Innovative Mobile Technologies for Assessing and Enhancing Soldier Performance

Krista L. Ratwani, Courtney R. Dean
Aptima, Inc.
Washington, DC
kratwani@aptima.com, cdean@aptima.com

Scott Flanagan

Sophia Speira
Carthage, NC
scott@sophiaspeira.net

Camilla Knott, Frederick Diedrich
Aptima, Inc.
Washington, DC
ccknott@aptima.com, diedrich@aptima.com

Jennifer S. Tucker
Army Research Institute
Ft. Benning, GA
Jennifer.s.tucker.civ@mail.mil

ABSTRACT

A key element of the Army's Human Dimension Concept is the need to prepare Soldiers to thrive in conditions of uncertainty while they contend with ambiguous and amorphous threats. To prepare for such conditions, advanced talent management strategies are needed to facilitate Soldier development across the cognitive, physical, and social domains. Comprehensive talent management systems must ultimately leverage assessment tools to gather large amounts of data to enable a detailed determination of Soldier strengths and weaknesses and facilitate continuous learning. The question remains, however, about how best to achieve this goal. This paper reports lessons learned from research with the Army Reconnaissance Course (ARC), Ft. Benning, GA, to assess and track student performance over time in both performance outcomes (e.g., fundamental skills, understanding information needs) and leader attributes (e.g., anticipation, accountability). The final ARC performance assessment system included a mobile application to record student observations, a method to link those observations to key competencies, and a method for presenting trends over time. The trending method enabled student data to be aggregated across instructors and over classes to demonstrate larger changes in performance over time. In this paper, we present the methodology for developing this assessment system, results from an evaluation of the system, and reactions to employing the full assessment system during a course. The findings reflect the results from the in situ testing and use of the assessment system to include additional features which facilitate future utility and promote usability. Implications of the research are discussed to provide suggestions and future research questions to inform the creation of a comprehensive Soldier assessment system as the Army strives toward effective talent management strategies.

ABOUT THE AUTHORS

Dr. Krista Ratwani is a Senior Scientist and the Director of the Advanced Cognitive Training Systems Division at Aptima. She has experience in leader development and training, training evaluation and design, experimental design, and qualitative data analysis. She holds a Ph.D. in Industrial/Organizational Psychology from George Mason University.

Mr. Courtney Dean is a Senior Scientist and Lead for Aptima's A-Measure Product Team. He has expertise in measure development, job analysis and test construction and validation. Mr. Dean received his M.A. in Applied Psychology from the University of West Florida.

Mr. Scott Flanagan is a retired Special Forces Master Sergeant with 20 years of active duty service in the U.S. Army. He has 18 years of Special Operations experience assigned to the United States Army Special Operations Command (USASOC) at Fort Bragg, NC. Mr. Flanagan has worked with organizations such as the U.S. Army Reconnaissance Course to conceptualize, plan, and develop Combat Applications Training Course (CATC).

Dr. Camilla C. Knott is a Principal Scientist and Lead for the Instructional Strategy and Advanced Skill Development Capability at Aptima. Her expertise is in the areas of human performance measurement, human cognition, and training development and assessment. She holds a Ph.D. in Applied Experimental Psychology from the Catholic University of America.

Dr. Frederick Diedrich is President at Aptima. As President, Dr. Diedrich leads Aptima's scientific and technical divisions. He is also a Principal Cognitive Scientist with expertise in the areas of training, education, and human performance measurement. He holds a Ph.D. in Cognitive Science from Brown University.

Dr. Jennifer Tucker is a Senior Research Psychologist with the U.S. Army Research Institute (ARI) for the Behavioral and Social Sciences. For the past 11 years, she has specialized in the measurement and evaluation of military training and education programs. Her current research interests include systems thinking in organizational decision making and cognitive skills assessment and training. She holds a Ph.D. in Industrial/Organizational Psychology from Portland State University, Portland, OR.

