Technical Report 1297

Designing Adaptive Instructional Environments: Insights from Empirical Evidence

Paula J. Durlach
U.S. Army Research Institute

Jessica M. Ray
University of Central Florida
Consortium Research Fellows Program

October 2011

United States Army Research Institute
For the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

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

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

Barbara A. Black　　　　Michelle Sams

BARBARA A. BLACK, Ph.D.	**MICHELLE SAMS, Ph.D.**
Research Program Manager	**Director**
Training and Leader Development Division	

Technical review by

Jean Dyer, U.S. Army Research Institute
Cheryl Johnson, U.S. Army Research Institute

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) October 2011	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) June 2009 – June 2011
4. TITLE AND SUBTITLE Designing Adaptive Instructional Environments: Insights from Empirical Evidence		5a. CONTRACT OR GRANT NUMBER
		5b. PROGRAM ELEMENT NUMBER 633007
6. AUTHOR(S) Paula J. Durlach (U.S. Army Research Institute) and Jessica M. Ray (University of Central Florida)		5c. PROJECT NUMBER A792
		5d. TASK NUMBER 363
		5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences Orlando Research Unit ATTN: DAPE-ARI-IF 12350 RESEARCH PARKWAY ORLANDO, FL 32826-3276		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Highway Arlington, VA 22202-3926		10. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER Technical Report 1297
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.		
13. SUPPLEMENTARY NOTES Subject Matter POC: Paula J. Durlach		

14. ABSTRACT *(Maximum 200 words)*:

In adaptive instructional environments, instructional interventions and/or content can be adapted to an individual learner's competence level or other characteristics. The intention behind adaptation is to maintain the optimal level of challenge for each individual student, to provide support, and to correct misconceptions. This report reviews technology-based adaptive instructional procedures. To be included, an experiment had to describe a direct comparison of learning outcomes resulting from an adaptive system vs. a nonadaptive system, or one adaptive method vs. another. Many of the experiments used multiple adaptive interventions together, making it difficult to determine the relative contribution of the different types of adaptive interventions to the superior learning outcomes. Consequently, we were unable to conclude which adaptive techniques might be more effective than others. Nevertheless, based on our analysis, we suggest the following adaptive techniques as the most likely to provide learning payoffs: (1) Error-sensitive feedback, (2) Mastery Learning, (3) Adaptive spacing and repetition for drill-and-practice items, (4) Fading of worked examples for problem solving situations, or fading of demonstrations for behavioral tasks (such as in scenario-based simulations), and (5) Metacognitive prompting, both domain relevant and domain independent.

15. SUBJECT TERMS
ALC 2015, technology-based training, adaptive training, mastery, faded work examples, spacing metacognition feedback, intelligent tutor, student model

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	44	Ellen Kinzer Technical Publication Specialist 703-545-4225

Technical Report 1297

Designing Adaptive Instructional Environments: Insights from Empirical Evidence

Paula J. Durlach
U. S. Army Research Institute

Jessica M. Ray
University of Central Florida
Consortium Research Fellows Program

Technology-Based Training Research Unit
Stephen L. Goldberg, Chief

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

October 2011

Army Project Number 633007A792	**Personnel Performance and Training**

DESIGNING ADAPTIVE INSTRUCTIONAL ENVIRONMENTS: INSIGHTS FROM EMPIRICAL EVIDENCE

EXECUTIVE SUMMARY

Research Requirement:

As outlined in the Army Learning Concept 2015 (ALC 2015), Army training and education is undergoing a transformation to a learner-centric model. As this occurs, learning outside the classroom will play an increasingly key role. Innovative learning technologies and methods will be required to make self-directed learning effective and efficient. One of the items in the ALC 2015 Action Plan is: identify state-of-the-art adaptive training and digital tutor capabilities, and develop standards, protocols, and guidance on employing these capabilities in interactive multimedia (IMI) modules. This report identifies technology-based adaptive instructional procedures that should be considered for inclusion, based on analysis of empirical evidence. This report is also relevant to the ALC 2015 requirement: The Army requires the capability to develop adaptive digitized learning products that employ artificial intelligence/ digital tutors in order to tailor learning to the individual Soldiers' experience/knowledge level and provide a relevant and rigorous, yet consistent, learning outcomes.

Procedure:

We identified over 200 research papers of potential relevance to the issue of whether adaptive training technology provides superior learning outcomes to nonadaptive training technology. In adaptive training environments, instructional interventions and/or content is tailored to an individual learner's competence level or other characteristics, either through pretesting prior to training, or through ongoing periodic assessment during training (or both). From the original set of papers, we eliminated from further consideration papers which failed to meet our inclusion criteria for experimental design and outcome measures. The resulting 20 papers (1) met our inclusion criteria and (2) provided undisputable positive evidence of superior learning outcomes for adaptive vs. nonadaptive methods. The content of these papers was categorized for the types of adaptive methods used, so that conclusions could be made about the relative effectiveness of various adaptive methods.

Findings:

Analysis of the adaptive interventions used among the papers revealed several types. Many of the experiments combined multiple adaptive interventions together (i.e., more than one technique in the adaptive condition, but none in the nonadaptive condition). This made it difficult to determine the relative contribution of the different adaptive interventions to the superior learning outcomes. There failed to be any apparent relation between number of adaptive techniques used in a condition and effect size obtained. Likewise, most of the experiments used multiple sources of student data, making it difficult to identify which sources were best for adaptation. The most common sources of student data were performance measures captured during the instructional experience. ALC 2015 places an emphasis on pretesting in order to

adapt content to the individual learner's experience and competence level. We failed to identify any experiments with positive results, which used pretest data alone to adapt instruction. We therefore recommend caution in over-reliance on pretest data, as compared with performance data collected during the learning experience.

We conclude from our review that there is evidence for beneficial effects of adaptation; however, the nature of the empirical data prevent us from concluding which specific adaptive techniques work best for different learning contexts. Instructional design know-how for adaptive systems is not mature enough to enable the mass production of effective adaptive learning environments, without the input of experienced human designers who can make both qualitative and quantitative expert design judgments. Yet, the following adaptive techniques are likely to be ones that will support learning payoffs: (1) Error-sensitive feedback, (2) Mastery Learning, (3) Adaptive spacing and repetition for drill-and-practice items, (4) Fading of worked examples for problem solving situations, or fading of demonstrations for behavioral tasks (such as in scenario-based simulations), (5) Metacognitive prompting, both domain relevant and domain independent.

Utilization and Dissemination of Findings:

The results and recommendations presented here should be of interest to designers and developers of technology-based training and education, and personnel involved in the implementation of ALC 2015. The adaptive techniques recommended here should be considered when designing any future technology-based training and education. Future specifications for procurement of technology-based training and education should include requirements for adaptive techniques like those listed here.

This report has been sent to TRADOC Capability Managers for dL and for Army Training Information Systems. The results were briefed to Army Training Support Center in June 2011. A copy of this report has been posted on the Army Learning Concept 2015 website.

DESIGNING ADAPTIVE INSTRUCTIONAL ENVIRONMENTS: INSIGHTS FROM EMPIRICAL EVIDENCE

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

DESIGNING ADAPTIVE INSTRUCTIONAL ENVIRONMENTS: INSIGHTS FROM EMPIRICAL EVIDENCE

Introduction

As outlined in the Army Learning Concept 2015 (ALC 2015), Army training and education is undergoing a transformation to a learner-centric model. As this occurs, learning outside the classroom will play an increasingly key role. Innovative learning technologies and methods will be required to make self-directed learning effective and efficient. The Army requires the capability to develop adaptive digitized learning products that employ artificial intelligence and digital tutors in order to tailor learning to the individual Soldiers' experience and knowledge, and which provide relevant and rigorous training with consistent learning outcomes. One of the items in the ALC 2015 Action Plan is: identify state-of-the-art adaptive training and digital tutor capabilities, and develop standards, protocols, and guidance on employing these capabilities in interactive multimedia Instruction (IMI) modules. This report identifies technology-based adaptive instructional procedures that should be considered for inclusion, based on analysis of empirical evidence.

