two platoons consisted of 48 trainees each and one consisted of 47 trainees. The average reported age of the trainees was 21 years (*SD* = 3.86). Fifty-seven percent of the trainees reported a grade of E-1, 23% indicated a grade of E-2, 13% held a grade of E-3, and 7% held a grade of E-4. The majority of the trainees reported that their highest level of education was a high school diploma (57%), followed by some college (29%), Bachelor's degree (5%), GED (3%), Associate degree (3%), and graduate level work or degree (3%). Distribution of age, grade, and education across platoons was similar and did not result in any significant differences.

The assessment with the three platoons was conducted simultaneously in separate classrooms, each facilitated by a member of the research team and a DS. Across all platoons, the Drill Sergeants were asked to lead an Army Value-based discussion based on the selected CAPE

videos. Following the videos and discussion, the trainees completed a learning check test designed to assess knowledge retention based on the videos. Each of the three platoons were assigned to one of three unique conditions that varied in the assessment support tools provided. The control platoon was provided with the CAPE videos for viewing and discussion. Trainees in the control condition did not receive the video SJTs, peer evaluations, or clickers. Trainees in this condition completed paper versions of the learning check test, background questionnaire, and feedback survey. In the paper-based SJT condition, the platoon was provided with the CAPE videos and associated SJTs to help the Drill Sergeants generate discussion with trainees, but not clickers. In this condition, the trainees responded to the SJTs, learning check test, background questionnaire, and feedback survey on paper. In the clicker condition, the platoon used clickers to enter responses to the peer evaluations, SJTs, learning check test, background questionnaire, and feedback survey. The research team trained the clicker condition participants on how to use the clickers in the beginning of the session, and facilitated the use of the clicker software throughout the session while the Drill Sergeants led the values-based discussion.

Prior to the assessment, the participating Drill Sergeants were familiarized with the tools and clickers that would be used during the pilot. Minimal guidance was provided to the Drill Sergeants on expected facilitation approach using the provided tools. For the control condition, the Drill Sergeant was simply asked to lead the discussion as they normally would, using the provided videos. For the paper-based SJT and clicker conditions, the Drill Sergeants were asked to lead a discussion on the Army Values based on the provided videos and SJTs. In the clicker condition, the clicker software enabled real-time viewing of response breakdowns (i.e., percentage of trainees in the class who selected each option), which helped tailor the discussion. The research team emphasized that the questions were intended to supplement the training and could be used to focus the classroom conversation.

At the end of the session the trainees completed surveys and participated in a brief after action review facilitated by a member of the research team; the Drill Sergeant was not present for this portion of the session. An after action review was also conducted with the participating Drill Sergeants separately. Each session took approximately one and a half hours to complete.

Materials

Peer evaluations. Use of the clicker technology to expedite the current battalion paper-based peer evaluation process was tested with one platoon. Trainees were provided with the Army Values rubric and asked to identify the top three and bottom three members of the platoon on the extent to which they demonstrate the Army Values in BCT. As discussed, the rubric describes behavioral examples of each value based on Drill Sergeant expectations for trainees at their level. The three-point rating scale of the rubric (poor/below standard, average/at standard, excellent/above standard) aligns with a portion of the peer evaluation that was not tested in the pilot study due to time constraints (i.e., rating every member of the squad on each of the seven Army Values on a similar three-point scale). Using clickers, the raters selected the nominated individuals, and were asked to type a values-based rationale for why they were in the top or bottom. The trainees could also use the rubric to help generate meaningful comments as part of the peer evaluation process (see Appendix A).

Ethical decision-making SJTs. SJTs were developed for two CAPE training videos that highlighted personal stories narrated by Soldiers about their experience in basic training. Three SJTs were used per video, each consisting of an item stem describing the ethical dilemma, focal question, and four response options outlining courses of action ranging in degree of "correctness." SJTs were displayed one at a time on a PowerPoint slide and trainees were prompted to select a response via clickers in the clicker condition or on paper in the paper-based SJT condition. For the clicker condition, PowerPoint was integrated with the clicker software and the response breakdown results were displayed on the slide to the class once polling was closed. After trainees responded to a question (either on paper for the paper-based SJT condition or using clickers in the clicker condition), the Drill Sergeant led a values-based discussion before proceeding to the next question. Overall, the response options across the collective set of SJTs were intended to capture increasingly mature and complex considerations regarding moral reasoning and ethical decision-making in accordance with the methods outlined above. The breakdowns of responses for each question were used to evaluate the quality of the SJTs and identify areas for improvement in the item stems and response options.

Learning check test questions. To assess retention of the information presented in the videos, a variety of items were developed, including multiple-choice and true/false items. Seven items were presented for each of the two videos following the ethical decision-making question discussion. The learning check items gauged how well trainees understood the basic situation and retained what was discussed in the videos, compared to the more complex ethical decision-making SJTs. These items also served as attention checks (e.g., asking which Army Values were discussed). An example learning check item from the PVT Gress CAPE video on Respect was: "The main challenge that PVT Gress described in the video is having to work with a peer who…(a) is not a team player, (b) has failed PT multiple times, (c) refuses to follow orders, and (d) threatens to commit suicide." Responses to the learning check questions were not discussed in the class.

Trainee feedback surveys. A self-report survey was administered to trainees to gather quantitative data on their experiences and perceptions of the training. For the clicker condition, the survey consisted of 34 total items. Six items pertained to discussion quality (e.g., "The class discussion was engaging"). Seven items each pertained to SJT content (e.g., "The questions were too complex"), SJT utility (e.g., "The questions made me think more deeply about the Army Values discussed today"), clicker functionality (e.g., "The clickers were easy to use"), and clicker utility (e.g., "Seeing how other students responded to questions was valuable to me"). Trainees responded to each item using a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Different versions of the survey were administered across the three platoons (i.e., conditions). Namely, the control condition trainees provided responses only to the six discussion questions, whereas the paper-based SJT condition provided ratings about the discussion and questions (20 items), but excluded the 14 clicker-related items. The control and paper-based SJT condition trainees provided responses on paper, whereas the clickers were used to capture trainee responses in the clicker condition.

Trainee background questionnaire. At the end of the session, trainees responded to standard demographic questions including age, grade, and educational level. Clickers were used to capture trainee responses in the clicker condition, whereas the other two conditions completed the questionnaire on paper.

Locus of control. Following the background questions, trainees also responded to a six-item locus of control scale developed by Manley, Benavidez, and Dunn (2007). Clickers were used to capture trainee responses in the clicker condition, whereas the other two conditions completed the locus of control items on paper. Locus of control has been defined as an individual's general belief about the degree to which he/she has control over the environment (Rotter, 1966). It was included in this study as an exploratory variable given the relationships that has been demonstrated between internal locus of control and high levels of self-esteem/efficacy, motivation, positive work attitudes, performance, and ethical behavior in past research (Ng, Sorensen, & Eby, 2006; Trevino, 1986). A sample locus of control item included "When you encounter a situation at work or school that you are unhappy with, how often do you conclude there is little one can do about it?" Response options ranged from 1 (*not at all)* to 5 (*very often*). Higher scores on the measure reflected a more external locus of control (i.e., belief that the environment is not responsive to personal agency but rather fate or luck), and lower scores indicate a more internal locus of control (i.e., belief that one has control over his/her environment and success is due to one's own effort and abilities rather than external factors).

After action review questions. At the end of each session, a semi-structured interview was conducted to gather feedback from the trainees on their experience with the tools used in their condition. The discussion focused on the trainees' impressions of the training event, utility and usability of the assessment tools, and level of engagement in the discussion. Example discussion questions included "What are your impressions of the training event just conducted?" "Did you find the videos/questions asked to be useful for learning about the Army Values discussed today? If yes, how? If no, why not?" "How relevant were the videos/questions to your training?" "What did you like/not like about how the peer assessment was conducted today?" "What did you think about using the clicker technology in this setting for the purpose of peer assessments?" Following this discussion, a similar approach was followed to request feedback on these topics from the participating DS. Example discussion questions included "What are your impressions of the training event conducted today?" "What do you think were the trainees' reactions toward the videos, questions, clicker technology?" "What do you think the trainees' gained from today's event?" "What should be kept the same/done differently for the next round of assessments?"

Clicker technology. The platoon in the clicker condition used clicker technology, which allowed multiple choice and text-based responses to various questions. Immediate viewing of group-level results enabled discussion on ethical decision-making. Clickers were also used to gather responses to the peer evaluation in this platoon. The clicker software is integrated with an interactive PowerPoint presentation that creates polling slides and enables real-time viewing of results. The accompanying hardware used in the pilot was the QT2 RF Technology Clicker with QWERTY integrated keypad (shown in Figure 1). The keypad enabled digital entry of comments as part of the peer evaluation. Notably any similar technology to collect individual responses could be employed, but the system uses TurningPoint 5 technology as it had prior approval for use on Army information systems.

Character tracking tool. The clicker software data were exported in an Excel format, enabling integration into the character tracking tool. The tracking tool prototype was tested with the data collected in the pilot study to help identify ways to improve upon the features, displays,

and analytics, as well as to create a more seamless data integration process. The tool was tested and revised iteratively as new requirements and contextual constraints were uncovered.

Data Analysis

Data from the CAPE video SJTs, learning check items, peer evaluation, background questionnaire, and trainee surveys were entered into IBM SPSS Statistics 25 to clean and analyze. Basic descriptive statistics were reviewed, and independent samples *t*-tests and one-way ANOVAs were conducted to examine differences in outcomes across the three conditions.

Part 2 Results

SJT Performance

Significant mean-level differences in learning test performance were found between conditions, $F(2, 138) = 8.02$, $p = .001$. Post hoc comparisons using Tukey's HSD test indicated that participants in the clicker condition ($M = 12.42$, $SD = 1.57$) scored significantly lower than participants in the control ($M = 13.37$, $SD = .83$; $p < .001$) and paper-based SJT ($M = 13.04$, $SD = .98$; $p = .03$) conditions. No significant differences were found between participant scores in the control and paper-based SJT conditions ($p > .05$). There was not a significant difference in SJT performance between the paper-based SJT and clicker-based SJT conditions ($p > .05$).

Locus of Control

A more internal locus of control (unstandardized $b = .35$, $p = .01$) – but not age, education, or grade – significantly predicted learning test performance, although the overall variance explained was somewhat negligible, $F(4, 129) = 2.46$, $p = .048$, Adj. $R^2 = .04$. Having a more internal locus of control was also significantly related to more positive perceptions of discussion quality ($r = .45$, $p = .003$). These results indicate that trainees with more internal locus of control may believe that they have greater control over their learning and take action to more positively impact their learning experiences (e.g., engage in classroom discussion, ask questions, pay attention).

Trainee Feedback

Results from the trainee feedback survey suggest that trainee perceptions about the class discussion were positive. Significant group differences were found between conditions, $F(2, 137) = 4.67$, $p = .01$. Post hoc comparisons using Tukey's HSD test indicate that the mean rating on discussion quality and utility for the control condition ($M = 4.37$, $SD = .37$) was significantly higher than both the paper-based SJT ($M = 4.10$, $SD = .57$) and clicker conditions ($M = 4.10$, $SD = .49$). Significant differences were also found for perceptions of CAPE video relevance, engagement in the class discussion, and class pace. Specifically, trainees in the paper-based SJT condition perceived the videos as more relevant ($M = 4.22$, $SD = .81$) than in the clicker ($M = 3.77$, $SD = 1.08$) condition, $F(2, 136) = 3.62$, $p = .03$. Self-rated level of engagement in the discussion was higher in the control ($M = 4.58$, $SD = .50$) than both paper-based SJT ($M = 4.17$, $SD = .64$) and clicker ($M = 4.00$, $SD = .85$) conditions, $F(2, 136) = 8.66$, $p < .001$. Furthermore, trainees in the control condition ($M = 4.69$, $SD = .47$) rated the pace of the discussion more

favorably than participants in both the paper-based SJT (M = 4.32, SD = .78) and clicker (M = 4.26, SD = .74) conditions, $F(2, 135) = 5.27$, $p < .01$.

Although trainee perceptions were assessed on the discussion quality, question quality and utility, and clicker functionality and utility, only perceptions of question content quality (unstandardized $b = 2.21$, $p < .001$) significantly predicted knowledge retention (i.e., learning check test questions), $F(5, 40) = 5.48$, $p = .001$, Adj. $R^2 = .33$. That is, those trainees who perceived the SJT questions were of high quality performed better on the knowledge retention test. This points to the importance of meaningful and well-developed material-based questions for Soldier learning.

Overall, trainee feedback on the assessment tools and process was positive. As part of the after action review, trainees indicated that the ethical decision-making questions were thought-provoking, and promoted sharing of opinions and introspection; the videos seemed outdated but relevant, and helped springboard discussion with DS; and discussion helped build a shared understanding of the Army Values and how to apply them day-to-day. The trainees especially liked the training structure and opportunity to voice their opinions, see how their peers responded to the questions, and hear other perspectives. In the future, the trainees would like to see more challenging questions based on ethical dilemmas in an operational setting. The trainees in the clicker condition shared that the clickers were easy to use and useful especially for expediting the peer evaluation process. They also indicated that the Army Values rubric was helpful for providing constructive feedback to peers.

Drill Sergeant Feedback

The Drill Sergeants also liked the structure of the training but cautioned that the added tools may put additional load on trainees. They shared that the videos and questions provided insight into the minds of the trainees, especially the "gray man" (i.e., trainees who rarely speak up or "fly under the radar"). The Drill Sergeants requested being able to input their own questions into the clicker software or select questions from a set. The Drill Sergeants agreed that the clickers helped achieve full participation and were efficient for testing and peer evaluation; rather than tallying up responses by hand from paper forms, the clicker software provided an output of the raw data for immediate use. Drill Sergeants also emphasized the benefit of the Army Values rubric for trainees, and noted that it could help to enrich the peer evaluation comments. Upon viewing the tracking tool, the Drill Sergeants highlighted its prospective utility for identifying struggling individuals and providing programmatic insights. Although the Army Values are typically taught informally in the field, the Drill Sergeants indicated that they would use these tools to supplement current practices in a classroom setting if available.

Part 2 Discussion

The character assessment tools and methods developed in Part 1 were tested in Part 2 to measure character, facilitate Army Values discussion, and to standardize and expedite the current peer evaluation process. Findings from the learning test and SJT performance across conditions indicate that instruction and technology could theoretically interact to differentially impact training outcomes. For example, the clicker condition performed better on ethical decision-making compared to the paper-based SJT condition, whereas the control condition performed the

best on knowledge retention compared to the other two conditions. These results suggest that technology alone does not result in learning, and if not harnessed properly, it may in fact distract from learning. As such, technology implementation and the accompanying instructional strategy must align with the primary training objectives. This way, the facilitator can curate the learning experience for trainees in a manner that directly targets the intended outcomes. Using instructional strategies that bolster student confidence and provide the necessary scaffolds at the appropriate time during training can support transition to a more internal locus of control, which could contribute to the creation and proactive maintenance of positive learning experiences.