Innovative Mobile Technologies for Assessing and Enhancing Soldier Performance

Krista L. Ratwani, Courtney R. Dean
Aptima, Inc.
Washington, DC
kratwani@aptima.com, cdean@aptima.com

Scott Flanagan

Sophia Speira
Carthage, NC
scott@sophiaspeira.net

Camilla Knott, Frederick Diedrich
Aptima, Inc.
Washington, DC
ccknott@aptima.com, diedrich@aptima.com

Jennifer S. Tucker
Army Research Institute
Ft. Benning, GA
Jennifer.s.tucker.civ@mail.mil

INTRODUCTION

In today's world, technology and information are everywhere and are constantly evolving. Every military organization in the world is attempting to build the biggest, fastest and most effective weapon systems while gaining and exploiting information about enemy forces. However, technology and information are only one part of the winning equation. Without smart and adaptive personnel capable of making effective decisions within the bounds of mission command, technology and information cannot win in a complex world (Department of the Army, 2014a). The U.S. Army, through a focus on assessing and developing Soldier skills and attributes, has acknowledged that one of its competitive advantages over other militaries is its people. At its crux, the Army's Human Dimension strategy is about optimizing human performance through a focus on Soldier development along three dimensions – cognitive, physical, and social skills (Department of the Army, 2014b). The cognitive component relates to development of memory, perception, and judgment. This piece of the Human Dimension strategy is focused on accelerating learning by delivering learner-centric experiences. The physical component of course relates to the physical fitness of Soldiers but also includes the concept of resilience. Finally, the social component is focused on developing Soldiers who interact effectively with others, embody the Army values and behave in morally and ethically responsible manners. As evident by these three components, the Army is not only concentrated on developing personnel who are tactically and technically competent. Although that element will never go away, Soldier development is about ensuring total readiness; therefore, a comprehensive approach to development must be taken to ensure that the cognitive, physical and social components are all targeted throughout one's career.

Thus, one objective of the Human Dimension strategy is successful talent management. Talent management is "systematic planning for the right number and type of people to meet the Army's needs at all levels and at all times so that the majority of them are **employed optimally**" (Department of the Army, 2015, p. 4). "Optimal employment" is extremely difficult and requires that Soldier development be continuously tracked so as to have an accurate record of not only assignments, but also strengths and weaknesses displayed within each of those assignments. Each assignment should build upon the previous to ensure that every Soldier is given the opportunity to effectively utilize past experiences. In addition, within each assignment, Soldiers must be provided with meaningful opportunities and associated feedback that allows them to continue to grow. By adopting such a model, meaning and motivation will not only be provided to every Soldier, but the Army can also be ensured that "the right person is in the right assignment at the right time" (Department of the Army, 2014b, p. 12). Critical to ensuring the success of this strategy is assessment and evaluation. Assessment must focus on all components of the Human Dimension concept (cognitive, physical, and social) and provide useful, actionable data to instructors and leaders so that progress can be tracked over time and throughout their careers.

The Army Talent Management Strategy (Department of the Army, 2015) explicitly dictates two near-term needs related to assessment that must be met: *"Identify, measure, and track the social, cognitive, and physical indicators required to assess performance and potential"* and *"Leverage scientific research to provide unbiased and relevant feedback on the baselines, attributes, and behaviors of individuals in order to enable continuous improvement"* (p. 16). The research conducted and described in this paper represents one method for helping the Army begin to meet those needs.

USE CASE: THE ARMY RECONNAISSANCE COURSE

One example of an Army course that is focused on developing the cognitive, physical, and social components of a Soldier is the Army Reconnaissance Course (ARC). The ARC aims to develop confident and agile Soldiers capable of operating under unpredictable combat and training situations. Specifically, the ARC has two sets of outcomes it tracks for students: outcomes associated with technical and tactical reconnaissance skills (e.g., planning and conducting a route reconnaissance), and leader attributes such as critical thinking, problem solving, and adaptability. One challenge that ARC instructors face, however, is in reliably observing and assessing those outcomes. Without effective diagnostic assessment tools, it is nearly impossible for ARC instructors to (a) accurately and consistently assess students across multiple dimensions (tactical and technical skills, as well as leader attributes) and (b) use those assessments to craft the next learning experience that will ultimately enable growth and development. Although the leader attributes are well-defined within the ARC curriculum, there is still a degree of subjectivity associated with the assessment of a competency such as confidence. Also, instructors who are observing and rating student performance rotate out during course exercises, creating a need to communicate the levels at which students are performing. For example, to enable continuity, when instructor A is replaced by instructor B during a training exercise, instructor B needs to understand instructor A's observations and assessments of student performance during instructor A's shift. Finally, within the ARC, the main mechanism for tracking student progress is through paper-based assessments; a method is needed to increase the reliability and sustainability of those assessments and relieve some of the burden created through a paper-based approach. Tracking progress on the leader attributes and other relevant performance outcomes is important so that course events can be adjusted to continuously challenge and develop the students based on prior performance.

