Effectiveness Evaluation Tools and Methods for Adaptive Training and Education in Support of the US Army Learning Model: Research Outline

by Joan H Johnston, Greg Goodwin, Jason Moss, Robert Sottilare, Scott Ososky, Deeja Cruz, and Arthur Graesser
NOTICES

Disclaimers

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

Citation of manufacturer’s or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.
Effectiveness Evaluation Tools and Methods for Adaptive Training and Education in Support of the US Army Learning Model: Research Outline

by Joan H Johnston, Greg Goodwin, Jason Moss, and Robert Sottilare

*Human Research and Engineering Directorate, ARL*

Scott Ososky

*Oak Ridge Associated Universities, Oak Ridge, TN*

Deeja Cruz

*University of Central Florida, Orlando, FL*

Arthur Graesser

*University of Memphis Institute for Intelligent Systems, Memphis, TN*

Approved for public release; distribution is unlimited.
Effectiveness Evaluation Tools and Methods for Adaptive Training and Education in Support of the US Army Learning Model: Research Outline

While human tutoring and mentoring are common teaching tools, current US Army standards for training and education are group instruction and classroom training, also known as “one-to-many” instruction. Recently, the US Army has placed significant emphasis on self-regulated learning (SRL) methods to augment institutional training where Soldiers will be largely responsible for managing their own learning. In support of the US Army Learning Model and to provide affordable, tailored SRL training and educational capabilities for the US Army, the US Army Research Laboratory is investigating and developing adaptive tools and methods to largely automate the authoring (creation), delivery of instruction, and evaluation of computer-regulated training and education capabilities. A major goal within this research program is to reduce the time and skill required to author, deliver, and evaluate adaptive technologies to make them usable by a larger segment of the training and educational community. This research includes 6 interdependent research vectors: individual learner and unit modeling, instructional management principles, domain modeling, authoring tools and methods, and evaluation tools and methods. This report (one of 6 interdependent research outlines) focuses on effectiveness evaluation research for adaptive training and education with the goal of determining the individual, training, and organizational characteristics that influence the adaptive tutoring process before, during, and after training.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>v</td>
</tr>
<tr>
<td>Preface</td>
<td>vi</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Research Goals and Objectives</td>
<td>2</td>
</tr>
<tr>
<td>3. Background</td>
<td>2</td>
</tr>
<tr>
<td>3.1 Self-Regulated Learning and the US Army Learning Model</td>
<td>3</td>
</tr>
<tr>
<td>3.2 Motivation for Research</td>
<td>4</td>
</tr>
<tr>
<td>3.3 Adaptive Training and Education Definitions</td>
<td>5</td>
</tr>
<tr>
<td>4. US Army Requirements for Adaptive Training and Education Systems and Effectiveness Evaluation Tools and Methods</td>
<td>8</td>
</tr>
<tr>
<td>4.1 Adaptive Training and Education Systems and Effectiveness Evaluation Tools and Methods</td>
<td>8</td>
</tr>
<tr>
<td>4.2 Big Data and Effectiveness Evaluation Tools and Methods</td>
<td>8</td>
</tr>
<tr>
<td>4.3 Training at the Point of Need and Effectiveness Evaluation Tools and Methods</td>
<td>9</td>
</tr>
<tr>
<td>4.4 Artificial Intelligence (AI) Capabilities and Effectiveness Evaluation Tools and Methods</td>
<td>10</td>
</tr>
<tr>
<td>5. Understanding the Dimensions of Effectiveness Evaluation Tools and Methods</td>
<td>10</td>
</tr>
<tr>
<td>5.1 Integrated Model of Training Evaluation and Effectiveness</td>
<td>11</td>
</tr>
<tr>
<td>5.2 Instructional Content and Design</td>
<td>12</td>
</tr>
<tr>
<td>5.3 Changes in Learners</td>
<td>13</td>
</tr>
<tr>
<td>5.4 Organizational Benefits</td>
<td>13</td>
</tr>
<tr>
<td>5.5 Implications for Adaptive Tutoring and Instructional Effectiveness Research</td>
<td>14</td>
</tr>
<tr>
<td>6. Instructional Effectiveness Research Goals and Challenges</td>
<td>14</td>
</tr>
</tbody>
</table>
6.1 Adaptive Tutoring for Cognitive, Affective, Psychomotor, and Social Skills 14
6.2 Adaptive Tutoring in Kinetic Learning Environments 15
6.3 Environmental factors 16

7. Interdependencies with Other Adaptive Training Research Vectors 16
7.1 Learner Modeling, Instructional Management, and Training Effectiveness 17
7.2 Automated Instruction and Training Effectiveness 17
7.3 Authoring Methods and Tools and Training Effectiveness 17
7.4 Domain Modeling and Training Effectiveness 17
7.5 Architecture and Training Effectiveness 18

8. Conclusions 18

9. References 19

Bibliography 23

List of Symbols, Abbreviations, and Acronyms 25

Distribution List 26
List of Figures

Fig. 1  Adaptive training interaction ...............................................................11
Fig. 2  Model of training evaluation and effectiveness (adapted from Alvarez et al. 2004) .................................................................12
Fig. 3  Adaptive training research vectors.....................................................16
Preface

This report is 1 of 6 interdependent research outlines in the Adaptive Training research program. Portions of this text, which originated in ARL-SR-0325,¹ appear in all 6 reports to ensure that readers get the same cross-cutting information.