One-on-one education and training by a human mentor is the epitome of adaptive instruction, and has been shown to be superior to traditional classroom-based approaches (e.g., Bausell, Moody & Walzl, 1972; Bloom, 1984). An ideal human tutor combines what they know about the student, about effective instructional strategies, and about the domain, to flexibly adapt during each teaching moment. The challenge for a software tutor is to represent and employ similarly rich knowledge and behavioral flexibility. Creating such a software tutor can be a tremendous undertaking. Attempts to do so have required multi-skilled teams of personnel, conducting iterative research over a period of years (e.g., Graesser et al., 2004; Koedinger& Aleven, 2007; Koedinger & Anderson, 1998; VanLehn et al., 2005). These artificially intelligent tutoring systems (ITS) typically consist of several component models, which interact to control the student experience. These components correspond to the knowledge used by human tutors: a student model—knowledge about the student, a pedagogical model—a set of instructional strategies and behaviors, a domain model—knowledge about the subject being taught, and an expert model—knowledge of how to solve problems in the domain.

If the goal is to improve learning outcomes from software-based education and training, then one might ask, across the different adaptive software systems that have been developed, what has been their success in improving learning outcomes, and are there specific common features across systems which have proven successful? The purpose of this paper is to present the results of such an analysis. In their review of computer-based adaptive learning environments, Vandervaetere, Desmet, and Clarebout (2011) stated that there was considerable variation in system design and sparse data related to empirical effectiveness with respect to enhancing learning outcomes. While they enumerated various techniques that have been used, they did not provide a detailed cross-walk of these against evidence. This review endeavors to accomplish this. Vandervaetere, et al. (2011) defined adaptive learning environments as those which accommodate the different learning needs and abilities of different learners. Similarly, Shute and Zapata-Rivera (2008) define adaptivity as the capability of a system to alter its behavior according to learner needs and other characteristics. Landsberg, et al. (2010) offer a somewhat more detailed definition: *"training interventions whose content can be tailored to an individual learner's aptitudes, learning preferences, or styles prior to training and that can be adjusted, either in real time or at the end of a training session, to reflect the learner's on-task performance" (p. 9).*

There is good evidence that ITS produce benefits when used to supplement regular classroom instruction (e.g., Koedinger, Anderson, Hadley, & Mark, 1997). There is also very good evidence that adaptive software systems produce learning (e.g., Anderson, et al., 1995; Graesser et al., 2004; VanLehn et al., 2005). These facts are not in doubt; but neither are they the question addressed here. Our question is concerned with the comparison of adaptive to nonadaptive technology-based learning environments: is there evidence for the benefits of adaptation when all other factors are held constant? Thus, we are seeking evidence, not that ITS produce benefits when used to supplement regular classroom instruction, but that they provide greater benefits in this regard than a parallel nonadaptive system. Likewise, we are seeking evidence, not that adaptive educational software produces learning, but rather that it produces superior learning compared to parallel nonadaptive software. We identified over 200 papers on adaptive educational systems; however, only a small subset of these addressed this specific question.

While examining these papers and considering how to organize our findings, we found it necessary to get more specific about the definition of "adaptive." Interactive systems alter their behavior based on what the user does; so clearly, interactive systems are a superset of adaptive systems. It is fairly well-established that interactivity supports learning to the extent that it focuses cognitive processing on the central concepts and principles to be learned (Chi, 2009; Renkl & Atkinson, 2007). Such focusing can be effective in improving learning outcomes by taking into account the nature of student errors, rather than just whether the student made an error; but is that adaptive?

This issue can be understood more clearly through the use a concrete example. Imagine computer-based instruction intended to teach four concepts (let's call them A, B, C, and D). Students are given a description of a situation and have to indicate whether the situation exemplifies A, B, C, or D, with the item presented on each question chosen randomly from a supply of examples. The student provides an answer and receives immediate feedback, correct or incorrect, and then is presented with the next item. This case is unambiguous: interactive, but not adaptive. Now consider a slightly modified procedure. Let's suppose if the student answers incorrectly, the system presents an explanation of why the choice selected was incorrect; so, a student erring by selecting B (when the correct choice was A) would get an explanation of the difference between an A and a B, whereas a student erring by answering C would get an explanation of the difference between an A and a C. It is well-established, that providing feedback like this, which tries to repair errors in understanding, improves learning, compared to accuracy information alone (e.g., Azevedo & Bernard, 1995; Gouli, et al., 2006, Jaehnig & Miller, 2007; McKendree, 1990); but it is not entirely clear whether this should be considered adaptive, because it does not explicitly use information about individual student differences (only about answer differences). We have adopted the policy of calling this type of interactivity local adaptation. It takes into account the fact that students can be incorrect in different ways and tailors the feedback provided specifically to those different ways. However, it does so taking into account only a single response on a single individual item. Hence, the adaptation occurs on local information only. This can be contrasted with model-based adaptation, in which even richer information about student differences is used to adapt content.

A further modification of the example will illustrate model-based adaptation (schematized in Figure 1). Now imagine that there is a database (the student model) that maintains a record of student performance, in the form of moving averages of accuracy on each concept, and in the form of a history of the order in which items from the different concepts have been presented. The system presents feedback based on local information, just like in the previous example; but in addition, it selects the next item by considering information in the database. Let's imagine that after each response, the moving averages and the sequence history are updated, and then used by an algorithm to select the next item. The algorithm is based on a theory of learning, taking into account concept accuracies (for that student) and concept spacing (i.e., how long since the student was presented with an item from each concept). Thus, the selection of the next item is model-based, requiring the historical data kept in the student model (for description of a specific retention algorithm, and the theory behind it, see Metzler-Baddeley & Baddeley, 2009). Note, the decision – how to adapt—is also model-based (i.e., the algorithm).

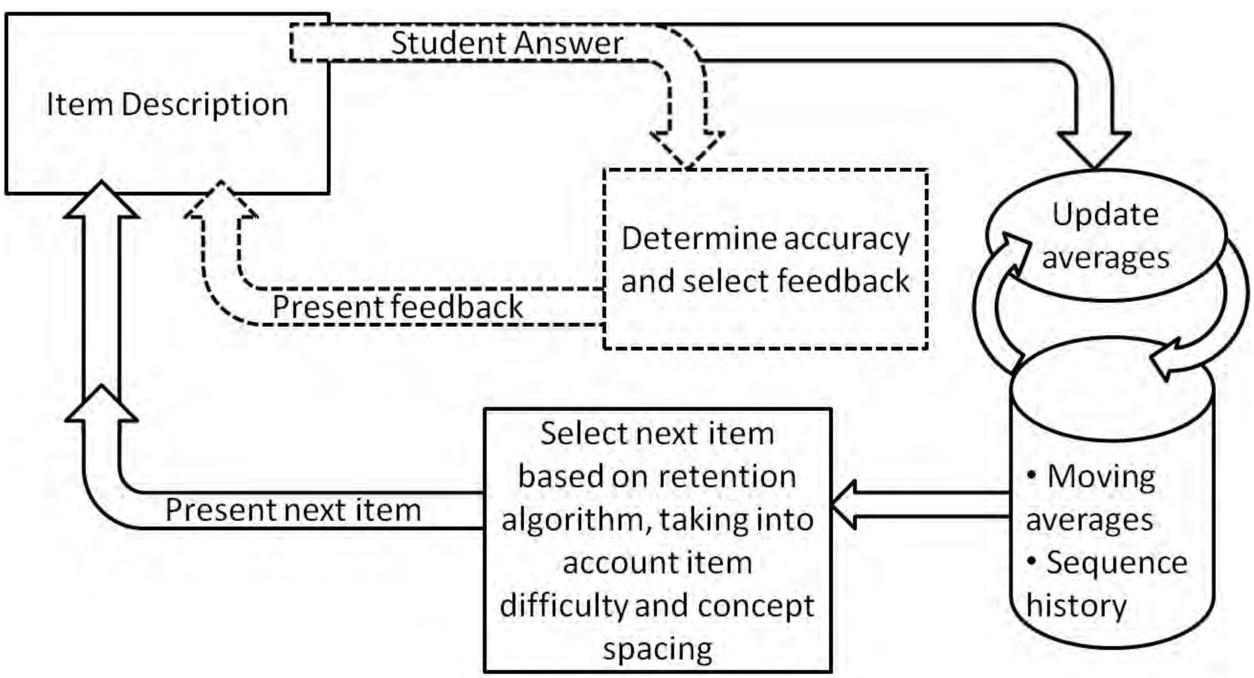

Figure 1. Schematic illustration of a concept training system with local adaptation selecting feedback (dashed circuit), and model-based adaptation selecting the next item.