The goals of the present research were as follows: (a) analyze the behaviors associated with the targeted leader attributes and develop objective behaviorally-based measures within the context of reconnaissance missions; and (b) develop a technology-supported performance assessment system to allow instructors to more easily track student performance (including the leader attributes) throughout the course. The analysis of the behaviors associated with the leader attributes can be found in Knott et al. (2014) and an example measure is showcased in Figure 1. Therefore, the remainder of this paper reports the research methodology that was employed to develop the performance assessment system and the results from an evaluation that was conducted to determine the effectiveness of the system in assessing and tracking student performance. More broadly, implications for how the performance assessment system can help the Army meet the intent of the Human Dimension strategy are discussed.

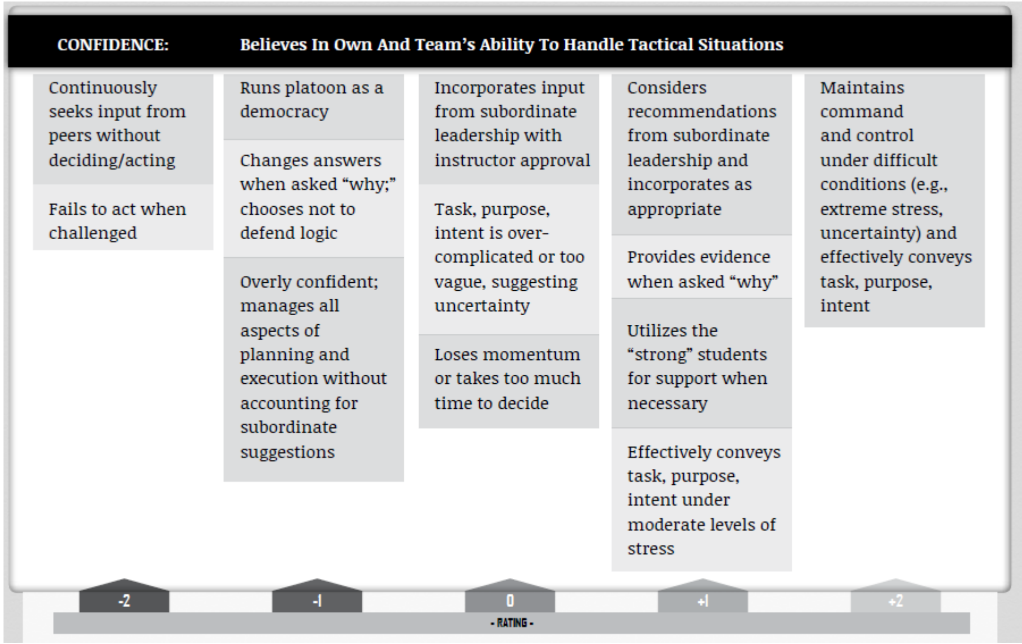

Figure 1. Example Leader Attribute Measure.

IDENTIFYING ASSESSMENT CHALLENGES AND REQUIREMENTS

Participants and Procedures

To determine how best to assess and track student performance throughout the course, a series of four workshops were conducted with ARC instructors; four to six instructors participated in each workshop, with several instructors participating in multiple workshops. The primary goal of the first workshop was to develop an understanding of the instructors' challenges and requirements associated with assessing students. Workshop discussion centered on the structure of the course and on identifying what instructors must do on a day-to-day basis to document and report on student progression. Several specific challenges and associated requirements were derived. First, the course is broken into phases, with field training exercises occurring every seven to nine days. Although the instructors were primarily concerned with assessment during those field training exercises, student assessment should also occur continuously throughout the course. The need for continuous assessment highlighted a challenge in that the instructor(s) assessing a student in the classroom was not necessarily the same as the instructor(s) assessing a student in the field. Therefore, there was a requirement to quickly and accurately share information with one another to enable continuity within and across students. Second, the instructors discussed the need to quickly make observations while in the field. Essential to being able to continually teach students was the need to not be buried in assessments during critical training moments. Therefore, a key to instructors facilitating student development through an assessment system was a separation of observation and assessment. In order to enable accurate assessments at a later time, the instructors were interested in being able to quickly capture video of key student behaviors, as well as make observations using both voice and text inputs. Such a thorough yet easy-to-use observation feature would provide instructors with the data needed to make assessments while on a break during the field exercise or back in their offices at the end of the day. Third, several times throughout the course, the instructors meet to discuss the progress of the students and are sometimes challenged to make effective decisions about students during those meetings. Therefore, the observations and assessments made with any assessment system should provide useful inputs to those meetings to facilitate meaningful discussions about student performance and enable tracking over time. A need was identified for simple graphical displays to enable sub-group comparisons as well as filtering of results by student, group assigned to within the class, class number, and year. Finally, although the instructors were in agreement that a digital solution would help them make more effective student assessments, they also expressed a desire to be able to create a packet of information on each student by the end of the course. Therefore, digital assessments that mimicked the current paper assessments and could be printed to hard copy would be most useful.