1. Introduction

Training and education tools and methods must be of sufficient intelligence to understand the needs of individual learners and units of learners, to mitigate negative learner states, and to guide and tailor instruction in real time to optimize learning. These tools and methods must also be affordable, effective, and easy to access and use. These requirements are enablers of the US Army Learning Model (ALM), which includes an emphasis on self-regulated learning (SRL) where Soldiers are expected to manage their own learning and career development through the growth of metacognitive self-assessment (e.g., reflection) and motivational skills (Butler and Winne 1995). While SRL skills are difficult to train and develop, support may be provided to the learner through “adaptive training technologies” (tools and methods), which may be focused to guide learning and reinforce SRL principles.

To support ALM, the US Army Research Laboratory (ARL) has developed a program of research called “adaptive training”, which includes 6 interdependent research areas or vectors: individual learner and unit modeling, instructional management (IM) principles, domain modeling, authoring tools and methods, evaluation tools and methods, and architectural and ontological support for adaptive training. The reports documenting these vectors expand the scope of the adaptive tutoring research described in ARL-SR-0284 (Sottilare 2013) to support ALM requirements in the mid- and long-term evolution of training and educational technology: the Synthetic Training Environment and the Future Holistic Training Environment for Live and Synthetic.

This report (one of 6 interdependent research outlines) focuses on Instructional Effectiveness research for adaptive training and education. Today, most intelligent tutoring systems (ITSs)—an adaptive training tool to support one-to-one computer-based instruction—effectively train well-defined domains in mathematics, physics, and software programming. Since Soldiers operate in more complex, dynamic, and ill-defined domains, it is necessary to expand the scope and effectiveness of adaptive training tools and methods to support training and education in these militarily-relevant domains. Evaluation tools and methods are crucial for establishing instructional effectiveness, which is the study of “the individual, [instruction], and organizational characteristics that influence the [instructional] process before, during, and after training [/education]” (Alvarez et al. 2004, p. 389). For clarity and brevity, we use the term adaptive instruction to encompass training and education. Adaptive instruction has the potential to influence and foster these characteristics to increase learning, self-efficacy, training/educational performance, and transfer to operational readiness.
2. Research Goals and Objectives

The goal of the research described in this report is to determine the characteristics of the individual, the instruction, and the organization that influence the adaptive tutoring process before, during, and after instruction as prescribed by the ALM. The primary objectives of this research are intended to provide guidelines, best practices, tools, models, and methods in support of this research goal:

- Understand and model the pretraining, during training, and posttraining assessments and the relationships among them to determine the instructional evaluation factors needed to establish the effectiveness of adaptive tutors.
- Understand and model relevant characteristics of individuals, instructional environments, and organizations to understand their influence on the effectiveness of adaptive tutoring before, during, and after tutoring.

This report examines the background and requirements for adaptive instructional capabilities along with research challenges, dimensions of instructional effectiveness, desired end states, and interdependencies with other adaptive training research vectors.

3. Background

While human tutoring and mentoring are common teaching tools, current US Army standards for training and education are group instruction and classroom training—also known as one-to-many instruction. Group instruction and classroom training have been generally focused on acquiring and applying knowledge in proxies for live training environments (e.g., desktop simulations, virtual simulations, constructive simulations, and serious games).

Classroom training, especially for complex topics, is often taught as a series of lists that the instructor goes through in a linear fashion (Schneider et al. 2013). This approach puts a heavy burden on the learner to build mental models and make conceptual connections. Using this instructional methodology may lead to varying degrees of success due to individual differences in skills, traits, and/or preferences.

Adaptive training technologies have the potential to improve the learning process, prior to, during, and after training.
Small group instruction in live environments has also been used to assess application of knowledge and the development of skills. A standard feedback mechanism for US Army training is the after-action review (AAR) where significant decision points and actions are captured for small group discussion that is conducted after the completion of an instructional event to help capture teachable moments and to aid Soldiers in reflecting on their recent training and educational experiences.

