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

Research Report 1913

Science of Human Measures Workshop: Summary and Conclusions

**Gregory A. Goodwin
Jennifer S. Tucker
Jean L. Dyer
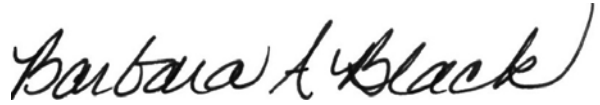Jacquelyn Randolph**
U.S. Army Research Institute

October 2009

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

**A Directorate of the Department of the Army
Deputy Chief of Staff, G1**

Authorized and approved for distribution:

BARBARA A. BLACK, Ph.D.
Research Program Manager
Training and Leader Development
  Division

MICHELLE SAMS, Ph.D.
Director

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) October 2009	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) January 2009 - August 2009	
4. TITLE AND SUBTITLE Science of Human Measures Workshop: Summary and Conclusions		5a. CONTRACT OR GRANT NUMBER W91WAW-09-C-0046	
		5b. PROGRAM ELEMENT NUMBER 633007	
6. AUTHOR(S) Gregory A. Goodwin, Jennifer, S. Tucker, Jean L. Dyer, and Jacquelyn Randolph (U.S. Army Research Institute)		5c. PROJECT NUMBER A792	
		5d. TASK NUMBER 359	
		5e. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute ATTN: DAPE-ARI-IJ P.O. Box 52086 Fort Benning, GA 31995-2086		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Highway Arlington, VA 22202-3926		10. MONITOR ACRONYM ARI	
		11. MONITOR REPORT NUMBER Research Report 1913	

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter POC: Dr. Gregory A. Goodwin

14. ABSTRACT *(Maximum 200 words)*:

The U.S. Army Research Institute for the Behavioral and Social Sciences hosted a workshop on human measurement. The workshop consisted of four panels that discussed assessment of attitudes and aptitudes, mental agility, individual performance, and new training programs. The workshop began with a plenary session with keynote addresses. Each panel was led by a retired general officer and a leading academic or industry researcher. With regard to measuring attitudes and aptitudes, key topics discussed included developing better ways to identify highly qualified individuals from among those who would otherwise be ineligible for service and developing better measures of Soldier and Family well-being. To develop mental agility measurements, panelists suggested building a model based on critical incidents of operational experience, developing measures to assess the critical skills identified in the model, and linking the measures to performance. Regarding the measurement of individual performance, panelists discussed re-scoping initial entry training to train and measure attributes like teamwork, initiative, and accountability in addition to basic combat skills and tasks. Finally, with regard to assessing new training programs, panelists discussed the many challenges in conducting quality assessments of new institutional courses, new equipment training, and unit training. At the end of the workshop, the co-leaders briefed the conclusions of their panels to an invited audience of Army leaders.

15. SUBJECT TERMS

Assessment, attitudes, aptitudes, Soldier well-being, mental agility, cognitive readiness, individual performance, new training programs, new equipment training, unit training

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON Ellen Kinzer Technical Publication Specialist (703) 602-8049
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	60	

Research Report 1913

Science of Human Measures Workshop: Summary and Conclusions

**Gregory A. Goodwin, Jennifer S. Tucker, Jean L. Dyer,
and Jacquelyn Randolph**
U. S. Army Research Institute

**Fort Benning Research Unit
Scott E. Graham, Chief**

**U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926**

October 2009

ACKNOWLEDGMENT

SCIENCE OF HUMAN MEASURES WORKSHOP: SUMMARY AND CONCLUSIONS

<u>EXECUTIVE SUMMARY</u>

Research Requirement:

The U.S. Army Research Institute for the Behavioral and Social Sciences hosted a workshop on human measurement in May, 2009. The purpose of the workshop was to examine how advances in the science of human measurement can better support personnel assessment, training, and leader development. The crux of workshop was four 20-person panels composed of active duty and retired officers and noncommissioned officers, Department of the Army civilians, and researchers from Department of Defense, academia, and industry who discussed measurement needs, adequacy of current approaches, high-payoff approaches, and unresolved measurement issues

Procedure:

The workshop opened with addresses by Army leaders and prominent academic researchers. Following this plenary session, measurement topics were discussed in four panels. The topics of the four panels were: the measurement of Soldier attitudes and aptitudes, the measurement of mental agility, the measurement of individual performance, and the measurement of new training programs. Each panel had two leaders: one, a retired general officer and the other, a leading academic or industry researcher. At the end of the workshop, the panel co-leaders presented the conclusions of each panel.

Findings:

Panel 1 - Assessing Attitudes and Aptitudes: This panel discussed the need for a more holistic selection process that assesses not only what individuals can do (aptitude) but also what they want to do (desires) and what they will do (motivation). Such an approach should be effective in identifying individuals who have a high likelihood of success in the Army, but who would otherwise be rejected based on current selection standards. To develop such assessment tools, it was recommended that the utility of non-cognitive measures be explored. Examples of such measures include the Assessment of Individual Motivation, the Tier Two Assessment Screen, and the Tailored Adaptive Personality Assessment System. The panel also discussed a more systemic approach for assessing Soldier and Family well-being, as to enable respondents to articulate needs for assistance and support. Additionally, the panel discussed the need for methodologies that reduce time needed to transmit survey responses to Army leaders.

Panel 2 - Assessing Mental Agility: The panel suggested that measures of mental agility should be based on a theoretical model derived from both cognitive science and critical incidents of operational experience.   The measures should be linked to performance indicators, and ongoing assessments should be tailored to the needs of Army Leaders. As the Army has already implemented some methods to develop mental agility, a fruitful next step would be to use an existing training approach to develop the model and validate the measures.

Panel 3 - Assessing Individual Proficiency. The panel discussed the need to change the training delivered in initial entry training from one that is only focused on the performance of basic individual tasks to one that is also focused on fostering the development of both Soldier skills and Soldier attributes like accountability, initiative, teamwork, and problem solving. Receiving units of initial entry training graduates have indicated that the development of these attributes is as important as the development of individual Soldier skills. Exemplar measures were described for teamwork and problem solving. The panel also discussed the need to shift away from measuring training inputs and processes (e.g., total number of hours of instruction) to student outcomes (e.g., proficiency levels).

Panel 4. Assessing New Training Programs. The panel focused on three areas where training assessments typically occur. The areas were: new training and education programs, training on new equipment, and unit training. The panel stressed the importance of defining the purpose for which measures are designed, the need for assessment to be multifaceted, and the need to identify the right things to measure. They noted that new equipment training is rarely formally assessed. They also noted the importance of assessing not just whether the Soldiers learned the skills being taught in a course, but whether they can apply those skills in real-world settings. This panel also discussed the difficulty units have training their core METL while preparing to deploy.

Utilization and Dissemination of Findings:

At the conclusion of the workshop, the panel co-leaders from each panel briefed their panel's findings to invited Army leadership including representatives from the Office of the Deputy Chief of Staff, G3/5/7 and G1; the Army Capabilities Integration Center; Accessions Command, and the Office of the Vice Chief of Staff of the Army. After the workshop, a summary of the final briefing was made available to all the workshop attendees and their sponsoring organizations. The workshop informed both researchers and Army leaders. Researchers gained insight into the measurement needs of the Army and Army leaders gained a better understanding of the measurement capabilities available to them.

SCIENCE OF HUMAN MEASURES WORKSHOP: SUMMARY AND CONCLUSIONS

CONTENTS

CONTENTS (continued)

Science of Human Measures Workshop: Summary and Conclusions

Overview of the Workshop

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) hosted a workshop on human measurement in Newport News, VA in May of 2009. The purpose of the workshop was to examine how advances in the science of human measurement can better support personnel assessment, training, and leader development. The crux of workshop was four 20-person panels composed of active duty and retired officers and noncommissioned officers (NCOs), Department of the Army civilians, and researchers from Department of Defense (DoD), academia, and industry who discussed measurement needs, adequacy of current approaches, new high-payoff approaches, and unresolved measurement issues.

The workshop opened with a plenary session followed one and a half days of discussion on the four panels. The topics of the four panels were: the measurement of Soldier attitudes and aptitudes, the measurement of mental agility, the measurement of individual performance, and the measurement of new training programs. Each panel was led by a retired general officer and a leading academic or industry researcher, with an ARI researcher as panel coordinator. At the end of the workshop, the panel leaders briefed the conclusions of each panel to an audience of invited leaders from across the DoD. The list of attendees, panel leaders, and invited guests for the brief-out session can be found in Appendix A.

In this report, the plenary session addresses have been summarized along with the major conclusions of each of the four workshop panels. Opinions, interpretations, conclusions, and recommendations in this report are those of the panel participants and are not necessarily endorsed by the U.S. Army or ARI.

Summary of Plenary Session Briefings

The Science of Human Measures: Anticipating and Meeting Army Needs

Dr. Michelle Sams
Director, ARI

ARI's mission is to improve Soldier, leader, unit, and organizational performance through behavioral and social science research focused on personnel, training, and leader development. This is achieved through interdependent research and development programs that target various aspects of this mission (e.g., leader development, training) as well as supporting programs (i.e., personnel surveys, occupational analysis). ARI fosters research partnerships with academia and industry to help achieve its mission. It also frames its work within the Army's cyclical readiness model of sustain - prepare - transform -reset. While the Army is naturally interested in all aspects of human capabilities, ARI's mission pertains to human aptitudes, attitudes, knowledge, and skills, but not to mental or physical health.

The Army uses human measures to support multiple assessment and prediction problems, starting first with using pre-enlistment aptitude and attitude measures to select and classify new Soldiers who will succeed in the Army. Assessing success through measures of individual Soldier knowledge, skills, and attitudes in turn predicts operational performance of individuals, teams, and units. To support all of this, the Army needs reliable, valid, and practical measures. Practical measures must show incremental value over existing measures, be cost-effective, and be easy to administer and analyze. Summarized below are historical highlights, and current challenges associated with measures related to selection and classification, and training and leader development.

Selection and Classification. Selection and classification measures currently used by the Army include medical examinations, education credentials, criminal records, aptitude tests (Armed Services Vocational Aptitude Battery [ASVAB], specialty tests such as the Defense Language Aptitude Battery), and temperament screens (Assessment of Individual Motivation [AIM] and Tailored Adaptive Personality Assessment System [TAPAS]). Scientific contributions to Army selection and classification needs have been significant, starting most notably with the Army Alpha and Beta selection tests developed to support the massive selection requirement during World War I. More recent examples include computerized adaptive testing in the 1990s and the introduction of temperament measures in the 2000s.

The Army selection and classification system is being challenged by changing demographics in the youth market and the fact that the percentage of the youth population that meets Army enlistment standards is decreasing. While the ASVAB is a highly valid predictor of trainability, knowledge, and job skills, it is less predictive of attrition, positive attitude, and intention to remain in the Army through and beyond the first enlistment term.

There are five selection and classification areas in particular need of research and development at this time.

- Integrated "whole person" assessment of cognitive and non-cognitive attributes

- Assessment of cognitive potential (e.g., mental agility, cognitive complexity)

- Assessment of temperament (e.g., motivation, dependability)

- Improving predictors of best career path and best person-job match

- Improving measures available for validation of selection metrics

The last point deserves special emphasis because identification, development, and administration of suitable indices of Soldier success are required to evaluate promising selection and classification measures.

Training and leader development. Army training and leader development focuses not only on knowledge and skills (e.g., basic combat skills, command and control), but also on attitudes reflected in such areas as Army values and cross-cultural competence. The Army uses a variety of training methods and adapts these to multiple environments, including live and virtual settings. Training success for individual Soldiers ranges from fairly objective indices (e.g., program completion) to somewhat more subjective measures (e.g., leader and peer ratings).

Historical highlights in the advancement of Army training methods through scientific advances include advances in real-time measurement, skill building techniques, and situational awareness and decision-making skills. Over the years, the Army's training requirements have evolved from those needed to support a standardized schoolhouse training model to those that can support the infusion of training opportunities and tools both within and outside of the schoolhouse.

Current challenges in training include increasingly complex missions that result in increasing and ever-evolving training requirements and the Army's high operational tempo (OPTEMPO), which decreases the time and resources available for training. While existing training measures are well-established for some traditional Army skills, such as gunnery time and accuracy, the Army is less well-prepared to measure emerging soft skills such as negotiation and cross-cultural skills.

The following training and leader development areas in which the Army needs improved training methods and measures:

- Turning civilians to Soldiers (basic combat training [BCT]),

- Measuring unit training and collective performance,

3

- Developing leader skills (e.g., influence, team building, complex organizations),

- Developing cross-cultural competence,

- Determining skill decay and methods to sustain skills,

- Rapidly turning lessons learned into training,

- Assessing the tactical employment of new technologies, and

- Effectively using virtual and game-based simulations.