Based on these challenges and requirements identified in the first workshop, a series of static mockups of a mobile field tool were produced and reviewed during the second and third workshops. The second workshop featured an initial set of mockups representing key interface elements of a mobile field tool. These elements were presented and described to instructors. The mockups were presented in order of expected workflow to allow the instructors to conceptualize use of the tool to collect data and complete assessments for their students. During the presentation, instructors were encouraged to ask questions and provide comments with respect to the mockups and the tool concepts. The third workshop's goal was to finalize interface and functional design ideas through a revised set of mockups. The revisions reflected feedback received during the second workshop. Again, the mockups were presented and described in a fashion consistent with the anticipated workflow, and participant instructors were encouraged to comment or question the approach.

The fourth workshop had two goals. The first was to demonstrate a working prototype of a mobile field tool and verify that its function and features aligned with expectations of the instructors. The second goal was to introduce mockups and storyboards of an integrated database which would allow for tracking performance over time. Until this workshop, the database had only been discussed with respect to the challenges and requirements identified in the previous workshops. During this workshop, instructors were invited to try out the mobile tool and provide comment. The database was presented in the same fashion as mockups for the mobile tool in prior workshops, and instructors were encouraged to question, comment and otherwise guide revisions or changes to the tool and database.

Results

The results of the workshop led to the development of a performance assessment system termed the ARC Performance Assessment Toolkit (ARC-PAT) with two main components: a mobile observer-based measurement tool (the ARC-Field Tool [FT]) and an integrated database to enable tracking. The system was designed to provide mobile, digital data capture solutions with easy to use interfaces that reduced redundancies and overall workload.

The ARC-FT represents a digitized version of the notebook carried by instructors for formative assessment and the paper assessment form required for summative assessment of student achievement. To enhance both formative (e.g., during After Action Reviews) and summative assessments, the system enables instructors to record student behavior by taking photos and videos or by using voice-text (see Figure 2). Instructors can then attach those observations to assessments of expected course outcomes and the leader attributes (as described above) in both classroom and field events (Figure 3). Ultimately, the ARC-FT enables ARC instructors to capture critical performance metrics for students that reflect learning and progression within the ARC; a simple trending interface within the mobile tool provides instructors with data on students over the duration of the course (Figure 4).

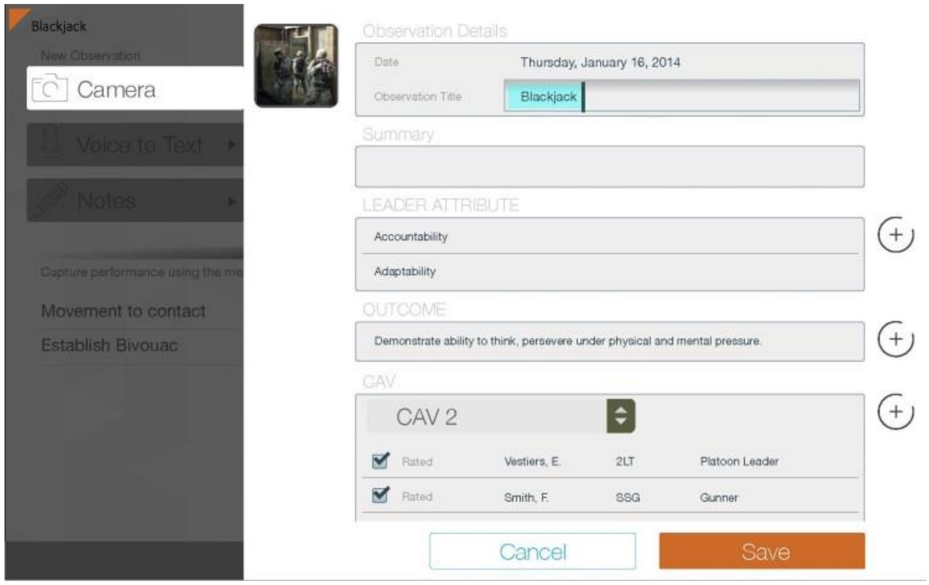

Figure 2. The Observation Feature of the ARC-FT.