Both classroom training and small group instruction are manpower intensive, requiring teachers, mentors, and support staff to guide the Soldier’s experience. Today, ITSs primarily guide learner training and education for cognitive tasks in well-defined domains (e.g., problem solving and decision-making tasks in mathematics and physics). Soldiers tend to perform cognitive, affective, psychomotor, and social tasks in both well-defined (e.g., building clearing) and ill-defined domains (e.g., leadership, resource allocation). ITSs generally provide static training (e.g., sitting at a desktop computer to train on a serious game) that falls short in matching the dynamic nature of many US Army operational tasks (e.g., psychomotor tasks) and thereby reduce opportunities to develop and transfer skills to the operational environment. Research is needed to determine the effect of individual learner characteristics, instructional techniques, and organizational characteristics that influence the adaptive tutoring process before, during, and after instruction.

Current Army standards for training and education are group instruction and classroom training, also known as one-to-many instruction. Group instruction and classroom training have been generally focused on acquiring knowledge and applying knowledge in proxies for live training environments (e.g., desktop simulations, virtual simulations, constructive simulations, and serious games). The effectiveness of these tools and methods is in question with respect to learning, performance, retention, transfer, and more specifically, with respect to enhancing SRL skills.

### 3.1 Self-Regulated Learning and the US Army Learning Model

In 2011, the US Army placed significant emphasis on the development of SRL skills with the expectation that new methods of instruction (e.g., ITSs) would augment institutional instruction (i.e., classroom and small group experiences). One-to-one human tutoring has been shown to be significantly more effective than one-to-many instructional methods (e.g., traditional classroom instruction: Bloom 1984; VanLehn 2011). However, it is neither practical nor affordable to have one expert human tutor to mentor each Soldier in the US Army for every required
operational task. This alone signals the need for capabilities to support one-to-one tailored instructional experiences.

Additionally, under the ALM, Soldiers are largely responsible for managing their own learning, but SRL skills are difficult to train and develop (Butler and Winne 1995; Azevedo et al. 2009; Graesser and McNamara 2010). We anticipate that adaptive training tools and methods will fill this gap and will provide personalized guidance to acquire, apply, retain, and transfer knowledge and skills to the operational environment. This signals the need for a computer-regulated learning strategy to augment missing SRL skills; however, adaptive training technologies must first become affordable, sufficiently adaptive, easy to use, and, of course, effective for this strategy to be realized.

3.2 Motivation for Research

A promising alternative to one-to-one human tutoring is one-to-one adaptive training tools that include ITSs. Meta-analyses and reviews support the claim that ITS technologies routinely improve learning over classroom teaching, reading texts, and/or other traditional learning methods. These meta-analyses normally report effect sizes (sigma [σ]), which refers to the difference between the ITS condition and a control condition in standard deviation units. The reported meta-analyses show positive effect sizes that vary from σ = 0.05 (Dynarsky et al. 2007) to σ = 1.08 (Dodds and Fletcher 2004), but most hover between σ = 0.40 and σ = 0.80 (Fletcher 2003; VanLehn 2011; Graesser et al. 2012; Steenbergen-Hu and Cooper 2013, 2014). Our current best meta-meta estimate from all of these meta-analyses is σ = 0.60. This performance is comparable to human tutoring, which varies from between σ = 0.20 and σ = 1.00 (Cohen et al. 1982; Graesser et al. 2011), depending on the expertise of the tutor. Human tutors have not varied greatly from ITSs in direct comparisons between ITS and trained human tutors (VanLehn et al. 2007; VanLehn 2011; Olney et al. 2012).

Graesser et al. (in press) are convinced that some subject matters will show higher effect sizes than others when comparing any intervention (e.g., computer trainers, human tutors, group learning) to a control. It is difficult to obtain high effect sizes for literacy and numeracy because these skills are ubiquitous in everyday life and habits are automatized. For example, Ritter et al. (2007) reported that the Cognitive Tutor for mathematics has shown an effect size of σ = 0.30–0.40 in environments with minimal control over instructors. Human interventions to improve basic reading skills typically report an effect size of σ = 0.20. In contrast, when the student starts essentially from ground zero, such as many subject matters in science and technology, then effect sizes are expected to be more robust. ITSs show effect sizes
of $\sigma = 0.60-2.00$ in the subject matters of physics (VanLehn et al. 2005; VanLehn 2011), computer literacy (Graesser et al. 2004; Graesser et al. 2012), biology (Olney et al. 2012), and scientific reasoning (Millis et al. 2011; Halpern et al. 2012). As a notable example, the Digital Tutor (Fletcher and Morrison 2012) improves information technology by an effect size as high as $\sigma = 3.70$ for knowledge and $\sigma = 1.10$ for skills. The effect sizes attributed to improved instruction and improved domain knowledge have not been separated in this analysis. Such large effect sizes would never be expected in basic literacy and numeracy.

Overall, these are promising results and equate to an increase of about a letter grade improvement over traditional classroom instruction. While ITSs are a promising technology to support adaptive training for individuals in well-defined domains like mathematics, physics, and computer programming, the US Army requires the ability to develop and exercise Soldier skills in more ill-defined domains (e.g., leadership) and at the unit level (e.g., collaborative learning and team training). Developing and maintaining the ability to make effective decisions under stress and in complex environments is also desirable.