U.S. Army Human Capital Strategy: Challenges for Human Measures

LTG Michael Rochelle
Deputy Chief of Staff, G-1, U.S. Army

The Army is out of balance. The demands of operations over the past several years have led to accelerated equipment wear out and stress on Soldiers and their families. The current organization of the Army will not allow it to respond to further increases in demand. To address this, the Army has developed a plan called the Army Force Generation Process (ARFORGEN). ARFORGEN anticipates restoring balance by 2011 using four imperatives.

The first of these imperatives is *Sustain*. By sustain, the Army seeks to sustain the quality of the All-Volunteer Force – including civilians. This means recruiting and retaining the right people. It also means increasing the quality of life and care of Soldiers, their families, and Army civilians.

The second imperative is *Prepare*. This is accomplished by readying Soldiers and their units to succeed. This, in turn, is accomplished by focusing on force sizing and structure and adapting training at all levels to achieve this goal.

The third imperative is *Reset*. This imperative is concerned with restoring deployed units to a level of personnel and equipment readiness that facilitates training for future missions. We should think of this as *Recover*, *Refit*, and *Renew* Soldiers and their families.

The fourth imperative is *Transform*. The Army faces diverse challenges ahead. To transform the force requires modular reorganization, operationalizing the Reserve Component, restationing our forces, and transforming leader development.

The Army needs a strategy to break-down silos that, until now, have prevented it from developing an enterprise-wide perspective. To that end, the Army has developed a comprehensive human capital strategy (HCS) to help it to better respond to current operations in Iraq and Afghanistan as well as future wars.

This strategy is based on three initiatives. The first is *Competency-Based Occupational Planning*. The second is *Performance Based Management*, and the third is *Enhanced Opportunity for Personal and Professional Growth*. If the Army can successfully develop and implement these initiatives through policies, programs, and processes across the total Army, then it will have the right people with the right skills at the right time and place.

The HCS will be applied across the total Army including the Regular and Reserve Components as well as the civilian corps. At the center of the HCS is making strategic personnel decisions. Recruiting, hiring, assigning, and developing personnel must be made with the larger strategic viewpoint of the Army in mind.

This HCS has a few challenges to address. How can we assess the complete performance potential of individuals? We are not only concerned with Soldiers, but we also must include the

civilian corps, which constitutes a larger portion of the Army now than in the past. How can we measure and develop adaptability and agility (e.g., mental flexibility, cross-cultural skills)? How can we assess levels of competency of individuals to improve training? How can we assess training programs to ensure their success and continued improvement? How can we better assess new/modified training programs to effect continuous improvement?

One other trait that needs to be assessed is *Resiliency*. How do we assess civilians and Soldiers who have the resiliency to make it through a very long and enduring conflict? This trait is vital for Soldiers and civilians to be able to do what our country has asked of them.

Selecting, Developing, and Retaining Soldiers with Heart

LTG Benjamin Freakley
Commanding General, U.S. Army Accessions Command

Because the Army plans to support 10 active divisions throughout the next decade, it is important to expand the measurement of Soldier values to include a focus on *resiliency*. The reality in which the Army currently finds itself is challenging. Soldiers are sent on multiple deployments. Their planned "dwell time" between deployments is often structured with intense training events (which reduces available family time) or is severely reduced when they are transferred to another unit that is about to deploy. This reality puts considerable stress on the Soldiers. Soldiers with greater resiliency are better able to cope with the stress. Particularly with the All-Volunteer Force, motivations and values are critical to continued effectiveness. However, without resiliency, motivations and values will not be sufficient.

Today's society does not really contribute to the development of resiliency. Parents sometimes do too much to shelter their children. The terms "helicopter parents" and "snow plow parents" are often used to depict parents who go to great lengths to remove barriers for their children, even as they approach an age that should require self-sufficiency. However, when injected into a combat environment, Soldiers need to be able to think, respond, and be adaptive in order to be effective. That environment is characterized by an enemy who fights from within the population; by ever-present and unsympathetic media coverage, by a need to collaborate and negotiate with multinational allies, and by technologies that will enhance Soldier effectiveness only if the Soldiers empower the technology.

The challenge is that considering the enlistment standards for education, criminal activity, and physical health, only 28% of young adults (ages 17–24) are eligible to enlist. When the Armed Forces Qualification Test (AFQT) criteria are included, only 20% are eligible to enlist. What is needed is a way to open the gates, so that those who will be good Soldiers despite health, education, and AFQT weaknesses, are qualified for service.

How do we identify those individuals? How do we measure the heart of the Soldier? The answer is not simply in the use of waivers. We should not have to waiver highly qualified individuals; we should be able to select them. This need exists not only for enlisted Soldiers, but also for officers and civilians. Work being done to develop non-cognitive measures and measures like the TAPAS, for example, is adding important predictive value to recruit testing by measuring "will do" rather than just "can do."

We need to look more closely at measuring Soldiers after selection and assignment. Right now we are focused on measuring talent or potential and then developing individuals without any further measurement. Measures need to be developed to help the Army better develop NCOs and officers. Further research will help us to understand whether our measures should be used again at multiple points throughout the Soldier's early development. It will also help us understand performance and what factors lead to good performance. This research will have the added benefit of helping us to reverse engineer to improve recruiting, selection, and training.

These measures will be essential as we attempt to transform Initial Entry Training (IET) to employ a team approach to Soldier development. The focus on outcomes-based training and education, where Soldiers are trained in how to think and how to apply what they have learned, is part of that transformation. Measures will help us to monitor that transformation and fine-tune it to maximize its effectiveness.

Research to better measure the performance of NCOs is needed. There is not a good understanding of who is being promoted to NCO and why. There is not enough work on NCO performance and how it develops. Contributing selfless service (i.e., going to combat) also should be rewarded, as that would serve to reinforce Army values.

We need research on the motivations and values of 12-year-olds today so that we will understand what will motivate them to enlist when they are 17-year-olds. The Army does not know enough about their motivations and values, or about the motivations and values of their parents. Incentives like selecting first unit of assignment may not be an incentive anymore. Research is needed so that we can be proactive and not reactive to the wants of future recruits.

We need to continue to measure values and motivation throughout the careers of Soldiers so that we can do a better job of retaining them. The current set of incentives may not align with their values. Incentives like paying for graduate school, choosing a unit of assignment, changing to a different branch may not be the right incentives. Although cash incentives are currently in use, they are not going to work indefinitely (the bonuses will continue to grow until they are insupportable), and it is not clear what other incentives would be effective.

In closing, if the Army is going to find ways to access, attract, and retain individuals who are currently ineligible to enlist (70% of the youth population) but who are inclined towards enlistment, there will need to be an increasing use of non-cognitive predictors of future performance. Additionally, recognition of the challenges and stress of combat deployments has created a need for research on resiliency and adaptability.

There are two opportunities where research and development on measuring human performance can support the Army's mission:

- providing an understanding of how the transition from the safe environment to the Army environment can best be achieved; and

- providing the tools for selecting, developing, and retaining individuals who are trainable, resilient and adaptive, and have leader potential.

8

Assessment Essential to Training Effectiveness and Readiness

Dr. Eva Baker
Co-director of the Center for Research on Evaluation Standards and Student Testing
at the University of Los Angeles

There are four essential components to effective training assessment: developing measures with good psychometric properties, specifying the purposes of training, specifying what is to be assessed and how, and developing performance-based assessment. These four components of successful training assessment are discussed below.

Psychometrics of assessment. For a measurement to be an effective and compelling indicator of success, it should have empirically supported psychometric properties. A solid assessment should also incorporate content related to training processes and outcomes, although judgment should guide the most useful emphasis given the training program or system at hand. Furthermore, assessments should exhibit good psychometric properties for making decisions at multiple levels, including decisions at the individual, team, unit, program, or platform level. In sum, psychometric rigor is the foundation upon which effective assessment is developed and must be considered across multiple domains.

Specifying the purposes of training. A second essential component of effective training assessment is specifying the purpose of training. Once the purpose of training is specified, one can and must design assessment of training into the training program. Too often, assessment is added as an afterthought. Training is developed before much if any thought is put into how to assess its effectiveness. When testing is developed this way, the assessment is usually trivial and rarely uniform.

Designing assessment, in alignment with training purposes, must be done prior to designing any other aspect of the training program. Training assessment may require very different strategies, depending on whether assessment is for training objectives in schools and on-the-job training, simulations and games, or e-learning. Furthermore, designing assessment into training is critical whether the assessment is at the individual or unit levels (e.g., selection, placement) or at the higher systems or programs levels (e.g., system monitoring, policy formulation, policy planning).

Specifying what to assess. The third essential component of training assessment is specifying what is to be assessed. Regarding critical Army needs, there are five primary components that may be trained, albeit with varying difficulty: domain knowledge, teamwork, problem-solving, communication, and self-regulation. The Army needs to look at each of these components and embed them in the relevant domain(s) or context(s). Embedding these in the relevant domains is important because this facilitates generalization of learning across situations and problems. In essence, we should be able to use a given assessment and/or training program, regardless of how the problems in theatre morph, because we have specified critical skills that are trained such that they transfer (or generalize) across situations.

There are several critical 21st Century skills that the Army should and can assess. These include adaptive (or agile) problem-solving, situation awareness and risk assessment, decision-making (i.e., the skill to compile information and make decisions under stress), self-management (i.e., the skill to maintain focus in the face of stress), teamwork, learning to learn (i.e., the skill to figure out what one needs to know and how to acquire that information), communication, and, most critically, the application of knowledge to new settings and situations.

Cognitive readiness is a higher-level skill that requires several of the critical 21st Century Army skills. Cognitive readiness is defined as "the mental preparation an individual needs to establish and sustain competent performance in the complex and unpredictable environment of modern warfare" and is, therefore, one of the most important skills needed in the current Army (Fletcher, 2004). Reacting to the aftermath of an improvised explosive device (IED) deployment is an example of a common situation that requires adaptability and cognitive readiness. How does this event change strategy? How do the personnel (e.g., officers, Soldiers) decide how to interact with each other and respond to the IED deployment? It is precisely the frequency of complex and changing events such as IED deployment that makes cognitive readiness such a critical skill.

We must go further than identifying critical Army skills to also consider the trainability of the various facets of these skills. Some of these facets may be relatively easy to train while others are more difficult to train. For example, cognitive readiness factors that are difficult to train include adaptive expertise and critical thinking, whereas factors that are relatively easy to train include adaptive problem-solving, teamwork, and metacognition. An essential step in the design and assessment of training is to focus on those factors that are trainable.

Performance based assessment. The final essential component of training assessment is considering the level at which training is assessed. Assessment should include the performance of trained skills. Though this seems rather obvious, performance-based assessment can be difficult to execute. For this reason, it is sometimes excluded in favor of more proximal assessments of training (e.g., trainee satisfaction with training program). In the case of complex, adaptive skills, such as cognitive readiness, the assessment procedures must be challenging and complex to reflect the complexity of the skills being trained.

Although embedding challenge and complexity into assessment can be difficult, there are several guidelines that should be considered to enhance the complexity of assessment. First, assessments must incorporate realistic situations. Second, assessments must use situations with changing components. Simulations and field exercises are good examples of assessments that follow these guidelines. Furthermore, technology and tools that can model complex skills (e.g., integrated mapping design) will be important in assessing complex skills from a performance perspective. Clearly, it will be essential to move beyond typical check lists or After Action Reviews (AARs) to more sophisticated criteria of training performance

In closing, training programs will not be effective unless assessment includes these four critical components. Effective training is critical to force readiness. If training is to produce Army personnel with the necessary 21st Century skills, personnel must be not only technically prepared but also prepared to react to complex targets, adapting their performance as conditions

constantly shift. Finally, the Army must produce evidence for assessment effectiveness at various levels, including individuals and larger units, as well as for different purposes and situations.

How we Know what we Know about Soldiers' Attitudes and Behaviors

Dr. David Segal
Director of the Center for Research on Military Organization at the University of Maryland

The Army got into the personnel assessment business during the First World War, but it wasn't until the Second World War that Army leaders realized the importance of social research in revealing information about Soldier attitudes. The science of survey development was not very far along at that time, however. Problems included non-random sampling of survey recipients and the presence of strong demand characteristics.

One particularly bad example of sociological research from this period is that of Samuel Marshall (Chambers, 2003). Marshall reported that less than 25% of US combat Soldiers fired their weapon in battle despite being in direct contact with the enemy. Marshall argued that the Army should revise its training to increase the willingness of Soldiers to engage the enemy. Subsequent researchers have raised questions about the empirical support for this statistic (e.g., Chambers, 2003; Spiller, 1988). Specifically, Marshall's data were collected through after-action, group interviews with enlisted men, which would sometimes occur weeks after the event. In addition, there was no evidence of any statistical analysis nor did Marshall keep field notes. This absence of any evidence has raised significant doubt about the accuracy of Marshall's conclusions.