Preliminary findings also indicate that while the assessment tools encouraged thought and introspection in learners, reaping the benefits of these tools ultimately depended on the facilitator. There were substantial differences in facilitation style across DS. It was noted by researchers observing that the control condition Drill Sergeant was highly experienced and proficient in leading engaging and meaningful discussions. In contrast, there was minimal interaction between Drill Sergeant and trainees in the clicker condition, as the Drill Sergeant used the response option breakdown charts provided by the clicker software to differentiate wrong from right courses of action (i.e., focus of discussion was on performance) rather than to generate rich discussion on underlying issue (e.g., "why did you answer that way"). As such, sufficient guidance on how to appropriately harness the technology to guide instructional technique and lead an engaging, interactive discussion is critical for learning in any setting. Instructors could be trained on effective questioning, probing, and elaboration techniques based on group-level responses, avoiding singling out individuals who selected the least optimal response, and tailoring the discussion to align with the level of understanding and needs of the class (e.g., relating material to day-to-day activities).

Overall, the end-users had positive reactions toward the assessment tools and process, with some caveats. The CAPE videos and SJT questions can serve as a springboard for discussion and help create an environment that promotes sharing among trainees. There was clear utility of clickers for peer evaluations, but further investigation is needed to determine utility for facilitating class discussion. The added level of structure introduced by the tools was positively received; however, flexibility of use is valued among instructors. Optimally, this approach, especially when combined with clicker technology that enables real-time viewing of trainee responses, allows Drill Sergeants to tailor their instruction to the needs of the classroom and therefore reduce the time needed to grasp the material. If Drill Sergeants are provided the option to add their own questions to the clicker software, a resource guide must be provided that describes the process for generating high-quality items and response options. Accordingly, instructions in Appendix C describe basic item writing guidelines, best practices, and instructions on how to write effective character scenarios.

Part 3 Method: Longitudinal Evaluation of the Assessment System and Approach

Based on the results and feedback from Part 2, the peer evaluations and video-based SJTs were refined. For instance, a selection of operational and BCT-focused videos was employed. In addition, the research team adjusted guidance provided to the trainees on clicker use and assessment process, and expedited the assessment events. Given these changes, the next part of this research focused on longitudinally assessing character with one company across an entire

cycle of BCT to more rigorously track individual and unit-level performance. These results further informed the refinement of the assessment tools.

Participants and Procedure

Assessments were conducted at three time points approximately three weeks apart during Red (beginning), White (middle), and Blue (end) phases of BCT. Each phase of BCT was approximately three weeks long. A total of 156 trainees in four platoons participated in the assessments. The platoons ranged from 38 to 40 trainees. Assessment with each platoon was conducted simultaneously in four separate classrooms, and was facilitated by a member of the research team, and in some cases, a DS. Although assessment was conducted with the same company at each time point, movement in and out of the platoons or company was noted due to absence, opting out of the research, or attrition from BCT. A total of 136 trainees participated in all three assessments, 16 participated in two, and seven participated in only one. The average reported age of the participating trainees across the three time points was 22.35 (*SD* = 3.79). Approximately fifty-five percent (54.6%) of the trainees reported their grade to be E-1, 16.9% indicated a grade of E-2, 16.2% held a grade of E-3, and 12.3% held a grade of E-4. The majority of the trainees reported that their highest level of education was high school diploma (39.7%), followed by some college (33.8%), Bachelor's degree (11.0%), Associate degree (7.1%), GED (7.1%), and graduate level work or degree (1.3%). Age, grade, and education distributions were not significantly different across platoons.

Trainee responses on peer evaluations, video SJTs, and background questionnaires were captured using the clicker technology described in Part 2. The research team trained the participants to use the clickers in the beginning of the session, and facilitated the use of the clicker software throughout the session. The trainees then evaluated designated peers via clickers following the approved battalion peer evaluation approach. A similar, but refined, process was used in Part 3 compared to Part 2. As part of the peer evaluation, trainees rated each member of their squad on the Army Values and identified the top and bottom three members of their platoon. Trainees were also asked to comment on the squad members they rated and the trainees they nominated for top and bottom three. The trainees were asked to reference the Army Values rubric during this process to help provide accurate, actionable, and meaningful feedback to peers. After completing the peer evaluation, the trainees viewed selected CAPE videos, and responded, via clickers, to SJTs pertaining to the situation discussed in the videos. At their discretion, the Drill Sergeants facilitated a discussion based on the videos and trainee responses to the SJT questions. In some cases, the Drill Sergeant was not present at all during the assessment, and it was administered by researchers. At the end of the session, the trainees provided responses to a background questionnaire using the clickers, and verbally shared feedback about the assessment tools and process.

Materials

Army Values rubric. The Army Values rubric used in Part 2 was provided to trainees in the beginning of the session (see Appendix A). The rubric describes behavioral examples of each Army Value based on Drill Sergeant expectations for trainees at their level. The behavioral examples spanned three proficiency categories ranging from poor to excellent.

Peer Army Values ratings. While referencing the Army Values rubric, trainees provided ratings for each member of their squad on a scale of 1 to 3 reflecting the degree to which the ratee exhibits each Army Value (Loyalty, Duty, Respect, Selfless Service, Honor, Integrity, and Personal Courage) in daily life, and provided comments to justify their rating. For analytic purposes, the ratings for each ratee were averaged across all raters to arrive at a single composite score for each Army Value as well as an overall Army Value score for each individual. Separate scores were created per phase (i.e., Red, White, Blue) for individual values, as well as an overall score for each phase. Higher scores indicated more favorable ratings on the Army Values.

Peer top and bottom ranking. As in Part 2, trainees identified the top three and bottom three members of the platoon based on the extent to which they demonstrate the Army Values in BCT, and entered a rationale for why those individuals were selected. A general comment about experiences in BCT was also requested as part of the peer evaluation. The number of times an individual was ranked in the top and bottom of the platoon was tallied per phase and overall across phases.

Peer trust questions. The trainees were also asked to respond *yes* or *no* to the following two trust-related questions for each member of their squad: "Would you trust this Soldier to do his job and duties as a member of your unit in a combat zone or forward deployed environment?" and "Would you trust this Soldier as the person immediately next to you to do his job and duties in a firefight?" Percent of peers who reported *yes* to both questions (i.e., would trust the Soldier being rated) was calculated and averaged at each phase to create a score for each phase, and averaged across phases for a total score.

Ethical decision-making SJTs. A total of 18 SJTs, developed in Part 1 and refined in Part 2, were used. The SJTs were based on nine CAPE videos identified by Drill Sergeants and the research team. Each video comprised two segments, and an SJT was associated with each segment of each video. SJTs were presented for each video segment in a sequential manner and trainees were asked to respond using clickers as described in Part 2. Responses to the SJTs were scored using the key developed in Part 1 and refined in Part 2, where the higher the score, the more optimal the selected response. See Appendix D for the full set of SJTs used in Part 3 along with the scoring key.

Trainee background questionnaire and locus of control. The background questionnaire and locus of control scale used in Part 2 were also used in Part 3.

After action review questions. After the assessments were complete, the research team followed the same semi-structured interview protocol as in the previous phase to gather feedback from the trainees on their experience with the assessment tools. The discussion specifically focused on the trainees' impressions of the training event, utility and usability of the tools, and level of engagement with the assessment process. The Drill Sergeants were not present for the after action review with the trainees. Following this discussion, a similar approach was followed to request feedback from the participating Drill Sergeants if time allowed.

Data Analysis

Following the assessment, the research team exported the raw clicker data from the peer evaluations, SJTs, demographics, and locus of control to Excel. The peer and SJT assessment results were imported into a version of the Character Tracking Tool for additional testing and refinement. The raw data were also entered into IBM SPSS Statistics 25 for further analysis. Descriptive statistics were computed as well as a series of bivariate correlations to examine relationships among focal variables. Given the within-subjects design of the study, repeated measures ANOVAs and correlations were conducted to examine relationships between scores at different measurement time points, as well as within-subjects differences on key variables across Red, White, and Blue Phases. Multiple regression analyses were also conducted to assess the relationship between variables across phases of training.

Part 3 Results

Relationships between Focal Variables

Table 2 presents descriptive statistics, correlations, and reliability estimates (in parentheses) for the focal variables for Red, White, and Blue Phase assessment data.

Overall, older trainees held higher grades (r_s = .37, p < .001), were more educated (r_s = .43, p < .001), were rated more favorably by peers on the Army Values (r_s = .26, p = .002), received more votes for Top 3 in the platoon based on the Army Values (r_s = .23, p = .004), and performed better on the SJTs based on the basic training CAPE videos (r_s = .17, p = .05). Grade was also significantly correlated to education level (r_s = .63, p < .001), peer ratings on the Army Values (r_s = .25, p = .002), and number of votes received for Bottom 3 in the platoon based on the Army Values (r_s = -.17, p = .03). Furthermore, trainees with higher levels of education were rated more favorably on the Army Values (r_s = .34, p < .001), received more votes for Top 3 in the platoon (r_s = .20, p = .01), received fewer votes for Bottom 3 in the platoon (r_s = -.19, p = .02), and were trusted by a greater proportion of peers in an operational context (r_s = .21, p = .01). In contrast, trainees with a more external locus of control were rated less favorably by their peers on the Army Values (r_s = -.29, p = .002), received fewer votes for Top 3 (r_s = -.26, p = .01), and received more votes for Bottom 3 (r_s = .24, p = .01).

Correlates of Top 3 and Bottom 3 votes were examined to determine if distinct underlying factors predict an individual's decision for whether a peer is considered at the top or bottom. High ratings on the Army Values are strongly correlated with trust in an operational context (r_s = .83, p < .001). Furthermore, trust is negatively correlated with being nominated for the Bottom 3 (r_s = -.68, p < .001) and positively correlated with being nominated for the Top 3 (r_s = .55, p < .001).

It is important to note that the magnitude of relationships between the focal variables varied across platoons. In one platoon, SJT performance on the basic training CAPE videos was significantly related to peer ratings on the Army Values (r_s = .42, p = .01) and trust (r_s = .48, p = .003). In another platoon, external locus of control was negatively correlated with grade (r_s = -.41, p = .03) and trust by peers (r_s = -.43, p = .02), and positively correlated with Bottom 3 votes (r_s = .45, p = .02). In a third platoon, older trainees performed significantly better on the SJT video questions overall (r_s = .46, p = .01) and received more votes for Top 3 in the platoon (r_s =

.37, p = .02). Also in this platoon, SJT performance on the operational videos was significantly related to educational level (r_s = .46, p = .01) and grade (r_s = .40, p = .02), and external locus of control was negatively related to Top 3 votes (r_s = -.40, p = .04). In a fourth platoon, trainees holding higher grades received significantly better peer ratings on the Army Values (r_s = .38 p = .02) and fewer Bottom 3 votes (r_s = -.37, p = .03). In this platoon, educational level was also negatively related to Bottom 3 votes (r_s = -.33, p = .05).

Table 2

Descriptive Statistics, Correlations, and Reliability Estimates for Focal Variables Based on Red, White, and Blue Phase Assessment Data

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
1. Age	22.35	3.79										
2. Grade	1.86	1.10	.37***									
3. Education level	1.66	.69	.43***	.63***								
4. Locus of control	2.25	.45	-.09	-.18+	-.13	(.61)						
5. Overall SJT performance	3.38	.23	.20*	.03	.07	-.06	(.40)					
6. SJT performance on basic videos	3.50	.28	.17*	-.05	.07	-.004	.70***	(.41)				
7. SJT performance on operational videos	3.22	.34	.04	.09	.06	-.11	.77***	.16+	(.26)			
8. Army Values rating	2.44	.33	.26**	.25**	34***	-.29**	.12	.16+	-.03	(.98)		
9. Top 3 votes	7.93	14.09	.23**	.13	.20*	-.26**	.11	.14+	-.01	.63***		
10. Bottom 3 votes	7.53	12.91	-.09	-.17*	-.19*	.24**	-.04	-.12	.04	-.66***	-.38**	
11. Trust	78.47	21.99	.10	.09	.21**	-.17+	.10	.16+	-.02	.83***	.55***	-.68***

Note. *N* ranges from 114 to 156 due to pairwise deletion. Spearman correlation coefficients are reported. Cronbach alpha reliability coefficients are along the diagonal where applicable. **Grade** was coded as 1-4 to represent grade levels ranging from E-1 to E-4. **Education level** was coded as 1 (GED/High School Diploma), 2 (Some College/Associate degree), and 3 (Bachelor's degree/Graduate level). **SJT performance** represents trainee responses to the ethical-decision making questions based on the CAPE videos (higher scores reflect more advanced levels of performance across the four response options). Separate scores were created for SJT performance on the basic training and operational videos. Higher **locus of control** scores indicate a more external locus on a 5-point scale. **Army Values** peer ratings variable reflects the average of all ratings provided by peers on a 3-point scale. **Top/Bottom 3 votes** are the total number of times an individual was selected by peers as Top 3/Bottom 3 in the platoon based on the Army Values. The **trust** variable reflects the percentage of peers who responded "*Yes*" to the two trust questions in the peer evaluation. Reliability estimates for locus of control and peer ratings (in parentheses) were comparable across the three measurement time points. $^{+}$ p < .10; * p < .05; ** p < .01; *** p < .001

Examining the relationships at the BCT phase level showed that a higher grade was significantly correlated with having a more internal locus of control (r_s = -.21, p = .02) at the end of training (i.e., Blue Phase). Similarly, higher levels of education and grade were significantly related to more favorable peer ratings at the end of training (r_s = .27, p = .001 and r_s = .23, p =

.01, respectively). Notably, higher performance on the Red Phase video-based SJTs was significantly related to Blue Phase peer ratings on the Army Values (r_s = .18, p = .03) and trust (r_s = .19, p = .03).

Group Differences

Platoons. Platoons differed significantly only on operational video SJT performance, where one platoon was significantly lower than the rest, $F(3, 128) = 6.93$, $p < .001$. Mean levels on demographics were generally comparable across platoons, although the range on locus of control was greater than the other variables, with more outliers falling toward the external end of the continuum. These results imply group-level, instructional, or other third variable effects that require a more in-depth analysis. The presence and quality of DS-led discussion based on the Army Values may have also contributed to these differences.

Educational level. One-way ANOVA results showed significant differences among educational levels on the Army Values ratings, $F(2, 143) = 7.95$, p = .001; Top 3 votes, $F(2, 41.13) = 5.17$, p = .01, and trust, $F(2, 52.17) = 4.66$, p = .01.[2] Post hoc analyses revealed that trainees whose highest level of education was GED/high school diploma received significantly less favorable ratings on the Army Values (n = 68; M = 2.34, SD = .34) compared to trainees with some college/Associate degree (n = 59; M = 2.50, SD = .30; p = .01), and Bachelor's degree and above (n = 19; M = 2.63, SD = .28; p = .002). With respect to Top 3 votes, only the difference between trainees with a GED/high school diploma (M = 4.42, SD = 7.32) and some college/Associate degree was statistically significant (M = 9.90, SD = 16.43; p = .04). The post hoc results for trust were similar in that only the difference between trainees with a GED/high school diploma (M = 72.53, SD = 24.69) and some college/Associate degree was statistically significant (M = 83.71, SD = 17.69; p = .01). The differences between some college/Associate degree and Bachelor's degree and above were not statistically significant ($p > .05$).