Adaptive systems by their nature require additional content and complexity to support tailored learning for each user and, as a consequence, have a very high development cost—a major barrier to adoption by the US Army. Adaptive systems are also insufficiently adaptive to support tailored self-regulated training and educational experiences across a broad spectrum of military tasks as required by the ALM. Today, few ITS authoring tools are generalized across all of the domains requiring training, and no evaluation criteria or standards have been developed to promote reuse and interoperability among ITSs (Sottilare et al. 2012). In other words, current adaptive systems are not yet intelligent enough to support the tailored instruction required by the US Army in the breadth of domains being trained; but there is a stable foundation of 50 years of science on which to grow an adaptive training and education capability for the US Army.

### 3.3 Adaptive Training and Education Definitions

In support of the ALM and affordable adaptive training and educational capabilities for the US Army, ARL is investigating and developing adaptive tools and methods. A desired end-state is the automation of authoring (creation) processes, instruction, and evaluation of computer-regulated training and education capabilities to help build SRL skills and support mixed-initiative interaction. A major goal within this research program is to reduce the time/cost and knowledge/skill required to author, deliver, and evaluate adaptive technologies to make them usable by a larger segment of the US Army training and educational community.
Adaptive training and education research includes elements of adaptive tutoring, distributed learning, virtual humans, and training effectiveness evaluation. For additional details on research specific to ITSs, refer to ARL-SR-0284 (Sottilare 2013). Definitions are provided for this section to distinguish between adaptive training and education elements and also to highlight their relationships:

- **Adaptive Tutoring**: Also known as intelligent tutoring. Tailored instructional methods to provide one-to-one and one-to-many computer-guided experiences focused on optimizing learning, comprehension, performance, retention, reasoning, and transfer of knowledge and acquired skills to the operational environment.

- **Adaptive Tutoring Systems**: Also known as ITSs. The mechanism or technologies (tools and methods) to provide tailored training and educational experiences; adaptive tutoring systems respond to changing states in the learner and changing conditions in the training environment to optimize learning; adaptive tutoring systems anticipate and recognize teachable moments.

- **Virtual Humans**: Artificially intelligent visual representations of people that simulate or emulate cognitive, affective, physical, and social processes.

- **Distributed Learning**: Concurrent distribution of training and educational content to multiple users at the point of need in which content is intelligently selected to support learning, increased performance, and long-term competency in selected domains.

- **Training/Learning Effectiveness**: Evaluation of the impact of training and educational tools and methods on usability, learning, comprehension, performance, retention, reasoning, and transfer of knowledge and acquired skills to the operational environment.

- **Adaptive Training and Education Systems**: A convergence of ITSs and external training and education capabilities (e.g., serious games, virtual humans, simulations) to support engaging experiences with reduced need for authoring (Sottilare 2015).

- **Generalized Intelligent Framework for Tutoring** (Sottilare et al. 2012, 2013): An open-source modular architecture whose goals are to reduce the cost and skill required for authoring adaptive training and educational systems, to automate instructional delivery and management, and to develop and standardize tools for the evaluation of adaptive training and educational technologies.
Adaptive training and education research at ARL is being conducted across 6 interdependent research vectors: individual learner and unit modeling; IM principles; domain modeling, authoring tools and methods; evaluation tools and methods; and architectural and ontological support. This report (1 of 6 interdependent research outlines) focuses on Instructional Effectiveness research for adaptive training systems with the goal of guiding learning in militarily relevant training and educational domains.

Soldiers operate in a variety of complex, dynamic, ill-defined domains where their ability to persevere in the face of adversity, adapt to their situation, collaborate, and think critically is key to the successful completion of their assigned missions. To develop and exercise these skills, it is paramount for Soldiers to train in challenging environments. Presently, these few challenging training environments have been largely provided through manpower-intensive methods or systems with little ability to adapt instruction to support their learning needs. To illustrate this point, Franke (2011) asserts that through the use of case study examples, instruction can provide the pedagogical foundation for decision making under uncertainty. However, this approach is limited in implementation by the expanse of potential cases that would need to be consistently updated and maintained to support large populations like the US Army.