While the distinction between local adaptation and model-based adaption seems obvious to us now, it was not when we started reviewing the literature. Once we recognized it, we struggled to distinguish clearly the properties of interactive learning environments that are not locally adaptive vs. ones that are. The standard levels of IMI do not really address this issue. These levels describe progressively greater degrees of interaction between the learner and the software, ranging from Level I, in which the learner is a passive recipient of information, to Level IV, in which the learner is immersed in a lifelike simulation. They do not address different instructional strategies, and consequently, do not separately classify software that provides correct/incorrect feedback only, vs. software that accompanies such feedback with material intended to repair student errors (example 1 vs. example 2 above). Rather than bore the reader with our mental machinations about possible definitions, suffice it to say that we decided to concentrate this review on model-based adaptation, thus relieving us of the burden of having to review the entire IMI literature.

Search Protocols and Inclusion/Exclusion Criteria

We selected five web-based databases: *PsycInfo*, *Academic Search Premier*, *Web of Knowledge*, *Defense Technical Information Center* (DTIC), and the *Interservice/Industry, Simulation and Education Conference* database to search for peer-reviewed papers that had been published since 1985. Within each of these databases we searched using a combination of the following terms: “Intelligent Tutoring System”; “Adaptive Training”; “Computer-Assisted Instruction” + ”Adaptive”; “Computer” + “Learning”; and “Computer” + ”Adaptive.” The number of papers identified was 181. Each of these was examined to see if the paper contained a direct comparison of learning outcomes resulting from an adaptive technology vs. a nonadaptive technology (here, using adaptive in undifferentiated sense), or a comparison of two or more adaptive technology implementations. To be retained for review, the comparison needed to involve two or more systems which were as alike as possible, save for the adaptive variable. So for example, an experiment comparing the results of classroom teaching with vs. without supplemental ITS use would not be included. Nor would one comparing learning in a traditional classroom vs. learning with an ITS; however, an experiment comparing learning from nonadaptive computer-based practice vs. adaptive computer-based practice would be included. In other words, all features of the supplemental practice had to be the same except the adaptation. In addition, in order to be included, the experiment had to have a measure of learning gains, assessed either immediately after training or after a period of retention. So for example, experiments that solicited student feedback about the learning environments, but did not directly assess learning gains, were not included (e.g., Moundridou & Virvou, 2002). We also required that the measure of learning gains be taken outside of the learning environment itself, to avoid the possibility that gains were due to learning about the system itself, as opposed to knowledge acquisition in the target domain.

Having identified only 17 papers meeting our criteria, the papers were analyzed more deeply and their references were used to locate additional papers, not necessarily identified in the initial database search. In turn, relevant references from new papers were collected, and so on. This was also the point at which we decided to distinguish local and model-based adaptation. Consequently, our search became more targeted on finding evidence about model-based adaptation and we therefore did not include newly found papers for which the abstract clearly indicated strictly local adaptation after this point. Several adaptive systems used both model-based and local adaptation. For example, Suraweera & Mitrovic (2004) found superior learning with their ITS compared to a nonadaptive version. The ITS contained both model-based and local adaptation, and the nonadaptive version had neither. For experiments such as this, when local and model-based adaptation were confounded, we placed it in the model-based category.

At this point, we also decided to disqualify experiments with matching/mismatching procedures. In matching/mismatching procedures, one experimental condition (matched) receives adaptations intended to be optimal for some student trait (e.g., cognitive style or learning style), whereas another condition receives adaptations deliberately intended to clash with the trait (mismatched). This procedure is typical of experiments examining aptitude by treatment interactions (see Paschler, McDaniel, Rohrer, & Bjork, 2008), and does use information about individual differences to make instructional decisions; however, it does not include a condition in which individual difference information is simply ignored (nonadaptive). Assuming that learning outcomes are superior in the matched than the mismatched condition, the problem is that this experimental design does not provide a baseline. That is, one cannot distinguish whether the matched condition produces better outcomes than a nonadaptive baseline condition, or whether the mismatched condition produces worse outcomes than the baseline (or both).

In summary, we ended up with two groups of papers, one for which the experimental manipulation involved local adaptations only, and one for which the manipulation involved model-based adaptations or a combination of model-based and local adaptations. Note that for the local-only group, some functions of the system may have used a student model; but, not for the manipulation that distinguished the experimental conditions. For example, both conditions might have selected content sequence based on a student model of mastery; but the type of feedback provided differed as a result of local information (e.g., Aleven & Koedinger, 2002). For the local adaptations, we acknowledge that our collection is in no way exhaustive; but, we nevertheless think our findings are worthy of presentation. For the combined category, we feel more confident that the collection is a relatively thorough review of the existing literature.

We will only discuss papers which met all our inclusion criteria, and found a statistically significant improvement in learning outcomes from adaptive vs. parallel nonadaptive training technology, or variants of adaptive strategies. This is because it is impossible to make a conclusion one way or another on the basis of failure to find a significant difference (Dallal, 2007). A failure to find a difference can be caused by factors other than the ineffectiveness of the manipulation of interest. In training effectiveness evaluation, for example, an effect may fail to be evident if the assessment measure lacks sufficient sensitivity. It takes a much more sensitive test to measure different degrees of learning than it does to measure whether any learning occurred at all.

For some of these “null result” papers, the researchers did find some evidence favoring the adaptive manipulations by conducting *post hoc* comparisons, which were not in their original analysis plan (e.g., Conati & VanLehn, 2000; Kavcic, 2004; Lane, & VanLehn, 2005). We retained these papers if the statistical techniques were appropriate for *post hoc* comparisons and an alternative interpretation for the *post hoc* results (i.e., alternative to the authors’) was not obvious.

We found no papers reporting significantly poorer learning outcomes from adaptive vs. nonadaptive systems. There were a few cases in which the students took longer to complete their work in the adaptive learning environments than the nonadaptive ones, with no statistically significant compensatory gains in learning outcomes (e.g., Goetzfried & Hannafin, 1985; VanLehn, et al., 2007).

Benefits of Local Adaptation

As previously stated, we did not attempt a thorough review of local adaptation, as that could potentially cover any form of IMI. Depending on the definition of adaptation, it could include passive forms of learning where the only form of interaction is pressing a “Next” button. Even insisting on a higher level of interaction, it could still encompass the literature on different methods of providing feedback (for a relatively recent review of this literature, see Jaehnig & Miller, 2007). Table 1 presents a summary of the strictly locally adaptive experiments, which were discovered during our literature analysis that contained positive evidence, and we deemed especially innovative and in keeping with the spirit of what it means to be adaptive (not just interactive). Each of the experiments uses a different form of adaptation. While no pattern of adaptive strategies immediately pops out, there is an underlying theme suggested by four of the papers in this collection (all but Park & Tennyson, 1986): Students benefit from support on self-evaluation and self-explanation. Self-evaluation refers to assessing one’s own knowledge (Do I understand? Did I make a mistake?); and following on from that, taking steps to remediate oneself or locate errors and self-correct. Self-explanation is a particular strategy of self-evaluation. It refers to explaining to oneself some aspect of the learning material (e.g., putting the information in one’s own words, or reasoning out why Y follows from X). It is a way of checking whether something is really understood. Several studies have found that learning is more effective when students explain examples to themselves, and this has come to be referred to as the self-explanation effect (Chi, Bassok, Lewis, Reimann, & Glaser , 1989; Chi, de Leeuw, Chui, & Lavancher, 1994; Johnson & Mayer, 2010). VanLehn and Jones (1993) reasoned that self-explanation causes students to uncover gaps in their knowledge and then fill them. Unfortunately, many students do not spontaneously engage in this behavior, and thus require encouragement.