Grade. Significant differences were also found among grades on basic training video SJT performance, $F(3, 131) = 2.85$, p = .04, and the Army Values ratings, $F(3, 142) = 4.00$, p = .01. Post hoc analyses using Tukey's HSD demonstrated that trainees who held a grade of E-4 performed significantly better on the basic training video SJTs (n = 16; M = 3.60, SD = .21) than E-2s (n = 22; M = 3.37, SD = .32; p = .05). E-4s also received significantly more favorable peer ratings (M = 2.67, SD = .26) than E-1s (M = 2.38, SD = .33; p = .004).

Closer Examination of Ethical Decision-Making Performance

Basic training vs. operational video SJTs. Interestingly, performance on the basic training video SJTs was not significantly related to performance on the operational video SJTs (r_s = .16, p = .07). Performance on the operational videos (M = 3.86, SD = .49) was significantly

[2] The assumption of homogeneity of variance was violated for Top 3/Bottom 3 votes and trust, as indicated by significant Levene's statistic; therefore, the Welch's F and Games-Howell post hoc test results were reported for one-way ANOVA analyses involving educational level.

lower than on the basic training videos (M = 4.46, SD = .34), $t(123) = 11.75$, $p < .001$. This could be due to the difference in complexity of the situation depicted in the videos and could suggest that trainees need to draw on different types of knowledge and experiences in order to respond to the ethical dilemma.

Preliminary evidence for SJT validity. Multiple regression analysis results showed that, in Red Phase, both basic training video SJT performance (unstandardized b = .26, p = .01) and locus of control (unstandardized b = -.25, $p < .001$) significantly predicted peer ratings on the Army Values, $F(3, 124) = 8.51$, $p < .001$, Adj. R^2 = .15. Basic training video SJT performance (unstandardized b = 19.97, p = .003) and locus of control (unstandardized b = -20.29, $p < .001$) also significantly predicted peer trust, $F(5, 122) = 6.90$, $p < .001$, Adj. R^2 = .19. Age was not a significant predictor of peer ratings or trust in Red Phase ($p > .05$).

To determine whether SJT performance predicted peer perceptions above and beyond other study variables, hierarchical linear regression analyses were conducted. As shown in Table 3, in Red Phase, performance on the SJTs predicted peer ratings (unstandardized b = .26, p = .01) above and beyond locus of control and age, $F(3, 124) = 8.51$, $p < .001$, ΔR^2 = .04. SJT performance also predicted peer trust perceptions in Red Phase (unstandardized b = 19.97, p = .003) above and beyond locus of control and age, $F(3, 124) = 11.19$, $p < .001$, ΔR^2 = .06.

Table 3

Multiple Regression Analysis Results of Video-Based SJT Performance Predicting Peer Evaluations

Variable	Red Phase				White Phase			
	Peer Ratings		Peer Trust		Peer Ratings		Peer Trust	
	B	ΔR^2	*B*	ΔR^2	*B*	ΔR^2	*B*	ΔR^2
Step 1		12.7%***		15.5%***		5.3%*		6.8%**
Trainee age	-.01 (.01)		-.02 (.53)		.02 (.01)*		-.10 (.51)	
Trainee LoC	-.25 (.07)***		-20.24 (4.23)***		-.09 (.06)		-11.91 (4.00)**	
Step 2		4.4%**		5.8%**		8.1%**		12.2%***
Trainee age	.01 (.01)		-.22 (.52)		.01 (.01)+		-.17 (.49)	
Trainee LoC	-.25 (.06)***		-20.29 (4.10)***		-.09 (.06)		-12.00 (3.75)**	
Trainee ethical decision-making	.26 (.10)**		19.97 (5.22)**		.32 (.09)**		25.88 (6.01)***	
Total R^2		17.1%***		21.3%***		13.4%***		19.0%***

Note. N = 128. Unstandardized coefficients are reported. Standard errors are reported in parentheses. LoC = Locus of Control corresponding to each depicted phase, where higher scores indicate more external LoC; Trainee ethical decision-making refers to performance on the Red Phase Army Value CAPE video SJTs.
+ $p < .10$; ** $p < .01$; *** $p < .001$

Interestingly, Red Phase SJT performance also predicted White Phase peer ratings (unstandardized b = .32, p = .001), whereas ratee age and White Phase locus of control did not, $F(3, 123) = 6.37$, $p < .001$, ΔR^2 = .08. Red Phase SJT also predicted White Phase peer trust (unstandardized b = 25.88, $p < .001$) above and beyond locus of control, $F(3, 123) = 9.64$, $p < .001$, ΔR^2 = .12, whereas age did not ($p > .05$). White and Blue Phase video-based SJT

performance did not significantly predict peer evaluations ($p > .05$). However, Blue Phase locus of control (but not age or SJT performance) was a significantly predictor of Blue Phase peer ratings (unstandardized $b = -.19$, $p = .001$; $F(3, 116) = 5.08$, $p = .002$, Adj. $R^2 = .09$) and trust (unstandardized $b = -8.09$, $p = .04$; $F(3, 116) = 2.89$, $p = .04$, Adj. $R^2 = .05$). Collectively, these results provide some promising initial validity evidence for the video-based ethical-decision-making SJTs.

A closer examination of the relationship between SJT performance and peer evaluations revealed that the Red Phase basic training CAPE video questions on duty and the White Phase operational CAPE video questions on "An Ethical Pause" were strongly and positively correlated to peer evaluations across the three phases.

Results across Training Phases

Table 4 shows descriptive statistics for the focal variables across the three BCT training phases. Examination of mean-level trends indicates more internal locus of control, more favorable peer ratings, and improved SJT performance as training progressed.

Table 4

Descriptive Statistics for Focal Variables across Training Phases

Variable	Red Phase *M* (*SD*)	White Phase *M* (*SD*)	Blue Phase *M* (*SD*)
Locus of control	2.33 (.48)	2.25 (.55)	2.21 (.56)
Peer ratings on the Army Values	2.37 (.39)	2.43 (.41)	2.49 (.36)
Peer trust	78.36 (25.93)	75.98 (25.64)	78.22 (24.45)
Basic training video SJT performance	3.47 (.30)	3.52 (.73)	3.57 (.38)
Operational video SJT performance	N/A	3.11 (.44)	3.33 (.44)

Note. N ranges from 132 to 156. Means and standard deviations are reported. Higher locus of control scores indicate a more external locus on a 5-point scale. SJT performance represents trainee responses to the ethical-decision making questions based on the CAPE videos (higher scores reflect more advanced levels of performance across the four response options). Army Values peer ratings variable reflects the average of all ratings provided by peers on a 3-point scale. The trust variable reflects the percentage of peers who responded "Yes" to the two trust questions in the peer evaluation.

A repeated measures ANOVA with a Greenhouse-Geisser correction revealed that locus of control significantly differed across phases of BCT, $F(1.95, 220.51) = 3.34$, $p = .04$. However, post hoc analyses indicated that only the difference between Red and Blue Phase was statistically significant ($p = .02$). Peer ratings of Army Values also significantly differed across phases, $F(1.80, 261.50) = 12.61$, $p < .001$, and specifically, Red Phase peer ratings were significantly lower than those in White and Blue Phases ($p < .001$). The differences on peer trust across phases were not significant. Performance on the operational video SJTs significantly improved from White to Blue Phase (operational videos were not used in Red Phase), $t(126) =$

4.93, $p < .001$. Despite the upward mean-level trend, the differences in basic training video SJT performance were not statistically significant ($p > .05$).

The Army Values as predictors of peer perceptions. Of interest was also exploring the saliency of each of the seven Army Values in peer perceptions of trust and identifying someone as top or bottom of the platoon across the BCT phases. A series of multiple regression analyses were conducted to examine which values emerged as significant predictors within and across phases relative to peer trust and Top 3/Bottom 3 votes. The results are described in detail for each BCT phase.

Red phase. In Red Phase, trainees were rated the least favorably on personal courage ($M = 2.31$, $SD = .41$) and selfless service ($M = 2.28$, $SD = .42$), and most favorably on respect ($M = 2.44$, $SD = .41$) and integrity ($M = 2.42$, $SD = .38$). In this phase, out of the seven Army Values, ratings on duty (unstandardized $b = 15.92$, $p = .003$) and personal courage (unstandardized $b = 20.23$, $p < .001$) significantly predicted peer trust, $F(7, 142) = 70.98$, $p < .001$, Adj. $R^2 = .77$. Ratings on loyalty (unstandardized $b = -4.66$, $p = .03$) significantly predicted number of votes for Bottom 3 in the platoon, $F(7, 142) = 20.94$, $p < .001$, Adj. $R^2 = .48$. In fact, higher ratings of loyalty in Red Phase (unstandardized $b = 20.86$, $p = .04$) also predicted peer trust perceptions during White Phase, $F(7, 140) = 22.06$, $p < .001$, Adj. $R^2 = .50$.

White phase. In White Phase, loyalty (unstandardized $b = 31.17$, $p < .001$) and personal courage (unstandardized $b = 12.24$, $p = .04$) were the only significant predictors of peer trust perceptions, $F(7, 144) = 68.23$, $p < .001$, Adj. $R^2 = .76$. Respect (unstandardized $b = -4.07$, $p = .01$) significantly predicted Bottom 3 votes in this phase, $F(7, 144) = 23.32$, $p < .001$, Adj. $R^2 = .51$. Low ratings on respect in White Phase (unstandardized $b = -4.40$, $p = .05$) also positively predicted Bottom 3 votes in Blue Phase, $F(7, 144) = 9.54$, $p < .001$, Adj. $R^2 = .28$. Furthermore, White Phase ratings on loyalty (unstandardized $b = 25.21$, $p = .02$), duty (unstandardized $b = 21.26$, $p = .02$), and selfless service (unstandardized $b = 21.19$, $p = .01$) were significant predictors of peer trust perceptions in Blue Phase, $F(7, 140) = 19.16$, $p < .001$, Adj. $R^2 = .46$.

Blue phase. Out of the seven Army Values, only selfless service (unstandardized $b = 6.30$, $p = .01$) significantly predicted Top 3 votes, $F(7, 142) = 4.83$, $p < .001$, Adj. $R^2 = .15$; however, the overall variance that the peer ratings on the Army Values explained was much less in Blue Phase compared to other phases, meaning that there are other factors that play a bigger role in peer perceptions at that time (e.g., technical/tactical competence). Integrity (unstandardized $b = -3.78$, $p = .02$) was the only significant predictor of being voted in the Bottom 3 of the platoon in Blue Phase, $F(7, 142) = 11.70$, $p < .001$, Adj. $R^2 = .34$. Duty (unstandardized $b = 19.28$, $p = .01$) was the only significant predictor of trust in this phase, $F(7, 142) = 43.94$, $p < .001$, Adj. $R^2 = .67$. Interestingly, the greatest mean-level improvement as training progressed from Red to Blue Phase was in personal courage ($\Delta M = .19$) and selfless service ($\Delta M = .14$).

Overall, the results collectively showed that the Army Values differentially predicted peer trust perceptions and the decision to select a trainee for top or bottom of the platoon across the different phases of training. Early on, in Red Phase, ratings of duty and personal courage were related to measures of trust among peers, and low loyalty ratings predicted nominations for Bottom 3. In White Phase, loyalty and personal courage ratings predicted trust ratings, and low

ratings on respect predicted nominations for Bottom 3. Interestingly, only duty predicted trust in Blue Phase. In addition, high ratings on selfless service predicted Top 3 nominations, whereas low of integrity predicted Bottom 3 nominations. The results also indicated that ratings earlier in training predicted peer perceptions later on in training, thus highlighting the lasting impact of values-related impressions on the formation and maintenance of trust among peers throughout BCT.

Drill Sergeant and Trainee Feedback

The feedback obtained from the trainees and Drill Sergeants across the four platoons was largely positive. Specifically, they liked the clickers for the peer evaluations as the process was quick and efficient. The videos, questions, and described issues were perceived to be relevant, inspiring, and in line with trainee experiences in BCT. The trainees indicated that they would like to see more videos and SJTs in the future, were engaged in the video-based discussion, and expressed interest in receiving peer evaluation results right away. Trainees especially benefitted from seeing how the other members of the platoon responded to the SJTs; the resulting discussion helped trainees to better understand others' perspectives. The Army Values rubric that accompanied the peer evaluations was helpful in terms of generating meaningful, actionable, and formative comments about peers.

Exploratory Analyses

Preliminary analysis of individual trajectories. Of additional interest to BCT unit leadership was an individual's path as training progressed, especially for those who struggled initially. Hence, given the longitudinal nature of the data, an exploration of the top and bottom 15 trainees (approximately the top/bottom 10 percent) in the company based on mean peer ratings during Red Phase was conducted. As shown in Figure 5, some trainees started out stronger than others in Red Phase based on peer evaluations, and ratings tended to change over time in unique ways.

Relative to those in the bottom, the top trainees tended to be older, more advanced in grade and education, and have a more internal locus of control. This group tended to score better on peer ratings of loyalty and duty, and as training progressed, they stayed highly rated by their peers throughout BCT. As a group, the bottom 15 trainees tended to have their lowest ratings in selfless service. Some trainees in this group dramatically improved whereas others did not. Additionally, those trainees who improved throughout BCT showed a greater tendency toward a more internal locus of control over time. Those trainees who struggled in Red Phase, improved in White Phase, but then struggled again in Blue Phase tended to shift toward a more internal locus of control in White Phase, but then reverted to a more external locus of control as training progressed in Blue Phase. This pattern of results may have been due to the nature of the activities in the training curriculum that may foster shifts in trainee confidence and self-efficacy. For example, successful performance on the White Phase exercises, which as a whole are more individualistic in nature (e.g., marksmanship), may give struggling trainees a boost that is captured by a more internal locus of control. As training progresses, however, the positive impact of this success may be short-lived and the trainees may return to their baseline in Blue Phase. Thus, these data suggest that innovations in instructional strategies might focus on how

to sustain and continue to amplify the positive impact of training events on Soldier confidence, motivation, and growth.

