Assessing Character in Army Initial Entry Training

Tatiana H. Toumbeva
Aptima, Inc.
Woburn, MA
ttoumbeva@aptima.com

Frederick J. Diedrich, Scott M. Flanagan
Sophia Speira
Carthage, NC
frederick.diedrich@gmail.com, scott@sophiaspeira.com

Jennifer S. Tucker, Elizabeth R. Uhl
Army Research Institute
Ft. Benning, GA
jennifer.s.tucker.civ@mail.mil, elizabeth.r.uhl.civ@mail.mil

Andrew Naber, Kristy Reynolds, Krista L. Ratwani
Aptima, Inc.
Woburn, MA
anaber@aptima.com, kreynolds@aptima.com, kratwani@aptima.com

Giovanna Ubillus, Charles Nocker, Corrine M. Gerard
PricewaterhouseCoopers Public Sector, LLP
Washington, DC
giovanna.ubillus@pwcus.com, charles.nocker@pwcus.com, corrine.m.gerard@pwcus.com

ABSTRACT

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 Ethic, 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 development of agile and adaptive Soldiers inculcated in the Army Ethic continues during 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 activity. A key challenge is assessing character in a manner that facilitates Soldier development while also enhancing leader understanding of training effectiveness. This research showcases an assessment strategy and tools developed and tested for capturing and tracking character in Basic Combat Training (BCT), a component of IET. Specifically, trainee responses on peer evaluations and ethical decision-making questions based on videos showcasing the Army Values were digitally captured using audience response clicker technology. An Excel-based tracking tool was also developed to automatically store and analyze assessment data for insights into programmatic trends. A quasi-experimental study was conducted with three platoons as an initial test of the assessment tools. Preliminary findings indicate that instruction and technology interact to differentially impact learning outcomes (knowledge, critical thinking). Technology alone does not promote learning; as such, its implementation 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. The assessment tool development process, study results, and practical implications (assessment tool integration, generalizability to other contexts) are presented and discussed.

ABOUT THE AUTHORS

Tatiana H. Toumbeva is a Scientist in the Learning and Training Systems Division at Aptima with expertise in quantitative and qualitative methodologies, training evaluation, assessment tool development and validation, and occupational health psychology. She holds a Ph.D. in Industrial-Organizational Psychology from Bowling Green State University, and M.A. in Mental Health Counseling from Boston University.

Frederick J. Diedrich is a Consulting Principal Scientist whose work focuses on methods of instruction and assessment designed to deliberately support student development across a wide range of desired outcomes. His work focuses on issues such as the creation of assessment tools to grow leader attributes, the influence of instructional design on social skills development, and the creation of assessment tools for team feedback. Dr. Diedrich holds a Ph.D. in Cognitive Science and a M.S. in Experimental Psychology from Brown University, and a B.A. in Psychology from Hamilton College.

Scott M. Flanagan is a retired Special Forces Master Sergeant with 20 years of active duty service in the U.S. Army and 18 years of Special Operations experience assigned to the U.S. Army Special Operations Command (USASOC) at Ft. Bragg, NC. He has worked with organizations such as the U.S. Army Reconnaissance Course to conceptualize, plan, and develop competencies and attributes associated with the Army Leadership Requirements Model.

Jennifer S. Tucker is the Chief of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) Fort Benning Research Unit, GA, specializing in the areas of leader assessment and development and human performance. Current research interests include developing innovative methods that assess, track, and provide feedback on critical cognitive, social, and other attributes that are needed to optimize Leader and Soldier performance across their careers. She holds a Ph.D. in Systems Science: Industrial/Organizational Psychology from Portland State University, Portland, OR.

Elizabeth R. Uhl is a Senior Research Psychologist at the U.S. Army Research Institute for the Social and Behavioral Sciences - Fort Benning Research Unit. Dr. Uhl's current research interests include talent assessment and management, such as the use of peer evaluations for formative feedback. Dr. Uhl holds a Ph.D. in General Psychology from the University of Texas at El Paso.

