Development and Evaluation of Mobile Adaptive Training Technologies

Rodney Long
Army Research Laboratory
Orlando, FL
rodney.a.long3.civ@mail.mil

Jessie Hyland, Joanne Barnieu
ICFI International
Fairfax, VA
Jessie.Hyland@icfi.com,
Joanne.Barnieu@icfi.com

ABSTRACT

This research involves development and evaluation of adaptive training strategies. The current prototype mobile training technology was designed to include an adaptive feature within one instructional cycle (at the Terminal Objective or Module level). Based on pre-assessment questions, the learner's sequence within the instructional cycle will be adapted accordingly. Should certain learning objectives be considered mastered via determined pre-requisite knowledge, those associated lessons will be collapsed in the curriculum yet accessible if desired by the learner. Additionally, after completing a lesson, the learner can receive guidance from the system if he or she answers the intermittent assessment item (Check on Learning) incorrectly and is also prompted to return to the lesson content should the guidance not assist the learner in answering the assessment item. To examine the effectiveness of the adaptive training, we conducted an experiment to compare students receiving the adaptive version of the prototype (i.e., treatment condition) versus those receiving the non-adaptive version (i.e., control condition). Specifically, participants were compared on the following dimensions: learner reactions; training efficiency; and training effectiveness. On learner reactions, we found some preliminary descriptive evidence that participants in the treatment condition were more engaged and held more favorable perceptions of the training adaptability than those in the control condition but the difference was not statistically significant. Evidence on the potential unintended negative consequences of the adaptive training was inconclusive. On training efficiency, we found the adaptive training to be more efficient than its non-adaptive counterpart, as expected. On training effectiveness, participants in the treatment condition performed as well as their counterparts in the control condition on an independent hands-on performance test. Our findings are encouraging but highlight the need to continue robust research in tandem with the development and integration of new technologies in order to realize the full potential of adaptive training.

ABOUT THE AUTHORS

Rodney Long is a Science and Technology Manager at the Army Research Laboratory in Florida and is currently conducting research in adaptive training technologies. Mr. Long has a wide range of simulation and training experience spanning 28 years in the Department of Defense (DoD) and has a Bachelor's Degree in Computer Engineering from the University of South Carolina and Master's degree in Industrial Engineering from the University of Central Florida.

Jessie Hyland, Senior Associate at ICF International, has over 9 years of experience in applied research and human capital consulting with a primary focus on training assessment development and validation.

Joanne Barnieu, Senior Technical Specialist at ICF International, has over 22 years of experience in instructional design with focus on game-based training for Defense-related projects.

Development and Evaluation of Mobile Adaptive Training Technologies

Rodney Long
Army Research Laboratory
Orlando, FL
rodney.a.long3.civ@mail.mil

Jessie Hyland, Joanne Barnieu
ICFI International
Fairfax, VA
Jessie.Hyland@icfi.com,
Joanne.Barnieu@icfi.com

INTRODUCTION

This paper describes research in the use of adaptive training strategies, in support of the Army Learning Model. To support our research, a prototype automated self-directed (i.e., non-instructor led) mobile training application for a military radio was designed to include an adaptive feature within one instructional cycle (at the Terminal Objective or Module level). Based on pre-assessment questions, the learner's sequence within the instructional cycle is adapted accordingly. Should certain learning objectives be considered mastered via demonstrated pre-requisite knowledge, those associated lessons are collapsed in the curriculum, yet accessible if desired by the learner. Additionally, after completing a lesson, the learner receives guidance from the system if he or she answers the assessment item (Check on Learning) incorrectly and is also prompted to return to the lesson content should the guidance not assist the learner in answering the assessment item correctly the second attempt. To examine the effectiveness of the adaptive training technologies developed in the current project, we conducted an experiment to compare students receiving the adaptive version of the prototype versus those receiving the non-adaptive version. Based on the literature review, we proposed a series of research questions to be answered through the evaluation study. In this paper, we describe our findings in response to the research questions based on results from the experiment.