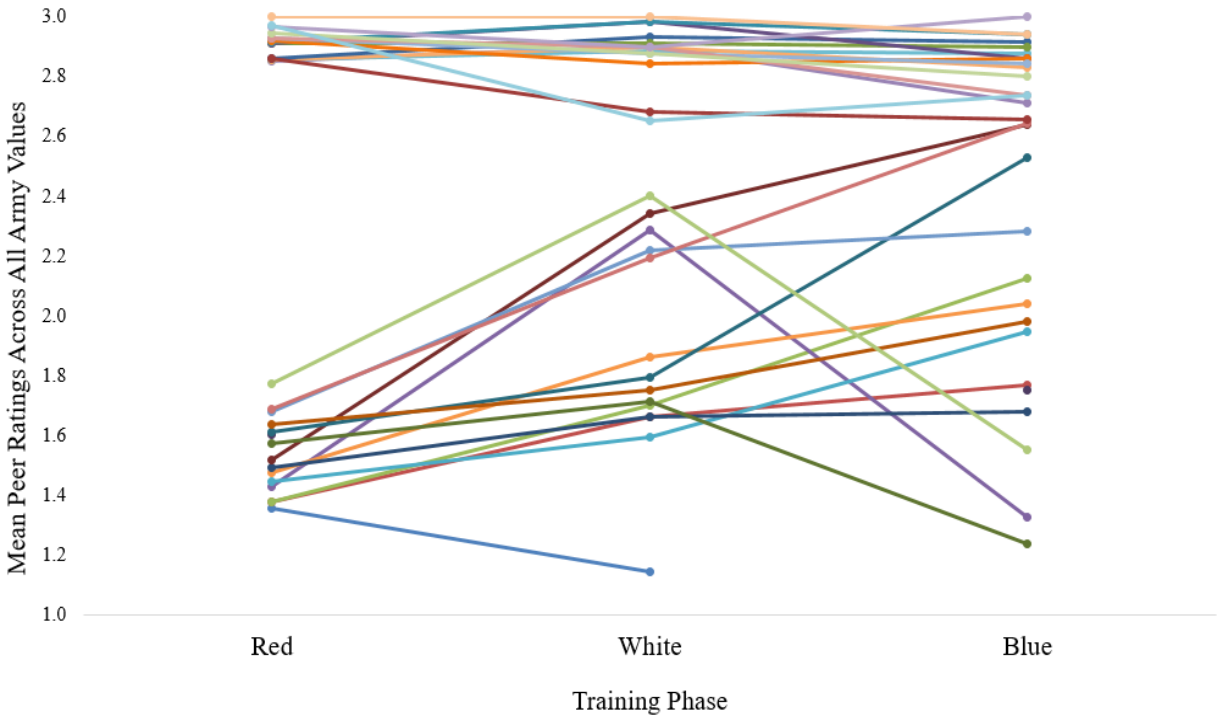


Figure 5. Mean peer ratings across phases for top and bottom 15 trainees from Red Phase. Each line represents an individual. The group of lines toward the top of the graph represent the top 15. The group of lines toward the bottom of the graph represent the bottom 15.

Closer examination of individual trajectories via clustering. Building on the overall insights gained from the primary analyses, as well as those from exploration of the trends associated with the top and bottom trainees, additional analyses were conducted on the entire company in order to examine trends in the larger data set. The goal of these analyses was to further examine the factors that might influence growth, while seeking insights that might guide future investigation and intervention strategies. Accordingly, as a step toward understanding the different paths trainees take through BCT, a clustering algorithm was implemented on the data. The goal of the clustering algorithm was to find groups of trainees with similar data characteristics to each other, or more generally, similar paths through BCT. This process enabled an in-depth exploration of the groups and identification of shared characteristics.

Method. Before conducting this analysis, the data were preprocessed. In particular, this analysis was limited only to the trainees who participated in all data collection sessions. Hence, any trainees who did not attend all three phases or who missed an entire segment of peer evaluations, locus of control, or SJTs were excluded from the analysis. This resulted in excluding 33 trainees, and leaving 123 participants in the clustering process. The demographics for the 123 participants included in the cluster analyses are shown in Table 5. For these participants, there were still 139 missing entries across 34 items and 66 participants. To impute these missing data, two strategies were employed. For age, the average age across all participants was substituted for missing entries. For the remaining variables, missing entries

were replaced with the average of all participants on that question/item and that participant's average on other questions in the same category. Ultimately, for this analysis, the relevant data features (i.e., variables) included were platoon, grade, and education (all categorical). In addition, the analyses included age, locus of control, peer ratings on the Army Values, SJT performance, Top and Bottom 3 votes, and trust percentage. Since these features all have different scales and clustering is sensitive to scaling, the last preprocessing step was to normalize each data feature on a scale of 0-1.

In order to cluster participants, a k-prototypes clustering algorithm was used (Huang, 1997). The k-prototypes algorithm is a modification of the more commonly used k-means clustering algorithm that accounts for the nature of categorical data. Informally, the k-prototypes algorithm follows four steps: (1) initialization – choose k random points in the feature space as centroids (cluster centers), (2) find the closest centroid for each point and label that point as belonging to that corresponding cluster, (3) recalculate the centroid for each cluster as the average of all points in that cluster, and (4) repeat steps 2 and 3 until clusters converge (no points switch clusters at step 2). The output of this method is a cluster label for each participant as well as the centroid (or average) for each cluster. In the analyses presented below, 10 clusters were selected as a reasonable balance between cluster variation (small number of clusters) and cluster sparsity (size). Based on these clusters, characteristics that distinguished the groups and outcomes within them were identified. To facilitate this analysis and enhance the interpretability of the results, those items within each cluster that stood out from the overall mean for the entire data set by 0.6 of one standard deviation were flagged.

Results. The overall means and standard deviations for each element used in the cluster analysis are shown in Table 5 for the complete set of 123 participants.

Table 5

Summary Descriptive Statistics for Complete Dataset used in Cluster Analyses

Item	Mean	*SD*	Item	Percent
Age (years)	22.33	3.71	PLT 1	29%
Red Peer	2.38	0.38	PLT 2	21%
White Peer	2.47	0.35	PLT 3	27%
Blue Peer	2.48	0.34	PLT 4	23%
Red Top 3	2.63	5.38	Grade (E-1)	58%
White Top 3	2.93	5.53	Grade (E-2)	15%
Blue Top 3	2.88	4.88	Grade (E-3)	16%
Red Bottom 3	2.46	5.39	Grade (E-4)	11%
White Bottom 3	2.10	4.15	Education 1	47%
Blue Bottom 3	2.20	4.24	Education 2	42%
Red Trust (%)	79.81	25.21	Education 3	11%
White Trust (%)	78.70	22.70		
Blue Trust (%)	78.47	23.79		
Red LoC	2.33	0.48		
White LoC	2.26	0.51		
Blue LoC	2.20	0.56		
Red SJT	4.43	0.39		
White SJT	3.95	0.55		
Blue SJT	4.21	0.42		

Note. Peer = Mean rating across all Army Values on a 3-point scale where higher is better; Top 3 = number of times rated top three in platoon; Bottom 3 = Number of times rated bottom 3 in platoon; Trust = Percent peers answering yes to the trust questions; LoC = Locus of Control, where lower scores correspond to more internal LoC; SJTs = Mean response on a 5-point scale aligning with the developed SJT scoring key, where higher is better; PLT = platoon (four total platoons were used in the analyses)*;* Education: 1 = GED/high school diploma, 2 = some college/Associate degree, 3 = Bachelor's degree/Graduate level. *SD* = standard deviation.

Overall, these analyses revealed several interesting findings that reinforce and extend insights addressed in the analyses presented above. Descriptions of the clusters are shown in Table 6 including the size of each cluster and mean peer ratings by training phase (bolded and underlined when beyond the 0.6 *SD* threshold), followed by a short summary of characteristics that exceed the 0.6 *SD* threshold along with the proportions for education, grade, and platoon.

Table 6

Summary of Each Cluster Based on Distinguishing Characteristics

Cluster	*N*	Red Peer Mean	White Peer Mean	Blue Peer Mean	Distinguishing Characteristics
0	8	**1.78 (↓)**	**1.92 (↓)**	**1.97 (↓)**	Red Bottom 3 (13.75↑), Red Trust (41.50%↓), White Bottom 3 (15.88↑), White Trust (43.38%↓), Blue Bottom 3 (15.88↑), Blue Trust (42.82%↓), PLT (0.62↑, 0.01↓, 0.12↓, 0.25↑), Grade (0.50↓, 0.12↓, 0.25↑, 0.13↑), Education (0.50↑, 0.38↓, 0.12↑)
1	6	**1.60 (↓)**	**1.99 (↓)**	**1.60 (↓)**	Red Bottom 3 (15.33↑), Red Trust (21.25%↓), White Trust (29.00%↓), Blue Trust (22.61%↓), Red LoC (3.09↑), White LoC (2.73↑), Blue LoC (2.79↑), PLT (0.00↓, 0.17↓, 0.50↑, 0.33↑), Grade (0.83↑, 0.17↑, 0.00↓, 0.00↓), Education (0.83↑, 0.17↓, 0.00↓)
2	18	2.47	2.58	2.57	White LoC (2.75↑), White SJT (3.52↓), PLT (0.28↓, 0.50↑, 0.17↓, 0.05↓), Grade (0.61, 0.11, 0.22, 0.06↓), Education (0.11↓, 0.83↑, 0.06↓)
3	13	2.48	2.54	2.58	Red LoC (2.61↑), Blue LoC (2.67↑), PLT (0.85↑, 0.00↓, 0.15↓, 0.00↓), Grade (0.85↑, 0.00↓, 0.08↓, 0.08↓), Education (0.77↑, 0.15↓, 0.08↓)
4	17	2.58	2.61	**2.70 (↑)**	Red LoC (1.94↓), White LoC (1.83↓), Blue LoC (1.77↓), Blue SJT (3.90↓), PLT (0.18↓, 0.59↑, 0.18↓, 0.05↓), Grade (0.29↓, 0.41↑, 0.24↑, 0.06↓), Education (0.24↓, 0.71↑, 0.06↓)
5	16	2.52	2.58	2.60	Age (19.4↓), White Trust (92.56%↑), PLT (0.06↓, 0.00↓, 0.25↓, 0.69↑), Grade (0.75↑, 0.06↓, 0.19↑, 0.00↓), Education (0.62↑, 0.38↓, 0.00↓)
6	14	**2.69 (↑)**	**2.79 (↑)**	**2.69 (↑)**	Age (26.07↑), Red Top 3 (5.86↑), White Trust (95.29%↑), Red LoC (1.97↓), White LoC (1.84↓), Blue LoC (1.85↓), Blue SJT (4.47↑), PLT (0.07↓, 0.10↓, 0.79↑, 0.14↓), Grade (0.29↓, 0.00↓, 0.21↑, 0.50↑), Education (0.14↓, 0.43↑, 0.43↑)
7	5	**2.84 (↑)**	**2.90 (↑)**	**2.85 (↑)**	Red Top 3 (19.0↑), Red Trust (100.00%↑), White Top 3 (25.60↑), White Trust (95.40%↑), Blue Top 3 (22.80↑), Blue Trust (97.32%↑), White LoC (1.93↓), PLT (0.20↓, 0.20↓, 0.20↓, 0.40↑), Grade (0.40↓, 0.20↑, 0.00↓, 0.40↑), Education (0.00↓, 0.60↑, 0.40↑)
8	14	**2.00 (↓)**	**2.09 (↓)**	2.34	Age (19.14↓), Red Trust (55.04%↓), White Trust (57.82%↓), Red SJT (3.82↓), White SJT (3.60↓), PLT (0.50↑, 0.14↓, 0.21↓, 0.15↓), Grade (0.64↑, 0.21↑, 0.14↑, 0.01↓), Education (0.79↑, 0.21↓, 0.00↓)
9	12	2.28	2.36	2.33	Blue Trust (54.34%↓), PLT (0.17↓, 0.25↑, 0.16↓, 0.42↑), Grade (0.67↑, 0.17↑, 0.08↓, 0.08↓), Education (0.83↑, 0.08↓, 0.08↓)

Note. Bolded and underlined mean peer ratings for each phase were greater than 0.6 *SD* from means for overall sample. Distinguishing characteristics includes anything additional that meets criteria of beyond 0.6 *SD* from means for overall sample with mean for cluster in parentheses. Arrows indicate if the cluster was (↑) higher or (↓) lower than the sample mean. Complete proportions are provided for PLT (in order from 1 to 4), grade (in order from E-1 to E-4), and education (in order from group 1 to 3). *N* = number of participants in cluster.

Consistent with the analyses of the top and bottom groups noted above, clusters of participants were identified that exhibited different patterns across variables, suggesting a set of different but common paths through BCT. For instance, with respect to peer ratings on the Army Values, clusters 6 and 7 were rated high throughout, whereas cluster 4 also rated highly by Blue Phase. In contrast, clusters 0 and 1 tended to be rated low throughout, whereas cluster 8 was rated low initially but was in the more typical range by Blue Phase. Overall, it is notable that from Red to Blue Phase, 8 of the 10 clusters improved at least a minimal amount over time. Likewise, only clusters 0 and 1 remained below 2.00 in mean peer ratings by Blue Phase, reflecting growth over time to acceptable levels in most of the clusters.

Looking more closely at the three higher rated clusters, cluster 7 was notable in that, as indicated, the trainees in this cluster were rated very highly by peers throughout BCT. Upon examining the distinguishing characteristics for this group (Table 6, right column), it is apparent that this group stood out in terms of education (100% were some college or beyond) and grade (60% were beyond E-1). Similarly, cluster 6 also stood out for high peer ratings throughout, and was also distinguished by education (86% had some college or beyond) and grade (71% were beyond E-1). These two clusters also tended to stand out for Top 3 votes, trust ratings, and more internal LoC. Cluster 6 also stood out for being older, having higher performance on Blue Phase SJTs, and they tended to disproportionally be in one of the four PLTs. Cluster 4 stood out for peer ratings by Blue Phase, and was also notably high on education (77% had some college or beyond) and grade (71% were beyond E-1), and for predominantly being in another one of the four PLTs. Interestingly, cluster 4 also stood out for more internal LoC. Overall, looking at these three top performing clusters, two of three stood out for being more internal in LoC in Red Phase, three of three in White Phase, and two of three in Blue Phase. Collectively, education and internal LoC were hallmarks of these higher rated clusters.

In contrast, looking at the lower rated clusters, clusters 0 and 1 tended to receive notably low peer ratings throughout BCT. Cluster 1 stood out for external LoC throughout, and low trust percentages. They also tended to be lower in education (100% had at most some college) and grade (100% were E-2 or lower). Cluster 0 also stood out for low trust ratings and bottom 3 votes. For this group, 88% were at most some college and 62% were E-2 or lower. Cluster 8 started out lower in peer ratings but was typical by Blue Phase, when trust ratings were also more typical. This group tended to have more difficulty with the SJTs in Red and White Phase as well. They also tended to be lower in age, education (100% had at most some college) and grade (85% were E-2 or lower). Notably, among these lower rated clusters, 8 and 0 tended to improve over time in peer ratings, whereas cluster 1 had an upside down u-shaped pattern for peer ratings, once again suggesting issues that may separate continued growth versus lack of growth following White Phase for some participants.

In total, although variable at the level of clusters, when combined with the overall results, these data collectively suggest interesting relationships between age, education, and LoC. With respect to LoC, each of the high performing clusters stood out for internal LoC during at least one BCT phase. In contrast, among the lower performing clusters, one cluster was notably external in LoC, whereas the others were more typical, and in fact, among the moderately performing clusters, all were either typical or with a more external LoC across phases. In terms of education, clear trends appeared in that all of the higher performing clusters had predominantly some college or beyond, whereas the lower performing clusters tended to have

almost no one with a Bachelor's degree or beyond. Finally, among the youngest three clusters (8, 5, and 4), only cluster 4 stood out for high peer ratings and that was not until Blue Phase. In contrast, amongst the oldest three clusters (3, 7, and 6), both clusters 6 and 7 received high peer ratings across phases, whereas cluster 3 (the youngest of the three) was more typical.