Andrew Naber is a scientist at Aptima, whose expertise and research interests focus on individual and team training and performance measurement, leadership, and culture. With his colleagues, Andrew has developed individual and team training systems, job and decision aids, and novel performance measures that support missions, leaders, and teams. Andrew's work appears in Human Factors, Journal of Applied Social Psychology, Human Performance, Personality and Individual Differences, and the Encyclopaedia of Industrial Organizational Psychology. Andrew holds a Ph.D. in Industrial-Organizational Psychology from Texas A&M University.

Kristy Reynolds is a Scientist in the Learning and Training Systems Division at Aptima. She has expertise in the areas of knowledge elicitation, job/task analysis, competency modeling, training needs analysis, training design and evaluation, and performance assessment. She also specializes in building advanced Excel-based tools to help facilitate, enhance, and/or streamline training-related functions. Kristy holds an M.S. in Industrial/Organizational Psychology from University of Nebraska Omaha, a B.S. in Psychology from Montana State University Billings, and an A.A.S. in Computer Programming and Application Development from Montana State University Billings—City College.

Krista L. Ratwani is a Principal Scientist and Director of the Learning and Training Systems Division at Aptima with experience in leader development and training, training evaluation and design, survey development, and qualitative data analysis. She holds a Ph.D. and M.A. in Industrial-Organizational Psychology from George Mason University, and a B.A. in Psychology from Monmouth University.

Giovanna Ubillus is a Manager in the National Security Practice at PricewaterhouseCoopers Public Sector LLP with over 10 years of experience delivering strategic communications and change management services to Department of Defense clients. She holds a B.A. in Liberal Studies from Georgetown University.

Charlie Nocker is an Experienced Associate in the National Security Practice at PricewaterhouseCoopers Public Sector LLP. He holds an M.A. in International Relations from the Johns Hopkins University School of Advanced International Studies and a B.S. in Foreign Service from Georgetown University.

Corrine Gerard is a Director in the National Security Practice at PricewaterhouseCoopers Public Sector LLP with over 10 years of experience delivering process improvement, strategy, communication, and change management services to Department of Defense clients. She holds a M.S. in Organizational Development and Knowledge Management from George Mason University, and a B.A. in Government and International Politics from George Mason University.

Assessing Character in Army Initial Entry Training

Tatiana H. Toumbeva
Aptima, Inc.
Woburn, MA
ttoumbeva@aptima.com

Frederick J. Diedrich, Scott M. Flanagan
Sophia Speira
Carthage, NC
frederick.diedrich@gmail.com,
scott@sophiaspeira.com

Jennifer S. Tucker, Elizabeth R. Uhl
Army Research Institute
Ft. Benning, GA
jennifer.s.tucker.civ@mail.mil,
elizabeth.r.uhl.civ@mail.mil

Andrew Naber, Kristy Reynolds, Krista L. Ratwani
Aptima, Inc.
Woburn, MA
anaber@aptima.com,
kreynolds@aptima.com,
kratwani@aptima.com

Giovanna Ubillus, Charles Nocker, Corrine M. Gerard
PricewaterhouseCoopers Public Sector, LLP
Washington, DC
giovanna.ubillus@pwcus.com,
charles.nocker@pwcus.com,
corrine.m.gerard@pwcus.com

INTRODUCTION

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, motivated by the legal and moral foundations of the Army Ethic, 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, right 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. Individuals demonstrate character when they correctly identify the ethical implications of a decision and act in accordance with the Army Ethic (ADRP-1; U.S. Department of the Army, 2015a).

Toward this end, internalizing the Army Ethic (i.e., the guiding laws, values, and beliefs of the Army Profession) as personal beliefs is essential for development as a member of the Army. Over time and across all activities, to develop character in others, leaders must embody the seven Army Values (Loyalty, Duty, Respect, Selfless Service, Honor, Integrity, Personal Courage), which are fundamental to the Army Ethic. Similarly, they must look for character development opportunities that exist in day-to-day operations, and communicate expectations that others also embody the Army Ethic inclusive of the Army Values. When Soldiers fall short, counsel must be provided on the gaps between personal values and those 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 therefore 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 The Army's Framework for Character Development, 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 (DS), 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; individuals 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). The culture of trust in the Army is built upon a foundation of learning to live the Army Values, and more generally the Army Ethic as habits of daily living that permeate all activities. In this manner, character is just as critical an objective as Soldiers learning to operate their weapons, navigate, and attain physical fitness. For instance, Soldiers must employ their weapons in the context of respect, integrity, and honor. 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 must also support Soldier learning while enhancing leader situational understanding of training effectiveness that ultimately informs the Army's larger character development initiative.