LITERATURE REVIEW AND RESEARCH QUESTIONS

Adaptive training provides trainees with training content that is matched with: 1) their performance level on a pre-assessment; 2) their current on-task level of knowledge, proficiency, or skill; 3) an individual difference; or 4) a combination of the three (Landsberg et al., 2012). Through this matching process, training is adapted to better suit the learner. This is beneficial for two reasons. First, presenting trainees with training content that is matched to their aptitude or ability level provides trainees with an appropriate level of challenge, leading to optimal levels of learning performance (Orvis, Horn, Belanich, 2008; Vygotsky, 1978; Yerkes & Dodson, 1908). This is thought to be the result of trainee engagement and other relevant learner reactions (e.g., satisfaction, motivation, self-efficacy, persistence). Second, adapting training could potentially lead to less instruction/training time, especially for high performers, thereby making training more efficient. If trainees are only presented content pertinent to them, they will spend less time on already mastered content (i.e., lessons collapsed based on a module pre-assessment) (Durlach & Spain, 2012; Landsberg et al., 2012; Park & Tennyson, 1980; Romero, Ventura, Gibaja, Hervas, & Romero, 2006). Alternatively, providing poor performing trainees with a remedial training path during adaptive training would make the training intervention longer, but likely more effective, than a non-adaptive training environment. When remedial training is used, trainees do not progress to a subsequent training domain or module until they have demonstrated a certain level of proficiency with the material. Thus, adaptive training is expected to be more effective for learners at all ability levels. Below, we present relevant literature for understanding potential relationships between adaptive training and learner reactions, training efficiency, and training performance.

Learner Reactions

Prior research has demonstrated the importance of learner reactions such as trainee satisfaction (Sitzmann et al., 2008), perceptions of training utility (Giangreco, Sebastiano, & Peccei, 2009), and perceptions of training challenge (Tulis & Fulmer, 2013). Learner reactions impact important factors related to training including motivation, self-efficacy, and engagement (Sitzmann, Brown, Casper, Ely, & Zimmerman, 2008), which, in turn, impact training performance (Pintrich, 2000, Sitzmann & Ely, 2011). Although research has demonstrated that trainees prefer a challenging environment (Belanich, Sibley, & Orvis, 2008) and that learner reactions predict post-training motivation and post-training self-efficacy (Sitzmann et al. 2008), not much is known concerning the ability of adaptive training to impact learner reactions such as motivation and engagement. It is assumed that presenting trainees with training that best suits their needs would lead to higher levels of engagement, higher levels of motivation, and fewer mal-adaptive responses elicited by individuals who find a task too challenging (decreased persistence and effort). The current training context relates to the assembly, programming, configuration, operation, and troubleshooting of the Joint Tactical Radio System (JTRS) enhanced Multiband Inter/Intra Team Radio (MBITR) (JEM), a combat radio. In this context, trainees may be at various levels of proficiency with the JEM and similar radio equipment. We would expect that expert students would be more engaged with the training if they are able to skip concepts on which they are already knowledgeable. Similarly, we would expect that novice students may be more motivated about the training if they receive remediation to assist them through challenging topics.

In addition to evaluating the overall learner reactions to the adaptive training prototype, it is also important to understand how learners evaluate particular adaptive training strategies within a course or a curriculum. One of these important strategies is performance feedback (Azevedo & Bernard, 1995; Jaehnig & Miller, 2007). Feedback provides learners with information regarding their current proficiency levels and allows them to improve their knowledge and skills through discerning the appropriateness of their actions and responses.

In an effort to create an integrative framework of training adaptation for instructional technology, Durlach and Spain (2012) created the framework for instructional technology (FIT). The purpose of FIT was to present a taxonomy of training adaptation to help progress adaptive training research by clarifying adaptation types and implementations. Durlach and Spain (2012) posited that providing researchers with such a framework would allow for the comparison of numerous adaptive techniques for the purpose of discerning adaptation effectiveness. FIT comprises four adaptive techniques:

1) *Corrective feedback* - focuses on how the training system should respond to incorrect responses
2) *Support* - concerned with methods for providing guidance, cues, and prompts to trainees during training
3) *Micro-sequencing* - concerned with methods of remediation for trainees not mastering content
4) *Macro-sequencing* - deals with methods of progression for trainees demonstrating content mastery

For the current adaptive training prototype, learners are first assessed at the course module level with a set of pre-assessment items (i.e., two items per lesson) to gauge their current knowledge level of the module training content. It should be noted that each lesson is aligned with one enabling learning objective, thus the two assessment items only cover content at that specific level of granularity. These pre-assessment items were vetted and validated by subject matter experts (SMEs) to ensure they are sufficient to capture proficiency in a particular lesson. If a learner answers both questions correctly for certain lessons, those lessons are collapsed. The learner receives an adaptive module and is notified that lessons are collapsed due to the pre-assessment performance; however, he or she is able to access those collapsed lessons if desired. According to the FIT, this progression method is classified as a test-out, macro-sequencing adaptation (Durlach & Spain, 2012). Alternatively, if a learner misses any questions assessing content knowledge for a specific lesson, this lesson becomes visible in the module and is required to be completed. Learners are presented with an interim assessment item following the completion of each lesson. Following the submission of a response, a learner who has not answered the question correctly is presented with both a hint and prompted to answer the question again. If the learner provides an incorrect answer the second time, a link to return to the lesson content for a second review is provided. At the end of the second review of the lesson content, if the learner still answers the question incorrectly, the correct answer will be provided. Fixed hints represent problem-based guidance that is intended to support the learner in correctly answering assessment items by prompting him or her to contemplate the content further or in a different manner. Thus, the training prototype created for this research consists of support, micro-sequencing, and macro-sequencing techniques.