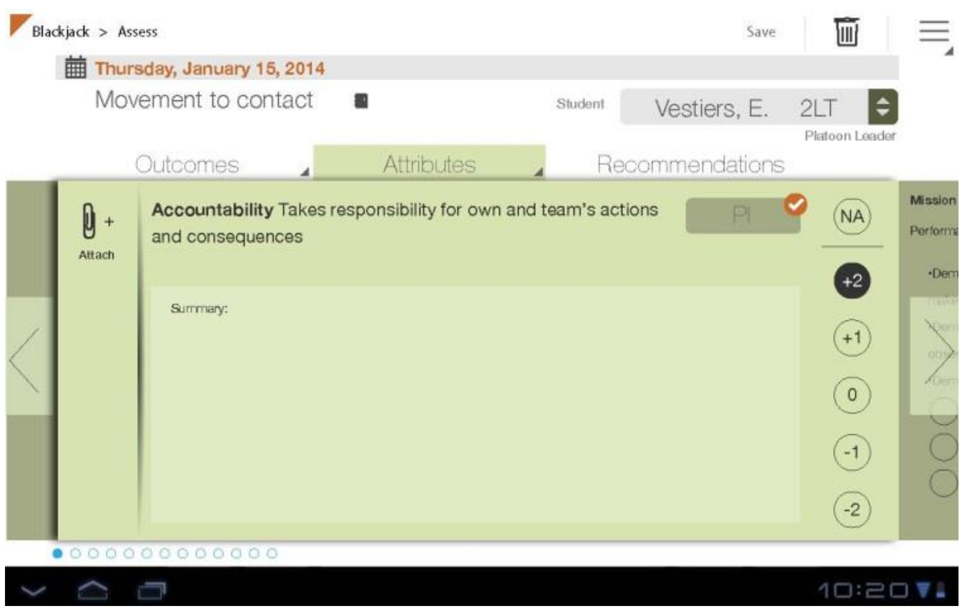

Figure 3. The Assessment Feature of the ARC-FT.

The Integrated Database was developed to store and manage student performance data. The user interface was designed to support the display and review of performance data across students, units, classes, and even years, and to aid in the review of data to identify trends and patterns. In addition, the database was set up such that student performance data, which mirrored the paper assessment forms, could be printed and stored according to current best practices. Instructors can access the database on a desktop or laptop computer and can pass data between the database and the FT through a wireless local area network (LAN). A diagram of the final ARC-PAT architecture is shown in Figure 5.

Figure 4. The Trending Feature of the ARC-FT.

Figure 5. Architecture Diagram of the ARC-PAT.

PERFORMANCE ASSESSMENT SYSTEM EVALUATION

Two evaluations of the performance assessment system were conducted. The first was a supported field test evaluation and the second was an unsupported leave-behind. During both evaluation events, instructors used the system to facilitate student performance assessment during graded course events. The primary goal of both evaluations was to conduct an initial evaluation of the ability of the ARC-PAT to support data collection and measurement during trainee assessment, and ultimately support Soldier development. In order for an assessment system to be efficient and effective, an instructor/unit leader must be able to use the system in an intuitive and easy manner while observing student behaviors. For research purposes, usability was defined as a *measure of the effectiveness, efficiency, and satisfaction with which specified users can achieve goals in a particular environment* (Bevan, Kirakowski, & Maissel, 1991). The term effectiveness refers to the utility of ARC-PAT in collecting, summarizing, measuring, and reviewing behaviors during training. Efficiency and satisfaction were operationalized as the users' ability to perform these tasks quickly and relatively error free. A secondary goal of the evaluation was to identify opportunities for enhancements based on usability issues or shortfalls in the system's functionality.

Participants and Procedure

First field test. Seven ARC instructors participated in the first field test, with some overlap across the instructors involved in the requirements-gathering workshops and the evaluation. This first field test was focused solely on evaluating the ARC-FT, not the integrated database. Prior to the start of the evaluation, instructors were given access to the tool, and brief training occurred on both the functionality of the tool and the measures. During the training, a written user guide was provided that described the intended use of the tool and identified the main operations and features that facilitate its use. Following the training, the instructors were encouraged to familiarize themselves with the functions of the tool prior to using it during the field training exercises and other graded course events.

This first, fully supported test consisted of embedding members of the research team with the ARC instructors during one of the ARC's field training exercises. Although some technical and user support was provided during the event, support was restricted to answering instructor questions and providing assistance as requested, rather than demonstrating the tool's potential capabilities. This strategy allowed the instructors to identify problem areas that needed to be addressed.