As noted previously, adaptive systems like ITSs have been shown to be effective in promoting learning in primarily static (e.g., learners seated at desktop computers) instructional settings within relatively simple, well-defined domains (e.g., mathematics, physics) for individual learners. For our purposes, static instruction includes cognitive, affective, or social training tasks where a desktop computer delivers instruction and where the physical movement of the learner is limited to activities that can be conducted while seated. For example, static instruction can effectively support cognitive tasks involving decision making and problem solving but is less effective for training tasks involving motion and perception (e.g., land navigation and marksmanship). Ideally, we desire portable adaptive instructional capabilities to go with Soldiers to support training and education at their point of need across a wide spectrum of US Army operational tasks. Research is needed to develop tools and methods to support broader domain modeling, which is representative of the full spectrum of US Army operational tasks. Standards, interoperability, and automation (e.g., automated scenario generation) (Zook et al. 2012) will likely play a significant role in making adaptive training practical. In this way adaptive training technologies will have the greatest impact on organizational learning in the US Army.
4. US Army Requirements for Adaptive Training and Education Systems and Effectiveness Evaluation Tools and Methods

The Army Science and Technology community uses Warfighter Outcomes (WFOs) as the authoritative source for identifying Warfighter needs. WFOs are used to share research and future technology solutions. In the training and education domain, the adaptive training and education research program is targeting 4 specific requirements to support the evolution of US Army training: adaptive training and education systems; big data; training at the point of need; and artificial intelligence.

4.1 Adaptive Training and Education Systems and Effectiveness Evaluation Tools and Methods

The primary gap to be addressed under this US Army requirement is the lack of adaptive systems (e.g., intelligent tutors) to support individual and collective (team or unit) training. The US Army needs an adaptive training and education capability that is persistent and easy to use/access with minimal startup time. There are also requirements to automate an informal AAR (also known as a postexercise critique) to reduce the time and skills needed to produce the AAR and improve its focus and quality. Another line of thought notes that the artificial intelligence in ITSs could be used to facilitate rapid mission planning and course-of-action analyses as a job aid in operational contexts.

The major connection between the adaptive training and education requirement and the effectiveness evaluation research vector is the need to understand how to ensure adaptive tutors are effective in relevant military domains. Ensuring the effectiveness of adaptive tutors will enable the identification of affordable solutions to support training in more complex, ill-defined, and dynamic domains that can result in more efficient transfer of knowledge and skills.

4.2 Big Data and Effectiveness Evaluation Tools and Methods

The primary gap to be addressed under this US Army requirement is that there is a lack of capability to handle and process large amounts of structured and unstructured data (also referred to as “big data”). One capability needed is a structured data analytics program linking individual data (e.g., achievements) to required long-term competencies in military occupational specialties (MOSs). This would allow Soldiers to understand where they rank in terms of experiences and achievements among other Soldiers in their MOS. It would also allow the US Army to identify specific experiences among successful Soldiers in that MOS and provide
a model for other Soldiers in that MOS to follow. The data could also be used by
course managers and instructors to continuously improve instruction and the mental
models of both human and computer-based instructors. Finally, data collected on
trainee learning and performance during adaptive training experiences could be
used to facilitate Unit Training Management where unit commanders would have
access to empirical data to support unit training decisions.

The major connection between the Army’s big data requirement and effectiveness
evaluation research is the capability for acquiring learner data, determining learner
(and team) states, and interpreting data in the instructional environment (e.g.,
serious game or simulation) to understand the relationships among pretraining,
during training, and posttraining evaluations and the factors (individual, training,
organizational) that influence them.

4.3 Training at the Point of Need and Effectiveness Evaluation
Tools and Methods

The primary gap to be addressed under this US Army requirement is the lack of an
easily accessible, persistent, cost-effective, and low-overhead training
environment. A capability is needed to bring training to Soldiers instead of Soldiers
going to fixed training locations. This point-of-need training capability would be
easily distributed, web based, and built upon open-enterprise architecture in the
cloud. US Army training and educational opportunities would be available on
demand anywhere and anytime. However, the delivery mechanism (e.g., laptop
computer, mobile device, and smart glasses) for adaptive training is critical in
determining the limitations of the domain model scope and complexity. For
example, it may be extremely difficult to train all the complexities of a psychomotor
task in a desktop computer setting.

The major connection between point-of-need training and instructional effectiveness
is the opportunity to increase opportunities for learning, and improve performance,
retention, and transfer. The learning environment requirements will need to be
understood to mitigate various problems that could positively or negatively influence
learning. What is the impact on instructional effectiveness if a trainee is required to
operate without Internet connectivity depending on access to a cloud-based network?
If a Soldier decides to take a 2-h training course while traveling and knows that
Internet connectivity will be intermittent, she might decide to download the course
to her device and take it off-line. This learning environment may or may not be
effective for the particular learning objectives. Effective learner evaluation methods
and understanding the individual, training, and organizational factors that influence
adaptive tutoring in point-of-need environments may not be known.
4.4 Artificial Intelligence (AI) Capabilities and Effectiveness Evaluation Tools and Methods

The primary gap to be addressed under this US Army requirement is that the Army lacks an automated capability to replicate the complexity and uncertainty of the operational environment. This gap specifically points to the lack of adaptiveness in virtual humans, ITSs, and other training capabilities. This gap leads to Soldiers developing training-response strategies that result in less challenging training over time along with lower engagement and lower levels of learning and transfer of skills to more challenging operational environments.