Over the years, social and behavioral research methodology has come a long way. Sampling and survey techniques have advanced, as have the statistical tools available to understand and interpret the data. Nevertheless, despite the sometimes less than optimal research methodology employed by early behavioral scientists, it is important to examine the historical trends in attitudes and opinions of Soldiers. Clearly, any such examination should look carefully at the research methodology used by all researchers.

Early research on cohesion. Samuel Stouffer examined combat motivation following WWII and published his findings in two volumes referred to as the American Soldier studies (Stouffer, Lumsdaine, et a.,1949; Stouffer, Suchman, et al., 1949,). The American Soldier studies surveyed about a half million veterans of WWII. Although the sampling techniques were not perfect, the findings shed light on the role of cohesion during WWII. Interestingly, the conclusion most often drawn from his work is often taken out of context. Stouffer's results are often quoted as showing that Soldiers fought for the solidarity of the unit rather than for idealistic causes. What often fails to get mentioned is that solidarity of the unit was not the most cited reason. In fact, the data show that for NCOs, the primary motivator was actually very pragmatic to get the job done and go home. Interestingly, officers thought that it was their leadership that kept the Soldiers going.

Roger Little examined the role of cohesion in combat units during the Korean War. He was embedded with an Army Rifle Company for over a year (November, 1953-February 1953), and he wrote a book entitled "Buddy Relations and Combat Performance," published in 1964, that focused on the importance of the two-man buddy team. He found that it was these dyadic

relationships that were the building block of the social fabric of the Army rather than larger cliques. He wrote that although often at odds with authority, the buddy system increased effectiveness by establishing boundaries on acceptable performance. He also noted that when officers were serving on the line with subordinates, they developed greater solidarity with their men and subsequently also supported deviations from the norm of the larger organization while still remaining loyal to that organization.

Early research on leadership. General Eisenhower was a major consumer of the research of behavioral science. One of the things he was interested in was how Soldiers felt about their leaders because he was commanding a conscript Army. He asked the American Soldier team to conduct a study. The team asked a sample of 2,827 enlisted men about their attitudes toward their leaders. The findings revealed general dissatisfaction in the Officer corps. As a result of these findings, Eisenhower wrote a letter to MG Taylor, superintendent of West Point, instructing him to add courses on practical behavioral science to prepare cadets to be better leaders. Subsequently, West Point started teaching leadership as an academic subject.

Also at that time, the Board on Officer-Enlisted Man Relationships published a report, known as the "Doolittle Report," that came to similar conclusions. In it, the board found that the causes of poor relationships between commissioned and enlisted personnel were traceable to two main factors: undeniably poor leadership on the part of a small percentage of those in positions of responsibility, and a system that permits and encourages a wide official and social gap between the commissioned and enlisted personnel. The Board recommended that every officer candidate (regardless of commissioning source) be instructed in command responsibility, personnel management, and human relations (Stouffer et al., 1949).

Early research on race relations. Samuel Stouffer also (1949a, 1949b) investigated questions about attitude on the racial integration of Army units during WWII. It is not clear whose idea it was to ask Soldiers what they thought about being integrated. One view was that it was the idea of the researchers. The other view is that Army Leaders did not want to integrate the Army and hoped to find data to support abandoning integration. In fact, the Army fought to maintain racial segregation and used the data to show that white Soldiers preferred to serve in segregated units.

The responses of veterans of this era reflected differing attitudes from blacks and whites. Whites predominantly favored a segregated Army, whereas blacks were as likely to favor integration as segregation. By the time of the Korean War, survey results showed that whites were evenly divided between favoring integration and segregation, whereas blacks strongly favored integration (Bogart, Leo, et al., 1969).

In conclusion, the history of social science research in the military reveals some important lessons:

- The results of this research can have important implications for military policy.
- Research findings are not always aligned with expectations, common opinion, or the desires of policymakers.

- Careful data collection and documentation are critical to preserve the legitimacy of the research.
- It is important for military sociologists to go where the Soldiers go to better understand the issues that affect them.

Summary of Panel 1 Discussion: Assessing Attitudes and Aptitudes

The purpose of this panel was to discuss measures and methods for assessing aptitudes and attitudes for four key areas: enlisted and officer characteristics, enlisted and officer selection standards, Soldier and Family well-being, and Army commitment and retention. Because of the breadth of the topics, the discussion on aptitude measures was focused primarily on enlisted Soldier characteristics and enlisted Soldier standards and the discussion on attitude measures was focused primarily on Soldier and Family well-being.

Assessing Aptitude

The military services have valid and reliable measures of applicants' cognitive aptitude ("can do"), but lack systemic ways to identify applicants' work-related desires ("want to do") or motivation ("will do"). For nearly a century, the military services have used measures of cognitive aptitude to identify qualified applicants. The ASVAB was introduced in 1976 as a set of cognitive tests designed, in part, to assist in matching recruits to the military occupations for which they were best suited. A portion of the ASVAB, known as the AFQT, is used for selection. The ASVAB tests have been found to provide reliable and valid (when correctly normed) measures of recruit potential. However, these measures have not been as powerful in predicting recruits' motivation to perform or person-job fit, both of which may be predictive of successful performance.

After years of war in Afghanistan and Iraq, and when the majority of American youth fail to meet minimal DoD or service entry requirements, the Army has been prompted to reconsider its definition of "quality" and has initiated research on the development of measures that would enable the identification of potentially successful applicants among those who might otherwise have been denied entry. Historically, the Army defined quality enlisted applicants as those with a high school diploma who scored in the top AFQT categories (Cat I-IIIA). This definition of quality is based entirely upon cognitive aptitude and academic success. Recent research has demonstrated that other non-cognitive or personality measures which predict applicants' motivation to perform, training performance, or person-job fit are also useful indicators of quality and would enhance the Army's ability to identify quality applicants. Such measures tap applicants' "will do" or "want to do" qualities.

Several such measures are either in use or in development. The AIM has been used by the Army for several years to identify quality applicants among non-high school graduates. The Tier Two Attrition Screen (TTAS) enables the identification of non-High School graduates who have a high potential for adapting to Army life and completing their initial term of enlistment. The TAPAS measures 13 facets of personality which generally represent the "Big 5" dimensions of personality (openness, extraversion, agreeableness, conscientiousness, neuroticism/emotional stability). Initial research has shown that TAPAS reliably and validly predicts "can do" (by predicting Advanced Individual Training grades, training graduation, and job knowledge) and "will do" (by predicting APFT scores, job effort, likelihood of indiscipline, and attrition). TAPAS can be used to screen out applicants with low motivation and/or a high likelihood of attrition while screening in highly motivated applicants who, based on research results obtained

in IET, perform at least as well as applicants in the next higher AFQT category. Limited implementation of TAPAS at Military Entrance Processing Stations began in May 2009.

In addition to the above, there are other non-cognitive measures that have the potential to improve the Army's ability to identify quality applicants. Biodata (which reflects an individual's background and experience), situational judgment tests, and measures of attitude may further improve the Army's ability to identify applicants who are likely to successfully complete training and become quality Soldiers. Other non-cognitive dimensions which may impact performance but require further investigation include measures of team orientation and measures which predict resilience.

In sum, the use of non-cognitive measures like AIM and TAPAS to supplement existing measures of cognitive aptitude such as the ASVAB have enabled the Army to expand its recruiting pool by identifying potentially high-performing applicants who might otherwise have been screened out. The panel also noted that current methods of identifying job requirements are limited and inefficient for selection purposes. Improved front end analyses are needed to support development of outcome measures which are in turn needed to determine the effectiveness of the selection and classification measures. Improved person-job match methods are also needed. These methods should incorporate non-cognitive as well as cognitive measures. Interest inventories may be particularly useful for person-job matching.

Assessing Attitudes

Soldiers' and Family members' perceptions of and satisfaction with life and work in the Army influence key outcomes such as career plans and commitment, yet a comprehensive model of well-being remains elusive. Soldier and Family well-being are critical because: (1) DoD policy requires that personnel and their families be provided a quality of life that reflects the high standards and pride of the Nation they defend, and (2) aspects of well-being are predictive of critical outcomes such as retention, deployability, and accession.[1] Although well-being has been defined as a state of physical, mental, and social health, research within DoD and the services has primarily focused on only limited aspects of well-being. For example, much research has focused on the impact of combat on the clinical component of well being (e.g., see Office of the Surgeon Multinational Iraq, and Office of the Surgeon General, 2006). What is needed is research that spans physical, clinical, and social components of well-being to create a more comprehensive model for decision-makers.

In both the Army and DoD, there are ongoing efforts are underway to integrate attitudinal (i.e., survey), objective, and clinical data into a "dashboard" that can be used to broadly indicate Soldier and Family well-being. Such efforts may include attitudinal measures (such as, satisfaction with family support services, satisfaction with the Army way of life, and spouse support for retention), objective measures (such as, divorce rates and suicides rates) and clinical measures (such as, experiences or rates of Post Traumatic Stress Disorder or Traumatic Brain Injury), all of which are expected to be generally indicative of well-being.

---

[1] Well-being may indirectly affect accessions through its influence on the likelihood that current service members and their Families will recommend the Army/military to others.

It will be important for these efforts to also consider the importance of Family well-being to the well-being of Soldiers. DoD policy directs that family support systems be responsive to the needs of service members and their Families. Surveys done to date contain an abundance of questions on Soldier and Family awareness of and satisfaction with various family support services, but lack detail about the specific kinds of assistance families actually need. Moreover, existing surveys have not investigated how Soldiers and Family members make decisions about which resources they will use to meet their needs. And, although some decisions about support services may be made at the installation level to reflect what Soldiers and Families need at a particular geographic location, Army-wide surveys are not designed to inform installation-level decisions. Consequently, existing survey data may not provide "actionable" information and may have limited utility for policy makers who must make resource decisions. Instead, survey data may be simply viewed as one of many factors to consider when broadly assessing Army support for Soldiers and Families.

To yield data on family well-being that can inform decision making, the research process must begin with an explication of what it is Soldiers and Families say they need. To this end, a survey could be fielded that first asks respondents to identify specific problems/needs, and then uses a series of cascading lists to enable respondents to "drill down" and, for each problem/need, identify available resources, reasons why particular resources were used/not used, satisfaction with used resources, and projected outcomes. This type of approach may prove useful in future efforts to more fully understand Soldiers' and Families' needs and the cognitive processes underlying decisions about how to meet those needs.

Another challenge that potentially reduces the utility of attitudinal measures to inform decision making is the length of time required to plan, collect, and analyze survey data in accordance with sound scientific methods. Policy makers need timely information to inform decision making, yet survey findings may not be available when they could be of greatest use. Convening panels in which selected panel members agree ahead of time to respond to surveys may be one way to reduce the time needed for planning and data collection while still assuring representative samples.[2]

Panel Conclusions and Recommendations

Assessing Soldier Aptitudes. Cognitive ability, non-cognitive attributes, skills and knowledge obtained through training, and experience together contribute to what makes a "quality" Soldier. Although reliable measures of cognitive ability have been used for decades, the Army has only more recently begun the development and limited implementation of non-cognitive screening measures. Other concerns involved the availability of useful front end analysis information, the need for institutional outcome measures and improved person-job match methods, and the need to predict team performance and resilience. Therefore, it is recommended that the Army:

---

[2] The Defense Manpower Data Center has largely avoided the use of panels. However, as part of Project First Term, ARI established a panel of about 70,000 Soldiers who were followed longitudinally throughout their first term.

(1) Take a holistic approach when accessing applicants for entry by supplementing indicators of cognitive aptitude with indicators of applicants' desire and motivation to succeed.

(2) Accelerate implementation of TAPAS to supplement the ASVAB.

(3) Develop improved job analysis, performance measurement, and person-job match methods.

(4) Investigate the use of interest inventories for classification.

(5) Examine the meaning and prediction of resilience and team orientation.

(6) Develop an Army-wide integrated database that contains attrition, training, and job performance data.

Assessing Soldier Attitudes. The Army and DoD have found that retention and other important outcomes are influenced by perceptions of or satisfaction with aspects of Army work and life. There has also been recent interest in integrating attitudinal, objective, and clinical measures to reflect the well-being of the force. However, survey data assessing satisfaction with support services may provide limited insight into Soldiers' and Family members' actual problems/needs, as well as why and how they attempt to address these problems/needs. Therefore, it is recommended that the Army:

(1) Continue efforts to develop a comprehensive measure of well-being.

(2) Continue examining the predictors and correlates of Soldier and Family well-being.

(3) Design survey items and instruments which enable respondents to articulate their particular needs for assistance/support.

(4) Explore methodologies that reduce the time required to plan, collect, and analyze survey data, thereby enabling the timely transmittal of survey research to Army policy makers.