Measures. Shortly after the field training exercise, instructors were asked to complete a usability questionnaire and also respond to a set of more specific usability and utility statements; both questionnaires were aimed at identifying the extent to which the tool sufficiently met a set of key design principles and thus achieved the goals of effectiveness, efficiency, and satisfaction. The usability questionnaire was a standard set of 10 general statements about the tool (e.g., *I think that I would like to use the ARC-FT frequently* and *I found the various functions in the ARC-FT to be well integrated*) to which respondents agreed or disagreed on a 4-point scale (Strongly Disagree, Disagree, Agree, Strongly Agree). The results of this questionnaire provided data on the general approval or disapproval of the tool but did not point to specific elements of the tool that may be desirable or undesirable. More specific queries were featured in the usability and utility statements.

The usability and utility statements were based on heuristics derived from research by Nielsen and Mack (1994) and Tognazzini (2003) that represent tool usability. The 26 items were structured as statements that were framed positively or negatively for the ARC-FT. Participants were provided a document with the items and asked to check agree or disagree with each of the statements (e.g., *I knew exactly where to go within the tool to capture an assessment* and *The interface had too much information on it*). These statements allowed for a more fine-grained assessment of the tool by asking questions about specific features, as opposed to the more general assessment provided by the standard usability questionnaire.

Second field test. The second evaluation consisted of a minimally supported leave-behind with the complete ARC-PAT available for instructor use (i.e., both the ARC-FT and the integrated database). ARC-PAT was left with the instructors for one full class cycle, or about 40 days. Five instructors participated in this evaluation, with some overlap in instructors across the two evaluation efforts. During this time, no researchers were present; however, routine check-ins were performed to ensure that the system was operating as intended and to answer instructors' questions. Instructors used the system as much or as little as they desired with no encouragement or additional instruction offered by the research team. The goal of this evaluation was to identify any outstanding problems that needed to be addressed prior to the end of the research. Additionally, the realistic conditions of the leave-behind permitted a more rigorous test of the assessment system's capabilities, functions, and interactivity with the instructors.

Measures. Similar to the first supported evaluation, instructors completed questionnaires following the evaluation. The same utility questionnaire and usability and utility statements were used to evaluate the ARC-FT. In addition, similar questionnaires were developed specifically to assess the integrated database. For the usability and utility statements, there were 22 items generated to assess specific features of the database.

Results

Responses to the 10-item usability questionnaire were positive on all accounts. For the first field test, the mean rating across 10 questions was 3.65 (SD = .56) on a scale of 4. For the second field test, the mean rating for the FT was 3.36 (SD = .63) and 2.92 (SD = .97) for the integrated database. Item-level means for both field tests are shown in Table 1. Based on the ratings from the usability scale, it appears that the largest challenge with the FT was in learning the system and becoming familiar with the features and functionalities. As instructors use the FT more often, it is anticipated that those ratings would increase, and instructors will feel more comfortable with it. As demonstrated by the means, the instructors reported lower ratings for the database compared with the FT. As expanded upon below, the instructors did not use the database as much as the FT, and thus, the lower ratings are not completely unexpected.

Table 1. Responses to Standard Usability Questions

	Field Test 1: ARC-FT		Field Test 2: ARC-FT		Field Test 2: Database	
Statement	n	Mean	n	Mean	n	Mean
1. I think that I would like to use these display concepts frequently.	7	3.71	5	3.40	3	3.00
2. I found the system[+] to be unnecessarily complex.*	7	1.29	5	2.00	3	2.00
3. I thought the system was easy to use.	7	3.71	5	3.20	3	3.00
4. I think that I would need the support of a technical person to be able to use this system.*	7	1.29	5	1.60	3	2.00
5. I found the various functions in this system to be well integrated.	7	3.57	5	3.40	2	3.00
6. I thought there was too much inconsistency in this system.*	7	1.57	5	1.60	2	2.00
7. I would imagine that most people would learn to use this system very quickly.	7	3.71	5	3.60	3	3.00
8. I found the system to be very cumbersome to use.*	6	1.33	5	1.80	3	2.00
9. I felt confident using the system.	7	3.86	5	3.60	3	2.67
10. I needed to learn a lot of things before I could get going with this system.*	7	1.57	5	1.60	3	2.33

[+] For the questionnaire about the database, the word "system" was replaced with "integrated database" in all questions
* Reverse coded items

Based on the modified Nielsen and Mack (1994) measure, responses to the usability and utility statements were also largely positive. Users were provided with two options for responding to usability statements (Agree, Disagree). Percent agreement was calculated and reported with respect to Agree. Because one intent of using this measure was to identify specific features and functionalities that needed to be changed, the research team set a strict "passing" criterion of 75%; anything less than that standard represented a feature that may be in need of revision. The 75% criterion was set so as to ensure a strict level of scrutiny, leading to the identification of better interactions and interface elements in the final development iteration.