Part 3 Discussion

Overall, Part 3 findings revealed that peer evaluations were predicted, at least in part, by age, education, grade, and locus of control; older, educated trainees, of higher grades, and with more internal locus of control were evaluated more favorably by their peers. Furthermore, the results suggest that events in BCT bolster locus of control in trainees, as well as help to instill the Army Values in trainees. Specifically, results showed that trainees with a more internal, as opposed to external, locus of control (i.e., belief that consequences are a direct result of their own actions rather than beyond their control) were perceived as more trustworthy and rated more favorably by their peers on the Army Values. As training progressed, trainees tended to became increasingly more internal in their locus of control. Given this trend, it may be argued that the curriculum in basic training helps to prepare trainees for their careers in the Army by equipping them with valuable knowledge, skills, and abilities – not only technical and tactical skills but also enabling trainees to feel in control of their learning experiences and acclimate more successfully into the Army Profession and Ethic. Future research could further explore the relationship between locus of control and adherence to the Army Values, and whether training that is targeted at developing a more internal locus of control enhances character and translates to more ethical decision-making.

Promising preliminary evidence of incremental validity was provided for the SJTs as well as for lagged effects where ethical decision-making ability in earlier phases may be related to outcomes in later phases. In other words, the empirical evidence suggests that the SJTs have the potential to differentiate among levels of moral/ethical reasoning and degree of inculcation with the Army Values as indicated by peer evaluation results. It also appears that the Army Values are differentially salient in a rater's decision-making regarding peer evaluation criteria. The results also show that how trainees are percieved in the beginning of training (i.e., first impressions) has short and long term influence on how they are perceived by their peers.

It is important to note, however, that the results based on SJT performance across phases should be interpreted with caution given that the video and question content differed across phases. In fact, ethical decision-making ability in operational vs. non-operational contexts may be separate constructs and have a different nomological network given that performance on each is differentially related to other variables. Thus, the primary benefit of the SJTs lies in their formative nature when used for DS-led values-based discussion.

The cluster analysis results suggest that some advantage exists for combinations of trainee age, internal locus of control, education, and grade. In general, a more targeted investigation is warranted to identify how these factors interact to contribute to performance, and how they might be useful for tailoring training in the future. For instance, informal observations from Drill Sergeants noted that older trainees often benefit from their education because they can "rise above" the challenges of BCT and act as role models for their peers. This in turn points to the potential benefit of strategic pairing of battle buddies (i.e., pairing individuals likely to do

well and those likely to struggle based on variables such as those explored here). More generally, although preliminary, these results are promising and point toward further investigations and/or interventions facilitated by data-driven insights.

General Discussion

The present research describes the development and testing of a solution for assessing and training character at a critical point in a Soldier's career – basic training. Specifically, character assessment tools and methods were developed in Part 1, tested as part of a quasi-experimental pilot study in Part 2, and evaluated longitudinally in Part 3. The overall goals of this research were to (a) create valid and reliable tools that measure character, (b) enable Army Values discussion, (c) standardize and expedite the peer evaluation process within basic training, (d) provide an efficient method to digitally capture assessment data, (e) develop a useful and usable tracking tool for easy management of character assessment data for programmatic insights, and (f) examine the different developmental paths individuals take as they progress through BCT.

The assessment tools and methods were iteratively tested and refined at all stages of the research to maximize their utility and sustainability in the target environment. To facilitate implementation, both training and accompanying materials were developed. A final training was conducted with end users on the assessment tools and process. User guides were also developed that outline how to use and manage the character tracking tool, video-based SJT questions, and clicker software and hardware. Literature-backed guidance was also provided to Drill Sergeants on writing effective ethical decision-making questions (Appendix C) and using clicker technology to facilitate discussion (Appendix E).

Developing and implementing tools involving ethical and moral dilemmas can be challenging. To mitigate this challenge, the tools presented here are flexible enough to support future scenario development across media types. Although these ethical decision-making items began with critical incidents provided within CAPE videos, this does not preclude alternate sources for critical incidents. Identifying other training materials to use outside of CAPE videos is also viable—including those from instructors. The procedures described herein and as part of the user guides may also be used to develop scenarios and learning checks with other materials including personal experiences, critical incidents from Soldier panels, and other Army training materials. But, it is important to identify incidents viable for scenario development that offer similar level of detail and quality to CAPE videos. The user guides and flexibility of the tools and process also enable end users to continue to refine and adapt the approach in a manner that continues to meet training objectives in the future.

Overall, moral education programs demonstrate significant potential for growth in moral reasoning through discussion (Nucci, 2001). Notably, a clear advantage of video-based ethical decision-making questions in the form of SJTs in any training context is that trainee experience with the simulation can foster personal development. Additionally, the mere act of taking character assessment as part of the training curriculum may have some developmental benefit for the trainee. However, the full potential of the character tools lies in individualized feedback opportunities for the trainees. Although in the current research, SJTs, peer evaluations, tracking tool, and clickers were used to assess character at the group level, the assessment data could be

used to provide customized, targeted formative feedback to the individual learner and maximize not only collective, but also individual learning processes and outcomes. The need for such individualized formative feedback is evident in the differences in Soldier trajectories through BCT. These tools can also benefit instructor development as well, such as by tailoring effective instructional approaches to lead meaningful values-based discussions with the class. In doing so, instructors can help foster confidence, cohesion, and positive learning experiences for Soldiers especially as they embark on their careers in the Army. Furthermore, the assessment data can help inform command team leader development policies and programs. Though the explicit focus of the current research was on the Army basic combat training environment, similar assessment tools could be developed and employed to support learning in any instructional context.

References

Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston, MA: Kent Publishing Co.

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

Center for the Army Profession and Ethic (2017). *The Army's framework for character development* [White paper]. Retrieved March 15, 2018, from http://data.cape.army.mil/web/repository/white-papers/armys-framework-for-character-development-white-paper.pdf

Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions* (2nd ed.). New York, NY: Routledge.

Huang, Z. (1997). *Clustering large data sets with mixed numeric and categorical values.* In Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference, Singapore.

Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), *Cognitive development and epistemology* (pp. 151–284). New York: Academic Press.

Manley, G. G., Benavidez, J., & Dunn, K. (2007). Development of a personality biodata measure to predict ethical decision making. *Journal of Managerial Psychology, 22*, 664-682.

Motowidlo, S. J., Hanson, M. A., & Crafts, J. L. (1997). Low-fidelity simulations. In D. L. Whetzel, G. R. Wheaton (Eds.), *Applied Measurement Methods in Industrial Psychology.* Palo Alto, CA: Consulting Psychologists Press.

Muraca-Grabowski, S. T. (2015). *Assessing Army Values in new Soldiers: Development of a behaviorally anchored peer evaluation form.* Ft. Eustis, VA: U.S. Army Center for Initial Military Training, Research and Analysis Directorate. (DTIC No. ADA616440).

Ng, T. W. H., Sorensen, K. L., & Eby, L. T. (2006). Locus of control at work: A meta-analysis. *Journal of Organizational Behavior, 27*, 1057-1087.

Nucci, L. (2001). *Education in the moral domain.* Cambridge, England: University Press.

Rest, J. R., Narvaez, D., Bebeau, M. J., & Thoma, S. J. (1999). *Post conventional moral thinking: A neo-Kohlbergian approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*, 609.

Smith, P. C., & Kendall, L. M. (1963). Retranslation and expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*, 149-155.

Trevino, L. K. (1986). Ethical decision making in organizations: A person-situation interactionist model. *Academy of Management Review*, *11*, 601-617.

U.S. Department of the Army (2013). *Army Leader Development Strategy*. Retrieved March 15, 2018, from http://data.cape.army.mil/web/character-development-project/repository/alds-2013.pdf

U.S. Department of the Army (2015a). *The Army profession* (ADRP-1). Washington, DC: Author.

U.S. Department of the Army (2015b). *Leader development* (Field Manual No. 6-22). Washington, DC: Author.

Appendix A

Army Values Rubric for Peer Evaluations

Army Value	3-Excellent/Above Standard	2-Average/At Standard	1- Poor/Below Standard
Loyalty	Has self squared away, helps battle buddy, supports leader, helps others in and beyond unit regardless of situation and even if it makes self look bad (e.g., when not socially acceptable/ popular with peers); Not afraid to tell someone if s/he is doing something wrong; Takes responsibility for fellow Soldiers	Has self squared away and supports leader; Helps battle buddy but may be inconsistent (e.g., in low risk situations, when socially acceptable to peers); Sometimes shows initiative and willingness to help	Does not help out peers and/or support leader (e.g., is disobedient); Has a "me" mindset; Lacks concern for battle buddy when s/he is struggling; Not a team player (e.g., slacks off)
Duty	Consistently shows great discipline, accountability for self and others, and efficiently accomplishes assigned tasks; Prepares for and organizes ahead by having a 24-hour mindset; Helps others out even when not in an assigned leadership role (e.g., polices peers; makes sure everyone is ready and on time, puts information out to peers); Is a big team player and always in the mud with peers	Generally shows discipline (e.g., has proper gear, no uniform violations, no talking); Has personal accountability; Accomplishes assigned task but does not go above and beyond; Does just enough to look like part of the team; Delegates but does not assist	Lacks discipline about personal appearance and wellbeing (e.g., fails to shave, uniform violations, careless); Moves without a sense of urgency; Unprepared for upcoming task (e.g., gear not ready); Lacks personal accountability; Chooses easy path even if not right one except when Drill Sergeant is watching
Respect	Consistently demonstrates respect toward superiors and peers through words and actions even when Drill Sergeant is not looking; Treats everyone the same and as one would want to be treated (e.g., Golden Rule); Boosts morale and brings peers together as a team; Identifies and proactively tries to resolve and prevent problems (e.g., formation of cliques)	Generally gives respect but may be inconsistent (e.g., when Drill Sergeant is present, keeps score, not aware of effect on bystanders); Follows customs and courtesies appropriately; Is willing to accept diversity but is a follower who could drift either way; Avoids confrontation to the point it is a problem (e.g., won't stand up for the little guy)	Demands but does not give respect (e.g., fails to use proper titles; acts entitled); Struggles to accept diversity; May fail to listen or be argumentative; Brings peers down and makes trouble; Creates a toxic environment
Selfless Service	Puts everyone else before self; Completes task/course quickly and goes back to help others along, coach others, and boost morale; Attributes success to "we" not "me"; Does not look for accolades or glory; Volunteers regardless of how bad the task	Completes tasks/course and willing to make sacrifices but may complain; Helps others in front of Drill Sergeant or when convenient; Goes above and beyond but wants to be recognized; Looks for incentives; Volunteers when beneficial to self or when inconvenience is minimal	Has an "all about me" mentality (e.g., does not work well with or help others); Purposefully fails to complete task/course (e.g., gives up); Avoids responsibility (e.g., asks why every single time, always has excuses, fails to volunteer)
Honor	Has the habit of abiding by the Army Values in all aspects of daily living; Demonstrates the ability and resolve to successfully perform assigned tasks despite adversity	Has the habit of abiding by the Army Values in all aspects of daily living; Demonstrates the ability and resolve to successfully perform assigned tasks under standard conditions	Fails to always abide by all the Army Values
Integrity	Pauses and thinks through options and makes the right ethical decisions even under extreme pressure and regardless of context (e.g., conflicts between own beliefs and mission/Army values); Does not lie, steal, and cheat; Is honest even in the face of consequences; Makes ethical decisions based on higher order principles or values (e.g., makes a reasoned argument)	Generally makes the correct ethical decision but may be affected by conditions (e.g., in the presence of Drill Sergeant or may need prompting); May choose ethical battles based on anticipated social consequences; When in a bad situation, pauses, steps back, and thinks about it even if ultimately s/he ends up making the wrong decision; Does the right thing so s/he does not get in trouble	Lies, steals, cheats (e.g., takes food out of chow hall, brings contraband in bay); Lies in order to stay out of trouble; Does not follow rules; Takes others down when s/he gets in trouble; Loses it when under stress; Follows unlawful order without thinking

Army Value	3-Excellent/Above Standard	2-Average/At Standard	1- Poor/Below Standard
Personal Courage	Is first to volunteer for or lead difficult tasks; Acknowledges fear even in front of others; Overcomes adversity even in extreme conditions; Chooses hard right over easy wrong; Stands up for the little guy even when in the minority; Makes an effort to correct peers in a professional manner; Escalates critical issues to Drill Sergeant for benefit of peer/unit regardless of consequence; Embraces diversity	Will attempt a difficult task but may not be the first to do it or may hesitate or do it slowly (e.g., go up eagle tower); Does not acknowledge s/he is afraid; Overcomes adversity under standard conditions; Only stands up for the little guy or cause when in the majority	Is unwilling to face his/her fears; Gives up in the face of adversity; Actively tries to avoid volunteering for or leading difficult tasks; Does not stand up for what s/he believes in or what s/he thinks is right

Appendix B

Overview of the Character Tracking Tool Tabs

Tool Tab	Description
Individual Dashboard	The Individual Dashboard is an interactive tab containing dynamic charts that can be used to view compiled individual-level results for a selected assessment period and trainee across BCT phases, including results from peer evaluation and video-based SJTs.
Squad Dashboard	The Squad Dashboard is an interactive tab containing dynamic charts that can be used to view compiled squad-level results for a selected assessment period and squad across BCT phases, including results from peer evaluation and video-based SJTs.
Platoon Dashboard	The Platoon Dashboard is an interactive tab containing dynamic charts that can be used to view compiled platoon-level results for a selected assessment period and platoon across BCT phases, including results from peer evaluation and video-based SJTs.
Cycle Dashboard	The Cycle Dashboard is an interactive tab containing a series of dynamic charts. The charts allow users to view higher-level trends and comparisons across training cycles for both peer evaluations and video-based SJTs.
Video Dashboard	The Video Dashboard is an interactive, dynamic tab that allows users to more closely examine the video-based SJTs. A score breakdown by squad and cycle is displayed for a selected question and platoon. An item response breakdown by cycle is provided for the selected question and platoon.
Roster IDs	The RosterID tab tracks the squad breakdown and movement of trainees across phases for each company, including any notes about the trainee during a given phase. The results import function and formulas within the tool rely on the official squad breakdown to perform properly.
Rank Summary	The Rank Summary tab provides a detailed summary of Top 3 / Bottom 3 in Platoon peer evaluation results after each BCT phase. This tab shows the squad and platoon rankings, which are based on the number of Top votes received during peer evaluations for each phase. This tab can be used to easily identify the trainees falling in the top, middle, or bottom of the squad/platoon for each phase.
Top 3 / Bottom 3 Summary	This tab provides a detailed summary of Top 3 / Bottom 3 in Platoon results for each squad over time (i.e., across Red, White, and Blue phases). This tab displays the total number of Top 3/Bottom 3 votes each trainee received across all three BCT phases.
Comments Summary	The Comments Summary tab displays aggregated comments for each trainee by phase, including comments on the Army Values ratings, Top 3 comments, and Bottom 3 comments. This tab also displays what platoon/squad the trainee was associated with for each phase, and whether each trainee was voted in the Top 3 or Bottom 3 for each phase.
Video Questions	The Video Questions & Answer Key tab displays the video-based SJTs associated with all CAPE videos used in this research. The tab lists the video/phase info, question text, response options, and answer key for each item. The answer key contains the points that are allotted for choosing each response option.
Other Tabs	The tool contains a 40+ hidden protected compiler tabs designed to import, compile, and store the raw results, and populate the summary tabs, interactive dashboards, and charts.