Summary of Panel 2 Discussion: Assessing Mental Agility

The stated purpose of Panel 2 was to discuss techniques for assessing mental agility and cognitive readiness. Discussions were to include the practical utility of such measures, the criteria against which to validate such measures, and how to scale them to various echelons. As the purpose for Panel 2 was quite broad, the panel leaders focused the discussions by first asking that the panel consider in their discussions several challenges in assessing mental agility to include

- defining "mental agility" and related key terms with an emphasis more on cognitive skills and less on emotional factors,
- reviewing current assessment techniques and available Army databases,
- determining Army needs for assessment,
- considering contextual factors (echelons, grades, climates),
- reviewing processes for developing mental agility, and
- suggesting high-payoff approaches to address Army needs.

Second, the panel leaders suggested that the panel use the following statement, "flexibility of mind, a tendency [*capacity?*] to anticipate or adapt to uncertain or changing situations" (U.S. Department of the Army, 2006) as a working definition of mental agility and its related key terms.

Third, the panel leaders divided the panel into three groups; each was assigned both general and unique questions to discuss. Group A answered questions relating to the topic of *what is to be assessed / what is mental agility.* Group B answered questions relating to *how to assess mental agility*, and Group C responded to questions of *how the Army can use assessments.* A summary of the panel's discussions is presented below (topics are not presented by group).

Army Problems/Needs

It was concluded that the long term health of the Army depends on retaining and developing individuals with a high level of cognitive skills. There needs to be an increasing capacity to deal with the complexities and uncertainty of the military operating environment and change at all organizational levels.

As mental agility has been identified as a critical factor in determining individual potential for effective Army service, strategies are needed for developing this cognitive skill. To do this, a better understanding is needed of the skills and abilities associated with mental agility. Many different types of cognitive (e.g., systems thinking, pattern recognition, situation awareness, critical thinking, spatial, creativity, hardiness, metacognitive), social (e.g., interpersonal), and affective (e.g., emotion regulation) skills and abilities are likely related to (indicators of) mental agility. Thus, a theoretical model is needed that will describe these relationships and guide the development of measures to assess and improve the effectiveness of strategies for individual and group development.

Adequacy of Current Approaches to Meet Army Needs

Currently, no single instrument or technique is available to assess mental agility. Further, there is no comprehensive database that tracks mental agility levels across the force or across the lifecycle.

For the purposes of this panel, the doctrinal definition of mental agility was used as the basis for the discussions. However, there are many overlapping terms in use in Army doctrine and research related to mental agility - adaptation, creativity, etc. Although some data have been collected on these different skills and psychological constructs, limited empirical work has examined their associations with mental agility.

New High Payoff Approaches

First, develop a model that is tailored to the Army needs of assessing and training mental agility. The best approach is to derive the model from two directions – bottom up and top town. The bottom-up perspective is needed to understand the skills related to mental agility; it builds on Soldiers' experiences by collecting critical incidents from operational experience. The top down perspective is needed to leverage findings from cognitive science theory.

The importance of the model is to identify the related components of mental agility. A broad sample of experts is needed to provide examples (critical incidents) of mental agility; experts need to explain why a particular skill, ability, or behavior was agile (e.g., examples of systemic thinking or creative problem-solving). Then, these incidents could be compared to a baseline or normative behavior to determine the degree of agility Soldiers are exhibiting. Although there are individual differences in cognitive ability, personality traits, and emotional attributes related to mental agility, the focus of this research should be to identify trainable skills associated with mental agility.

It is important to note that mental agility is only one of many other important cognitive skills. Leaders and Soldiers need to use judgment, critical thinking, etc. Data from critical incidents would offer researchers and Army leaders a better understanding of the skills and attributes necessary for effective performance.

A model of mental agility also should describe the nature of the relationships between breadth of experience and mental agility. Expertise likely plays a large role in the development and manifestation of agility; however, research is needed to determine whether a person can be "universally agile" and whether adaptive expertise crosses domains. Further, the model should inform Army leaders as to whether exposing Soldiers to broad experiences (e.g., varied deployments and/or leadership experiences) likely leads to increased agility, and if so, whether linking these experiences to certain points in Soldiers' careers is critical for developing agility.

Second, develop mental agility measures to assess the critical skills identified in the model. A good assessment of mental agility will involve a battery of tests. A matrix was proposed as a framework to determine which types of assessments would be most appropriate for different purposes. For example, many different types of measures could be developed (e.g.,

biodata measures, assessment centers, dynamic testing, simulations); however, they all would not be appropriate for selection, training, self-development, etc. A complete matrix of all possible metric types and their uses would provide researchers and Army leaders with a better understanding of what is available and appropriate for use in different assessment contexts.

Third, establish a link between the mental agility measures and performance. This is an important step for two reasons. First, to effectively use the measures for selection or development purposes, the measures need to be validated. The validation criteria (performance measures) should be selected based on a consensus of Army experts, and the research should demonstrate that changes in performance are due to differences in mental agility skills. Second, mental agility assessments that are coupled with development strategies are better able to demonstrate added training value and show that the measures actually matter – to answer the "so what" question.

Fourth, after the validation research, ongoing assessments should be tailored for specific career points. The assessment of mental agility should occur throughout the career of a Soldier so that changes and/or improvements in mental agility can be determined. As success at different levels might look different, the performance criteria associated with mental agility skills need to be determined for a range of career stages and Military Occupational Specialties.

Other Topics Discussed

The panel recognized some practical concerns regarding the implementation of widespread cognitive skills assessments. Specifically, the acceptance of widespread assessments of mental agility skills and the development of mental agility training programs has both logistical and cost implications. As it is a zero-sum game, Army leaders will need solid research findings that mental agility is a critical skill for effective performance to devote resources for these research endeavors.

Researchers and Army leaders also should consider the effects of organizational culture and climate on the application and development of mental agility. The local command climate can enhance or impede the development and application of mental agility to a significant extent. Although command climate clearly has a potent effect on mental agility, no program within the Army educational system focuses on building the command climate. Training programs within the schoolhouse are needed to teach leaders how to build effective climates and, in particular, climates that foster mental agility.

The panel concluded that entry levels of intellectual talent (to include mental agility skills) are adequate for meeting the 21st century operational demands on performance. The concern is how to best develop and retain individuals with high potential so that the force is equipped to handle the increasing complexities of the current operating environment COE.

Panel Conclusions and Recommendations

First, conduct a baseline assessment of the measurement techniques, instruments, and data related to mental agility currently available within and outside the Army. The assessment should include descriptions of the skills and abilities being measured and their expected associations with mental agility as well as information regarding relevancy of the measures to the Army and the resources needed for development and force-wide implementation. For example, although simulations provide a more realistic context than paper/pencil measures for scenario-based assessments, they may not be practical for large-scale assessment efforts.

The assessment also could include discussions with Army leaders at each echelon to find out what questions they want answered related to the measurement and training of mental agility skills throughout the force. For example, if Army leaders suggest that the use of the measures is for wide-scale accessions, then a more cost-effective approach would be needed. On the other hand, if the focus is on the development of mental agility in smaller groups, then more resources could be allocated to the development of the measures and training. Further, the findings could be used to select the specific indicators of mental agility on which to focus the efforts.

Second, use an existing training approach (that is a good fit to explicitly focus on the development of mental agility), such as an outcome-based training and education method or the small unit center of excellence approach, to develop the model and define and validate mental agility measures. The measures reflecting key mental agility constructs need to be based on the critical incidents and the model. Then, the face validity of the measures can be systematically evaluated using a panel of experts. Leveraging existing programs can be an effective way to both reduce the costs associated with the development of mental agility measures and training and gain high-level support for the research efforts.

Summary of Panel 3 Discussion: Assessing Individual Performance

Panel three discussed expanding measurement of Soldier proficiency in IET to include developing measures of intangible attributes and characteristics while refining task proficiency measures. The major conclusions of this panel are summarized below.

Measuring Soldier Attributes in Initial Entry Training

IET prepares new Soldiers for their first unit of assignment by giving them a basic set of knowledge skills, and abilities (KSAs). At the conclusion of the BCT portion of IET, Soldiers are proficient in using their assigned weapon and have demonstrated a minimum level of physical fitness. Additionally, they have received training on a wide range of topics including land navigation, first aid, hand to hand combat, drill and ceremony, and the Army values.

With few exceptions (e.g., rifle marksmanship and physical fitness), most outcomes collected during IET reflect easily quantifiable measurements such as, attendance, task completion, and hours of instruction rather than measures of task proficiency. Thus it is not known whether proficiency is attained on many KSA. Furthermore, there is somewhat limited development of Soldier attributes[3] like confidence, initiative, accountability, teamwork, and problem solving.

It was the opinion of those on the panel who have been involved in overseeing and delivering training in IET, that today's newest Soldiers are capable of more than is currently expected of them. Training and measurement should push Soldiers to demonstrate high levels of confidence, initiative, accountability, judgment and problem-solving ability in addition to greater proficiency at traditionally trained KSAs.

Placing greater emphasis in IET on the development of the Soldier attributes will require careful consideration of which attributes to emphasize. It will take time to develop a list that is broadly agreed upon; however, there was consensus that the list of attributes developed by the Directorate of Basic Combat Training at Fort Jackson is a good starting point. That list appears below:

- A proud team member, possessing the character and commitment to live the Army values and Warrior Ethos.

- Confident, adaptable, mentally agile, and accountable for own actions.

- Physically, mentally, spiritually, and emotionally ready to fight as a ground combatant.

---

[3] There was some discussion of what term would best describe these characteristics. A term that has often been applied is *intangibles*. However, several panelists said that this implies that the characteristics are unknown and even unmeasureable. Eventually some consensus arose that *attributes* was a suitable term, therefore the term "attributes" is used to denote characteristics like confidence, initiative, accountability, etc. In measurement language these may be considered latent or unobserved variables associated with broad internal states or desired characteristics.

- Master of critical combat skills and proficient in basic Soldier skills.

- A self-disciplined, willing, and adaptive thinker capable of solving problems commensurate with position.

Training on these attributes in IET cannot be developed until there are reliable, valid, and practical measures of them. Measurement instruments are necessary to allow training developers to determine whether training is effective, to enable instructors to provide feedback to individual Soldiers, and to fine tune the training techniques of instructors.

In the next section, an approach to measurement design is described. This approach, known as Evidence Centered Design (ECD), sees measures as more than questionnaires or tests and offers a framework for conceptualizing assessment and measurement that should be particularly useful for IET.

Evidence Centered Design

ECD was developed to provide a modern language for talking about assessment and measurement. Traditional concepts of assessment are centered around questions and testing. The developers of ECD saw this as a very limited means of assessment and wanted to create a comprehensive descriptive framework for conceptualizing it (Behrens, Mislevy, Bauer, Williamson, & Levy, 2004).

One of the key assertions of ECD is that assessment and measurement are distinct processes. Assessment is conceptualized as the end and testing, evaluating, and measuring are possible means to that end. Whereas assessment is a process of characterization (i.e., determining that someone is an expert or is qualified to do something) measurement is a process of quantifying traits, skills, aptitudes, etc. This view holds assessment as a broader process than "testing" which denotes the creation of specific tasks and circumstances to elicit

Assessment can depend on formal measurement or tests but also can be fed by informal observations. For example, a squad leader who has been working and training with his squad for a year probably has a very good understanding of (i.e., assessment of) the strengths and weaknesses of every member of the squad without having done any formal testing. On the other hand, formal measurement is necessary when numerous assessments must be done in a relatively short time-frame, such as at the end of a course.

ECD works backwards from the claims one wants to make about those being assessed. For example, to make claims about Soldiers' marksmanship abilities, statements need to be developed that describe what individuals at different marksmanship skill levels can do. Next, the evidence needed to make those claims must be identified. The final step is to determine how that evidence would be observed. In other words, decide what tasks, tests, etc., the person being evaluated must do to produce evidence to support claims about his or her ability.

Assessment under ECD involves four processes. First, a measurement/testing activity is selected. The activity can be formal or informal, specific or general. The activity should be

chosen because it is expected to provide evidence necessary to make an assessment. The second process is presentation whereby the examinee produces what is called a work product. The work product may be responses to a questionnaire or performance of some task. It might be the result of a contrived activity or the result of actual job performance.

The third process is evidence or feature identification. Here, the work product is examined and evaluated. The evaluation may be a simple correct/incorrect judgment or a complex multidimentional evaluation. For example, an operations order is likely to be evaluated on many dimensions such as its consideration of weather and terrain, effective fire control measures, synchronization of assets, etc.

In the final process, all the evidence is evaluated according to a measurement model which weights the data from the various measurements. If it is determined that further data is needed before an assessment (decision regarding Soldier abilities) can be rendered, then additional measurement/testing activities may be scheduled.

This four process model can be applied in a variety of settings including the administration of paper and pencil tests, performance in a simulator, and on-the-job assessments. In fact, modern portable electronic devices such as cell phones and personal digital assistants open up remarkable opportunities for measurement and assessment. These ubiquitous presence of these devices creates opportunities to quickly gather data from large samples almost anywhere.