The major connection between AI capabilities and instructional effectiveness involves the discovery and innovation of learner assessment and modeling techniques to predict the relationships among instructional effectiveness assessments, and the individual, training, and organizational variables that affect them. The ability to establish these predictive relationships will enable adaptive tutoring to be tailored to individual learning requirements, increasing the overall effectiveness of training.

AI-based capabilities in adaptive training and education systems may also support data acquisition (sensing), natural language, problem-solving strategies, and perceptual/interaction mechanisms in the adaptive tutor.

5. Understanding the Dimensions of Effectiveness Evaluation Tools and Methods

There are 4 typical elements that compose ITSs, a prime example of an adaptive training and education system: a learner or trainee model, an instructional or pedagogical model, a domain model, and some type of user interface. The domain model typically includes an expert or ideal student model by which the adaptive system measures/comparisons/contrasts the progress of the learner toward learning objectives. The domain model also includes the training environment, the training task, and all of the associated instructional actions (e.g., feedback, questions, hints, pumps, and prompts) that could possibly be delivered by the adaptive system for that particular training domain. Typical interaction between the learner, the training environment, and the adaptive system (tutoring agents) is shown in Fig. 1.
Over the last 4 decades, the US military has invested in research to design and develop adaptive tutoring systems to improve the elements of instructional effectiveness: learning, performance, retention, and transfer. The goal of the Army Learning Concept (ALC) 2015 is to embed adaptive tutors in a continuous learning environment (TRADOC 2011). The need for such a strategy is so that adaptive tutors can develop the 21st century competencies throughout a Soldier’s career. Cutting across the cognitive, affective, psychomotor, and social domains, they are tactical and technical competence; critical thinking and problem solving; adaptability and initiative; character and accountability; lifelong learner; comprehensive fitness; communication and engagement; cultural competence; and teamwork and collaboration. Next we describe a research-based model of instructional effectiveness that will allow us to define the research challenges and goals for developing adaptive tutors that will foster these competencies.

5.1 Integrated Model of Training Evaluation and Effectiveness

Alvarez et al. (2004) conducted an analytical review of empirical research studies (between 1993 and 2002) to produce an “Integrated Model of Training Evaluation and Effectiveness” (IMTEE), presented in Fig. 2. Their main objective was to identify the individual, training, and organizational variables that are empirically related to the targets of evaluation in instructional effectiveness studies. We envision this model may also be applied to educational domains and cover “instruction” for training and educational purposes. The authors defined the IMTEE model components as measures of training content and design, changes in learners,
and organizational benefits. The arrows in the model represent only the significant relationships (both causal and correlational) that were found among the model components. The one exception is organizational results because no studies (that fit the inclusion criteria) demonstrated a significant relationship between organizational results and any of the other measurement components or variables.

![Model of training evaluation and effectiveness (adapted from Alvarez et al. 2004)](image)

Alvarez et al. included the organizational results in the IMTEE on the strength of the theoretical argument for it, and recommended that future research focus on this weakness in the model. A more recent training transfer literature review by Burke and Hutchins (2007) had a similar conclusion. Alvarez et al. then determined the pretraining, during training, and posttraining variables (individual, training, and organizational) that were empirically related to each of the model components. We identified this model to use as a baseline for guiding research needs for adaptive tutoring.

### 5.2 Instructional Content and Design

Determining learner reactions (i.e., satisfaction) following instruction is probably the most common strategy for assessing effectiveness. However, Alvarez et al. determined there was no relationship between positive reactions to instruction and learning outcomes or transfer. Instead they found postinstructional reaction assessments of instructional utility and relevance were significantly related to learning and transfer, and that preinstructional self-efficacy and preinstructional motivation were individual variables that were positively related to reactions.

Research is needed to investigate how preinstructional self-efficacy and motivation influence utility and relevance reactions to adaptive tutors, and how these variables will affect authoring and automated tailoring of adaptive instructional strategies.
5.3 Changes in Learners

The postinstructional attitude *self-efficacy* was significantly related to cognitive learning, training performance, and transfer performance. Postinstructional self-efficacy was also positively related to instruction (learning principles that incorporate practice, behavior modeling, and feedback) and organizational (posttraining interventions that support skill transfer) characteristics. The individual variables of preinstructional self-efficacy, experience, and posttraining mastery orientation were significantly related to changes in posttraining self-efficacy.

Cognitive learning was significantly related to instructional performance and transfer performance. The individual variables of preinstructional self-efficacy, preinstructional motivation, and cognitive ability were positively correlated with cognitive learning, and age was negatively correlated with it.

Instructional performance (observable learner behaviors demonstrating knowledge or skill following instruction) was significantly related to transfer performance, and individual (experience) and instructional characteristics (learning principles) were positively related to it.

Research is needed to understand how an adaptive tutor can effectively model learner affect, cognitions, and instructional performance, and assess how those models are affected by measures of individual, instruction, and organizational characteristics.