While the FIT taxonomy helps to distinguish between variations of adaptive training, further research is needed in order to better understand the effects of these adaptations. In particular, it is unclear whether learners who receive modules with collapsed lessons experience negative reactions, such as finding the training to be disorganized or lacking a logical progression. For example, if a learner answered all questions on the JEM pre-assessment correctly except for a question related to attaching the antenna to the radio, the learner might perceive the training to be disjointed if he or she only takes the "Attaching the Antenna" module. If learners do experience negative reactions resulting from too many collapsed lessons, then further research is needed to investigate how these negative reactions can be mitigated by altering the adaptive training strategy or remediation strategy. To address these issues, we present the following research questions regarding learner reactions to adaptive training.

Training Efficiency

As previously stated, adaptive training has the potential to reduce overall training time and tailor the length of the training to the needs of the learner (Landsberg et al., 2012; Park & Tennyson, 1980; Romero, Ventura, Gibaja, Hervas, & Romero, 2006). The current adaptive training prototype utilizes both macro-sequencing and micro-sequencing techniques to adapt the training to suit the learners' needs. Specifically, if a learner can demonstrate strong skill proficiency or content knowledge for a particular lesson during a pre-assessment, he or she has the opportunity to pass over the associated training content. However, learners who are not demonstrating content mastery via interim assessment items will encounter guidance and have the opportunity to review the lesson content a second time, thus spending more time.

Training Effectiveness

Adaptive training is designed to tailor content in a dynamic way to best suit the needs of the learner; therefore, the goal is to present learners with content and material that is not too difficult, yet still challenging. Vygotsky (1978) referred to this targeted threshold as the zone of optimal development. While learners may become discouraged when material is too hard or bored when the material is too easy, the zone of optimal development refers to the area where the challenge is high enough to yield maximum improvement and knowledge gains. Adaptive training strives to present material that maintains the zone of optimal development for each learner.

Although over a century old, the Yerkes-Dodson optimum arousal theory is also relevant when considering the role of adaptive training in trainee performance (1908). Specifically, optimum arousal theory specifies that each individual has a level of stress/arousal that delivers optimum performance. Too little or too much stress/arousal will similarly produce lower relative performance. Durlach and Spain (2012) classified an instructional approach that provides trainees with moderately difficult content (i.e., adapting content based on a learner's level of knowledge or ability) as a mastery approach.

One of the main goals of training is to improve content knowledge, ability level, or some other on-the-job performance outcome (Kraiger, 2003). Assessments of training performance are typically used to evaluate the effectiveness of training with the assumption that training performance should relate to, and predict, performance for the trained behavior, knowledge or skill in an environment different from the environment where training occurred (i.e., transfer). In addition to research questions that seek to understand the relationships between adaptive training as a whole, characteristics of the training, learner reactions, and training efficiency, it is also important to understand the impact of adaptive training on job performance.

METHODOLOGICAL APPROACH

In this section, we describe the approach used for conducting the evaluation study of the adaptive training structure in comparison to the non-adaptive version of the training. To evaluate the effectiveness of the adaptive feature of the course structure, a controlled experiment was conducted utilizing two versions of the training prototype (adaptive and non-adaptive), with participants using the non-adaptive training serving as the control group.

Participants

A total of 40 Soldiers from Fort Gordon, Georgia participated in the experiment. Participants were recruited from the Regimental Non-Commission Officer Academy's (NCOA) Advanced Leadership Course (ALC) and Senior Leadership Course (SLC). ALC is a branch-specific course for Soldiers selected for promotion to staff sergeant that focuses on preparing leadership and technical skills required to effectively lead squad/platoon size units. SLC is also a branch–specific course that provides an opportunity for Soldiers selected for promotion to Sergeant First Class to acquire the leader, technical, and tactical skills, knowledge, and experience needed to lead platoon/company size units. Participants had an average tenure of 11 years in the military and had an average of 2-3 deployments. The majority[1] of the participants were from the military occupational specialties (MOS) of 25U (Signal Support Specialists; N=17) and 25S (Satellite Communication Systems Operator-Maintainers; N=14) with ranks from E5 through E7 and an average of 47 months in rank.

Evaluation Measures

Four different types of measures were developed and delivered to Soldiers during the experiment: a demographic questionnaire; a learner reaction survey; a multiple choice post-test; and a hands-on radio performance task assessment administered at the end of training.