Appendix C

Guidelines for Developing Scenario-Based Questions on Ethical Decision-Making

OVERVIEW

Complex ethical scenarios coupled with clicker technology can help Drill Sergeants to generate discussion with Soldiers and track their progress. This allows Drill Sergeants to focus on facilitating discussion while leaving assessment to the technology.

In this tool, each scenario is a short, low-fidelity simulation designed to elicit similar ethical judgment processes that would occur in operational settings. A scenario comprises a brief vignette and narrative from an operational setting or basic training situation, including any relevant details related to ethical issues, challenges, and concerns. After Soldiers review a scenario, they are presented with a number of exercises representing potential ethical dilemmas related to that scenario. With each exercise, the Soldier reviews a list of four possible responses to the ethical dilemma and then selects which behavior they believe they should engage in. The proportion of responses are visible to the Drill Sergeant and can be used to focus the discussion.



Scenarios, Exercises, and Responses.

There are five components that make up a complete scenario that Soldiers can engage with.

Vignette: A vignette refers to brief ethical dilemma that a Soldier may encounter. Vignettes are the raw material used to build training scenarios, exercises, and responses. Vignettes may come from a variety of sources: a short video, a written story, or even a critical incident from a Drill Sergeant's own experiences.

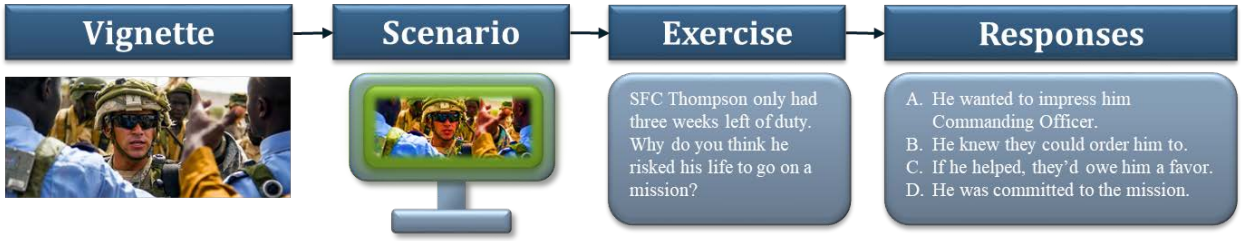
Scenario: A scenario is a brief vignette detailing an ethical dilemma that is turned into a training scenario. Although the scenario may focus on one primary ethical issue, a good scenario will likely be complex enough to comprise multiple ethical issues.

Exercise: An exercise is a brief statement or question. It will focus on a single ethical consideration or issue and four responses of how one may perceive and act on this issue.

Response: A single exercise has multiple responses. Soldiers select these responses selected behavior or value that a Soldier endorses in response to an exercise's prompt.

Learning Checks: Learning checks are simply questions to assess that a Soldier understands the details of the scenario. This ensures that Soldiers understand the details (people involved, plot points) enough to identify and discuss the ethical issues.

Drill Sergeants can help to build this tool further by finding new vignettes to turn into training scenarios. Turning a vignette into a scenario requires writing new exercises, responses, and learning checks. This task, however, can be quite challenging. Effective and plausible scenarios, exercises, and responses should facilitate meaningful group discussion and meet requirements for adequate assessment. This user guide will help Drill Sergeants to select viable vignettes and write quality exercises in order to use these tools more effectively and expand these materials for further use.



MORAL REASONING AND ETHICAL DEVELOPMENT

Moral development corresponds to stages, ascending from lower levels—emphasizing immediate concerns to one's self—and up through higher levels—emphasizing internalized ethical principles that apply to oneself and others (e.g., life, liberty, justice). Ethics are complex. Ethical principles may interact, conflict, or take precedence over one another given a particular situation. Principles comprise navigating between both formal Army Rules and Regulations as well as unexpected situations—situations that may entail complications, complexities, and contradictions that many civilians will never experience.

Remember that moral development is *not* moral behavior. Rather, moral development refers to the capacity to identify, consider, and evaluate the relevance of different ethical issues within a particular situation. As can be seen in the table below, greater moral development reflects an understanding of greater moral complexity in a situation and how we perceive our own actions within that situation. These moral development stages are the building blocks of the exercises. Each stage reflects a different response to an ethical dilemma that Soldiers might encounter.

MORAL DEVELOPMENT STAGES				
Increasing Levels of Cognitive Moral Development	**Stage**	**Orientation**	**Definition**	**Example**
	1	Punishment and obedience	Reasoning determined by individual receiving rewards and punishments.	"I should not steal because I will get in trouble."
	2	Instrumental relativist	Reasoning determined by exchanges between others regarding rewards and punishments.	"I should not steal because they may steal back from me – but if they steal from me first, then I would steal from them."
	3	Interpersonal concordance	Reasoning based on obedience to authority.	"I should not steal from someone, even if they steal from me, because others will think less of me."
	4	Law and order	Reasoning based on importance of doing "one's duty".	"I should not steal from someone, because it is their property."
	5	Social-contract	Reasoning based on good intentions, and obedience to authority, laws, and rules.	"I should not steal from someone, because we all rely on respecting the property of others."
	6	Universal ethical-principle	Reasoning through individual principles of conscience; including likely views of everyone affected by a moral decision.	"I should not steal, but sometimes, it is necessary to steal in order to help someone in a difficult situation."

Moral development refers to the fundamental conflict, between formal structures of society as well as the practice of virtue in everyday, face-to-face relationships. Pre-conventional reasoning judges the morality of an action based on its direct consequences—how does my behavior hurt or help me? Conventional reasoning focuses on society's views and expectations—how is my behavior perceived by others and directly affect them? Post-conventional reasoning on individual principles of conscience—how do my values and behavior impact, help or harm others, directly or indirectly?

SOURCING VIGNETTES AND DEVELOPING ETHICAL DILEMMA SCENARIOS

Discussion of ethical issues remains the best means of developing ethical reasoning in individuals. Discussion is not philosophical theorizing. It should not occur in the abstract, but be grounded in the complex ethical reality that individuals are likely to confront. Vignettes that provide realistic situations, contain sufficient detail for discussion, and are clear to understand, offer the best means of facilitating ethical discussion. Furthermore, they are most likely to be free of assessment contamination. Vignettes that are unfamiliar, overly complex, poorly written, or contain information that Soldiers are unlikely to know or understand introduces bias and harms assessment.

Identifying viable vignettes to produce quality scenarios form the foundation for generating productive discussion and plausible response options. The initial build of this tool exploited existing Center for the Army Profession and Ethic (CAPE) videos, available on their website

(http://cape.army.mil/). Further versions of this tool can expand to include additional CAPE videos. These videos are still available and relevant for these purposes.

A variety of sources for scenario are viable; and may contribute unique advantages as well as disadvantages. Options include popular war movies, news stories, or even personal experiences. Indeed, scenarios based on critical incidents obtained from subject matter experts are generally used to exhibit adequate realism and complexity.

Evaluating Effective Scenarios and Learning Resources to Use		
Possible Sources	**Advantages**	**Disadvantages**
CAPE Videos	• Realistic situations that encounter in Basic Combat Training and operational settings • Already developed by US Army CAPE • Contain associated discussion materials	• Limited in number • Test security
Movies	• Exciting • Motivating • Resonates with Soldiers	• Not all information may be available • Availability • May be too obvious
News Stories	• Important • Impactful • Realistic	• May not be relevant to Soldiers in Basic Combat Training. • May not contain sufficient detail
Personal Experiences	• Meaningful and engaging • Provide nuance and unique details	• May be too complex for a Basic Combat Training setting

It is important to remember that the purpose is identify vignettes that facilitate developing discussion-oriented exercises and responses. A single event may encapsulate some, all, or an unclear mix of Army Values. Scenarios drawn from operational settings and personal experiences offer rich details, but may be too complicated or advanced. In writing exercises (see more below), you may notice a variety of unique responses to these situations. For example, many CAPE videos included additional accompanying materials, including written descriptions, facilitator guides, and discussion questions. News stories on the same incident may contain unique information and more detail. Reviewing these materials can offer insights and ideas about exercises and responses.

In evaluating effective vignettes to use as scenarios, there are a number of key considerations to consider:

- Do these vignettes describe scenarios that Soldiers are likely to encounter in Basic Combat Training?
- Do these vignettes describe operational scenarios that Soldiers are likely to encounter in their immediate future?
- Do these vignettes provide adequate detail to write exercises and responses?
- Are these vignettes too difficult or complex for a Soldier in Basic Combat Training to respond to?
- Are these vignettes too easy or simple for sufficient discussion?
- Are these dilemmas representative of the Soldier's rank?

GENERATING SCENARIO EXERCISES

WRITING EXERCISES

After selecting a viable vignette to use as a scenario, exercises must be developed. Remember, scenarios and exercises do *not* need to perfectly reflect every detail in order to be effective. A vignette is only a grounded scenario to facilitate the discussion of ethical issues. An exercise focuses in on a particular ethical issue and possible responses. These ethical issues may or may not have actually occurred in the vignette itself—a detail can be highlighted or even changed in an exercise to make a new issue relevant, or more salient. Particular ethical issues may be changed to highlight unique aspects of the scenario. Exercises can present hypotheticals that may shift the focus of the scenario, maybe from fellow Soldiers to a non-combatant, or from oneself to a leader. Modifying the details or premises of a scenario may change, vary, or adapt the relevant ethical issues to provide thought-provoking exercises.

IDENTIFYING ETHICAL ISSUES

Ethical dilemmas will form the basis of each exercise, as well as responses. First, review the scenario, and identify the critical ethical issues faced by the Soldier in the video. These may include any number of issues throughout the moral development stages outlined above. Soldiers may be concerned about how the scenario could help or harm themselves, how leadership may interpret the challenge, or how an issue relates to Army Core Values. Remember, ethical issues are not always a single, clear choice, but may contain a variety of possible choices based on different ethical issues, values, and outcomes.

In fact, a challenging ethical dilemma often contains more than one ethical issue in the narrative that may be directly at odds with another ethical issue. For example, in a particular situation, the Army Cores Values of loyalty and duty may come into conflict. Or, Army Core Values may come into conflict with family or cultural values. Highlighting how these ethical issues compete, produce different outcomes, or be interpreted differently, make for complex ethical challenges and therefore a healthy discussion. Finally, consult any accompanying materials for additional context or thoughts. Variations on the ethical issues can be used to produce multiple exercises and responses.

Questions to consider when identifying ethical issues in a scenario:

- What is the core ethical dilemma? Is there more than one?
- Do these scenarios illustrate one (or more) Army Core Values? Which?
- How might the concerns of another Soldier affect this situation?
- What might other Soldiers do in this situation?
- How might leadership affect this situation?
- How might the inclusion of a non-combatant affect this situation?
- What external or situational concerns might change this situation?
- Are these exercises too easy or transparent to facilitate a challenging discussion?
- Are there any issues that could be changed or varied that might affect Soldiers' likely responses (e.g., consider response options such as, "What if…", "What about…", or "Imagine if…")?
- Are there additional resources related to scenarios (e.g., learning resources, discussion questions)?

GENERATING SCENARIO RESPONSES

In the abstract, the "correct" answers on a moral reasoning test are easy to describe, but challenging to develop. Responses may actually be degrees of ethical. In writing responses, Drill Sergeants should consider the various levels of moral development stages relevant to the particular situation. This includes understanding how individuals at various levels of moral development may interpret the issues relevant to a moral dilemma differently.

When generating possible responses, options flowing from and representing these stages are likely to illustrate the breadth of Soldiers' possible responses. Just because a Soldier operates from a lower moral reasoning stage does not mean that it is poorly justified. Even an unethical choice may be well considered—it may just be based on different underlying values. Nobody believes that they are unethical. Accordingly, responses that are justified using a misunderstood or misconstrued ethical principle offer better discussion.

USING MORAL DEVELOPMENT STAGES TO WRITE EXERCISES

Increasing Levels of Cognitive Moral Development	Stage	Orientation	Item-Writing Strategy	Example
	1	Punishment and obedience	"How can this issue hurt the Soldier who is telling the story?"	Cheating on a test would get the Soldier in trouble.
	2	Instrumental relativist	"How can this issue help the Soldier who is telling the story?"	The Soldier would be owed a favor for doing something nice for someone else.
	3	Interpersonal concordance	"Who might tell this Soldier not to do this?"	Soldier would help battle buddy to gain Drill Sergeant approval.
	4	Law and order	"What might others think about this situation, and how would this affect the Soldier?"	Soldier chooses to report an issue to Drill Sergeant because it is the Soldier's duty.
	5	Social-contract	"What other existing systems can be leveraged to do the right thing?"	Soldier's actions would impact unit cohesion.
	6	Universal ethical-principle	"What values are at play here and how can I find a way to behave in a manner consistent with them?"	Soldier's actions could impact others' perception of the U.S. Army.

Although moral development theory proposes six stages, more response options are not necessarily better. In contrast to even most publically available tests, three- or four-option exercises are actually better than five or more option responses. Too many responses may confuse students, slow down discussion, or even offer hints as to what the best response might be. Additionally, it can be difficult to generate many alternatives, and consequentially, the fifth or sixth responses are quite poor. Producing too many responses can lead to writing transparent responses that are too easy or not taken seriously. Remember, plausible distractors (less correct responses) are behaviors that a Soldier may actually engage in. Nevertheless, in some cases, it may be possible to generate

many plausible responses. In this case, an SME panel should help review and select the best four responses (described below).

In developing responses, consider:

- How might Soldiers realistically respond to this situation—regardless of particular ethical issues?
- How might Soldiers respond to the letter of a regulation but not its spirit?
- How might Soldiers consider their career progression, punishment, or rewards to these ethical issues?
- What other Soldiers in the scenario may see things from a unique perspective?
- What other individuals (e.g., noncombatants, local population) in the scenario may have unique perspectives?
- What unintended consequences might happen from these responses?

SCORING SCENARIO RESPONSES

In a traditional multiple-choice test, there is a single keyed correct answer and non-keyed alternatives—or distractors. Given this distribution, scoring is clear and binary: correct and incorrect. However, ethical dilemmas are often too complex for a single right and wrong choice.

An exercise may represent multiple Army Core Values—values that may compete with one another—and there may be a variety of unique behavioral responses. These responses are also complex and based on a variety of considerations. Responses may require a knowledge of rules and regulations, consider the motivations of various stakeholders, and pit individual versus societal values against one another. Soldiers must decide between and amongst these complex issues and various responses. In contrast from traditional multiple-choice tests, these sorts of decisions are not clear and binary, but *degrees of effectiveness*.

To represent these choices, responses may be scored as degrees of effective (or correct) on a sliding scale. The present scoring rubric are in increments of 25, from zero to 100. Some responses may be particularly poor (i.e., 0). Some may be somewhat justifiable, but not particularly ethical (i.e., 50). Multiple responses may be scored the same level. However, for scoring purposes, only one response is keyed as maximally right (i.e., 100). It is advisable to have someone else than the Drill Sergeant who initially wrote the items to assist, review, and discuss scoring keys (see Review Panel below).