5.4 Organizational Benefits

Completing the IMTEE model are organizational benefits that are reflected in learner transfer performance (e.g., on-the-job behavioral changes as a result of instruction) and other such quantifiable results as increased efficiency and morale, or quantity of outputs.

Transfer performance was operationalized as supervisor evaluations and postinstructional retests several months after training. Alvarez et al. found that learner reactions and the changes in learner variables were significantly related to transfer performance. Individual (preinstructional self-efficacy), instruction (learning principles), and organizational characteristics (postinstructional interventions, low transfer difficulty, and supportive transfer environment) were positively related to transfer performance.

The organizational benefits of adaptive tutors in the military are typically defined as Soldier and unit readiness, and combat effectiveness. Demonstrating the impact
of adaptive tutoring on readiness is by far the most critical to the US Army’s commitment to investing in it. Research is needed to determine how adaptive tutors are linked to effectively develop the 21st century Soldier competencies, how they contribute to Soldier readiness, and the methods, tools, and strategies required to measure the model components and variables that influence these components.

5.5 Implications for Adaptive Tutoring and Instructional Effectiveness Research

The Alvarez et al. findings spell out the important individual, instructional, and organizational factors that should be considered as independent variables in an adaptive tutoring research paradigm. Variations in self-efficacy, motivation, experience, cognitive ability, and age are the individual factors that have significant effects on pretraining attitudes and during training learner changes.

Variations in mastery orientation manipulations, learning principles, and high training difficulty levels are the training factors that have significant effects on learner changes during training. Variations in posttraining interventions, learning principles, and high training difficulty levels are the training factors that have significant effects on transfer performance. Variations in the postinstructional transfer environment (perceived supervisor support, mandatory instructional attendance, rewards for practice, and follow-up evaluations) are the organizational factors that have a significant effect on transfer performance.

6. Instructional Effectiveness Research Goals and Challenges

The foundational goal of adaptive training research at ARL is to model the perception, judgment, and behaviors of expert human tutors to support practical, effective, and affordable learning experiences guided by computer-based agents. To this end, 4 primary challenges to developing effective adaptive tutoring systems are discussed in this section.

6.1 Adaptive Tutoring for Cognitive, Affective, Psychomotor, and Social Skills

The first goal aims to address the “Changes in Learners” component of the IMTEE by conducting research to determine how well adaptive tutors train cognitive, affective, psychomotor, and social skills that matter to Soldier competency development (Sottilare 2013). The challenge in meeting this goal is that most of the intelligent tutors in existence today are model-tracing tutors that train simple procedural tasks for math, physics, and software programming. In contrast, the
Army Learning Concept 2015 vision is to employ a multitude of adaptive tutors that develop the 21st century competencies throughout a Soldier’s career. Adaptive tutors are expected to improve the knowledge and skills that underlie these competencies, to include higher-order cognitive skills (e.g., decision making and judgment), psychomotor skills (e.g., physical movement, coordination, and the use of the motor skills), affective skills (e.g., self-awareness and growth in attitudes and emotion management), and social skills (e.g., promoting group cohesion, interpersonal skills, teamwork, and cross-cultural competence).

Tutoring strategies probably differ as a function of developmental paths for each domain type, including retention and skill decay. Therefore, for each domain, research is needed to address each element in the IMTEE: determining effective training content and design, and determining valid measures of posttraining reactions, learner change, and organizational benefits. For each domain, research is needed to determine the individual, training, and organizational variables that affect each IMTEE component.

### 6.2 Adaptive Tutoring in Kinetic Learning Environments

The second goal targets all 3 model components by focusing on making training more relevant and more likely to change learners, result in skill retention, and transfer to the task, and to contribute to overall readiness. The goal is to conduct research on how well adaptive tutors can train in limited kinetic, enhanced kinetic, and operationally embedded environments (Sottilare 2013).

The challenge in meeting this goal is for adaptive tutors to represent tasks of varying dynamics, definition, and complexity. Specifically, having varying levels of kinetic interaction with the training system will present significant challenges. For example, decision-making and problem-solving tasks may be taught effectively in a limited kinetic mode along with tasks requiring physical orientation (e.g., land navigation). Enhanced kinetic environments support tasks where freedom of movement and a high degree of interaction with other learners may be critical to learning, retention, and transfer to the operational environment.

Building clearing and other team-based tasks may be taught more effectively in an enhanced kinetic mode. A fully kinetic mode would have adaptive tutoring embedded in the operational environment where a very high degree of freedom of movement and a high degree of interaction with the unit is critical to learning, retention, and transfer. Research is needed to determine reliable and valid measures and measurement strategies that assess acquisition, retention, remediation, and transfer of knowledge and skills within and across increasing levels of kinetic interaction.
learning environments. Research is needed to identify the individual, training, and organizational characteristics that have an impact on the IMTEE model components.