Demographic Questionnaire

The demographic questionnaire was designed to collect background information on each Soldier including time in military (years), rank, time in rank (months), military occupational specialty (MOS), and number of deployments. The questionnaire also included a number of items to assess Soldiers' prior experience with the JEM and similar handheld radios (i.e., the AN/PRC-148 Multiband Inter/Intra Team Radio; MBITR) including questions on participants experience using the JEM radio while deployed, perceived expertise with the JEM radio, experience loading Communication Security (COMSEC) keys, and confidence performing various tasks on the JEM radio. The format of these questions varied, some were open-ended, some included Likert-type response scales, and others included yes/no responses.

Learner Reactions Survey

Reactions of learners focus on dimensions such as engagement and satisfaction with the training, issues and challenges, and perceived utility. Learner reactions refer to what the trainees thought about the training from a general perspective, for example, whether they thought it was interesting or boring, whether the material was presented effectively, and whether they believed the content was useful and relevant to their job. These reactions are often a critical factor in the continued success of the training program. To respond to the majority of these items, learners would indicate their agreement with each statement on a 5-point Likert scale where 1=Strongly Disagree and 5=Strongly Agree. To adequately address the research questions proposed, the reaction measure includes items on usability, content and Instructional Systems Design (ISD), learner engagement and attention, perceived training adaptability, perceived learning, perceived learning transfer, satisfaction with training, technical issues that were encountered, and general comments.

Post-Test

In addition to the four pre-assessments presented at the beginning of the training modules, a paper-and-pencil post-test was used to capture the learners' knowledge at the end of the training. The post-test consists of 14 multiple-choice questions covering the four content areas in the training (i.e., each test consists of similar number of items from each content area).

Hands-on Performance Test

The radio performance test was a structured, timed, hands-on exercise that required participants to perform various tasks using an actual JEM radio. The measure consisted of four tasks based on the learning objectives covered in the training content. Each task consisted of setup instructions for the test administrator, equipment requirements, instructions for the test-taker, a scoring rubric and a time limit. Scoring of the work sample task included a behavioral checklist where Soldiers earned points for successfully completing procedures for each task (each task consisted of multiple procedures). In addition, several tasks included follow-up questions that were asked verbally by the test administrator. Each follow-up question was worth one point. These questions, and potential answers, were included in the scoring rubric. If soldiers failed to complete a task within the given time limit, the test

[1] The remainder of the participants were spread out across the MOS of 25B (Information Technology Specialist; N=3), 25N (Nodal Network Systems Operator-Maintainer; N=2), 25P (Microwave Systems Operator-Maintainer; N=2) and 25W (Telecommunications Operations Chief; N=2).

administrator awarded zero points for the remaining procedures under the task. In total, Soldiers could earn up to 20 points in completing the work sample task.

Procedures

Upon arrival to the lab, participants were given an introduction that outlines the steps they would be taking as part of the experiment. The research team then disseminated an informed consent and collected signatures. After that, participants were asked to complete the demographic questionnaire. They were then randomly assigned to one of the four training sessions that were held; two of them featured the adaptive training and the other two were non-adaptive. As described before, the adaptive version of the mobile training application has built-in macro and micro adaption strategies while the non-adaptive version features the same content without adaptation. The four sessions were held over two days with one of each condition on each day; one in the morning and one in the afternoon. The order of the training sessions on each day was alternated to counterbalance any extraneous effects due to the time of the day the training was held. Each participant was assigned one of 10 iPad stations. The researcher showed participants how to log in and then asked them to proceed through the self-directed training on the device. Following completion of the training, participants met with one of the researchers individually to complete the post-test and the hands-on performance assessment. The same experiment administration process was used for both conditions. The only difference between the adaptive and non-adaptive groups was whether macro and micro adaptation strategies were built in the training applications.

RESULTS

Equivalence of Groups

A key assumption in the experimental design employed in our study was that the treatment and control groups were comparable prior to treatment. That is, we assumed participants in each condition were similar to each other in their level of experience with the radio through random assignment. The experimental control becomes more effective when the similarity between the treatment and control groups is confirmed by pre-test scores (Campbell & Stanley, 1963, p. 47-48). The first step in our analysis was to check this assumption by examining the equivalence of the control and treatment groups. We compared scores between the two groups on the pre-test score from the training (i.e., a sum score was created for the four module pre-tests), participant experience in using the MBITR/JEM (i.e., whether the participant has experience using an MBITR or JEM and amount of experience in months), participant MBITR/JEM expertise (i.e., self-rated level of expertise using an MBITR/JEM on a 4-point Likert scale), and participant experience loading keys (i.e., self-rated level of experience loading keys on an MBITR/JEM on a 4-point Likert scale).