To address this gap, the research presented here showcases an assessment strategy and tools developed and tested for capturing and tracking character in the BCT environment. Working with a Battalion (BN) at Ft. Benning, GA, we focused on the creation of formative assessment tools. In contrast to summative tools, our focus was on creating assessments that could be used to provide feedback to guide learning (Bransford, Brown, & Cocking, 2000). Our emphasis was not to create 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, our starting assumption was that all Soldiers enter the Army from a variety of backgrounds with differing conceptions of what character means. Although not yet formally quantified, the role of BCT is to move Soldiers from whatever starting point they begin to a place where they internalize and live in accordance with how the Army sees character. Given this emphasis, and the focus of the BN on the Army Values in particular, our approach includes tools to enable consistent peer-based feedback regarding values-based behaviors as well as tools to assess reactions to situations involving ethical decision-making. Together, these tools were designed to provide developmental feedback to individual Soldiers and programmatic feedback to BN leadership as different instructional approaches to character assessment and development are pioneered in the future.

We begin by providing an overview of the assessment approach. Within this context, we then describe development and initial testing of various components of the solution. We present evidence concerning DS and Soldier reactions to the approach, as well as evidence concerning knowledge and critical thinking related to performance on ethical decision-making questions. We then conclude with a discussion of the implications and limits of the approach, and provide suggestions for how to harness the tools to support learning.

Overview of Assessment Approach

The primary components of the assessment system and approach are illustrated in Figure 1. In broad terms, from left to right, the system hinges on two sources of data (ethical decision-making questions and peer evaluations) that feed Soldier responses into a database and automatic tracking tool. Soldier responses to the ethical decision-making questions and peer evaluations are captured via Turning Technologies "clickers," which are off-the-shelf polling tools that can be used in a classroom setting. The system was built on existing and approved tools to enable use with multiple individuals at a single time given high instructional tempos. In the upcoming sections, we discuss the assessment system components in detail and describe how they were developed, or adapted, for use in this context.

Figure 1. Assessment System and Approach

Development of Ethical Decision-Making Questions

The ethical decision-making questions were created based on pre-existing Army Values videos from the Center for the Army Profession and Ethic (CAPE) that were frequently used by DS 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 Soldiers view the CAPE videos, the clicker

system presents the SJTs that pose short hypothetical moral scenarios. Soldiers are prompted to think critically about the scenarios in light of ethical decision-making, and respond using 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, paragraph 5-4). This notion stems from extant theories of the stages of moral development (Kohlberg, 1971; Rest, Narvaez, Bebeau, & Thoma, 1999). Cognitive moral development moves from lower levels emphasizing immediate concerns to one's self to higher levels, emphasizing individuals as independent moral agents who live by their own ethical principles (e.g., life, liberty, justice) that may interact, conflict, or take precedence over society's broader view (Kohlberg, 1971). Rest and colleagues (Rest et al., 1999) perceived Kohlberg's theory as primarily addressing the formal structures of society as well as the practice of virtue in everyday, face-to-face relationships. The basis of ethical decision-making, and the individual differences of moral reasoning among both individuals and Soldiers, certainly vary. Nevertheless, a Soldier's character is revealed in the decisions and behaviors that conform to the moral principles of the Army Ethic, and therefore by doctrine. 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). The ethical decision-making questions used in this study were developed based on cognitive theories of moral development, learning methods to maximize collective learning processes and outcomes for individual and group tasks, and best practices for item development. Broadly, moral development stages correspond to pre-conventional, conventional, and post-conventional stages, which formed the basis for developing the SJT response options.