Table 1

Positive Evidence for Improved Learning Outcomes with Local Adaptation. (see Appendix A for explanation of effect sizes)

Citation: Aleven & Koedinger (2002)	
Context	15-16 year old students used the Geometry Tutor to learn about angles, as a supplement to normal classes. N= 11 and 13 in the adaptive and nonadaptive conditions, respectively.
Measures of Learning	Pretest and Posttest, containing problems similar in form to those practiced with the tutor, and transfer problems, which required the same conceptual knowledge, but were presented in a new format. Besides solving problems, students had to justify their answers in terms of geometry definitions and theorems. Cohen’s *f* effect size for pretest to posttest gain, averaged across different problems = 0.46.

Basis for Adaptation	Ability of students to explain their problem solving steps.
Adaptation	In the adaptive (explanation) condition, if students were incorrect in explaining problem solving steps, they were given hints as to how to identify the correct explanation. 5 levels of hint were available, which became increasingly detailed. In the nonadaptive condition, students were not required to explain their steps. Note, both conditions applied student-model based mastery approach to problem selection and provided hints for problem steps.
Citation: Forbes-Riley & Litman (2011)	
Context	41 participants, who had never taken college physics, spent 20 – 40 minutes reading a physics text, then took a pretest. They then used the software to complete 5 qualitative physics problems and took a posttest.
Measures of Learning	26-item multiple choice pretest and posttest. Effect size on posttest scores as measured by Hedges g* = .86.
Basis for Adaptation	Uncertainty Note: a human performed speech recognition, natural language understanding, and uncertainty judgment.
Adaptation	The student was provided with additional tutoring content (automated) after every incorrect student answer and after every correct answer if uncertainty was detected. In the nonadaptive condition, the student was provided with additional tutoring content only after incorrect answers. Note: result likely not due to additional tutoring alone, as two other conditions also received extra tutoring but did not learn significantly better than the nonadaptive condition.
Citation: Kalyuga & Sweller (2004)	
Context	26 high school students participated in a 30 – 50 minute session, solving algebraic equations.
Measures of Learning	Pretest and posttest using rapid diagnostic testing procedure: Student had to provide their first step in solving a problem, rather than the whole solution. Scoring based both on correctness and how many mental steps contributed the first typed step. Cohen's *f* effect size = 0.46.
Basis for Adaptation	Results on rapid diagnostic testing both prior to and during problem solving.
Adaptation	Faded worked examples: Students were given problems that were partially solved, and had to supply the missing parts of the solution. The degree to which the first problem was solved depended on the individual's score on the initial rapid diagnostic test (the poorer the score, the more of the problem was already solved). Subsequently, it depended on problem solving performance and intermittent rapid diagnostic testing. Thus, the number of steps the student had to complete in each problem was gradually increased, based on ability to correctly complete preceding examples. Each student in the nonadaptive condition was yoked to a student in the adaptive condition; i.e., they received the same pattern of worked example fading as a participant in the adaptive condition.
Citation: Mathan & Koedinger (2005)	
Context	Adults with general computer experience, but spreadsheet novices, learned about using spreadsheets, over 3 sessions. During the first session, they were

	given 90 min of instruction and procedural practice. Sessions 2 and 3 involved procedural practice with different versions of the training software. The versions differed on how feedback on errors was given.
Measures of Learning	Session 2: Pre- and posttests involving practical problems and questions on conceptual understanding. Session 3: 8 days after Session 2, pre- and post- "transfer" tests with exercises calling upon cell-referencing skills in the context of a structurally complex spreadsheet. Experiment 1 effect sizes were problem solving--0.50, conceptual understanding-- 0.59, transfer--0.43, and retention--0.33. Experiment 2 effect sizes were problem solving--0.62, conceptual understanding—1.05, transfer--0.78, and retention--0.70. Method of calculating effect sizes not given.
Basis for Adaptation	This study compared the effect of 2 different ways of adapting to student errors while problem solving. Thus, the basis for adaptation was detection of an error.
Adaptation	In the immediate condition, the learner received feedback as soon as an incorrect formula was entered. Upon an error, they could try to correct the error on their own or ask for help. Help interactively guided learner to the solution. In the delayed condition, the learner was not notified of errors until they deemed the solution complete. At that point an error triggered feedback to check for errors; as the learner attempted to correct their solution, they were given feedback in the same manner as the immediate condition.
Citation: Park and Tennyson (1986)	
Context	72 11th grade social studies students learned about the psychology concepts: positive reinforcement, negative reinforcement, positive punishment, and negative punishment. All students received initial instruction on the concepts including example situations of each. During computer-based training, they were given a series of situations and were asked to indicate which concept was exemplified by the situation. Students continued training until they reached a criterion of 75% correct, adjusting for guessing.
Measures of Learning	Posttest: 24-item multiple choice test, given immediately after learning. A retention test given 1 week later, repeated the posttest and also required students to write definitions of each concept. Effect size as measured by Hedges g^* = 1.01 and 1.04 for the immediate and delayed multiple choice tests, respectively, and for definition writing = 1. 18.
Basis for Adaptation	This study compared the effect of 2 different ways of adapting to student errors. Thus, the basis for adaptation was detection of an error.
Adaptation	In both conditions, an error produced feedback, and the next example was either from the concept category of the correct answer or the concept category of the erroneous answer. The conditions differed by whether the example given after an error was presented as another question (interrogatory) or as remediation (expository). In the latter case, the concept and its definition were given along with the example. Students in the expository condition performed significantly better on all the measures of learning than those in the interrogatory condition.

Referring back to Table 1, the simplest adaptive intervention, that of Park and Tennyson (1986), provides evidence that remediation on errors improves final learning outcomes, compared to merely providing knowledge of results (correct vs. incorrect). This is a finding already clearly established in the feedback literature (e.g., Jaehnig & Miller, 2007). The results of Forbes-Riley and Litman (2011) build on this, showing that remediation is beneficial not only on errors, but also when the student is correct but uncertain. Presumably, if a student were self-evaluating while studying and were uncertain of an answer, they would (or should) provide themselves with remediation. Thus, the Forbes-Riley and Litman (2011) procedure could be viewed as supporting remediation that would follow self-evaluation. Two of the experiments directly demonstrated benefits from requiring students to self-evaluate, either by locating their own errors (Aleven & Koedinger, 2002) or by supplying explanations for problem solution steps (Mathan & Koedinger, 2005). Finally, in the fifth experiment (Kalyuga & Sweller, 2004), the beneficial procedure was the adaptive fading of worked examples in the context of solving algebraic expressions. Worked examples are step-by-step demonstrations of how to perform a task or solve a problem, and are commonly provided to novice learners in many contexts. Particularly for problem solving, they support self-explanation by providing the opportunity to reason through the rationale for each step without the additional burden of having to work out the solution for each step as well. The rationale for fading worked-examples is that as the student becomes more knowledgeable about reasoning and procedures, the burden of conducting the procedures should be shifted onto to them, essentially keeping the cognitive demands about the same throughout (Sweller & Cooper, 1985). The Kalyuga and Sweller (2004) paper showed that using student performance on the previous problem to govern the fading process results in better learning outcomes than fading according to an arbitrary schedule.