Writing plausible distractors is often challenging. Generating four responses that capture specific levels of effectiveness and are equally distributed along an "ethics spectrum" is particularly difficult. In many cases, there are various degrees of ethically justifiable responses and some of which may be equally defensible. Given the challenges of writing plausible distractors, flexibility in scoring allows for a means of correcting for these differences while maintaining quantitative assessment. That is, it is advisable to generate the best, most reasonable responses to scenarios, and then to determine their effectiveness, rather than trying to force a particular distribution of responses.

COMMON CHALLENGES

COMPLEXITY AND READABILITY

Ethical dilemmas are already challenging. Unnecessarily high-reading level or simply too much text, may confuse Soldiers without having anything to do with the scenario itself. This confounds

measurement of moral reasoning with reading ability. Scenarios can already be quite short. Soldiers may be tired, distracted, or simply miss an important detail. Write exercises and response options at an accessible reading level. This includes limiting the number of words in a sentence, number of sentences overall, and eliminating complex clauses where possible. Never use a big or uncommon word where a simple one will do.

TRANSPARENCY AND REALISM OF RESPONSES

When writing responses to exercises, it is important to balance the plausibility of the distractor responses with social desirability. *Social desirability bias* refers to the fact that respondents will often answer questions in a manner that will be viewed favorably by others. Some responses are simply more socially desirable than others—they are too obviously the correct response or an implausible distractor. This will artificially increase the likelihood that all Soldiers will select or de-select these responses. When all Soldiers select the same response because it is obvious, it does not provide much room for discussion or much insight into differences between Soldiers' opinions.

Responses that are too long or overtly mention an Army Core Value are likely to be high in social desirability and should be re-written. However, in the domain of character development, incorrect responses may also be too transparent. Responses may be very *low* in social desirability—and therefore often easy for respondents to eliminate as plausible. Violations of rules and regulations are obviously inappropriate responses. Particularly extreme violations or poorly rationalized behaviors may call attention to the response as implausible. Even when Soldiers knowingly break a rule, they are likely to rationalize and defend this behavior using a value or moral argument.

SHOULD VERSUS WOULD

When wording exercises—that is, the ethical questions posed to Soldiers—it is important to think about if questions are worded as "should" or "would" statements. When questions are worded as "should"—for example, "Given the chance to cheat, what should you do?"—respondents are primed to think in terms of right and wrong answers. In contrast, when questions are worded as "would", respondents are primed to think in terms of their personal behaviors and likelihood of engaging in that behavior. When writing *should* rather than *would* stems, it is also important to consider how the different response options may change the feel of the exercise as a whole.

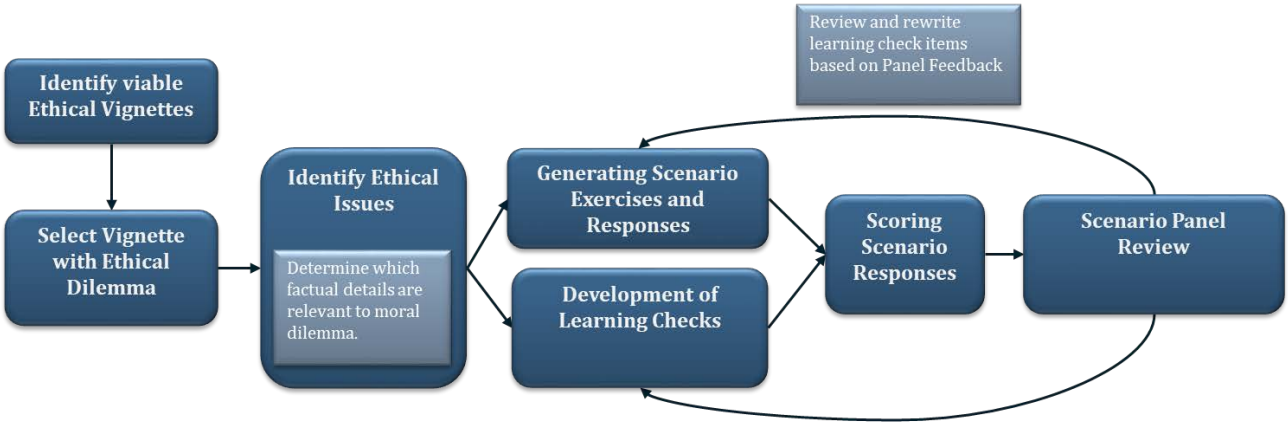
Primarily, these exercises should be used to assess whether or not Soldiers have learned and understand Army values. Soldiers are trained to know what they *should* do, but not try to guess what they *would* do in the future. Nevertheless, even though these exercises are written as "should" statements, discussing what a Soldier feels that they are likely to actually do can be useful. However, this will likely come about during discussion and does not need to be written into the exercises.

DRILL SERGEANT REVIEW PANEL

To ensure that scenarios are understandable, appropriate, and representative of the desired domain (i.e., either operational or basic training), each scenario should be reviewed by subject matter expert review panels. A review panel, similar to a focus group, of other Drill Sergeants acting can be a valuable resource to help vet new exercises and responses. It is best for the panel of SMEs to be a different group of people than those who originally generated the initial scenario exercises or responses.

Ideally, successive panels reviewing scenarios, exercises, and responses can provide insights at every stage of the process. SME panels can quickly eliminate poor or irrelevant exercises and revise the remainder. For example, separating scoring from exercise- and response-writing would ensure that each step is interpretable by a Soldier. This avoids confounding SMEs dedicated to exercise development with SMEs focusing on exercise evaluation.

This is particularly true if Drill Sergeants are pulling scenarios from their own experiences. Some details may be remembered or assumed, yet not included in the actual exercise. These details may contain important information for making ethical decisions. Exercises and responses may be idiosyncratic, or only reflect the choices that one single Drill Sergeants considered or observed. Additionally, a Drill Sergeant Review Panel offers unique experiences that may be able to expand on initially drafted items. They may generate different responses or offer unique perspectives. This can improve the quality of the responses. All exercises, responses, and scoring can be improved by a second review.



Exercise Writing Process Overview

LEARNING CHECK QUESTIONS

Before responding to the scenario or engaging in discussion, Soldiers must understand a sufficient amount of information from the scenario. Ensuring that Soldiers understand the necessary details of the story first, means that they will be able to focus on the ethical dilemmas, rather than simply trying to remember details. This may include the names or MOS of key characters, the context, and basic plot points.

Learning checks can be used to verify that the individual paid attention to the scenario and understood the essential information provided. This is necessary to confirm that Soldiers know the details relevant to ethical dilemmas, and are able to understand the scenario well enough to respond to exercises and participate in the discussion. Here, learning checks are *not* used to assess a particular level of learning, but to double-check that Soldiers know what they need to know in order participate. Ultimately, the point is to ensure that Soldiers remember and understand details, issues, and key take-aways from the situation so that they are able to converse about it.

EXERCISE WRITING FORMATTING

Like many good habits, good exercise writing rules are not necessarily commonly practiced. Indeed, even exercise banks from textbooks are often poor—and therefore usually require some revisions. Although basic exercise-writing rules may seem simple or obvious, minor errors offer clues as to the correct responses or give the appearance that assessments or training is careless or unfair. Or, minor errors may distract learners and lead to selecting incorrect responses for arbitrary reasons. Clearer exercises and responses will provide more accurate feedback about a Soldier's progress, and therefore allow Drill Sergeants to tailor instruction to meet their needs efficiently.

EXAMPLE EXERCISE FORMATS			
Exercise:	1) Jack uses a red-based polish to wax his car because the car's color is: or 1) Jack uses a red-based polish to wax his car because the car's color is…	2) Which of the following types of gasoline should Jack be using in her car?	3) Jack pays more auto insurance than Jill because she drives a _____ car.
Responses:	A. black. B. blue. C. red.	A. Diesel B. Regular unleaded C. Super unleaded	A. black B. blue C. red
Explanation:	The responses are not capitalized and end with a period to finish the sentence.	The responses are capitalized but do not have a period. These responses are independent but do not constitute a complete sentence.	The responses are *not* capitalized and do *not* have a period. These responses complete the sentence, but do not finish that sentence.

Checklist for Scenario Development	
Steps	**Items**
1. Sourcing Vignette	☐ Realistic? ☐ Sufficiently detailed? ☐ Complex enough for discussion? ☐ Too complex? ☐ Too operational? ☐ Meaningful? ☐ Engaging? ☐ Appropriate Rank?
2. Identifying Ethical Issues	☐ Identified core ethical dilemma? ☐ Identified all relevant ethical issues? ☐ Identified relevant Army Core Values? ☐ Identified perspectives of relevant individuals and groups? ☐ Reviewed all available accompanying materials? ☐ Identified all rules and regulations that apply?
3. Generating Scenario Responses	☐ Are there any adequate, but not optimal responses? ☐ Do responses reflect moral development stages? ☐ Considered possible hypothetical changes to the scenario? ☐ Considered concerns of other Soldiers? ☐ Considered Non-combatants? Local population? ☐ Considered unintended consequences? ☐ Considered long-term consequences? ☐ Are responses justified?
4. Scoring Responses	☐ Do responses and their scores actually vary along the 0-100 continuum? ☐ Can you articulate why each response is scored? ☐ Are any responses equally justifiable? Are they scored as such? ☐ Does the score require justifying information form an external source?
5. Reviewing Exercises and Responses	☐ Is reading level appropriate? ☐ Is anything unclear? ☐ Are items written as “should”, or moral imperatives?
6. Writing Learning check Items	☐ Are there details that are vital but easy to miss? ☐ Would Soldiers know who is who in the scenario? ☐ Would Soldiers know the relevant chain of command? ☐ Would Soldiers understand what happened in the scenario? ☐ Would Soldiers know all the military terms?
7. Drill Sergeant Review Panel	☐ Are these Drill Sergeants different from those who wrote the exercises? ☐ Are any responses too obvious or transparent? ☐ Are the distractors plausible or too obvious? ☐ Are the correct answers too high in social desirability? ☐ Do any responses offer hints to the correct answer (e.g., opposites)? ☐ Would Soldiers actually respond this way? ☐ Is the scoring appropriate? ☐ Are any alternate ways of perceiving responses?
8. Final Review	☐ Properly formatted exercises?

Appendix D

Video-Based Ethical Decision-Making Questions and Scoring Key

RED PHASE CAPE VIDEO ETHICAL DECISION-MAKING QUESTIONS

PVT Gress (Respect)

http://cape.army.mil/case-studies/vcs-single.php?id=57&title=pvt-gress

Item	Ethical Decision-Making Questions	A	B	C	D
1	PVT Gress describes a Soldier in her unit who isn't a team player and is disrespectful to others. What should you do in this situation? I should... (Part 1)	keep an open mind about this Soldier and give her opportunities to show that she can rise to the occasion as a team player.	never help PVT "My Way" if she ever needs it.	approach this Soldier in private and tell her that she is failing to earn the unit's respect.	tell other Soldiers about PVT "My Way's" behavior.
2	PVT Gress describes how her team members have come a long way since day one and have learned to trust and respect each other. What do you think "showing respect" means? I think it means… (Part 2)	telling everyone else what I think is going to happen, when it is going to happen, and why it is going to happen.	doing something for others without being asked first; just helping people out of a desire to make their lives easier.	helping people because the Commanding Officer expects it.	treating people the way I want to be treated, because it makes it easier for people to trust and respect each other.

PFC Smith (Duty)

http://cape.army.mil/case-studies/vcs-single.php?id=50&title=pfc-smith

Item	Ethical Decision-Making Questions	A	B	C	D
1	The student "Platoon Leader" heard about Soldiers using the phone before being allowed. You are one of the Soldiers who made a call home before the platoon's assigned time. If confronted by your platoon leader, what should you do? (Part 1)	Admit you made phone calls and work to regain trust and respect from your platoon.	Tell the Platoon Leader that you used the phone, but do not tell who else used the phone.	Say you did not use the phone.	Tell the Platoon Leader that you made phone calls and take responsibility for breaking the rules.
2	Your student "Platoon Leader" already punished the Soldiers that snuck off for a phone call. But then, you overhear two Soldiers in your platoon talking about whether or not they should confess. What should you do? (Part 2)	Do your duty and report them, even if they do not want to confess.	Encourage them to do their duty and admit their mistake.	It's already over, ignore them.	Don't snitch on my platoon members, if asked.

PV2 Cevidanes (Personal Courage)

http://cape.army.mil/case-studies/vcs-single.php?id=51&title=pv2-cevidanes

Item	Ethical Decision-Making Questions	A	B	C	D
1	PV2 Cevidanes knew that her father disagreed with her decision to serve in the Army. She also knew that her father would eventually find out she was enlisting. What should PV2 Cevidanes do? (Part 1)	Tell her father and accept the consequences for breaking the family's rules.	Avoid bringing up enlisting as long as she can.	Tell him that she understands his rules, but wants to do something that helps others.	She should be honest and tell her father she has enlisted.

2	You overhear a male Soldier in your unit say, "PV2 Cevidanes should have listened to her father and not enlisted. Women have no place in this Army." A few of the other Soldiers laugh and agree. How should you respond? (Part 2)	You pretend you did not hear the conversation.	Disagreeing will make them think less of you, so you laugh it off.	You must say something to your Squad Leader because this will affect unit cohesion.	You tell them that no one deserves to be told they do not belong.

WHITE PHASE CAPE VIDEO ETHICAL DECISION-MAKING QUESTIONS

PVT Anzaroot (Integrity)

http://cape.army.mil/case-studies/vcs-single.php?id=55&title=pvt-anzaroot

Item	Ethical Decision-Making Questions	A	B	C	D
1	PVT Anzaroot's platoon found out that other platoons are cheating during the BRM competition. What do you think PVT Anzaroot's platoon should do in this situation? PVT Anzaroot's platoon should…(Part 1)	not cheat, because this could get them in trouble.	not cheat, because other platoons would get angry and upset.	not cheat, because it could hurt trust among their platoon members.	not cheat, because cheating is wrong.
2	Imagine that you are in PVT Anzaroot's platoon. How do you think your platoon's decision to *not* cheat would influence other platoons? The rest of the platoons should…(Part 2)	get angry and upset.	view your platoon as role models.	stop cheating.	report cheating to their leader.

An Ethical Pause (Operational Video)

http://cape.army.mil/case-studies/vcs-single.php?id=4&title=an-ethical-pause

Item	Ethical Decision-Making Questions	A	B	C	D
1	SFC Thompson only had 3-weeks left of duty. Why should she risk her life to go on a mission? (Part 1)	To impress her Commanding Officer.	Because she knew they could order her.	Because she would be owed a favor.	Because she was committed to the mission.
2	SFC Thompson chose *not* to push through, against the SOP, to rescue the Turkish driver. How should her fellow Soldiers react to her actions? They should… (Part 2)	think that she's willing to risk U.S. Soldiers for Turkish drivers.	see what SFC Thompson is willing to do for others.	feel more confident in SFC Thompson's leadership.	be more likely to support their Turkish drivers.