There are various approaches used to develop SJTs across a variety of jobs, contexts, and competencies (cf., Motowidlo et al., 1997). An SJT item includes multiple responses that may be considered degrees of "correct," as understood by subject matter experts (SMEs) to be the best response among the alternatives. The ethical decision-making questions were developed based on several requirements that aligned with existing practices and available technology. The questions 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 three interrelated item development methods: (a) moral reasoning development, (b) SJT, and (c) discussion-based training. In the future, this process could be replicated with some adaptations by DS to produce similar items (e.g., based on other videos, personal experiences). Optimally, this approach, especially when combined with clicker technology that enables real-time viewing of trainee responses, allows DS to tailor their instructional approach to the needs of the classroom and therefore reduces the time needed to grasp the material.

Producing the questions followed an iterative process of development, review, and refinement. Based on the selected CAPE videos and accompanying facilitator materials, critical ethical challenge(s) depicted in the video were identified. Some videos contained more than one ethical challenge 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. In the abstract, the "correct" answers on a moral reasoning test are easy to describe, but challenging to develop. In writing responses, the researchers considered the various levels of moral development stages relevant to the particular situation. This includes understanding that individuals at different levels of moral development may interpret the issues relevant to a moral dilemma differently. The total SJT item comprised a synopsis of the video, a brief recounting of the relevant details of the ethical challenge, the final question, and four behavioral responses to that item. 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 would you do in this situation? I would... (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* (the most correct response), *(b) never help PVT "My Way" if she ever needs it* (the least correct response), *(c) approach this Soldier in private and tell her that her attitude goes against the Army Value of respect, and (d) tell other Soldiers about PVT "My Way's" behavior*. Multiple items were developed for a subset of videos prioritized by the DS. The item stem questions employed knowledge-based (i.e., what should you do?) and behavioral tendency prompts (i.e., what would you do?). Based on guidance from the DS and unit leadership, a set of CAPE videos and questions were selected that aligned with the learning objectives at each phase of BCT.

Although there are a variety of Army Values that may be more or less relevant to a particular scenario, individual items were not meant to reflect any particular value. The dilemmas that Soldiers face in practice, and the particular scenarios available, require Soldiers to apply Army Values in totality. That is, Soldiers bring a set of values that may

be individually more or less relevant to a complex moral/ethical situation. Given the complexity, a Soldier's character allows him/her to appropriately weigh the values that correspond to the particular situation. Indeed, a single SJT item may assess a number of constructs simultaneously: knowledge of rules and regulations, 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.

Development of an Army Values Rubric for Peer Evaluations

For peer evaluation, the system requires Soldiers to rate and comment on peers based on the Army Values using clickers. Given that only short definitions of each Army Value accompanied the peer evaluation form used by the BN, a rubric was developed to help create a shared mental model about how the values can manifest behaviorally in BCT. The goal of the rubric was to supplement the peer evaluation process and improve the utility and usability of the peer evaluation results. The rubric was created in the form of a behaviorally anchored rating scale (BARS). BARS consist of brief, specific, observable behaviors (i.e., behavioral anchors) at different proficiency levels that prompt the rater to think about the extent to which Soldiers act in accordance with each Army Value. When used for peer evaluations, the BARS can provide Soldiers 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. BARS have been shown to 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 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 DS as in the current study). Based on guidance from DS, such measures should focus on the Army Values only as they are fundamental in character development and lie at the core of BCT. A series of focus groups were conducted with several DS at a time 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). Following an iterative process, several cycles of anchor development, review by SMEs, and revision were completed 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 and DS understand programmatic trends, identify struggling Soldiers, and optimize training. 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., DS vs. unit leadership). Among the requirements gathered were unit-level composition and maximum number of (a) Soldiers in

squad, (b) cycles per year, (c) DS per squad, (d) peer evaluations sessions per cycle, (e) peer rating questions per evaluation session, and (f) CAPE video questions across phases of BCT. The current and future peer and video assessment objectives and needs were discussed with DS 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 question data can be exported from the clicker software in Excel. Because of the built-in features of the tool, minimal end-user effort is required. With a simple click, the clicker software output is automatically transformed into a data entry error free format that is immediately imported into the tracking tool. The results can be viewed in real time following upload. Standardization of the peer and CAPE 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 cycle-over-cycle (see Figure 2).