This model also indicates that measurement instruments may be a set of questions or individuals like instructors or leaders. Just as it is important to determine the validity and reliability of paper-and pencil tests, it is important to determine the validity and reliability of instructor- (e.g., drill sergeants) and leader-ratings and/or observations.

Changing Measurement and Training in IET

Any attempt to change IET to enhance the development and measurement of selected Soldier attributes will face a number of challenges but also will have some opportunities. As already mentioned, there is no universally agreed upon list of attributes that should be emphasized in IET. Nor is there agreement on the appropriate level of training that should be done. There is no clearly defined training methodology that specifically addresses many of these attributes, and measuring them has always been difficult.

There are also cultural and institutional impediments. Specifically, instructors are trained to impart knowledge and skills, but they have little training on imparting attributes, so cadre training will also need to be changed. Current training is also very process- and resource-focused rather than being learner- and outcome- focused. The training focus must be driven by the outcome of the training rather than the process of training. Instructors will need to be trained to adapt their training to insure that proficiency is the critical measure of success rather than the

delivery of so many hours of instruction, for example. But, this presents yet another challenge to changing IET, that of predicting the resource requirements of this new approach to training. [4]

On the other hand, several opportunities exist for the adoption of this new approach. First, the Directorate of BCT at Fort Jackson has recognized a need to better develop these attributes in IET Soldiers as evidenced by the list presented earlier. Second, techniques for training these attributes during rifle marksmanship instruction have been developed by the Asymmetric Warfare Group. These techniques can be used as a model for training on other KSAs. Third, many combat-experienced drill sergeants recognize the importance of developing these attributes in the Soldiers they are training. Fourth, volunteer Soldiers are highly motivated and eager to be led. Finally, Army schools have a culture of being willing to examine programs and processes and make changes when needed.

Failure to identify specific and general attributes required of entry level Soldiers and to modify training to ensure those attributes have been instilled risks graduating Soldiers who may have specific KSAs but lack essential Soldier attributes. It was the consensus of the panel members that unless defined and systematically measured, the desired Soldier attributes will only be developed in an uneven, sporadic way.

Developing Measures for Soldier Attributes

Developing reliable, valid, and user friendly measures of Soldier attributes will take time but the process must begin by addressing three fundamental questions:

- What is the purpose of the measure?
- What are the levels of this attribute, and to what degree of precision do they need to be measured?
- What are the key performance indicators that characterize a Soldier at each level?

Knowing the purpose of the measurement is important because it is unlikely that any one measure will be optimal for all situations. For example, a six-hour long questionnaire might allow fine distinctions to be made between individuals, but this would be an impractical measure if the purpose of the assessment was only to provide quick feedback to Soldiers during training. Purposes of measurement during IET include determining eligibility for graduation, providing developmental feedback, selecting individuals for special assignments, or providing instructors with feedback regarding their training effectiveness.

Determining the precision of the instrument is important because a precise instrument is apt to require significant time and other resources. Although it might seem desirable to measure every attribute as precisely as possible, this is not always practical. The precision of the

---

[4] Although it is difficult to predict the resource requirements needed to train the specified attributes , it was the experience of panel members who had implemented some form of this training at Ft Benning, Ft Jackson and USMA, that current resources are generally more than adequate. With training time being the most critical training resource, it was panel members' experience that simply making clear to the trainers that an attribute would be measured was most often sufficient to ensure that it was trained and reinforced adequately during the conduct of essential skills training.

measurement should be only as high as is the purpose of the measurement dictates. For example, a lengthy battery of tests might be appropriate for selecting individuals to an elite unit, but it would be impractical in other contexts.

Once the levels of the attribute and precision of the measure are established, key performance indicators of those levels can be identified. Experienced instructors and other subject matter experts should be used to develop potential performance indicators. In the language of ECD, this has to do with identifying the claims one wants to make about individual performance at each level and which evidence to collect (performance indicators) to support those claims.

With performance indicators identified, measurement instruments can be empirically developed, validated, and refined. Research and experience will determine the performance indicators that are the most indicative of the attribute being measured, the most reliable ways to record those behaviors, and the conditions necessary for observing them. As noted above, training also will need to be developed to insure that Drill Sergeants and other evaluators know how to use the measures that are developed.

Prototype Measures for Selected Soldier Attributes

To illustrate how these three fundamental questions might be answered, examples based on two Soldier attributes are described below. The first Soldier attribute discussed is teamwork and the second is problem solving.

Teamwork is a critical attribute for any Soldier and should be developed throughout his or her career. To train or develop teamwork (or any other attribute) in IET, it is not necessary to add training time. Instead, Drill Sergeants should be shown how to observe teamwork in all of a Soldier's activities. For example, on the rifle range, in the dining facility, in the barracks, on a road march, does the Soldier encourage or put down his/her fellow Soldiers, does he/she try to enhance unit performance or only his/her own performance?

The panel identified multiple purposes for measuring teamwork including providing feedback to cadre, helping individual Soldiers become better team members, and serving as a graduation requirement. Each of these purposes requires a different level of precision. As a graduation requirement, only the worst performers need to be identified. On the other hand, if the purpose is to provide feedback to the cadre or for individual Soldier development, the feedback should be more nuanced yet could be informal and qualitative.

There are many performance indicators for teamwork in IET. One example is during rifle marksmanship training when a Soldier serves as a peer mentor. Drill Sergeants should note whether the mentor takes his job seriously, pays attention to the behavior of the shooter, and provides encouraging feedback to the shooter. A second example is during formation, Drill Sergeants should notice whether Soldiers deliberately try to sabotage their team (e.g., by arriving late) or whether or not they attempt to encourage and coach their team.

Identifying and measuring behaviors that are indicative of teamwork throughout IET will convey to IET Soldiers that teamwork is a critical Soldier attribute. As IET Soldiers learn the value of teamwork and the many ways in which it is exhibited, they will assimilate this attribute and become more proficient team members.

The second Soldier attribute illustrated here is problem solving. While this can be conceptualized as an academic skill, the panel believed that this attribute extends beyond simply being able to solve abstract problems. It also includes identifying problems in multiple dimensions, thinking of possible solutions to those problems, weighing the solutions and choosing appropriate courses of action, and finally showing the initiative to solve the problems.

The purposes of measuring this attribute include facilitating self-development, developing individual and group flexibility and responsiveness, and assisting cadre in improving their instruction. Panelists did not think that problem solving should be a graduation requirement. Regardless of the purpose, it was recommended that this attribute be measured at three levels: inadequate - someone who fails to recognize and solve problems or conversely someone who sees problems everywhere and has difficulty prioritizing them; adequate - someone who identifies and comes up with good solutions to problems appropriate for an IET Soldier; and superior - someone who identifies and comes up with excellent solutions on par with more advanced Soldiers. A challenge will be to identify the level of problem solving appropriate for an IET Soldier.

A Soldier performing at the lowest level would be one who must constantly be told exactly what to do or who is easily thwarted in his/her activities. A Soldier performing at an adequate level would be one who can accomplish most tasks given to him/her and who clearly makes an effort to overcome challenges. A Soldier at a superior level would be one who quickly grasps complex problems and shows exceptional ability to develop solutions. As can be seen, problem solving involves intellectual as well as motivational .

As with teamwork, there are many opportunities for IET Soldiers to exhibit and develop problem solving skills. For example, problem solving can generally be observed whenever Soldiers are being trained on a new task. Good problem solvers won't necessarily perform the task perfectly the first time, but they show an understanding of what they did wrong and have some grasp of how to correct their mistakes. Poor problem solvers will persistently make the same errors and will make little effort to correct them.

Problems facing Soldiers in IET exist in many dimensions from purely intellectual to interpersonal. To solve those problems, Soldiers must be adept at dealing with interpersonal relationships, their own fears and emotions, their physical limitations, and finally intellectual challenges. Paper and pencil questionnaires are inadequate to measure these kinds of problem solving skills. Good measurements will include performance measures in a variety of contexts and conditions. It will also involve training Drill Sergeants to provide effective feedback to their Soldiers so that they can work to improve their own problem solving abilities.

In the future, battlefield simulation may provide new opportunities to observe and assess Soldiers' problem solving skills. Continued development of new technologies for observing and

recording Soldier activity in such circumstances may provide additional opportunity for data collection, assessment and Soldier performance and attribute shaping.

One final note, instituting new measurements in IET may have both positive and negative effects on the way training is executed. On the positive side, the cadre and unit leaders will be more aware of how they are training the Soldier attributes that are measured. This greater emphasis on measurement should cause graduates to more clearly exhibit these attributes. On the negative side, there will be tremendous pressure on the cadre and leaders to train to the measurements rather than the intended student outcomes. Because of this, it will be necessary to carefully design the new assessment process and corresponding measurement instruments and allocate sufficient resources to train the cadre about the process. Periodic evaluations should be conducted to insure the intended effects are achieved.

Training Knowledge, Skills, and Abilities

The panel also spent time discussing the training of KSAs. Panel members agreed that training needs to focus more on understanding why things are done rather than just what to do. Put another way, training should focus on "how to think" not "what to think." In the past, Soldiers had time to develop this understanding in their assigned units, but given the current operational tempo, this is no longer possible. In short, there must be a change in the way KSAs are trained in IET.

For example, in basic rifle marksmanship training, new Soldiers must zero their assigned weapons. This requires them to make adjustments to their sights so that their rounds hit the aim-point. It is a critical skill that they will need to exercise over and over again. Making these sight adjustments correctly requires them to understand the relationship between sight position and the trajectory of the round. It also requires them to be able to look at a shot group and determine the center of mass. Rather than teaching these skills to new Soldiers so that they can understand how to adjust their sights, Drill Sergeants typically do it for them, telling them how many "clicks" to turn each adjustment knob. Although this is more efficient from the standpoint of getting a unit's weapons zeroed in the shortest possible time, it denies Soldiers an important learning opportunity.

The emphasis on teaching "how to think" must begin in IET reinforced throughout a Soldier's career. Although much of IET is Soldier-skill and task based, there is ample opportunity to develop the flexible and adaptive thinking inherent in "how to think." This reflects a paradigm shift from the way that IET is currently executed.

In his white paper, BG(R) Schwitters noted organizational and leadership changes that will also need to take place:

> Creating an organizational environment in IET in which Soldiers practice "how to think" will require different leadership skills than most current leaders have been exposed to. It will require that leaders, particularly at lower echelons, become comfortable with more ambiguity and are skilled at different methods of assessing unit capability. They must tolerate certain types of risk, in many cases much

more than is currently accepted. The organizational climate must also accept that the best ideas and solutions will most often be found among those that will be responsible for executing them. The means to find, act on, and credit those ideas must be honed. Leadership characteristics such as over-centralization of decision-making and micro management, which are undesired in any situation, will markedly reduce the extent to which agile and flexible thinking is present in an organization.

Developing Soldiers with an adequate understanding of underlying principles and concepts (the "why" of what we desire to do) inevitably requires their leadership and trainers to possess that knowledge themselves as well as the skill to deliberately and thoroughly pass that knowledge on. Thus, Drill Sergeant School and other courses that prepare instructors will need to make adjustments to better prepare instructors to train the principles and concepts behind KSAs.

Panel Conclusions and Recommendations

The transition of a civilian into a Soldier entails more than just the acquisition of knowledge, skills, and abilities. It also entails developing attributes that are consistent with and that support Army values and can be described as qualities that are the "essence" of a Soldier. Unless those attributes are better defined and measured to an appropriate degree, the Army can have no confidence that Soldiers leaving IET will have attained them. Transitioning the way IET is conducted to ensure essential Soldier attributes are achieved will require changes in IET trainer preparation and training methodology but little if any additional time, and other resources. To successfully accomplish this transition, the panel had the following recommendations:

- Develop a comprehensive Soldier model.   Both Soldier attributes and Soldier skills (critical KSAs) need to be included to provide units with high quality Soldiers.   Measuring and training Soldier attributes should not diminish the training of basic Soldier skills, but should support a comprehensive Soldier model that includes both critical KSAs as well as attributes. It will take time to fine tune IET to produce Soldiers with the proper balance of KSAs and attributes but a good start has already been made and is currently being utilized.

- Determine the most appropriate uses for Soldier attribute assessments. Measurements may be used to provide developmental feedback, fine-tune a program of instruction, or serve as a graduation requirement. The measurement design will always be dependent on the purpose of the measurement.

- Develop measurements that fit their purposes. Rarely is a single measurement method ideal for all purposes. Measures must be developed that are practical and effective for the purposes they serve.

- Develop a system for monitoring and refining measurements. Over time, instructor turnover and changes in the program of instruction will cause measurements to become obsolete and be misused. In addition, trainers will always attempt to train to the measurement. Care must be taken to ensure that the impact of the measurement is positive

(e.g., the focus of training to move away from the process to the outcome)[5] A program to monitor and refine measurements should be a part of any effort to implement new measurements.