Two sets of statistical analyses were conducted to test for the pre-treatment equivalence between the two conditions. First, a series of t-tests were employed to compare the group means of pre-test scores, participant experience and expertise with the MBITR/JEM, and participant experience in loading COMSEC keys. No significant differences between the two groups on these variables were found.

Second, the condition assignment (i.e., control vs. treatment) was regressed on a number of demographic variables through a binomial logistic regression model to examine whether the condition assignment could be explained by any systematic differences between the groups. Demographic variables entered into the model included rank, number of deployments, MBITR/JEM experience while deployed, experience and expertise with the MBITR/JEM, and experience in loading keys. As expected, neither the overall regression model nor any of the individual predictors were found to be statistically significant, thus supporting the random assignment assumption.

In addition to establishing equivalence between the two conditions, we also examined predictors of the pre-test score to further explore any systematic differences existing in the overall sample prior to treatment. A linear regression analysis revealed that the only significant predictor was whether the participant had previous experience with MBITR/JEM, explaining 55.9% of variance in pre-test scores ($F = 48.15$, $p < .001$). As expected, participants with previous MBITR/JEM experience scored significantly better on the pre-test ($\beta = .75$, $p < .001$). This was to be expected because we intended to create a sample covering a range of experience levels in order to test the full effect of the adaptive training. Through random assignment, we ensured a roughly equal mix of novices and experts in each condition. As mentioned previously, there was no statistically significant difference in pre-test score or MBITR/JEM expertise between the two conditions. Table 1 shows a breakdown of the numbers of participants with and without prior experience and the subgroup pre-test scores by condition. This further supports the conclusion that our sampling and condition assignment strategy was adequate in meeting the assumptions required of the experimental design.

Table 1. N sizes by Experience with MBITR/JEM and Experimental Condition and Subgroup Pre-Test Scores

	Number of Participants		Mean (SD)
	Non-Adaptive (N = 20)	Adaptive (N = 20)	Pre-Test Score
Experience with MBITR/JEM (Yes)	11	12	60.74 (6.82)
Experience with MBITR/JEM (No)	9	8	45.94 (6.46)

After validating the pre-treatment equivalence between the two conditions, we analyzed the data to address each of the research questions. Specifically, we examined differences between the experimental and control groups in regard to learner reactions (RQ1), training efficiency (RQ4), and training effectiveness based on the knowledge test (RQ5), and hands-on performance (RQ6). In addition, to address RQ2 and RQ3, learners in the experimental group who were able to pass over mastered lessons were identified, and their responses were compared to learners who took all or almost all of the lessons. The goal is to investigate any potential negative consequences that may result from collapsed lessons. Analyses included descriptive statistics, correlations, and t-tests. Qualitative data regarding learners' reactions to the training were coded to identify any emerging themes. In the remainder of this section, we present the results by evaluation outcomes and research questions.

Learner Reactions

Independent sample t-tests were conducted to compare between participants in the adaptive and non-adaptive conditions on the means of all survey dimensions and items. None of the comparisons yielded any statistically significant results except for one item: "I was able to access the training with minimum assistance." The adaptive group (M = 4.53, Standard Deviation (SD) = 1.02) scored significantly higher (t = 1.5, p < .05) than the non-adaptive group (M = 3.75, SD = 1.33). However, this is likely due to a network server failure that occurred during the administration of the training. While there were such incidents in both treatment conditions, the frequency of occurrence was higher in the non-adaptive condition. When asked about the technical issues encountered during the training, five participants[2] in the non-adaptive group commented on the network failure compared to three in the adaptive group.

Since the majority of the survey items were not found to be statistically different between the two groups, a summary of the descriptive statistics are presented in Table 2 in the format of aggregated percent favorable at the group level instead of the means and standard deviations. The percent favorable statistics are more easily interpretable than the means and can be used for comparison between the two groups although we were unable to rule out the possibility that the differences are due to chance. Based on the percent favorable results, participants' reactions to the training, whether it was the adaptive or non-adaptive version, were generally positive. Overall, there was no clear evidence of systematic differences in participant reactions between the two conditions. On the dimensions of Usability, Comfort with Technology, Perceived Learning, Perceived Learning Transfer and Satisfaction, there was no clear evidence that one group of participants reacted more positively than the other. However, on the dimensions of engagement and perceived adaptability, results were slightly clearer that participants

[2] Another two participants in the non-adaptive group also commented on other technical difficulties unrelated to the network issues, whereas there were no other comments on technical issues from the adaptive group.

in the adaptive group held more favorable perceptions. Conversely, participants in the non-adaptive group rated the dimension of content/ISD more favorably.