In summary, at the surface level, the collection of papers in Table 1 may seem rather heterogeneous; however, there is an underlying current indicating benefits for adaptation aimed at supporting student self-explanation and self-evaluation. These activities foster a deeper understanding of the conceptual aspects to be learned (the "why" as well as the "how"). Thus, for future adaptive training technology development, including adaptive support for self-explanation and self-evaluation appears to be a strategy worth including.

Benefits of Model-Based Adaptation (or Combined Model-Based and Local Adaptation)

Table 2 presents a summary of experiments with positive evidence for improved learning outcomes for student model-based adaptation or combined model-based and local adaptation. Table 3 summarizes the entries in Table 2 in terms of the different types of adaptive interventions that distinguished the adaptive from the comparison conditions. These will be briefly explained, before examining the evidence.

Mastery, Level of Detail or Difficulty

Mastery refers to the technique of tailoring the content to the student's current level of understanding. Students are not allowed to advance to the next level or module until they master the content of the current one. They are given additional instruction or practice until they do. Traditionally, the mastery learning technique gates advancement through the materials; however, a variation of the traditional approach is to adjust the instructional content in addition to gating advancement. For example, Tseng, et al. (2008) had three ways of presenting content: “Easy,” with very detailed content, including a review of prerequisites as well as new basic concepts, “Middle,” with detailed descriptions of the new basic concepts but only the most relevant prerequisites, and “Difficult,” with only brief descriptions of basic concepts and some advanced concepts. The version presented on module N+1, depended on student performance on module N, with the difficulty set higher for better performing students. In addition, if a student failed to pass an end-of-module test, they redid the module at a lower level of difficulty (if available). We’ve labeled this type of variation on the mastery learning technique “level of detail or difficulty.”

During Problem Guidance

There are various automated methods of providing guidance to a student in the midst of an exercise. In some systems, help must be explicitly requested by the student. In other systems, a hint is performance-triggered. It might be provided after some period of time without the correct action in a simulation; or after an incorrect answer in response to a problem step. One common method of providing guidance is to have multiple hints available for the same issue. Each successive hint is more directive than the previous, with the last one, “the bottom-out hint”, providing the correct response. Guidance can be based on local information only, or use information from a student-model. Wood and Wood (1999) suggested that the more knowledgeable the student, the more abstract the hint should be; the less knowledgeable, the more detailed. In addition, there is some evidence (*post hoc* only) suggesting that unsolicited help might be better for some students, whereas requested help might be better for other students, depending on the student’s motivation and/or ability to self-evaluate. Thus, the guidance can be adaptive in terms of when to give it; but, then nonadaptive thereafter (i.e., the sequence of potential hints is fixed). Alternatively, it can be adaptive both as to when to give it, and how to give it. Because these alternatives have not been rigorously compared, we have grouped them into one category.

Tutoring Dialogs

As previously mentioned, self-explanation has been shown to be a highly important element of learning (Chi et al., 1994). For this reason, a number of automated tutoring systems currently use natural language processing techniques to engage students in interactive dialogues, which prompt students to elaborate on answers, trying to approach an automated version of Socratic dialog. These tutoring dialogs often supply during-problem-solving guidance and motivational support. One example is the CIRCSIM tutor (Zhou et al., 1999), which helps students learn circulation principles. Another is AutoTutor, which helps students learn physics (VanLehn, et al., 2007). Different systems use different techniques to model the dialog process and to compose the automated tutor’s side of the dialogue. Despite differences in models, the aim is generally the same, which is to get the student to reason and apply principles in the context of solving problems in the training domain.

Error-sensitive Feedback

As discussed earlier, error-sensitive feedback refers to feedback that provides information relevant to the specific error made. So, rather than just whether an input was correct or incorrect, or what the correct response was, the feedback is aimed at repairing student misunderstandings, with information about why the student's response was erroneous. Some systems include "bug libraries," containing common student misconceptions. These help the software to diagnose the nature of a student error, and supply tailored corrective information.

Self-correction

Self-correction refers to encouraging students to locate and correct their own errors. Rather than being told as soon as an error is committed, feedback may not occur until several problem steps or actions have been taken. Upon feedback regarding a solution flaw, the student is required to attempt to correct the solution themselves. If they cannot, guidance on locating and fixing errors may be provided.

Fading Worked Examples

As discussed earlier, worked examples are step-by-step demonstrations of how to perform a task or solve a problem. Fading worked examples refers to an instructional technique in which the amount of the example that is solved is gradually reduced. The student is required to complete the unsolved steps. Over time, the student goes from reviewing completely worked out examples, to solving entire problems. The process can also include requiring students to justify solution components.

Hyperlink annotation and Direct Navigation Support

These are adaptive techniques used in adaptive educational hypermedia systems. Educational hypermedia systems use graphics, audio, video, plain text, and hyperlinks to create a non-linear medium for instruction. Adaptive navigation support techniques are used to guide users through hyperspace by annotating links or making direct next-link suggestions, based on the goals, knowledge, and other characteristics of an individual user (Brusilovsky, 2003). Hyperlink annotation refers to the technique of annotating hyperlinks (usually with colors) to indicate something about the material at the linked site. For example, if a student has not completed the prerequisite learning to understand the material at the link, it might be presented in red. Alternatively, the link itself might be disabled if the student is not prepared to go there (link hiding). Direct navigation guidance refers to recommending the link a student should go to next. In some systems, the student must follow this direction, in others it is only a suggestion.

Metacognitive Prompts

Metacognitive prompts encourage students to carry out specific metacognitive activities while learning. These include activities like self-explanation and self-evaluation discussed in the previous section. The prompts are intended to focus learners' attention on their own mental activities while learning. We have already talked about tutorial dialogs, which are intended to get students to self-reflect and elaborate, in the context of a discussion about solving a problem. We use the term metacognitive prompts in Table 3 to refer to domain-independent prompts (such as asking, did you understand the main point of the last paragraph?).

Spacing and Repetition of Domain Problems

This technique incorporates what is known about learning and memory from laboratory-based studies, in which the same activity occurs repeatedly (e.g., memorizing vocabulary meanings). Spacing refers to the finding that once an item is mastered, retention can be maintained best by increasing the time (or spacing) between subsequent repetitions (within a practice session). Repetition refers to the finding that more difficult items require more repetitions to learn than easier ones do. So, a student with a history of erring on item A 50% of the time and item B 75% of the time will be given more repetitions of item A than B.

Other

One experiment used multiple other techniques involving content presentation order, feedback detail, guidance style, and organizational tools, based on an assessment of cognitive style, specifically whether the student was judged to be an analytic (field independent) or holistic learner (field dependent). Because the manipulation involved multiple elements, and none of the other experiments used any of these, we have simply labeled this as other.

Examination of the Evidence

With its columns explained, we can now turn to a discussion of the contents of Table 3. The purpose of Table 3 is to make it easier to see the potential causes of learning benefits across the experiments. The rows represent each experiment listed in Table 2, the columns represent different adaptive techniques potentially responsible for the experimental results, and the shaded cells represent the adaptive techniques that distinguished the adaptive vs. nonadaptive conditions in each experiment. If an adaptive technique is not represented by a shaded cell in Table 3, it does not necessarily mean that it was not employed. It may have been employed in both conditions, and thus could not be responsible for the observed effects. For example, in the Salden et al. experiment (2009, 2010) students in both the adaptive and the nonadaptive conditions were required to explain their problem solving steps (note: these two papers present the same data set). Self-explanation was not listed as a column in Table 3, because it did not differentiate conditions.