- Take advantage of new technology to improve measurement accuracy and reduce workload. Technological advances in telecommunications are rapidly increasing the ease with which measurements can be conducted. The Army should remain poised to take advantage of these opportunities.

In closing, it should be mentioned that defining the essential Soldier attributes and developing ways to measure them is necessary work, but equally important, is the leadership that is needed to direct the overall effort. For this to happen, leadership, measurement experts, and other interested parties must come together to discuss the details of such a transformation. It is hoped that the summary of this panel will form a foundation for that dialogue.

---

[5] Here again, it was the experience of panel members who had attempted to define desired Soldier attributes and deliberately incorporate their development into IET that training was considerably improved - away from measuring simple observables such as hours trained or time spent at a task toward real, positive changes in Soldier attitude and attributes. This did not come at the expense of developing Soldier KSAs but seemed to have a synergistic effect by blending skills training with attribute development and reinforcement.

Summary of Panel 4 Discussion: Assessing New Training Programs

The panel focused on three areas where training assessments are typically conducted in Army settings. The areas were: new training and educational programs, training on new equipment, and unit proficiency at home station. The orientation was on identifying general needs, requirements, and approaches rather than identifying specific measurement techniques, as techniques depend on the unique objectives of each assessment.

Panel Leader Discussion Papers

The discussion papers by the two panel leaders served as a framework for much of the panel's discussion. These papers approached the measurement and assessment issues from different, yet complementary, perspectives. Dr. Pellegrino's paper discussed the issue of assessment validity, which in turn has implications for the measures used in the assessment. Clearly, individuals designing assessments must fully understand what should be measured and why. What follows are selected quotations from that paper.

> … we are interested in an appraisal of outcomes and obtaining evidence that allows us to make some practice and/or policy decision that is actionable.

> Assessment … is a very carefully constructed process designed to gather evidence with respect to some particular, well-defined objective or outcome. …. the term assessment is a broad description of a set of data collection activities that can range from informal to formal observations (tests), which in turn can produce simple or complex measures and scores (measurement). One goal to keep in mind is that we "should measure what we understand, and understand what we measure.

> Validity is the "holy grail" of assessment since there is no single index of an assessment's validity. Instead validity involves multiple components and must be based on a complex argument structure in which various pieces of evidence must be assembled to support the validity of the inferences drawn from a given assessment. One of the most critical aspects of validity is "construct" validity – the extent to which an assessment measures that which it purports to measure.

> Related to the idea of validity is the notion that assessments are conducted to fulfill certain purposes and that they are executed in differing contexts. An assessment is valid only to the extent that it is appropriate for its intended purpose and context of use.

> Assessment design is not about jumping from poorly defined constructs to tasks or performances that seem to have "face validity." This typically leads to poor and inadequate assessment. Rather, assessment design requires a careful thinking through of what claims and evidence one is seeking about persons, programs, etc. and then designing the situation to provide the relevant evidence.

… contemporary theory about the nature of assessment and its design alerts us to the fact that the generation of more valid assessments will require two things. The first is an expanded knowledge base about domains of competence and performance, about tasks and their affordances, and about measurement models that appropriately match different forms of data. The second requirement is multidisciplinary teams that work together rather than separately to bring to bear models of domain knowledge and skill (competence), tasks, and analysis methods in the process of designing and developing assessments.

MG(R) Ernst's paper also focused on the need for good assessments, keeping potential stakeholders in mind, and how the assessment information is used in making decisions. His paper specifically addressed military settings, the impact of the current operational environment on military training within schools and units, with corresponding needs to assess that training, both what it is intended to accomplish and what it is not accomplishing. Below are excerpts from MG(R) Ernst's paper.

The current operational environment's operational tempo (COE OPTEMPO) has necessitated significant changes in training in both units and schools.

Examples of the changes that are occurring and the associated challenges were cited in the paper.

The current operational environment (COE) has changed the training and leader development landscape considerably, if not drastically. This can best be seen in how the Army has articulated training the engaged force in the context of the ARFORGEN Model[6] by compressing the cycles of ARFORGEN … into a Train/Reset Cycle of 18 weeks. The implications of ARFORGEN compression can easily be seen beginning with the Chief of Staff of the Army's Training and Leader Development Guidance. This guidance has significant impact across all training and modifies many or most Training and Doctrine Command (TRADOC) courses, collective unit training and new equipment training. Some pertinent highlights from that guidance are: Units with 18 months or less dwell time between deployments are to focus on their Directed Mission Essential Task List (DMETL) vice their Core METL (CMETL). Leader development courses are to focus on full spectrum operations and impart fundamental major contingency operations skills on officers and noncommissioned officers.

Thus the time frame for [most] training and leader development is 18 weeks. This has resulted in the shortening of many TRADOC courses and a significant amount of functional courses taught via MTTs[7] at unit locations. … All of this must be reconciled into unit collective training schedules. Exacerbating this for institutional training and leader development, is demand for inserting lessons from the COE into now compressed programs of instruction. To further compound the time demand on the Reset/Train Cycle is that new equipment

---

[6] ARFORGEN stands for Army Force Generation. It consists of a three-phase unit readiness cycle. The three phases are Reset/Train, Ready, and Available (Deploy).
[7] Mobile Training Teams

fielding and training occurs here as well, and it is axiomatic that during dwell all units receive new or modified equipment.

Buried within the collective unit focus on DMETL is that some … unit's DMETL is significantly different from their table of organization and equipment (TOE) mission and CMETL, requiring them to reorganize and or leave their TOE equipment behind. Notable examples are field artillery and armor units.

The real problem is that nearly all training now falls into the significantly modified, virtually new category. Therefore, even current measurement begs revalidation. Even if this is done, is it enough or is there something else? Suggest that the answer is that this is not enough and there must be an additional, and perhaps, more important measurement.

Compromises to training due to COE demands suggest that measuring what is not done is critical to assessing the impact and informing decisions for post COE training and leader development. Let's return to our example of an artillery battalion employed in an entirely different role and without its TOE equipment. It would be fairly easy to quantify CMETL needs and necessary resources to get a battalion back to T1 readiness against CMETL. However, this would miss the years of growing leaders by two or more rank levels who missed the "school of practice" part of their branch education, exacerbated by less CMETL related subjects/time in their "school of theory" time. This example likely applies to many, if not most, types of units in the Army.

It seems, therefore, [that] measuring what is not trained is an added dynamic and necessary aspect of future measurement of training and leader development. This is likely a more difficult aspect of measurement. … The [view that] training and leader development ….[has] changed dramatically requires a review of how we Measure>Assess>Decide.

General Considerations

Certain themes cited by Panel 4 members applied to all three areas. Consistent with Dr. Pellegrino's discussion paper, the panel agreed that with any assessment it is critical to clarify the purpose -- to determine what different stakeholders require and how the information will be used. A full understanding requires dialogue between the requestor/user and those conducting the assessment. To avoid misuse of assessment findings, all parties need to understand that assessment validity is situation-specific. Thus communication of the results is also critical to the assessment process.

The use of a logic model was suggested as a general way of conceptualizing the range of assessments that may be needed and of determining what should be measured. A logic model attempts to lay out all the components of a larger system like training in terms of inputs, outputs and outcomes, including the logical and implied causal relationships among those components. The inputs include things like the resources allocated, while the outputs include activities like

specific courses, exercises, or programs as well who is supposed to participate or benefit. The desired outcomes can range from those that are proximal and close in time like skill or knowledge development, those more medium term like transfer to the operational field, and those that are more distal like an enhanced operational capability. The development of a logic model for training allows prioritization of the assessment enterprise and a clearer sense of the measurement needs. This in turn allows for designing measures of the various components and ensures continuity of measures and documentation linking analysis, design, process, and outcome. The logic model also forces you to articulate assumptions and expectations, and clarifies what you might want or need to assess for system monitoring and improvement.

For most assessments, the measures need to be multi-faceted. Examining just student outcomes or looking at training from the view point of the learner is typically not sufficient. There is a need to expand measures to determine what supervisors think and the extent to which transfer occurs. There is a need to determine if the training can be used as well as if it is effective. Tasks, conditions, and standards should be retained as a means of measuring proficiency at the individual and collective levels, but it is important to go beyond these sets of measures. It is also important to verify what training actually occurred, to identify variations in implementation and organizational barriers to performance. Often resource implications need to be identified.

Another general theme was the need to establish baselines and some means of measuring Soldier progress thereby providing a continual assessment. Although some baseline data are collected, the data are typically not maintained and saved. Longitudinal databases would facilitate determining the effects of changes in training and education. In addition, baseline information could facilitate more regular monitoring of training processes, potentially providing early indicators of training and educational gaps and of trends in performance regardless of whether substantial training changes occurred. And at certain levels of training, baseline information could serve as a diagnostic measurement tool.

Measurement feeds assessments, and in turn, assessments feed decisions. The outcomes are used by leaders at different levels for different purposes. At lower levels of command in the institution or the unit, assessment feedback is used to make decisions about the adequacy of training, and the need for retraining or additional training. Decisions at higher levels (to include the Department of the Army) are more focused on training resources and return on investment. Individuals involved in assessment design and in developing the associated measures should be cognizant of the different ways in which the findings may be used.

Training and Educational Programs

A commonly used approach to assess training and educational programs is Kirkpatrick's (1998) four-level model, particularly the first two levels. Kirkpatrick's model fits under the more generic logic model cited previously with respect to different levels of outcomes. The panel examined how Kirkpatrick's model is often applied in Army training assessments.

- The first two levels of Kirkpatrick's model, reaction and learning, are commonly used when comparing baseline and "new/modified" programs. Measuring effectiveness by just

these two levels was judged by the panel to be inadequate, although such measures do provide basic, necessary information. For Level 1 (reaction – by participants to the training), specific measures often focus on determining whether the actual instruction unfolded according to the program of instruction and an examination of the instructional quality. These data are typically collected through surveys or focus groups with instructor and learners. For Level 2 (learning – improvement in knowledge and skills), all Army school-collected data can be used.

- Levels 3 and 4, called behavior and results respectively, are more difficult to assess (proxy measures may be used), and often cannot be measured given the constraints of most assessments. However, the Panel agreed that the biggest pay off would be in also measuring Levels 3 and 4.[8]

    The panel recommended that TRADOC continue its programs that examine training issues and effectiveness. These included:

- The accreditation process of periodically examining institutional courses.

- The "Studies" program whereby Army schools and/or TRADOC can request a study of specific training issues.

- "Pilot" programs to examine the effectiveness of training modifications prior to formal implementation.

    Other approaches discussed by the panel were:

- The desirability of having longitudinal, historical, evidence databases for assessing new programs was stressed. When possible, natural-occurring events should be leveraged. For example, automated tools could be enhanced to facilitate this process, particularly for training that is already automated such as distance learning. Knowledge management modeled on the Warfighter forums was another recommended approach.

    Unresolved measurement issues cited were:

- There is a need to consider how the COE could impact the type of measures used in assessments.

- With new programs or courses, the validity of the content and the training objectives as well as whether the learning approaches actually work should be established prior to a formal assessment of effectiveness.

---

[8] Behavior (level 3) refers to determining the degree to which participants applied what they had learned when they are on the job – the extent to which behavior changes occurred because of the training program – often referred to as transfer. Results (level 4) refer to whether the targeted organization outcomes have been achieved.

- Programmatically, there are no established criteria for selecting training issues for the "Studies" program, an important consideration given the limited resources for such assessments. It was noted that you cannot address all the needs by means of "studies." Continuous assessments would help address many educational and training issues. Studies should be reserved for special analytic needs.

- Overall, current measurement techniques were viewed as being adequate. The central problem is that measurement is difficult and challenging, and not always possible to execute as desired. In addition, demand for such assessments often exceeds the capacity and resources to deliver.

- A specific institutional training area discussed was using Mobile Training Teams (MTTs) as a means of delivering NCO courses. MTTs were initiated because of the compression in training time for institutions and units. The NCO education system was designed to select, educate, and promote NCOs. Use of MTTs is a different paradigm and there is a need to know whether NCO professional development training is compromised because it is conducted "on the road" and whether the original intent is being met. MTTs are a good example of where multiple dimensions should be measured: impact of additional training load on units, potential for lower quality training due to training facilities and/or unit demands on students; consequences of narrowing of the student mix and less opportunity to share knowledge and experience from peers in other units which impacts small-group dynamics in seminar style education.

New System Training

The panel's focus on training of new systems covered the different varieties of this training that occurs during system development and initial fielding. For purposes of this report, all these varieties are referred to as New Equipment Training (NET).