For the first field test, 20 out of 26 items reflected the passing criteria of over 75% agreement. For the second evaluation, 25 out of 26 items had agreement levels of 75% or above for the FT. In both cases, the failing items pointed to revisions that should be made to the tool. For example, the instructors had a difficult time knowing if they had completed an observation, indicating that changes to the interface were potentially necessary. In both the first and second evaluations for the FT, users were not always aware of errors made or how to correct those errors. For the first evaluation, the three items with the highest agreement and the three with the lowest agreement are displayed in Table 2. Notably, several of those items had higher percent agreement during the second field test, indicating that perhaps as the instructors became more familiar with the tool over a longer period of time, they began to feel more comfortable with it. For example, during the second evaluation, 100% of the instructors who answered the question (n = 4) reported knowing when they made an error, as compared to only 29% during the first field test.

Table 2. Responses to Usability Statements during the First Field Test for the ARC-FT

Statement	n	Percent Agreement
I knew exactly where to go to capture an observation.	7	100%
The user interface supported my work style and allowed me to capture data in the way that I think is most effective.	7	100%
It was easy to navigate to different attributes and outcomes.	7	100%
I found that I made errors when I completed an assessment for a student.	6	50%
I hardly ever made an error when using the ARC-FT.	7	43%
When I made an error, I always knew it.	7	29%

The instructors did not rate the Integrated Database as highly as the FT, with only 12 out of 22 items receiving "passing" responses. Two caveats exist in regard to those data. First, the FT was more frequently used by instructors than the database. While the FT was carried by the instructors throughout the course, the database was used only after the data were collected by each rater on the FT. Thus, the database was possibly only accessed once during the class. Further, the instructors appeared to elect an unofficial Database Administrator during the second field test which significantly reduced the number of users. This point leads to the second caveat, which is that, given the low response rate, further testing of the database is needed to better understand its strengths and weaknesses. In addition, the database may be more useful to a course administrator, a senior instructor, or even the commanding officers overseeing the course. The instructors who interact with the students on a daily basis are likely more concerned with day to day activity versus a more integrated view that can be used to make higher level decisions and recommendations.

DISCUSSION

This paper describes research conducted to understand performance assessment challenges in relation to the Army's Human Dimension and Talent Management strategies within one Army course. Based on the research conducted, a performance assessment system was developed and then evaluated to enable instructors to more effectively and efficiently track student progress over time. Overall, feedback on the performance assessment system prototype was positive. ARC instructors reported use of the ARC-PAT and the intention to continue using it, as evidenced by its use at the ARC today. In addition, within the evaluation settings, the instructors reported relative ease of use. In general, the assessment system provided a digitized method to collect and store observations and assessments and to enable tracking of student achievement over time. Hence, this brief study indicated that the concept of a digitized tracking system has potential and appears to be feasible, at least within the institutional Army setting investigated in the work described here.

Based on this work, several key lessons learned emerged that inform use of such an assessment system within an institutional setting more generally. First, a key insight was the separation of observation and assessment functions. Although ideally assessments might be made in real time (to save time and/or be more timely), instructor feedback indicated that such a procedure was not possible or desirable. Simply put, an instructor's job is to teach, not only assess. Accordingly, a tracking tool must allow instructors to remain actively engaged with students and cannot require an instructor to be heads-down filling out lengthy forms. This requirement resulted in the solution demonstrated in the work reported here – instructors can make quick observations (videos, notes, pictures) that are attached to learning objectives (outcomes and leader attributes). Following instruction, at end of the day or at breaks, instructors can view those notes and make more formal assessments. The implication is clear – to be useful, a tracking tool must organize and record observations but not interfere with teacher-student interaction. Second, methods to enable continuity between instructors are essential. On a technological level, such sharing requires networking and access to common databases. However, beyond the technical issue, a critical challenge explored here is how to better enable multiple instructors, who observe the same students, to consistently observe and rate the students' performance. The findings demonstrate that clear observable behaviors linked to leader attributes is an effective way to enable less interpretation of concepts like confidence and initiative that are traditionally challenging to observe and measure. This method facilitated consistency and the development of mutually understandable student assessments across individual instructors. Third, results concerning less satisfaction with the database aspect of the ARC-PAT demonstrate the different assessment uses that are likely present even within one course. While instructors need specific information on individual students readily available, course managers and their commanders need methods to abstract data to determine course level trends as well as the ability to drill down to details when needed for individual students. This finding suggests a requirement for data aggregation tools at different levels of the organization.