Table 2

Positive Evidence for Improved Learning Outcomes with Student Model-based Adaptation. (See Appendix A for explanation of effect sizes, η^2 and η_p^2)

Citation: Anderson, Boyle, & Reiser, (1985)	
Context	Undergraduate students enrolled in a LISP programming course attended lectures and completed normal class assignments. In addition, they completed extra programming exercises, either with the aid of an ITS (N= 10), or on their own (N=10).
Measures of Learning	Results on course final exam. Not enough information given to calculate effect size.
Basis for Adaptation	Student programming steps during exercise completion were compared to steps produced by a cognitive model of LISP programming. A mismatch triggered adaptation. Besides producing the correct solution, the model also could recognize common errors. In addition, the program kept track of number of false starts to a solution.
Adaptation	When a mismatch occurred, the student was notified of an error and was required to correct it. If the type of error was recognized, diagnostic information (nature of the error) was provided. Upon student request or detection of a criterion number of false starts, student was guided through problem analysis. In the nonadaptive condition students received no guidance or feedback.
Citation: Chien, Yunnus, Ali, & Bakar (2008)	
Context	12 and 13 year olds learned about algebraic expressions. Instruction (30 min) was delivered by a commercially available computer-aided instruction (CAI) program, then students worked on exercises for five hours, spread over 8 school days. Total N = 62 (31 per group).
Measures of Learning	Gain in proficiency (posttest – pretest). Cohen's *f* effect size = 0.64.
Basis for Adaptation	Pretest performance and analysis of exercise solutions during practice.
Adaptation	Exercise selection, step by step exercise guidance, suggestions for improving performance. In the nonadaptive condition, students did the exercises with the CAI program, which provided correct vs. incorrect feedback only, on exercise solutions.
Citation: Corbalan, Kester, & van Merriënboer (2008)	
Context	First year students in vocational education in the health sciences completed learning tasks in the domain of dietetics, for entry into a lottery. N = 15 and 13 in the adaptive and nonadaptive conditions, respectively.
Measures of Learning	Conceptual knowledge test (paper and pencil), with 20 multiple choice questions, given one week after training. Proportion of variance accounted for by adaptive manipulations η_p^2= .087; Hedges g^* = .71.
Basis for Adaptation	Adaptation started at problem 3, using the data from problem 2. After each problem students answered 6 multiple choice questions. Scores on these questions were combined with score on problem performance to create a

	competence measure (C). Students also rated (1 to 7) “effort required to complete the task.” C and effort score were used to select the support level of next problem. The higher C and the lower effort, the bigger the decrease in support level. Advancement in problem difficulty occurred when a problem was completed successfully without support (support level 5).
Adaptation	Level of support provided with each problem and problem difficulty. Problems could be presented with one of 5 levels of support: (1) worked-out examples with solution steps and rationale, (2) worked-out examples with solution steps, (3) almost completed problems, (4) somewhat completed problems, (5) problems needing full completion. Problems could be presented at 5 levels of difficulty (defined by domain experts). Each participant in the nonadaptive condition was yoked to a participant in the adaptive condition (i.e., received same sequence of problems as one person in the adaptive condition).
Citation: Davidovic, Warren, & Trichina (2003)	
Context	Undergraduate students spent 20 – 60 minutes learning about recursion in JavaScript; students were prescreened for prerequisite knowledge of JavaScript and programming ability. Learning consisted of instruction, examples, and exercises. N per condition not provided. Experiment was not part of a class.
Measures of Learning	Gain in proficiency (posttest – pretest), as measured by ability to program two recursion problems. Not enough information to calculate effect size.
Basis for Adaptation	Pretest results, exercise solutions (multiple choice questions, phrase insertion, example structure exercises)
Adaptation	1. Hyperlink annotation* 2. Direct navigation guidance** 3. Two levels of hints to correct errors before giving correct answer. In nonadaptive condition, student chose navigation path without assistance, and were immediately given the correct answer upon an error.
Citation: Metzler-Baddeley & Baddeley (2009)	
Context	Memorization of Spanish vocabulary in a lab study. Each student was asked to memorize two sets of 35 Spanish-English word pairs. They were given a Spanish word and had to produce the English equivalent. Learning of each set occurred on different days, 2 weeks apart. N=32 university undergraduates.
Measures of Learning	Performance on test of training items presented with random order and spacing, both immediately after training and also 2 weeks later. Cohen’s *f* effect sizes = 0.96 and 0.90 on the immediate and delayed posttests, respectively. Note: there was a large forgetting effect (immediate test vs. delayed) which was substantially larger (about 20 words) than the effect of adaptation (about 5 words); Cohen’s *f* for forgetting = 11.26. Delay and adaptation did not interact.
Basis for Adaptation	Timing and quality of student response, combined with history of previous presentation pattern supplied to algorithm, which calculated optimum timing of next repetition to maximize retention, based on known characteristics of learning and forgetting.
Adaptation	Spacing between repetition of words and number of repetitions of each word-

	pair. In nonadaptive condition, spacing and repetitions were random.
Citation: Perrin, Dargue, & Banks (2003)	
Context	Biannual refresher training for employees on export control rules. Multimedia content was used to present instruction. Each block of instruction was followed by multiple choice questions (test). N= 25 per condition.
Measures of Learning	A posttest with 10 problem solving exercises was scored for accuracy and speed. These were converted z-scores and then averaged. Not enough information to compute effect size.
Basis for Adaptation	Types of errors made on interspersed multiple choice questions used to update scores on learning objectives.
Adaptation	There were 2 adaptive conditions. In the Mastery condition, an error on a test question would trigger re-presentation of content relevant to the correct choice. In the Loop-Back condition, an error would trigger presentation of content relevant to the incorrect choice. This remediation could be repeated up to 3 times. Successful completion of one section required for advancement to the next section. In the nonadaptive condition, test performance did not trigger remediation or affect advancement to the next section, although learners could choose to review material.
Citation: Pon-Barry, Schultz, Bratt, Clark, & Peters (2006)	
Context	In a lab study, participants learned about shipboard damage control by completing practical simulated problems assisted by an automated tutor. N= 20 per condition.
Measures of Learning	Pre and post-tests with 11 multiple choice questions. Calculation of effect size for learning gain was ambiguous: most conservative Hedges g^* = .52; least conservative = 1.02.
Basis for Adaptation	Correctness of responses to questions plus knowledge of previous dialog interactions.
Adaptation	In adaptive interactive tutoring dialogs, the tutor paraphrased correct answers and referred back to past dialog on incorrect answers. In the nonadaptive condition, the tutor acknowledged correct answers and provided hints upon incorrect answers.
Citation: Rosé, Jordan, Ringenberg, VanLehn, & Weinstein (2001)	
Context	10 undergraduates in a first year physics (5 in each condition) course completed the experiment, in which they worked on 8 physics problems using the system over the course of a 2-week period (self-paced).
Measures of Learning	Pretest and postest consisting of 34 multiple choice conceptual physics questions. One student in the control condition was matched with one student in the experimental condition (on pretest score and teacher) for purposes of analysis of posttest scores. Effect size reported = 0.90; method of calculation not reported.
Basis for Adaptation	Errors on problem solving steps and history of whether the same error had already been made in the session.
Adaptation	In the experimental group, each time a new error occurred in a session, it triggered an interactive tutorial dialog intended to help student better understand