"Out of Character" (Operational Video)

http://cape.army.mil/case-studies/vcs-single.php?id=30&title=out-of-character

Item	Ethical Decision-Making Questions	A	B	C	D
1	Imagine that you are in SFC Kolberg's situation: you have a lot of money in your hands from a suspected terrorist. What should you do in this situation? I should… (Part 1)	*not* take the money because I will likely be caught.	take the money and share it with my Team Leader.	*not* take the money because it would hurt relations with the locals.	take the money. He is after all a terrorist.
2	Imagine that after SFC Kolberg gave the Iraqi his money back, the Iraqi spit on him and walked back inside. How should you respond in SFC Kolberg's place? (Part 2)	Tell your platoon that it was still the right thing to do.	Report the Iraqi's behavior to the proper authorities.	Ignore it.	Take the money back from the ungrateful Iraqi.

BLUE PHASE CAPE VIDEO ETHICAL DECISION-MAKING QUESTIONS

Honorable Retaliation (Operational Video)

http://cape.army.mil/case-studies/vcs-single.php?id=19&title=honorable-retaliation

Item	Ethical Decision-Making Questions	A	B	C	D
1	Regardless of MOS, rank, and experience, the Army holds all Soldiers to the same ethical standards, but the Iraqi court system at the time was less consistent. Imagine that the Soldiers *knew* that the Sniper would go free. What should the squad do? (Part 1)	Kill the Sniper, it's the only way to get justice.	Plant evidence in the Sniper's house.	Bring the Sniper in; it was their mission to do so.	Bring the Sniper in, and coerce him into confessing before the Iraqis get him.
2	What if SSG Huntley's squad decided *not* to bring the Sniper back no matter what, and killed him in his house. What are the *worst* long-term consequences that could have occurred? The squad's actions could have…(Part 2)	resulted in the unit getting in trouble.	made the local population question Soldiers' respect for Iraqi law.	embarrassed the U.S. Army.	led to more dead Soldiers in vengeance.

PV2 Miller (Loyalty)

http://cape.army.mil/case-studies/vcs-single.php?id=53&title=pv2-miller

Item	Ethical Decision-Making Questions	A	B	C	D
1	Imagine that you are a Soldier in this platoon. You hear that this Soldier is struggling with separating from his fiancé back home and openly considering suicide. How should you respond? (Part 1)	Report the Soldier through proper channels, and then keep an eye out for him.	Report the Soldier through proper channels, and then avoid him afterwards.	Report the Soldier through proper channels, and recommend appropriate resources to get help.	Avoid this Soldier, because it's not your responsibility to handle.
2	Imagine that you are a Soldier in this platoon, and you see other Soldiers in your platoon making fun of the depressed Soldier. How should you respond to your fellow platoon members? (Part 2)	The Soldier is struggling so make an effort to help him.	Let the other Soldiers do what they want, unless the Platoon Sergeant says it's a problem.	Take the side of the struggling Soldier and give the others a taste of their own medicine.	Tell other Soldiers not to tease him.

Identity Crisis (Operational Video)

http://cape.army.mil/case-studies/vcs-single.php?id=20&title=identity-crisis

Item	Ethical Decision-Making Questions	A	B	C	D
1	SFC Carroll's team conducted a search of the vehicle and found the person dead, but there was no weapon. What should SFC Carroll's unit do next? (Part 1)	Continue on their mission—SFC Carroll said he saw a weapon.	Continue on their mission. It was a bad area and the person was probably bad too.	Secure the area and contact their commander about the situation.	Secure the area and search for the weapon again.
2	Before they found the weapon, SFC Carroll says, "I remember everything just froze for me. I had just directed an entire platoon worth of weaponry at a non-combatant." Why do you think he should be upset at the time? He should be upset because…(Part 2)	his platoon will have doubts about his future decisions.	this will likely harm his opportunity for promotion.	he will likely be investigated.	it was his decision, and now he has to live with it.

ADDITIONAL VIDEOS NOT USED IN PART 3 ASSESSMENTS

PFC Breland (Selfless Service; Suggested for White Phase)

http://cape.army.mil/case-studies/vcs-single.php?id=48&title=pfc-breland

Item	Ethical Decision-Making Questions	A	B	C	D
1	Imagine that PFC Breland is deciding whether or not to help her Battle Buddy. What should PFC Breland consider when making her decision? She should think whether…(Part 1)	other people can help instead.	her Drill Sergeant will allow her to help.	she is able to help.	it is worth her trouble to help.
2	Imagine that you are a member of PFC Breland's platoon. You helped one of your Battle Buddies who fell behind on a 2-mile run. Other platoon members mocked you about it. What should you do in this situation? I should… (Part 2)	listen to them; after all every Soldier is in it for him/herself.	ignore their comments, because sometimes others need help.	make sure that my Drill Sergeant saw that I helped my Battle Buddy.	report their mocking to my Drill Sergeant.

PV2 Montes (Honor; Suggested for Blue Phase)

http://cape.army.mil/case-studies/vcs-single.php?id=54&title=pv2-montes

Item	Ethical Decision-Making Questions	A	B	C	D
1	Imagine that you just saw the Drill Sergeant jump on the potentially live grenade to protect your squad. How should you respond to the PVT who just dropped the grenade? (Part 1)	Help the Private to learn from his/her mistake.	Ignore the Private. Follow the Drill Sergeant's instructions.	Tell your squad members to stay away from the Private in future live-fire training.	Tell your squad members about the Drill Sergeant's sacrifice, not the Private's mistake.
2	Imagine that you are at the grenade range and you see fellow Soldiers in your squad joking, shoving each other, and mishandling training (non-explosive) grenades. What should you consider? (Part 1)	Soldiers should take the safety of their fellow Soldiers seriously.	Soldiers should know training might save their squad members' lives when they deploy.	The platoon is having fun and improving cohesion.	The Drill Sergeant hasn't said not to, so it's fine.

Scoring Key (Incorrect = 0; Correct = 25, 50, or 75; Most Correct = 100)	A	B	C	D
Red-V1-Q1 (Gress, Respect)	75	0	100	25
Red-V1-Q2 (Gress, Respect)	0	75	25	100
Red-V2-Q1 (Smith, Duty)	100	50	0	75
Red-V2-Q2 (Smith, Duty)	75	100	0	25
Red-V3-Q1 (Cevidanes, Personal Courage)	50	0	100	75
Red-V3-Q2 (Cevidanes, Personal Courage)	0	0	75	100
White-V1-Q1 (Anzaroot, Integrity)	50	50	100	75
White-V1-Q2 (Anzaroot, Integrity)	0	100	75	50
White-V2-Q1 (Ethical Pause, Operational)	50	25	0	100
White-V2-Q2 (Ethical Pause, Operational)	0	75	100	50
First72-V1-Q1 (Out of Character)	25	0	100	0
First72-V1-Q2 (Out of Character)	100	50	50	0
Blue-V1-Q1 (Honorable Retaliation, Operational)	0	0	100	0
Blue-V1-Q2 (Honorable Retaliation, Operational)	25	100	50	75
Blue-V2-Q1 (Miller, Loyalty)	75	25	100	0
Blue-V2-Q2 (Miller, Loyalty)	100	0	25	75
Blue-V3-Q1 (Identity Crisis, Operational)	25	0	100	75
Blue-V3-Q2 (Identity Crisis, Operational)	75	25	25	100
Additional questions not used in the part 3 assessments				
White-V2-Q2 (Breland, Selfless Service)	25	50	100	25
White-V2-Q1 (Breland, Selfless Service)	0	100	25	50
Blue-V2-Q1 (Montes, Honor)	100	50	25	75
Blue-V2-Q3 (Montes, Honor)	75	100	0	0

Appendix E

Guidance from Academic Literature on Using Clicker Technology

Best practices and recommendations (for a review, see Caldwell, 2007)

- *Planning for clicker technology use:*
  - Know why you are using the clicker technology in class, and keep this in mind while writing questions.
  - Plan your grading system in advance. Make sure it aligns with your learning goals.
  - Plan in advance for how to deal with students whose clickers are forgotten, need batteries, or are broken: Use slips of paper, have students trade ID cards for clickers, or keep some "loaner" clickers on hand. Discourage perpetual freeloaders.
  - Before teaching your first course, watch another instructor who uses clicker technology.
  - Be aware that the first year of use requires extra time to prepare good questions.

- *Use of clickers for participation:*
  - If you want to increase participation, use clickers regularly and link clicker usage to course performance.
  - Use clickers especially with introductory courses to encourage participation, engagement, and accountability.
  - If you are requiring participation, expect an increase in noise and possibly some disengaged students who are participating only for points.

- *Use of clickers for communication with Students:*
  - Explain to students why you are using the system and what you expect students to gain from the experience in order to get them to support the idea, especially if you are using it for nontraditional activities like active learning.
  - Plan discussion time to respond to questions using clickers. Be willing to adapt your lesson plan according to the results you collect. Let students "learn from the discussion of right and wrong answers." This is considered vital by most researchers in the field.
  - If incorporating a class-wide discussion into your clicker use, be sure to summarize the discussion and explain the correct answer afterward.
  - Explain to students the purpose of "homework," and use clickers to hold them accountable.
  - If used for testing, discuss cheating with students, and clearly state that use of another student's clicker is unacceptable.

- *Use of clickers for peer learning:*
  - If using peer learning groups, limit group size to no more than four to six members. Students seem to prefer small-group discussions to class-wide discussions led by the instructor.

- *Use of clickers and grades & anxiety:*
  - If clicker scores are part of the course grade, make those scores accessible on a regular basis to reduce student anxiety. Consider showing students clicker scores from past semesters on the first day of class.
  - Give partial credit for any answer and full credit for correct answers to reduce anxiety and limit cheating. Consider dropping a few of the lowest clicker scores or selecting a portion of data at random.

- *Use of clickers and preventing wasted time & frustration:*
  - Spend some time in the first classes training students to use clickers.
  - Set up the system before class, and practice this before the semester begins.
  - If your clickers require a registration system, test it in advance.
  - Expect that a few students will intentionally press the wrong button, cast misleading votes, or delay voting to consume class time.
  - If possible, find a resource person, train a teaching assistant, or start an instructor support group to share best practices.

- *Other survival tips:*
  - Keep a positive attitude, and be willing to make a few mistakes as you learn. Consider this a chance to model the learning behavior you desire from your students.
  - Be willing to throw out or regrade a question that contains an error or is unclear.
  - Encourage students to discuss answers with each other. This increases peer learning and engagement.
  - Encourage class discussion of incorrect answers to reveal unclear wording; this can be especially important if you notice dramatic improvement in scores after peer discussion.
  - Consider building a library of clicker questions with colleagues, as too few of these exist in most fields.

- *Some useful goals for question design can be culled from the literature:*
  - Good clicker questions should address a specific learning goal, content goal, skill, or reinforce a specific belief about learning:
  - Questions can:
    - assess students' background, knowledge, or beliefs
    - make students aware of others' views or of their own
    - locate misconceptions and confusion
    - distinguish between related ideas
    - show parallels or connections between ideas
    - explore or apply ideas in a new context.

- *Some examples of questions recommended by the literature include:*
  - given a term or concept, identify the correct definition from a list, and vice versa

- given a graph, match it with the best description or interpretation, and vice versa
- match a method of analysis with an appropriate data set, and vice versa
- questions that link the general to the specific
- questions that share a familiar situation or example with several other questions
- questions that students cannot answer, to motivate discussion and curiosity before introducing a new topic
- questions that require ideas or steps to be sorted into order
- questions that list steps and ask "Which one least accurate? Which one is the most appropriate response? Why did you answer this way?"
- questions that apply a familiar idea to a new context.

- *To construct questions that support deeper understanding, it is helpful to:*
  - identify student misconceptions and include them as answers, plausibly phrased.
  - listen to students to find out how they think.
  - include answers that contain common misconceptions or errors.

- *Practical suggestions include:*
  - limit the number of answers to five or less, so that question is easy to read and consider
  - assess knowledge of jargon separately from concepts to ensure that each is addressed clearly and effectively
  - create wrong answers (distractors) that seem logical or plausible to prevent "strategizing" students from easily eliminating wrong answers
  - include "I don't know" as an answer choice to prevent guessing
  - plan to ask some questions twice to allow peer learning and build emotional investment. (Allow students to answer individually, but do not display the correct answer; then direct students to discuss the question with their peers and answer again.)

- *A 6-step model for use of clickers for formative assessment:*
  - The instructor constructs a lesson module and related assessments based on the perception of the students' readiness and prior knowledge (Stage 1).
  - The instructor presents the lesson module (Stage 2)
  - The instructor administers an assessment (Stage 3).
  - The instructor considers assessment results. The student considers the assessment results (Stage 4).
  - Dialogue between the instructor and the student begins (Stage 5). Depending on dialogue with the instructor, the student adjusts learning style or proceeds with current style.
  - Depending on the dialogue, the instructor adjusts teaching or proceeds to the next learning module (Stage 6).

Additional Resources for Clicker Technology Use in Classrooms

- Barber, M., & Njus, D. (2007). Clicker evolution: seeking intelligent design. *CBE—Life Sciences Education, 6*(1), 1-8.
- Beatty, I. (2004). Transforming student learning with classroom communication systems. *EDUCAUSE Center Applied Research (ECAR) Research Bulletin, 2004*(3), 1-13.
- Beatty, I. D., Gerace, W. J., Leonar, W. J., & Dufresne, R. J. (2006). Designing effective questions for classroom response system teaching. *American Journal of Physics, 74*(1), 31-39.
- Bruff, D. (2009). *Teaching with classroom response systems: Creating active learning environments*. San Francisco, CA: Jossey-Bass.
- Bruff, D. (2007). Clickers: A classroom innovation. *National Education Association Advocate, 25*(1), 5-8.
- Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE—Life Sciences Education, 6*(1), 9-20.
- Connor, E. (2009). Perceptions and uses of clicker technology. *Journal of Electronic Resources in Medical Libraries, 6*(1), 19-32.
- Czekanski, A. J., & Roux, D. M. P. (2013). *The use of clicker technology to evaluate short- and long-term concept retention.* In Proceedings of the American Society for Engineering Education Annual Conference. Retrieved from https://www.asee.org/public/conferences/140/papers/25153/view
- DeBourgh, G. A. (2008). Use of classroom "clickers" to promote acquisition of advanced reasoning skills. *Nurse Education in Practice, 8*(2), 76-87.
- Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning, 20*(2), 81-94.
- Frehywot, S., Vovides, Y., Talib, Z., Mikhail, N., Ross, H., Wohltjen, H., ... & Scott, J. (2013). E-learning in medical education in resource constrained low-and middle-income countries. *Human Resources for Health, 11*(1), 4.
- Han, J. H. (2014). Closing the missing links and opening the relationships among the factors: A literature review on the use of clicker technology using the 3P model. *Journal of Educational Technology & Society, 17*(4), 150.
- Han, J. H., & Finkelstein, A. (2013). Understanding the effects of professors' pedagogical development with Clicker Assessment and Feedback technologies and the impact on students' engagement and learning in higher education. *Computers & Education, 65*, 64-76.
- Stull, J., Varnum, S. J., Ducette, J., & Schiller, J. (2011). The many faces of formative assessment. *International Journal of Teaching and Learning in Higher Education, 23*(1), 30-39.