Figure 2. Character Tracking Tool Example Dynamic Tabs Depicting Hypothetical Data

PILOT STUDY METHOD

A pilot study was conducted to gather initial utility and usability feedback on the assessment tools (i.e., selected CAPE videos, ethical decision-making questions, Army Values rubric) and method for capturing peer and video question responses via clickers. The study also examined how conducive the technology and 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 rank of E-1, 23% indicated a rank of E-2, 13% held a rank of E-3, and 7% held a rank 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's

degree (3%), and graduate level work or degree (3%). There was a comparative breakdown of age, rank, and education across platoons.

The assessment with the three platoons was conducted simultaneously in separate classrooms, each facilitated by a DS and a member of the research team. Across all platoons, the DS were asked to lead an Army Value-based discussion based on the provided CAPE videos. Following the videos and discussion, a learning check test was administered to assess knowledge retention based on the videos. The three platoons were assigned to three different conditions that varied in the assessment support tools provided. One platoon served as the control and was provided with the CAPE videos for viewing and discussion. The control condition did not receive the ethical decision-making questions, peer evaluations, or clickers; trainees in this condition completed a paper-based learning check test, background (i.e., demographics) questionnaire and feedback survey. Another platoon was provided with the CAPE videos and associated ethical decision-making questions to help DS generate discussion with trainees, but not clickers; in this condition (paper-based SJT condition), the trainees responded to the ethical decision-making questions, learning check test, background questionnaire, and feedback survey on paper. A third platoon used clickers to enter responses to the peer evaluations, ethical decision-making questions, background questionnaire, and feedback survey (clicker condition). 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 DS led the values-based discussion.

Prior to the assessment, the participating DS were familiarized with the tools and clickers that would be used during the pilot. Minimal guidance was provided to the DS on expected facilitation approach using the provided tools. For the control condition, the DS was simply asked to lead the discussion as he normally would using the provided videos. For the SJT and clicker conditions, the DS was asked to lead a discussion on the Army Values based on the provided videos and ethical decision-making questions. 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. It was emphasized that the questions were intended to supplement the training and can be used to focus the classroom conversation.

At the end of the session, the trainees were asked to complete a feedback survey followed by a demographics questionnaire, and subsequently participate in a brief after action review (AAR) facilitated by a member of the research team; the DS was not present for this portion of the session. An AAR was also conducted with the participating DS separately. The session with each platoon took approximately one and a half hours to complete.

Measures

Peer Evaluations and Army Values Rubric

Use of the clicker technology to expedite the current BN paper-based peer evaluation process was tested with one platoon. Using clickers, trainees identified the top three and bottom three members of the platoon on the basis of embodying the Army Values in BCT, and entered comments justifying their selection. The Army Values rubric was provided to help trainees generate meaningful comments. As discussed, the rubric describes behavioral examples of each value based on DS expectations for Soldiers 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).

Ethical Decision-Making Questions based on CAPE Videos

The ethical decision-making questions employed two CAPE training videos from an existing set (see http://cape.army.mil/case-studies/) highlighting personal stories narrated by Soldiers about their experience in basic training. Three questions were used per video, each consisting of an SJT item stem describing the ethical dilemma, focal question, and four response options outlining courses of action ranging in degree of "correctness." Questions 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 the responses were collected on a question, the DS led a values-based discussion before proceeding to the next question. Overall, the response options across the collective set of questions were intended to capture increasingly mature and complex thoughts regarding moral reasoning/ethical decision-making. For instance, at the

highest levels the response options tended to address higher order principles, understanding of second order consequences and reciprocity, whereas the lower level responses tended to focus on issues related to self-interest and obedience to authority. For the purposes of the pilot, responses were dummy coded into 0 = least correct/incorrect, 1 = acceptable response, 2 = best/most correct response based on an answer key developed by the researchers. Higher scores on the SJT indicated better performance and suggest increasing levels of character (i.e., choosing the best response).