	the concept related to the error. In control group, students had evaluative feedback and non-interactive reference materials explaining all relevant concepts.
Citation: Salden, Aleven, Renkl, & Schwonke (2009); Salden, Aleven, Schwonke & Renkl (2010)	
Context	38 9th and 10th graders were paid to participate in a lab study during which they practiced 11 geometry problems concerning application of 4 theorems. At each problem step, all students had to choose an explanation (from a menu) for the step. All students received feedback on correctness of each step.
Measures of Learning	Immediate posttest and delayed (1-week) posttest without feedback. Proportion of variance accounted for as measured by η^2 = .09 and .08 or the immediate and delayed tests, respectively. Effect size as measured by Hedges *g** = .63 and .68 for the immediate and delayed tests, respectively.
Basis for Adaptation	Estimate of whether understanding of theorem was mastered based on explanations chosen for problem solution steps on previous problems.
Adaptation	All students initially received worked-out examples. Subsequently, completed steps in examples were gradually removed, either adaptively or according to a preset fixed sequence. For the adaptive condition, this fading was based on students' past performance on the concept relevant to the step; i.e., a threshold criterion for past performance determined if the step solution was presented or had to be provided by the student. For the nonadaptive condition, fading occurred according to a fixed schedule.
Citation: Schwonke, Hauser, Nuckles, & Renkl (2006)	
Context	In a single session, undergraduate psychology students learned about a social psychology phenomenon by reading text and then writing a "learning protocol," which is a written explanation of one's own learning processes and outcomes. They were paid for participation. N = 49 and 20 in the adaptive and nonadaptive conditions, respectively.
Measures of Learning	Knowledge posttest of facts in the text. Hedges *g**effect size = .49.
Basis for Adaptation	Prior to learning, participants completed a questionnaire concerning their use of learning strategies and knowledge of metacogntion. A student model based on these responses was used to select prompts during production and revision of learning protocols.
Adaptation	During revision of learning protocols, participants received prompts about what to think about and include (e.g., what were the main points?). In the adaptive condition, these prompts were based on learning strategy deficiencies identified in the pre-training questionnaire. In the nonadaptive condition, the prompts were selected randomly.
Citation: Suraweera & Mitrovic (2004)	
Context	62 university students enrolled in the course "Introduction to Databases" completed computer-based training on database design during a 2-hour session.
Measures of Learning	Pretest and Posttest, graded by a human blind to experimental treatment. Reported Cohen's *d* effect size = 0.63, but it was unspecified if this was for comparison of posttest scores or pre-to posttest gains. Calculation of Hedges *g**effect size on posttest scores only = 0.53.

Basis for Adaptation	Errors during database design used to select next problem so as to target student weaknesses; errors during each problem used to select feedback and hints.
Adaptation	Each problem had to be correctly completed before moving on. After attempting a problem the student could "submit it" and get feedback (correct/incorrect). If incorrect, the student could request hints intended to help them locate and correct errors. In the nonadaptive condition, students could view a correct solution after each problem, and could skip among the problems as desired.
Citation: Triantafillou, Pomportsis, Demetriadis, & Georgiadou, E. (2004)	
Context	4^{th}-year computer science undergraduates enrolled in computer science used a hypermedia environment to learn about multimedia technology. 36 used adaptive hypermedia and 30 used traditional hypermedia.
Measures of Learning	10-item open ended questions on pretest and posttest. Calculation of Hedges *g**effect size on posttest scores only = 0.58.
Basis for Adaptation	Prior knowledge, as measured by the pretest, and ongoing knowledge acquisition as measured by pages visited in the hypermedia environment. Also cognitive style (field dependence or independence) as measured prior to training using the Embedded Figures Test.
Adaptation	1. Hyperlink annotation* 2. Direct navigation guidance** 3. Tailored content presentation, feedback, guidance, and other organizational tools based on cognitive style; learners had the ability to alter several of these options. In the nonadaptive condition, none of the above were used.
Citation: Tseng, Chu, Hwang, & Tsai (2008)	
Context	Learning about mathematical sequences, divided into 4 units presented with a hypermedia system. Junior high students completed the experiment using the on-line materials. N= 32 and 30 in the adaptive and nonadaptive conditions, respectively.
Measures of Learning	Posttest. Calculation of Hedges *g**effect size on posttest scores = 0.79.
Basis for Adaptation	Pretest performance for unit 1; test results for the prior unit for units 2-4.
Adaptation	The content presented was adapted over 3 levels of content difficulty, where levels differed in both amount of detail and concepts to be learned (e.g., Easy = very detailed, prerequisite and basic concepts; Difficult = brief descriptions of basic concepts and some advanced concepts). The better performance the higher the level of difficulty used next. The nonadaptive condition received Middle version throughout.
Citation: Tsiriga & Virvou (2004)	
Context	Learning use of English passive phrasing by two groups (N= 51 each) of 5^{th} and 6^{th} graders in an authentic learning setting over 2 1-hour sessions of learning using an adaptive or nonadaptive hypermedia system. Content consisted of didactic instruction and exercises.
Measures of Learning	Performance on pretest vs. delayed posttest (delay not specified but implication was at least one day and at most 11 days). Items were similar to the exercises given during training: multiple choice, fill in the blank, and sentence transformation between active and passive. Calculation of Hedges *g**effect size

	on posttest scores = 0.45.
Basis for Adaptation	Student’s native language, student’s familiarity with other languages, pretest scores, student’s conscientiousness, mastery of learning objectives based on exercise performance, types of errors committed
Adaptation	1. Hyperlink annotation* 2. Direct navigation guidance** 3. Exercise selection 4. Feedback provided advice based on error diagnosis. In nonadaptive condition, linear progression shown through content highlighted; navigation path under student control; feedback specified only whether response was correct or incorrect.
Citation: Xu & Wang (2006)	
Context	Undergraduates completed four on-line chapters on introduction to Oracle databases, over four days. N= 117 and 111 in adaptive and nonadaptive conditions, respectively. It was not clear if this was part of a university course or conducted for research only.
Measures of Learning	End of chapter quizzes, and a final exam. Calculation of Cohen’s *f* effect size on final exam scores = 0.21; effect sizes on end of chapter quizzes were all smaller than this, ranging from 0.08 to 0.16.
Basis for Adaptation	Pretest, quiz results, time spent on instructional materials
Adaptation	Sequencing of instructional materials and learning activities; level of detail presented in instructional materials (low, medium, or high), automated feedback and guidance. Few details of nonadaptive condition provided; presumably, students chose their own instructional sequencing

* Hyperlink annotation refers to the technique of annotating hyperlinks (usually with colors) to indicate something about the material at the linked site. E.g., if a student has not completed the prerequisite learning to understand the material at the link, it might be presented in red. Alternatively, the link itself might be disabled if the student is not prepared to go there.
** Direct navigation guidance refers to recommending the content a student should go to next. In some systems, the student must follow this direction, in others it is only a suggestion.

Table 3

Potential Causes of the Beneficial Learning Outcomes for the Experiments Listed in Table 2

	Mastery, level of detail or difficulty	During problem guidance	Error-sensitive Feedback	Self-correction	Hyperlink anno-tation & Direct naviga-tion support	Fading worked examples	Meta-cognitive prompts	Spacing and repetition of domain problems	Other
Anderson, et al. (1985)									
Chien,, et al. (2008)									
Corbalan,, et al. (2008)									
Davidovic,, et al. (2003)									
Metzler-Baddeley and Baddeley (2009)									
Perrin,, et al. (2003)									
Pon-Barry, et al. (2006)		**Dialog***							
Rosé, et al. (2001)		**Dialog***							
Salden , et al. (2009/10)									
Schwonke,, et al. (2006)									
Suraweera & Mitrovic (2004)									
Triantafillou, et al. (2004)									
Tseng, et al. (2008)									
Tsiriga & Virvou (2004)									
Xu & Wang (2006)									
Totals	7	7	7	4	3	2	1	1	1

Dialog* indicates that guidance was given through tutorial dialogs.

It can be seen in Table 3, that for most of the experiments, the experimental conditions differed in multiple adaptive strategies. This makes it difficult to identify the impact of any specific adaptive strategy on the learning outcomes. Four of the experiments did use a single manipulation, however. For two of these, they are the only experiments that employed these techniques. Metzler-Baddeley and Baddeley (2009) used the spacing and repetition technique, and Schwonke, et al. (2006) used metacognitive prompts.

The spacing and repetition technique is suitable for training situations with many short "challenges," such as vocabulary learning, the domain Metzler-Baddeley and Baddeley (2009) were working in. Even though we only identified one experiment which employed this technique in the literature we searched, spacing and repetition have been well-investigated in cognitive psychology laboratory experiments (e.g., Atkinson, 1972; Kornell, et al., 2010; Pashler, Zarow, & Triplett, 2003; Woziak, & Gorzelanczyk, 1994). Based on this research, adaptive spacing and repetition should produce superior learning outcomes compared to random spacing and repetition in any "drill and practice" type of educational software. There is at least one commercial software product for self-training based on this technique (see http://www.super-memo.com/supermemo2008.html). Thus, although there is only one entry in our table applying this technique, there is a preponderance of evidence in the experimental literature backing up the effectiveness of this adaptive approach. Moreover, the effect size obtained from the Metzler-Baddely & Baddeley experiment (2009) was quite respectable, at 0.90.