At one end of the continuum is NET that goes directly to an operational unit when a new system is fielded. The equipment or system being trained could have previously undergone extensive Army testing for its contribution to a unit's warfighting capability and for its reliability, or very little to no Army testing. NET can occur while a unit is deployed, and there could be little prior experience with the NET program of instruction. The training objective of NET at this stage is to ensure that the unit as a whole and individuals within the unit are proficient with the system, can employ it effectively, can maintain it, and are prepared to execute their own sustainment training. The quality of NET can be very critical, particularly when conducted just prior to or during deployment, which allows little to no additional time for the unit to work with the system.

On the other hand, there is what can be called "early NET," that is, NET that occurs with the initial formal Army testing of a new Army system. This test is often a Limited User Test (LUT). Systems subject to such formal Army testing are typically major or large systems. Under these circumstances, NET is the first time Soldiers, leaders, and the unit have been exposed to and are trained on the system. The goal of NET in this case is to prepare all unit members, who are test players, to operate and employ the system sufficiently well during the test

to insure that the test is valid and results are not flawed in some way (e.g., results do not reflect training weaknesses but reflect system capabilities instead), and can be used to make valid decisions about new systems (Hawley, 2007/2008). A similar NET process occurs before a system's Initial Operational Test and Evaluation. The training programs and techniques used prior to such tests often become prototypes for later institutional and unit training programs, and consequently have a life beyond the test itself.

Regardless of the type of NET, the consequences of not being aware of NET strengths and weaknesses have long-term implications. Without assessments, training gaps are not identified. The importance of assessing NET has become more complicated and more critical in recent years with the increasing complexity of systems (Hawley, 2006/2007).

Despite the different types of NET, the fundamental measurement and assessment issues identified by the panel were remarkably similar. The consensus was that NET is typically not formally evaluated, regardless of when it occurs during system development. Moreover, often there is no opportunity to evaluate the training. In cases when NET is evaluated, these evaluations are not conducted early enough and are not comprehensive. For example, training feedback can simply be a report that the training was conducted, not an assessment of whether the training accomplished its intended purposes.

Assuming that NET is assessed, the panel had suggestions regarding what should be measured and approaches that could be used:

- NET assessments and measurement should include an operational component. Panel members indicated that often NET focuses on individual skills (e.g., "switchology") and does not address system employment by the unit. However, there is documentation that can serve as a basis for the development of system employment measures. Systems have an operational and organization plan, a concept of employment, which depicts how systems are to be employed on the battlefield. There are analyses indicating what operational gaps can be narrowed or eliminated with the system. When NET supports operational testing, the test plan is another relevant document. These documents can be leveraged to help identify what operational components should be measured and what would be appropriate exit criteria.

- Front-end analyses should be conducted to determine the level of proficiency desired for different types of NET, followed by preliminary assessments of the NET plan for consistency with the front-end analysis. For example, the NET prior to a LUT with a system under development requires a different level of individual and unit proficiency than NET associated with equipping a unit with a fully-developed "go-to-war" system. Also, as the system progresses, the scope and impact of NET broadens to include unit support and maintenance, sustainment training, etc. Can the unit use and maintain the system without the contractor? The measurement tools developed should be consistent with the level of proficiency required.

- Other aspects of NET that could be assessed and potential measurement approaches were cited. These points were consistent with one of the panel's general themes: the need to carefully examine the purpose of the assessment.

    o Develop measures of the extent to which leaders can transfer the concept of employment.
    o Assess the broader consequences of NET as often the focus of NET is too narrow. An example of this is to assess the impact on unit operations of using detailed personnel for systems, rather than dedicated operators.
    o Measure the impact of not getting the right people trained. When NET is conducted in conjunction with equipment fielding (vs. a system test), units often have difficulty in insuring all the appropriate people are trained.
    o Because systems are often fielded with deploying/deployed units (e.g., Rapid Fielding Initiative), it is important to assess the training given by deployed NET teams.
    o Develop measures of prerequisite skills as these skill levels can impact the effectiveness of NET.

- There could be a feedback on NET assessments and measures from early system development to fielding the first unit. This would provide a continuous development cycle for NET where training plans and measurements are refined based on lessons learned, ultimately improving follow-on NETs. In other words, NET could undergo a development cycle in concert with the system development cycle.

- Another topic raised by the panel is who should conduct the NET assessments. Panel members acknowledged that different agencies are officially responsible for monitoring the different types of NET. However, the consensus was that individuals doing the monitoring and conducting the assessment should be independent of the NET team and have system expertise and knowledge.

Unit Proficiency

Much of the panel's discussion on measuring unit proficiency related to concerns raised by MG(R) Ernst in his discussion paper. Input from panel members provided additional insights into the difficulties in assessing unit proficiency in today's environment. Experience to date has shown that during ARFORGEN Reset, the time is used to let Soldiers regain their balance, so time away from the unit is minimized and there is some local training. During Train, DMETL is stressed, almost to the exclusion of CMETL. Many Soldiers work outside their core area or CMETL. Total Reset and Train is 12 months. Although CMETL is vital, there is insufficient time to train to it. Time is spent at a Combat Training Center, and there can be required school training during this period as well.

Another condition identified by the panel that inhibits the unit's ability to measure proficiency is personnel turbulence. It was pointed out that stability of unit personnel, both Soldiers and leaders, is needed for training and for assessments. However, during Reset-Train, there is personnel turbulence, particularly with first-term Soldiers, which can be disruptive to the

training process. Also leaders need to be assigned and given time to stabilize in the Train phase. When these conditions do not exist, training and assessment of unit proficiency can be insufficient.

Much of the panel's dialogue centered on collective tasks, conditions and standards (TCS). The panel agreed that collective TCS should be retained. The following associated issues to be resolved and potential measurement approaches were major points of discussion:

- TCS related to DMETL are not standardized and consistent across the force. Currently, units develop their own TCS to support unit DMETL requirements not addressed by CMETL. [In this context, DMETL requirements refer to "new" tasks resulting from the COE.] There needs to be more effort in consolidating and standardizing these efforts, with a single source for DMETL standards at all levels of command.

- The question was raised regarding whether everyone agrees on what "right looks like" for both new and current collective tasks. It was suggested that a systematic examination/ investigation of what is actually happening in units (Brigade Combat Teams) when they are deployed would help address this issue. An accurate and complete picture would help determine if proposed DMETL TCS are appropriate and if current TCS are outdated. The suggested research approach was to insert a multi-disciplinary team of embedded observers in a unit, starting with the Reset phase and continuing into deployment; i.e., a longitudinal research effort to assess performance and proficiency.[9]

- To gain a better understanding of unit status and readiness, the gap between DMETL and CMETL proficiency should be measured.

- TCS are the minimum requirement; but they are not sufficient measures. For example, measures could be expanded to assess higher levels of proficiency, adaptive expertise, and transfer. AARs could be leveraged to be more objective. Commanders and leaders could be trained on using additional measures of proficiency.

- MTTs could be used to assist with DMETL training and with assessments of training effectiveness.

- Another way to examine DMETL training is to conduct post-deployment assessments of the pre-deployment training.

- A commander's assessment of unit readiness is important, but needs to be supplemented by data on technical skills and certification.

Following are other topics the panel examined which were not linked so directly to TCS:

---

[9] *Note*. This approach is consistent with what was done in World War II as described by Dr. David Segal, University of Maryland, in his keynote address at the workshop.

- How to establish good feedback loops from the field to the schoolhouses was discussed. Warfighter forums are part of this solution, but other means need to be identified. It was noted that the Combat Training Centers (CTCs) provide scenarios and performance-based feedback for DMETL.

- An alternative approach to the CTCs which is currently being tried is an exportable training capability where CTC personnel come to a unit. Examining the effectiveness and utility of this new approach would be valuable.

- Lastly, an overall gap analysis is needed of the cumulative effect of DMETL emphasis combined with the shortfall in TRADOC schools and courses on full spectrum operations and major combat operations.

Panel Conclusions and Recommendations

In summary, the underlying issue for Panel 4 was whether the right things are measured when assessments are conducted. The answer was "*sometimes*." We should retain the measures and assessment approaches that are basic to our understanding of training issues. There are many training areas where assessments are needed, but they either do not occur or the scope of the assessment is not comprehensive. Due to the stresses placed on training and education by the COE OPTEMPO, establishing baselines now could be a significant reference point for returning to a more "normal" training and education model. Improvements to assessment design and scope, plus expansion of the number of assessments and the capacity to design and conduct them, would substantially improve the findings provided to decision-makers.

Metrics Workshop Final Conclusions

There was no shortage of assessment needs expressed at the meeting. In his plenary address, LTG Rochelle asked for better ways to assess the performance potential of individuals in support of the Army's Human Capital Strategy. This strategy seeks to train and assign individuals based on their competencies and reward individuals for strong performance. LTG Freakley highlighted the need for more sensitive selection tools for identifying those who would have successful Army careers.

Both LTG Rochelle and LTG Freakley also spoke of the need to assess resiliency. The latter has become important because of the chronic stress felt by Soldiers and their families from multiple deployments. To better support families of Soldiers, Panel 1 recognized the need for a more systemic approach for assessing Soldier and Family well-being. They recommended collecting a range of measurement data such as satisfaction with family support services and divorce, suicide, post traumatic stress disorder, and traumatic brain injury rates.

Panel 1 also discussed the need to assess not just what Soldiers can do (i.e., aptitude), but what they want to do (i.e., desire), and what they will do (i.e., motivation). The development of non-cognitive measures was noted as a fruitful approach to meet this need.

Panel 2 discussed the need to assess mental agility and cognitive readiness, two attributes that are valuable in the current operational environment. They recommended that such measures be based on a theoretical model derived from critical incidents of operational experience.

Discussions on panels 3 and 4 focused on whether the right things are being assessed in IET, NET and unit training. For IET, members called for a need to assess attributes like initiative, accountability, problem solving, and teamwork. Regarding NET, panel 4 members pointed out the need to assess training effectiveness by how well the training translates to job performance. With respect to unit training, panelists discussed the need to of assess training gaps. As units focus training on DMETL there is a danger that Soldiers and leaders are maturing without an adequate understanding of their Branch's CMETL.

Although a wide range of topics were covered across the keynote addresses and panel discussions, some common threads emerged across these workshop sessions. These themes are summarized below.

Measurement Feeds Assessment

There was agreement in the panels that measurement is distinct from assessment. Although these two terms are often used interchangeably, assessment was seen as being something that goes beyond measurement or testing. As discussed in Panel 3, measurement is the act of quantifying characteristics whereas assessment is a process of characterization. Or as described by Dr. Pellegrino on Panel 4, assessment is a process of gathering evidence through both formal and informal observations for a specified purpose. As MG(R) Ernst put it: Measurement feeds assessment and assessment in turn feeds decision-making.

Good Measurement Can Not Be Developed from Poorly Defined Constructs

The starting point for all good measurements is well-defined constructs. Constructs like mental agility, adaptability, teamwork, resiliency, and performance potential need to be clearly defined before large-scale research efforts are implemented to investigate them. Although certain construct names are widely used in the Army, they often have different meanings among organizations. If behavioral researchers rush to develop measures for these attributes before there is consensus on their meaning, the resulting measures will satisfy few people. Instead of fostering progress in the understanding of the construct's role in training and/or performance, research findings based on such measures will have little utility for decision makers.

Ideally, Army leaders would be convened to discuss the attributes and their related indicators prior to the implementation of related research projects. In spirit of ECD, the best approach would be for the leaders to carefully think through the claims they want to make about Soldier behaviors and student outcomes and then determine the best way to obtain the evidence to support those claims. Getting consensus on an abstract definition of an attribute is thus less important than consensus on the ways in which the attribute is manifested by Soldiers.

Good Measurement Includes More Than Just Questionnaires

To measure many of the constructs mentioned in this report, the panels agreed that questionnaires are not sufficient; multiple, different types of measures are required. For example, problem solving involves not just the ability to identify problems across many different domains and develop solutions for them but also the initiative to solve them, a multi-faceted tool is needed to adequately measure of all of these different dimensions. Similarly, a battery of measures, including behavioral attitudinal, and biographical measures were recommended to adequately measure mental agility. Constructs like teamwork, resiliency, and well-being will similarly require multiple types of measurement.

In his keynote address, LTG Freakley called for a battery of measures to select qualified individuals for service, including resiliency, propensity to achieve, and the motivations and values of those eligible to enlist. He emphasized the need to measure these throughout the careers of Soldiers. Panel 1 echoed this philosophy when it recommended taking a holistic approach to assessing applicants for entry into the Army.

Integrating the data across different measurement types will be challenging from a psychometric perspective. There always will be different ways to weight and combine the data. Determining the best balance of measures for the various attributes will take time. From a practical standpoint, combining multiple measures from a variety of contexts and time points may be too resource intensive. Researchers will have to be innovative in their use of existing measures and emerging technology to insure the assessment process and corresponding measurements are practical. Some innovative solutions emerged from the panels such as using the Internet, cell phones and other personal electronic devices for data collection.