Learning Check Test Questions based on CAPE Videos

To assess learning (i.e., knowledge retention based on information presented in the videos), multiple choice, true/false, yes/no, and agree/disagree items were developed. Seven items were presented per video following the CAPE video and ethical decision-making question discussion. The learning check items were more straightforward compared to the ethical decision-making questions, and gauged how well trainees understood the basic situation and retained what was discussed in the videos. These items also served as attention checks (e.g., 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 toward the training conducted as part of the pilot study. For the clicker condition, the survey consisted of 34 items, of which six pertained to the discussion quality (e.g., *The class discussion was engaging*); seven pertained to the ethical decision-making question content (e.g., *The questions were too complex*); seven questions pertaining to ethical decision-making question utility (e.g., *The questions made me think more deeply about the Army Values discussed today*); seven questions about clicker functionality (e.g., *The clickers were easy to use*); and seven questions about clicker utility (e.g., *Seeing how other students responded to questions was valuable to me*). Trainees were asked to respond 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), excluding 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

As part of a background questionnaire at the end of the session, trainees responded to standard demographic questions including age, rank, and educational level. Clickers were used to capture trainee responses in one platoon, whereas the other two platoons (i.e., control and paper-based SJT conditions) completed the questionnaire on paper.

After Action Review (AAR) Questions

At the end of the session, a semi-structured interview protocol was followed to gather feedback from the trainees on their experience with the CAPE videos, questions, and discussions, as well as the clickers and peer evaluation process, if applicable. The discussion focused on the trainees' impressions of the training event, utility and usability of the assessment tools, and experienced level of engagement with the discussion. Following this discussion, a similar approach was followed to request feedback on these topics from the participating DS.

Clicker Technology

For one platoon, data capture employed clicker technology that 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 (TurningPoint 5) 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.

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

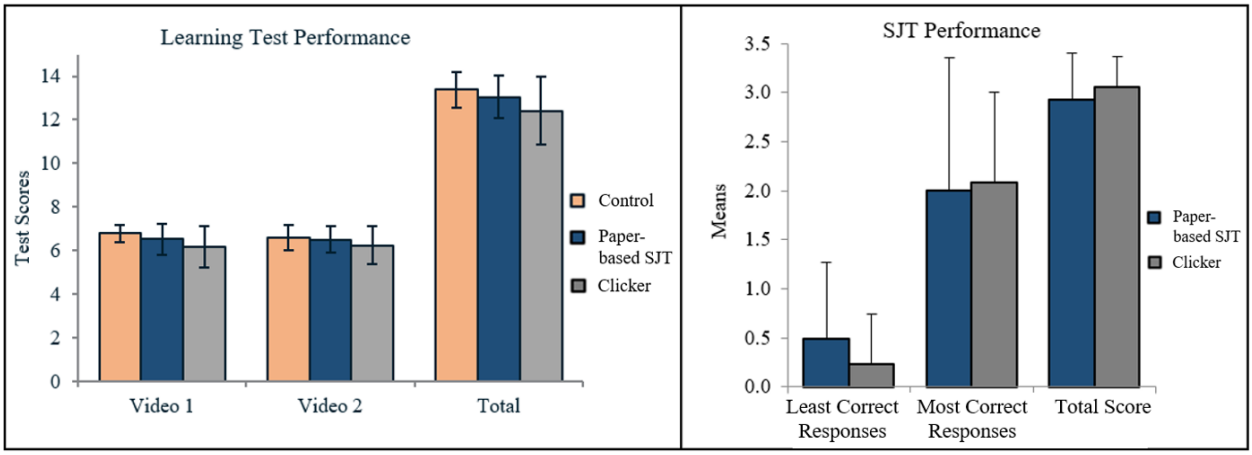
features, displays, and analytics, as well as create a more seamless data integration process. The tool was tested and revised iteratively as new requirements and contextual constraints were uncovered in the current and subsequent evaluation studies.

Data Analysis

Data from the CAPE video questions, 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 of interest across the three conditions.