6.3 Environmental factors

The third goal aims to address the “training content and design” component of the IMTEE by conducting research to determine the effect of environmental factors on delivering effective adaptive tutoring (Cohen et al. 2009). The challenge in meeting this goal will be mitigating the environmental factors that can hamper the effectiveness of deploying adaptive tutors throughout a Soldier’s career. Research must develop methods, tools, and strategies that will support employment and deployment (e.g., increase usability, positive human-machine/system interaction) of adaptive tutoring environments so they can accelerate and increase Soldier and unit readiness, and reduce development and sustainment costs of existing/future training systems.

7. Interdependencies with Other Adaptive Training Research Vectors

This section examines interdependencies between instructional effectiveness and the other 5 adaptive training research vectors (Fig. 3). This discussion forms the basis for the sequencing of research and ultimately bringing adaptive training capabilities into a state of practice.

Fig. 3 Adaptive training research vectors
7.1 Learner Modeling, Instructional Management, and Training Effectiveness

The instructional effectiveness research goals are inextricably linked to the Learner/Unit Modeling vectors. Valid and reliable pretraining, during training, and posttraining learner and team modeling provides the foundation for a sound evaluation of the effectiveness of an adaptive training program.

7.2 Automated Instruction and Training Effectiveness

Likewise, instructional effectiveness research goals are inextricably linked to the IM research goals. The IM research strategy for each skill domain type, and for varying types of kinetic learning environments, will enable the instructional effectiveness research goals to be addressed. Conversely, insights gained from instructional effectiveness measures and metrics will support the research goals of IM that are concerned with instructional optimization based on performance data.

7.3 Authoring Methods and Tools and Training Effectiveness

Authoring methods and tools research goals should be aligned with the instructional effectiveness goals for ensuring valid and reliable measures that are available to assess each component of IMTEE, and to develop valid and reliable measures of the individual, training, and organizational variables that affect the IMTEE components. One of the valuable things that can be accomplished with the centralized capture of both authoring data and training data is the evaluation of the link between authoring and effectiveness. Research questions such as “Does additional time spent authoring courses lead to larger effect sizes in learning?” can be evaluated “in the wild” without significantly burdening the author or the student. The automated capture of tool usage statistics allows for the continuous improvement of the tools, and the ability to steer authors toward the types of actions that make a difference for instruction. Additionally, it is possible to conduct research on the use of an authoring tool as part of a learning activity, as “learning by teaching” has been shown to be effective.

7.4 Domain Modeling and Training Effectiveness

The instructional effectiveness research goals are inextricably linked with the domain modeling research goals of developing accurate models of skill domains, accurate selection of instructional strategies, and extending training to military task domains.
7.5 Architecture and Training Effectiveness

The architecture research goals are mutually dependent on the instructional effectiveness research outputs, which are dependent on the architectural data capture. Instructional effectiveness and architectural requirements are driven for the tools and technologies required to gather and store data determined to be relevant to instructional effectiveness measures, translate those data into models of instructional effectiveness, leverage instructional effectiveness models to inform instructional strategy selection, and provide guidance for automating and/or semi-automating the processes to present actionable results to the various stakeholders (e.g., training facilitators, subject-matter experts, students) associated with a particular adaptive training course. The primary goals for these overlapping efforts are to ensure clean longitudinal data capture, storage, accessibility, and analysis. The joint goal of this program would be to analyze both the short- and long-term effectiveness of various approaches and to open new research in areas such as retention and refresher training delivery.

8. Conclusions

This report describes a research-based model of instructional effectiveness that allowed us to define the research challenges and goals for developing adaptive tutors that will foster Soldier learning and performance. This report outlines research objectives for determining how individual, training, and organizational characteristics influence the adaptive tutoring process before, during, and after training.
9. References


## List of Symbols, Abbreviations, and Acronyms

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAR</td>
<td>after-action review</td>
<td></td>
</tr>
<tr>
<td>AI</td>
<td>artificial intelligence</td>
<td></td>
</tr>
<tr>
<td>ALM</td>
<td>US Army Learning Model</td>
<td></td>
</tr>
<tr>
<td>ARL</td>
<td>US Army Research Laboratory</td>
<td></td>
</tr>
<tr>
<td>IM</td>
<td>instructional management</td>
<td></td>
</tr>
<tr>
<td>IMTEE</td>
<td>integrated model of training evaluation and effectiveness</td>
<td></td>
</tr>
<tr>
<td>ITS</td>
<td>intelligent tutoring system</td>
<td></td>
</tr>
<tr>
<td>MOS</td>
<td>military occupational specialty</td>
<td></td>
</tr>
<tr>
<td>SRL</td>
<td>self-regulated learning</td>
<td></td>
</tr>
<tr>
<td>WFO</td>
<td>Warfighter outcome</td>
<td></td>
</tr>
</tbody>
</table>
INTENTIONALLY LEFT BLANK.