Finally, it was stressed that data from outcome measures need to be routinely collected and maintained in longitudinal databases across TRADOC and Forces Command. Accessible,

reliable longitudinal databases would decrease the measurement burden on the Army by reducing redundant data collection efforts and would reduce the time needed to gather data for researchers and decision-makers.

Maintaining the Army's Human Measurement Capability

As indicated by Drs. Sams and Segal in their keynote addresses, there have been many advances in the science of human measurement since the Army began selecting and classifying Soldiers during World War I. The Army's human measurement capability has grown to include the evaluation of programs, the assessment of KSAs in many different training contexts, equipment usability, and unit readiness. Information gained from these measures is used to facilitate individual development, to measure attitudes and opinions that shape Army policies, and to understand how Soldiers perform in combat.

Clearly the Army benefits at many levels from having a human measurement capability, but this capability requires more than just behavioral science to implement effectively. In order for human measurement to improve Army processes and products, there must be consensus among stakeholders regarding the constructs being measured, Army leadership must support and enforce the use of the measures, and processes must be in place to insure that the measures continue to serve the purposes for which they were designed. Maintaining the Army's human measurement capability will therefore require continued collaboration between Army leadership, measurement scientists, and all other relevant stakeholders.

References

Behrens, J. T., Mislevy, R. J., Bauer, M., Williamson, D. M., & Levy, R. (2004). Introduction to Evidence Centered Design and lessons learned from its application in a global e-learning program. *International Journal of Testing*, *4*(4), 295-301.

Bogart, Leo, et al., (1969). *Social research and the desegregation of the U.S. Army.* Chicago: Markham Publishing Co.

Chambers, J.W. (2003, Autumn). *S.L.A. Marshall's Men Against Fire: New evidence regarding fire ratios. Parameters, 33*, 113-121.

Fletcher, J. D. (2004). *Cognitive readiness: Preparing for the unexpected.* Alexandria, VA: Institute for Defense Analyses. (DTIC No. ADA458683)

Hawley, J. K. (2006/2007, December/January). Training and testing: A complex and uneasy relationship. *International Test and Evaluation Association Journal, 27(4)*, 34-40.

Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels* (2nd ed.). San Francisco: Barrett-Koehler.

Little, R. W. (1964). Buddy relations and combat performance. In M. Janowitz (Ed.), *The new military: Changing patterns of organization*. New York: Russell Sage Foundation.

Office of the Surgeon Multinational Force-Iraq, and Office of the Surgeon General (2006). *Mental health advisory team (MHAT IV). Operation Iraqi Freedom 05-07final report.* Retrieved from http://www.armymedicine.army.mil/reports/mhat/mhat_iv/mhat-iv.cfm

Shils, E., & Janowitz, M. (1948). Cohesion and disintegration in the Wehrmacht in World War II. *Public Opinion Quarterly*, *12*, 280-315.

Spiller, R. J. (1988). S.L.A. Marshall and the rate of fire. *Royal United Services Institute Journal*, *133(1)*, 63-71.

Stouffer, S. A., Lumsdaine, A.A., Lumsdaine, M. H., Williams, R.M. Jr., Smith, M. B., Janis, I.L., et al. (1949b)**.** *The American Soldier: Combat and its aftermath* (Volume II). Princeton, NJ: Princeton University Press.

Stouffer, S. A., Suchman, E.A., DeVinney, L. C., Star, S. A., & Williams, R. M. Jr. (1949a)**.** *The American Soldier: Adjustment during Army Life* (Volume I). Princeton, NJ: Princeton University Press.

U.S. Department of the Army (2006). *Army leadership* (FM 6-22). Washington, D.C.: Author.

Williams, R. M. (1989). The American Soldier: An assessment, several wars later. *Public Opinion Quarterly*, *53*, 155-174.

Acronyms

AAR	After Action Review
AFQT	Armed Forces Qualification Test
AIM	Assessment of Individual Motivation
ARFORGEN	Army Force Generation
ARI	U.S. Army Research Institute for the Behavioral and Social Sciences
ASVAB	Armed Services Vocational Aptitude Battery
BCT	Basic combat training
CMETL	Core Mission Essential Task List
COE	Current operating environment
CTC	Combat Training Center
DA	Department of the Army
DMETL	Directed Mission Essential Task List
DoD	Department of Defense
ECD	Evidence Centered Design
HCS	Human capital strategy
IED	Improvised explosive device
IET	Initial Entry Training
KSA	Knowledge skills, and abilities
MTT	Mobile Training Teams
NCO	Noncommissioned officer
NET	New Equipment Training
OPTEMPO	Operational tempo
TAPAS	Tailored Adaptive Personality Assessment System
TCS	Tasks, conditions and standards
TOE	Table of organization and equipment
TRADOC	Training and Doctrine Command

Appendix A

Workshop Participants, Panel Leaders, Keynote Speakers and Brief out Guests

List of Participants

Workshop Host

Dr. Michelle Sams
U.S. Army Research Institute for the Behavorial and Social Sciences

Workshop Chair

Dr. Scott Graham
U.S. Army Research Institute for the Behavorial and Social Sciences

Keynote Speakers

Dr. Michelle Sams
U.S. Army Research Institute for the Behavorial and Social Sciences

LTG Michael Rochelle
U.S. Army, G-1

LTG Benjamin Freakley
U.S. Army Accessions Command

Brief-Out Session Guests

BG Richard Longo
U.S. Army, Office of the Deputy Chief of Staff, G-3/5/7

BG Frank Turner
U.S. Army Accessions Command

COL Steven Chandler
Army Capabilities Integration Center

Mr. Robert Seger
U.S. Army Training and Doctrine Command

Dr. Kathy Quinkert
Office of the Vice Chief of Staff, Army

Workshop Sponsor

LTG Michael Rochelle
U.S. Army, Deputy Chief of Staff, G-1

Workshop Coordinator

Dr. Gregory Goodwin
U.S. Army Research Institute for the Behavorial and Social Sciences

Dr. Eva Baker
National Center for Research of Evaluation, Standards, & Student Testing (CRESST) at the University of Los Angeles

Dr. David Segal
Center for Research on Military Organization at the University of Maryland

Dr. Cary Chabalowski
Office of the Assistant Secretary of the Army (Acquisition, Logistics and Technology)

Dr. Michael Drillings
U.S. Army, Office of the Deputy Chief of Staff, G-1

SMA (R) Jack Tilley
Independent Consultant

SMA (R) Julius (Bill) Gates
U.S. Army Research Institute for the Behavorial and Social Sciences

Panel 1: Assessing Attitudes and Aptitudes

LTG (R) Frederick E. Vollrath (Co-Leader)
Independent Consultant

Dr. Bernard Rostker (Co-Leader)
The RAND Corporation

Dr. Jacquelyn S. Randolph (ARI Coordinator)
U.S. Army Research Institute for the Behavorial and Social Sciences

Dr. Shonna Waters (Recorder)
HumRRO

Dr. David Alderton
U.S. Navy Personnel Research, Studies, & Technology (NPRST)

Dr. Jane Arabian
Office of the Secretary of Defense (Personnel and Readiness)

Dr. Richard Fafara
U.S. Army Family and Morale Welfare and Recreation Command

Dr. Tonia Heffner
U.S. Army Research Institute for the Behavorial and Social Sciences

Dr. Pat Kyllonen
Educational Testing Service

Dr. Peter Legree
U.S. Army Research Institute for the Behavorial and Social Sciences

Dr. Rachael Mapes
Defense Manpower Data Center

Dr. Laura Miller
The RAND Corporation

Dr. Bruce Orvis
The RAND Corporation

Dr. Ken Pearlman
Independent Consultant

Dr. Michael Rumsey
U.S. Army Research Institute for the Behavorial and Social Sciences

COL Jeff Schamburg
U.S. Army Accessions Command

Dr. Neal Schmitt
Michigan State University

Dr. David Segal
University of Maryland

LTC Val Siegfried
U.S. Army, Office of the Deputy Chief of Staff, G1

Dr. Robert Sinclair
Clemson University

Dr. Naomi Verdugo
Office of the Assistant Secretary of the Army (Manpower and Reserve Affairs)

Dr. Len White
U.S. Army Research Institute for the Behavorial and Social Sciences

Panel 2: Assessing Mental Agility

LTG (R) Walter Ulmer, Jr. (Co-Leader)
Independent Consultant

Dr. Jean MacMillan (Co-Leader)
Aptima, Inc.

Dr. Jennifer Tucker (ARI Coordinator)
U.S. Army Research Institute for the
Behavorial and Social Sciences

Dr. Amy Hooper (Recorder)
HumRRO

Dr. Allison Abbe
U.S. Army Research Institute for the
Behavorial and Social Sciences

COL Paul Bartone
National Defense University

Dr. Craig Bullis
U.S. Army War College

Dr. Jon Fallesen
Center for Army Leadership

Dr. Bert Huggins
U.S. Army Cadet Command

Ms. Arwen Hunter
U.S. Army Research Institute for the
Behavorial and Social Sciences

COL Thomas Kolditz
United States Military Academy

Dr. James Lussier
U.S. Army Research Institute for the
Behavorial and Social Sciences

Dr. Richard Marcy
Center for Creative Leadership

Dr. Michael Matthews
United States Military Academy

Dr. John Morrison
Institute for Defense Analyses (IDA)

Dr. Rose Mueller-Hanson
Personnel Decisions Research Institute
(PDRI)

Dr. Jennifer Riley
SA Technologies

Dr. Steve Stemler
Wesleyan University

MAJ (R) Donald Vandergriff
U.S. Army Capabilities Integration Center

LTC (Ph.D.) Kristin Woolley
U.S. Army JFK Special Warfare Center &
School

Dr. Michelle Zbylut
U.S. Army Research Institute for the
Behavorial and Social Sciences

Panel 3: Assessing Individual Performance

BG (R) James Schwitters (Co-Leader)
Independent Consultant

Dr. John Behrens (Co-Leader)
Cisco Systems, Inc.

Dr. Gregory Goodwin (ARI Coordinator)
*U.S. Army Research Institute for the
Behavorial and Social Sciences*

Dr. Karen Moriarty (Recorder)
HumRRO

Ms. Louise Banton
U.S. Army Training and Doctrine Command

Dr. Jim Belanich
*U.S. Army Research Institute for the
Behavorial and Social Sciences*

Dr. Winston (Wink) Bennett
U.S. Air Force Research Laboratory

Dr. Marshell (Glenn) Cobb
*U.S. Army Research Institute for the
Behavorial and Social Sciences*

COL (R) Chuck Cornwell
Directorate of Basic Combat Training

CSM (R) Morgan Darwin
Asymmetric Warfare Group

CSM (R) Jim DePriest
U.S. Army, III Corps

Dr. Frederick Diedrich
Aptima, Inc.

Dr. Tom Duffy
Indiana University

SMA (R) Julius (Bill) Gates
*U.S. Army Research Institute for the
Behavorial and Social Sciences*

COL Casey Haskins
United States Military Academy

Ms. Pam Hicks
U.S. Army Training and Doctrine Command

Mr. Elrin Hundley
U.S. Army Training and Doctrine Command

Dr. Pamela Raymer
U.S. Army Management Staff College

SMA (R) Jack Tilley
Independent Consultant

Panel 4: Assessing New Training Programs

MG (R) Carl Ernst (Co-Leader)
Independent Consultant

Dr. James Pellegrino (Co-Leader)
University of Illinois at Chicago

Dr. Jean Dyer (ARI Coordinator)
U.S. Army Research Institute for the
Behavorial and Social Sciences

Ms. Charlotte Campbell (Recorder)
HumRRO

Ms. Marta Bailey
U.S. Army Training and Doctrine Command

Dr. Barbara Black
U.S. Army Research Institute for the
Behavorial and Social Sciences

Dr. Anna Cianciolo
Command Performance Research, Inc.

COL (R) Ron Corson
JRC Research & Analysis Inc.

Mr. Harry Crumling
U.S. Army, Office of the Deputy Chief of
Staff, G-3/5/7

Dr. Michael Drillings
U.S. Army, Office of the Deputy Chief of Staff,
G-1

Dr. Stephen Goldberg
U.S. Army Research Institute for the
Behavorial and Social Sciences

Mr. Luciano Iorizzo
U.S. Army Training Support Center

COL (R) Hank Kinnison
The Wexford Group International
(WGI-CACI Inc)

Mr. D.E. MacAllister
TRADOC - QAO

Dr. Gary Riccio
The Wexford Group International
(WGI-CACI Inc)

MAJ Walter Richter
U.S. Army, III Corps

Dr. Scott Shadrick
U.S. Army Research Institute for the
Behavorial and Social Sciences

Dr. Diana Tierney
U.S. Army Research Institute for the
Behavorial and Social Sciences

COL Casey Wardynski
U.S. Military Academy

Dr. Brenda Wenzel
U.S. Army Training and Doctrine Command
Analysis Center

COL Martin Wilson
U.S. Army, III Corps