PILOT STUDY RESULTS

Significant mean-level differences in learning test performance for the two CAPE videos were found between conditions, F (2, 138) = 11.06, p = .001. Post hoc comparisons using Tukey's HSD test indicated that (a) participants in the control (M = 13.37; SD = .83) and SJT (M = 13.04; SD = .98) conditions scored significantly higher than in the clicker condition (M = 12.42; SD = 1.57). No significant differences were found between participant scores in the control and paper-based SJT conditions (p > .05). An examination of mean-level trends suggests that participants in the control condition outperformed those in the paper-based SJT and clicker conditions, and participants in the clicker condition performed the worst on the learning check test. Furthermore, no significant mean-level differences in SJT (i.e., ethical decision-making question) performance was found between the paper-based SJT and clicker conditions (p > .05); however, mean-level trends indicate that individuals in the clicker condition performed better on the ethical decision-making questions than individuals in the paper-based SJT condition. Test performance results for each condition are shown in Figure 3.



Figures 3. Test Performance across Conditions

Based on survey feedback, trainee perceptions about the class discussion was positive. Significant group differences were found between conditions, F (2, 137) = 4.67, p = .01. Post hoc comparisons using Tukey's HSD test indicated 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 in mean ratings were also found for CAPE video relevance perceptions, 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 = 4.22; SD =.81) than both paper-based SJT and clicker (M = 3.77; SD = 1.08) conditions, F (2, 136) = 8.66, p < .001. Furthermore, trainees reported that the pace in the control condition (M = 4.69; SD = .47) was better than in both paper-based SJT (M = 4.32; SD = .78) and clicker (M = 4.26; SD = .74) conditions, F (2, 135) = 5.27, p < .01.

Overall, trainee feedback on the assessment tools and process was positive. As part of the AAR, trainees indicated that the (a) ethical decision-making questions were thought-provoking, and promoted sharing of opinions and introspection, (b) videos seemed outdated but relevant, and helped springboard discussion with DS, and (c) 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.

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

DISCUSSION

The present research describes the development and initial evaluation of a solution for assessing and training character at a critical point in a Soldier's career – basic training. Specifically, tools and methods were tested as part of a pilot study to measure character, facilitate Army Values discussion, and standardize as well as expedite the current peer evaluation process. Findings pertaining to 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 the best on ethical decision-making whereas the control condition performed the best on knowledge retention. Furthermore, lower scores on the learning check questions in the clicker condition may have been an artifact of the technology. 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.

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 that the control condition DS was highly experienced and proficient in leading engaging and meaningful discussions. In contrast, there was minimal interaction between DS and trainees in the clicker condition, as the DS used the response option breakdown charts provided by the clicker software to differentiate wrong from right course 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. There was clear utility of clickers for peer evaluations, but further investigation is needed to determine utility for facilitating class discussion. The CAPE videos and SJT questions can serve as springboard for discussion and help create an environment that promotes sharing among Soldiers. If DS are given 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. The added level of structure introduced by the tools was positively received; however, flexibility of use is valued among instructors. Additional testing is needed to fine-tune the tools and identify optimal ways to harness the technology in the context of BCT. For example, research currently underway is focusing on longitudinally assessing

character with an entire company at in the beginning, middle, and end of BCT to more rigorously track individual and unit-level performance. The results of that research will further inform the refinement of the assessment tools and process (i.e., CAPE video selection, ethical decision-making questions, peer evaluations, tracking tool, and use of clicker technology) and identify instructional approaches that will support Soldier learning and development across different phases of training.

Moral education programs demonstrate significant potential for growth in moral reasoning through discussion (Nucci, 2001). Notably, a clear advantage of SJTs in any training context is that trainee experience with the simulation can foster personal development. The act of taking the SJT as part of an assessment process may have some developmental benefit for the trainee. However, the full potential of SJTs for trainee development lies in individualized feedback opportunities for the trainees. Although in the current research, SJTs and other tools were used to produce discussion and 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. As such, while 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.

Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the US Army. The research described herein was sponsored by the US Army Research Institute for the Behavioral and Social Sciences, Department of the Army (Contract No. W911NF-16-F-0010).

ACKNOWLEDGEMENTS

We would like to thank the 194th Brigade 1-46th Infantry Battalion Command Team and Drill Sergeants who served as proponents of this work and provided valuable time, input, and feedback.

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.

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

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.

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).

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

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

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.

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

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, Headquarters (2015a). *The Army profession* (ADRP-1). Washington, DC: Author.

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