Table 2. Percent Favorable (Agree and Strongly Agree) on Learner Reaction Survey Items by Condition

	Dimension/Item	Adaptive[+] (N=19)	Non-Adaptive (N=20)
	Usability		
1.	The technology interface was easy to use.	90%	85%
2.	The technology allowed for easy review.	80%	70%
3.	I was able to access the training with minimum assistance.	85%	65%
4.	I was able to successfully operate the functionality within the training.	85%	90%
5.	I am satisfied with the technology interface.	85%	85%
	Comfort with Technology		
1.	I have used technology like this before.	60%	35%
2.	I am very familiar with this type of technology.	60%	45%
3.	I felt comfortable using the technology from the very beginning.	70%	80%
4.	It took me a while to get used to this type of technology.*	65%	70%
	Content/ISD		
1.	The course learning objectives were met.	60%	85%
2.	The design of the training was an effective way to present the subject matter.	65%	75%
3.	The material was presented in a logical sequence so that it helped me understand the subject matter.	80%	90%
4.	The media (i.e., graphics and animated sequences) appropriately illustrate the points being discussed.	70%	85%
5.	The design and presentation of material motivated me to learn.	35%	45%
6.	Overall, I am pleased with the way the training was presented.	65%	75%
	Engagement		
1.	I was engaged with the topic at hand throughout the training.	60%	55%
2.	Something about the training itself made it hard for me to focus.*	60%	45%
3.	I can recall most of what was taught during the training.	60%	50%
4.	I paid close attention to all the material throughout the training.	65%	45%
5.	I tried to access all training resources available to me, including optional content.	35%	35%
6.	The training kept me focused the entire time.	50%	40%
	Adaptability		
1.	The course content was a good fit for me as a learner.	60%	60%
2.	The training was well tailored to my knowledge and skill level on the subject matter.	55%	55%
3.	I did not have to waste time reviewing material I already knew.	65%	45%
4.	The training was too difficult for me.	5%	15%
5.	The training was too easy for me.	15%	25%
	Perceived Learning		
1.	Learning this material was fun.	30%	35%
2.	Overall, I have learned a lot from this training.	65%	75%

Table 2. Percent Favorable (Agree and Strongly Agree) on Learner Reaction Survey Items by Condition (continued)

	Dimension/Item	Adaptive[+] (N=19)	Non-Adaptive (N=20)
	Perceived Learning Transfer		
1.	It is clear to me that the people conducting the training understand how I will use what I learn.	70%	50%
2.	This training was relevant to my job in the Army.	65%	60%
3.	I believe the training will help me do my current job in the Army better.	35%	45%
4.	I learned something I can apply immediately to my work in the Army.	30%	30%
5.	I am prepared to train other Soldiers on what I learned in this training.	35%	40%
6.	I plan to use what I learned on my job in the Army.	50%	35%
7.	I am excited to apply the new skills I've learned on my job in the Army.	35%	25%
8.	I will be using the JEM on my job in the Army after the training.	15%	35%
9.	The training was of practical value to me.	55%	55%
	Satisfaction		
1.	I enjoyed this training program.	40%	45%
2.	My time on the training was well spent.	45%	50%
3.	I would recommend this training program to other Soldiers.	60%	65%

[+]One participant in the adaptive group only completed the first item on the survey.

As shown above in Table 2, participants in the adaptive condition showed less favorable attitudes on content/ISD than their counterparts in the non-adaptive condition. Most notably, 25% fewer participants (i.e., five participants in a group of 20) in the adaptive group believed the course learning objectives were met. However, when analyzing the write-in comments in the survey, no specific comments were found to specifically point to the content structure being disjointed in the adaptive condition. In fact, one participant in the adaptive condition indicated that the adaptive training was not confusing. It is thus unclear at this point whether the less favorable ratings on content/ISD by the adaptive group, if it is a meaningful difference between the two groups, indicate an unintended negative consequence of the adaptive module content sequencing.

Training Efficiency

Next, to examine the impact of the adaptive features on training efficiency, we conducted a t-test to compare the total training time spent by the adaptive group versus the non-adaptive group. Participants in the adaptive group, on average, tested out of approximately 14 out of 48 lessons. We found that participants in the adaptive group (M = 61.46, SD = 15.39) on average spent about 15 minutes less than the non-adaptive group (M = 76.05, SD = 15.35). This difference was statistically significant (t = 3.00, $p < .01$). Within the adaptive group, about half of the participants (N=11) received remediation while the other half did not (N = 9). However, participants who received remediation on average spent just over a minute less than those who did not. Therefore, we were unable to isolate the effect of the amount time spent on the training that was specifically due to remediation alone.