Building on these findings, there are several broader implications for the talent management of personnel within the Army and its link to the Human Dimension concept in general. While this research focused on the challenges and requirements of tracking students within a program of instruction, in the end, to enable Soldier-centered development in which experiences are truly tuned to Soldier needs, similar tracking will be required across a career, in and outside of the institutional Army. This challenge implies various critical issues that must be resolved in order for such a performance assessment system to be useful and successful. First, over a career, what should a trend line look like? For instance, presumably Soldiers should become better at problem solving as they progress in rank and experience. However, the problem solving expected for a new Soldier might be very different than that expected from a Sergeant First Class. In the first case, problem solving might be assessed though zeroing of a rifle in basic rifle marksmanship, whereas in the latter, the Sergeant might have to plan training to help his unit achieve objectives. In such a case, a problem solving trend might be flat, for while problem solving skills have improved, the problems have become more difficult as well. As a result, in order to understand trends across a Soldier's career, there must be ways to link to higher level competencies that are normalized in some manner. Second, in a related point, it is worth noting that development may not be linear and may occur in "stages." For instance, many models of human development posit stages where depending on the resolution of observations, development may look flat for periods of time separated by periods of rapid change (e.g., Piaget, 1952). The point here is that trends over a career may not always go up, but that does not mean that development is not occurring. What should Soldier development over a career really look like? The trend may be surprising, showing ups, downs, and little change depending on what and how development is measured. Finally, in a third related point, open questions remain about what level of detail is required for a system that tracks Soldier development over time. At one level, merely knowing that a Soldier participated in an experience (e.g., a mission, a program of instruction) tells you little about what that Soldier actually knows. On the other hand, while precise details on individual skills might be needed for formative feedback within a particular setting, clearly some level of abstraction is needed as one tracks a career. How can the fact that a Soldier needs more experience in problem solving for example be best represented so that representation is useful and actionable for his next commander? Innovations like Experience-API (e.g., Poeppelman, Ayers, Hruska, Long, Amburn, & Bink, 2013) must wrestle with this issue – the challenge lies not only in how to share information but what to share in the first place. The current research findings suggest that even within a course, problem solving must be abstracted in a manner to let other instructors know where to focus, captured in this case in the nature of anchors tied to descriptions of the leader attributes. The findings from ARC, however, also suggest a plausible model for how to begin to tackle this issue. As work continues, critical research and development is required to better understand how Soldiers actually do and should develop, and how best to support assessment through innovative methods and toolsets.

ACKNOWLEDGEMENTS

We would like to thank all of the leadership and instructors at the Army Reconnaissance Course for allowing us to conduct this research and providing valuable input and feedback. Specifically, we thank the ARC leadership as well as the many instructors who served as proponents of this work.

REFERENCES

Bevan, N., Kirakowski, J., & Maissel, J. (1991). What is usability? *Proceedings of the 4th International Conference on HCI.* Stuttgart, Germany.

Department of the Army (2015). *Principles and Functions for the Army's Talent Management Strategy*. Version 4.1. Draft Document.

Department of the Army (2014a). *The U.S. Army Operating Concept*. TRADOC Pam 525-3-1. Ft. Eustis, VA: Training and Doctrine Command.

Department of the Army (2014b). *The U.S. Army Human Dimension Concept*. TRADOC Pam 525-3-7. Ft. Eustis, VA: Training and Doctrine Command.

Knott, C.C., Flanagan, S., Bickley, W.R., Ratwani, K., Dean, C., & Diedrich, F. (2014). An Army Learning Model implementation: Challenges, successes, future directions. In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2014). Orlando, FL

Nielsen, J., & Mack, R. L. (1994). *Usability Inspection Methods.* John Wiley & Sons: New York.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International University Press.

Poeppelman, T.R., Ayers, J., Hruska, M., Long, R., Amburn, C., Bink, M. (2013). *Interoperable performance assessment using the experience API.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Tognazzini, B. (2003). http://www.asktog.com/basics/firstPrinciples.html. Retrieved June 12, 2009.