Training Effectiveness

Finally, we tested for any difference in training performance between the two experimental groups. Again, we utilized the t-test to compare scores from the adaptive and non-adaptive groups on the knowledge-based post-test and the hands-on performance test. Results are summarized in Table 3. None of the outcome measures were found to yield any statistically significant results between the two conditions. It should be noted that we did not include a pre-test that is a parallel form of the post-test in the study design because there are already four built-in pre-tests in the training. As a result, we were unable to compute a change score between the pre-test and the post-test. However, we were able to rely on the results from a previous evaluation study done on a suite of training technologies that included the non-adaptive version of the mobile training used for the current study (Lodato, Hyland, Mulvaney & Spain, 2014). In the previous evaluation study, the same post-test was used and participants in the treatment condition (i.e., those who went through the training prototype) scored 10.12 on average, which is comparable to the scores we found in the current study. Given that the sample used in the previous evaluation study was more experienced than the one in the current study and went through a more robust training program (i.e., the suite of training technologies that was evaluated in the previous study included the non-adaptive mobile training, a virtual classroom and an interactive collaborative scenario assessment), we are confident that the post-test scores resulted from the current study demonstrated that the training was effective.

Table 3 - Means, Standard Deviations and t-Values on Training Outcome Measures

	Mean (SD)		
	Non-Adaptive (N = 20)	Adaptive (N = 20)	*t*
Post-Test Score*	9.95 (1.90)	10.70 (2.11)	1.18
Task Completion Score**	12.85 (2.58)	12.60 (4.47)	-.22
Task Completion Time (in minutes)***	17.75 (4.14)	17.54 (4.09)	.16

*Score is out of a possible total of 14;
**Score is out of a possible total of 20;
***The limit for task completion time is 26 minutes.

DISCUSSION

The goal of the evaluation study was to investigate the extent to which adaptive training would impact learner reactions, training efficiency and training effectiveness. On the outcome of learner reactions, we expected learners taking the adaptive training to be more engaged and motivated. It is assumed that presenting training that best suits the learners' needs would lead to higher levels of engagement, higher levels of motivation, as well as fewer maladaptive responses elicited by individuals who find a task too challenging and subsequently demonstrate decreased persistence and effort. We were unable to find any statistically significant results in support of this hypothesis. However, we were able to show some preliminary descriptive evidence that participants in the adaptive condition were more engaged and held more favorable perceptions of the training adaptability. Compared to the non-adaptive group, four more participants (20%) in the adaptive group reported that they did not have to waste time reviewing material they already knew, two fewer participants (10%) found the training to be too difficult or too easy, and four more participants (20%) reported paying close attention to all the material throughout the training. In addition, we suspected that there might be potential unintended negative consequences of the adaptive training, such as seemingly disjointed or disorganized content sequencing due to the lesson collapsing feature. The evidence we found was inconclusive. While we did find that 25% fewer participants (i.e., five participants in a group of 20) in the adaptive group believed the course learning objectives were met than the non-adaptive group, we were unable to find any anecdotal evidence to further explain why. Contrarily, one participant alluded to the fact that the adaptive feature was not confusing, as one would have expected. Further research is warranted to better understand if and why the adaptive structure would lead to less positive perceptions of the training content/ISD and what could be done to alleviate such unintended consequences.

On the outcome of training efficiency, we found the adaptive training to be more efficient than its non-adaptive counterpart, as expected. The time savings by the adaptive prototype was about 15 minutes in an approximately one- to one-and-a-half-hour training; approximately a one-SD difference. We also suspected that the remediation adaptation strategy (i.e., micro-sequencing) in the adaptive training may diminish the training efficiency gained from the test-out adaptation strategy (i.e., macro-sequencing). We did not find any evidence in support of that. Participants who encountered the micro-sequencing as a result of providing incorrect answers to in-lesson interim assessment questions spent about the same amount of time on the training on average as those who answered all the check-on-learning questions correctly and were not provided the remediation adaptation.

Finally, on the outcome of training performance, we found the adaptive training to be at least as effective as the non-adaptive training. There was no statistically significant difference between the two groups on the knowledge-based post-test or the post-training hands-on performance assessment. While we did not use a pre-/post-test design that would allow us to directly compare the performance outcomes to a baseline, we can conclude with reasonable confidence that both versions of the training took effect because the post-test scores were comparable to those shown in previous evaluation studies that used the same post-test in a pre-/post-design.

Limitations

One limitation in this study was the small sample size. Due to practical and logistical constraints, we were only able to meet the minimum sample requirement that allows for adequate power to detect large effect sizes. We made every attempt to minimize Type II error by utilizing a true experiment design, including random assignment in a controlled lab setting. However, a larger sample size might have allowed us to reach more definitive conclusions on how the two conditions differed on learner reactions. In addition, if a larger sample was available, we would be able to further analyze any subgroup differences by experience levels within each condition. Another limitation was the short time lag between the training and the post-training assessments. While we were able to show the immediate effect of the training, it is unclear whether and how long any Knowledge, Skills, and Abilities (KSAs) acquisition can be retained. Future research should be conducted to examine how adaptive training may impact the sustainability of training outcomes.

CONCLUSION

In sum, the results of this preliminary study shed light on how adaptive training technologies can be used to improve training efficiency while maintaining training effectiveness. The current adaptive prototype is only showing the beginning of the tremendous potential of intelligent tutoring, among other training technology advances in support of the Army’s vision of an integrated learning environment. Our findings highlight the need to continue robust research in tandem with the development and integration of new technologies in order to realize the full potential of adaptive training.

REFERENCES

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 111-127.

Belanich, J., Sibley, D. E., & Orvis, K. L. (2004). *Instructional characteristics and motivational features of a pc-based game*. U.S. Army Research Institute for the Behavioral and Social Sciences Technical Report 1822, Arlington, VA.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research on teaching*. In N. L. Gage (Ed.), Handbook of research on teaching (pp. 171–246). Chicago, IL: Rand McNally.

Durlach, P. J., & Spain, R. D. (2012). Framework for instructional technology. In V. G. Duffy (Ed.), *Advances in applied human modeling and simulation* (pp. 222-231). Boca Raton, FL: CRC Press.

Giangreco, A., Sebastiano, A., & Peccei, R. (2009). Trainees' reactions to training: An analysis of the factors affecting overall satisfaction with training. *International Journal of Human Resource Management, 20*(1), 96-111. doi: 10.1080/09585190802528417

Jaehnig, W., & Miller, M. L. (2007). Feedback types in programmed instruction: A systematic review. *The Psychological Record, 57*(2), 219-232.

Klein, H. J. , Noe, R. A. , & Wang, C. (2006). Motivation to learn and course outcomes: The impact of delivery mode, learning goal orientation, and perceived barriers and enablers. *Personnel Psychology, 59*(3), 665-702. doi: 10.1111/j.1744-6570.2006.00050.x

Kraiger, K. (2003). Perspectives on training and development. In W. C. Borman, D. R. Ilgen & R. J. Klimoski (Eds.), *Handbook of psychology, volume 12, industrial and organizational psychology*. Hoboken, NJ: John Wiley & Sons.

Landsberg, Carla R., Astwood, Randy S., L, Townsend, Lisa N., Steinhauser, Natalie B., & Mercado, Alyssa D. (2012). Review of adaptive training system techniques. *Military Psychology (Taylor & Francis Ltd), 24*(2), 96-113. doi: 10.1080/08995605.2012.672903

Lodato, M., Hyland, J., Mulvaney, R., & Spain, R. (2014). Delivering Training Assessments in a Soldier-Centered Learning Environment: Year 2 Report. ARI Technical Report. Manuscript submitted for publication..

Orvis, K. A. , Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior, 24*(5), 2415-2433. doi: 10.1016/j.chb.2008.02.016

Park, O., & Tennyson, R. D. (1980). Adaptive design strategies for selecting number and presentation order of examples in coordinate concept acquisition. *Journal of Educational Psychology, 72*(3), 362-370. doi: 10.1037/0022-0663.72.3.362

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). New York: Academic Press.

Romero, C., Ventura, S., Gibaja, E. L. , Hervás, C., & Romero, F. (2006). Web-based adaptive training simulator system for cardiac life support. *Artificial Intelligence In Medicine, 38*(1), 67-78. doi: 10.1016/j.artmed.2006.01.002

Sinclair, J., Joy, M., Yin-Kim Yau, J., & Hagan, S. (2013). A practice-oriented review of learning objects. *IEEE Transactions on Learning Technology, 6*(2), 177-192.

Sitzmann, T., Brown, K. G., Casper, W. J, Ely, K., & Zimmerman, R. D. (2008). A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology, 93*(2), 280.

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological bulletin, 137*(3), 421.

Tulis, M., & Fulmer, S. M. (2013). Students' motivational and emotional experiences and their relationship to persistence during academic challenge in mathematics and reading. *Learning and Individual Differences, 27*, 35-46.

Valderrama, R. P., Ocaña, L. B., & Sheremetov, L. B. (2005). Development of intelligent reusable learning objects for web-based education systems. *Expert Systems with Applications, 28*(2), 273-283.

Vygotsky, L. . (1978). *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Yerkes, R., & Dodson, J. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative and Neurological Psychology, 18*, 459-482.