Metacognitive prompts in education are included to aid students in self-evaluation, self-explanation and self-regulation of learning processes. There is substantial evidence that these behaviors improve learning (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chui, & Lavancher, 1994; Johnson & Mayer, 2010), and that many students are negligent in performance of these activities (e.g., Winne & Nesbit, 2009). The Schwonke, et al. (2006) experiment illustrated that reminding students to engage in the metacognitive behaviors they are weakest in was especially effective. All students in their study received metacognitive prompts; but only the adaptive condition received prompts targeted at students' weaknesses. This adaptive prompting produced superior learning outcomes in the learning domain. In light of the evidence strongly pointing to the importance of metacognition in traditional education, and the evidence reviewed on self-evaluation and self-explanation in the prior section on local adaptation, it is sensible to infer from Schwonke et al.'s results (2006) that adaptive metacognitive prompts can also produce superior learning outcomes in technology-based educational environments. The most effective methods of implementing metacognitive prompting may require additional research, however. Schwonke et al.'s experiment covered a single lesson. When metacognitive prompting is applied over several lessons, adapting the prompts appropriately may require additional considerations above and beyond identified student weaknesses at the start of instruction (Nückles, Hübner, & Renkl, 2008).

For the other two papers that used a single adaptive technique (Salden, et al., 2009, 2010; Tseng, et al., 2008), at least one other paper in Tables 2 and 3 also used their technique. Salden, et al. (2009, 2010) demonstrated a learning benefit from adaptively fading worked examples, in the context of students solving geometry problems requiring the application of four different theorems. In the adaptive condition, transition from presenting a solved problem step vs. requiring the student to solve the step was based on an estimate of whether the student understood the relevant theorem. That estimate was based on whether the student previously was able to choose the right justification (from a menu) for an analogous step in previous examples. Students in this condition performed better on a posttest than students who had received fading of worked examples according to a fixed schedule. One other study in Tables 2 and 3 also faded worked examples adaptively (Corbalan, et al., 2008). That experiment used a somewhat different technique for adapting the fading, and combined its use with the mastery technique. Also, recall that one experiment discussed in the section on local adaptation used adaptive fading of worked examples. In that study, by Kalyuga and Sweller (2004), adaptation was based on student performance on the immediately preceding problem. In total, the evidence suggests that adaptive fading of worked examples can be a productive technique for enhancing learning outcomes.

Finally, the fourth experiment in Table 2 demonstrating benefits from a single adaptive technique was conducted by Tseng, et al. (2008). They used the mastery/level of detail or difficulty approach discussed above. The mastery technique, sometimes referred to as mastery learning or programmed instruction, has been shown to be effective in traditional classroom settings (Kulik, Kulik, & Bangert-Drowns; 1990). It can be seen in Table 3, that it is one of the most frequently used adaptive techniques among the experiments under consideration. The effectiveness of the mastery technique depends on identifying a logical progression in the domain material (e.g., that x needs to be learned before y), as well as the mastery criterion used. In other words, if 75% is considered mastery, it may have less of an effect on final learning outcomes than if 95% is considered mastery. It should also be noted that the mastery technique can only be effective it the performance measures used are valid and linked to the desired learning outcomes. Thus, close attention to the construction of assessment measures is essential.

For the seven experiments in Table 3 that used mastery, Tseng, et al. (2008) obtained the largest effect size (0.79). Effect sizes for the other experiments, which all combined mastery with other adaptive techniques, ranged from 0.21 (Xu & Wang, 2006) to 0.71 (Corbalan, et al., 2008). This demonstrates how the effect of applying the mastery technique can vary, depending on exactly how it is implemented. Indeed, this is an issue for all of the adaptive techniques. If it were not, then one would expect that the experiments that used multiple adaptive techniques would have higher effect sizes (if the effects of the different techniques were additive); but, this was not the case. There failed to be any apparent relation between the number of techniques used and effect size obtained (r = -0.23 when considering Table 2 alone, and -.19 when considering both Table 1 and Table 2). Furthermore, multiple approaches to meta-analysis across studies failed to identify any of the adaptive techniques as a significant predictor of effect size. Thus, we are unable to conclude from this type of analysis which adaptive technique may be more effective than another.

Basis of Adaptation

Table 4 summarizes the various data that were used as the basis of adaptation for the experiments listed in Tables 1 and 2. It can be seen that much of the input used for making adaptive decisions concerned student ability to answer questions or solve problems in the domain being taught, both prior to and during learning. It should be noted that none of the studies which used pretest data as a basis for adaptation used it as the sole basis; all of the experiments using pretest data also used during learning performance to make adaptive decisions. Thus, there is no evidence that adapting on the basis of pretest data alone produces benefits. Neither did we find any studies which addressed whether adapting on the basis of pretest plus during learning performance produces superior learning outcomes compared with during learning performance alone.

A common feature of many adaptive applications is the use of local error information to provide guidance during problem solving, and the use of model-based information to provide decisions about content sequencing (decisions about what content or problem to present next). Adaptive interventions provided during problem solving are sometimes referred to as micro-adaptive (Park & Lee, 2004), or the inner-loop (VanLehn, 2006), whereas those guiding sequencing of content have been referred to as macro-adaptive (Park & Lee, 2004), or the outer-loop (VanLehn, 2006). Inspection of Table 4 indicates that accuracy during pretesting of domain knowledge and during problem solving or question-answering (check on learning) is the most common basis for adaptation, although there are other parameters of student response which have been used (e.g., latency of response, certainty of response). Latency may be a particularly sensitive measure in the context of simulation-based task performance (e.g., Billings & Durlach, 2010).

A few of the studies used data besides domain-relevant performance, such as aptitude (e.g., language skills) or proclivity (e.g., cognitive style, conscientiousness). Analogous to the above discussion with respect to the adaptive interventions used, most of the experiments used multiple sources of student data, making it impossible to draw any firm conclusions with respect to the most discriminative sources. Neither can we make firm conclusions with respect to the adequacy of employing local student data only vs. model-based. Logically, model-based adaptive decisions should be superior, since they take into account more information. However, this will only be true in actuality to the extent that two conditions are met. First, the data used must be valid, and discriminating of student understanding with respect to the learning objectives (and the outcome measures if evaluating effectiveness). Second, the adaptive intervention selected, based on the data must be the right one, given the current state of the student. Having one of these without the other is not sufficient; both are required to produce improved learning outcomes (Brusilovsky, Karagiannidis, & Sampson, 2004). Thus, an adaptive training environment may fail to produce superior learning outcomes if the student model is good, but the adaptive intervention was implemented poorly, or if the adaptive interventions included were good, but the ability to determine when to intervene is faulty, because of inadequate or poorly conceived student models.

Table 4

Types of Data Used as the Basis for Adaptation for the Experiments Listed in Tables 1 and 2

Basis for Adaptation	Experiments	Total (out of 20)
Errors during a specific problem	Anderson, et al. (1985); Chien,, et al. (2008); Corbalan,, et al. (2008); Davidovic,, et al. (2003); Mathan & Koedinger (2005); Metzler-Baddeley and Baddeley (2009); Park and Tennyson (1986); Pon-Barry, et al. (2006); Rosé, et al. (2001); Suraweera & Mitrovic (2004); Tseng, et al. (2008); Tsiriga & Virvou (2004); Xu & Wang (2006)	13
Error patterns over time	Anderson, et al. (1985); Pon-Barry, et al. (2006); Rosé, et al. (2001); Suraweera & Mitrovic (2004); Tsiriga & Virvou (2004)	5
Pretest on domain knowledge	Chien,, et al. (2008); Davidovic,, et al. (2003); Kalyuga & Sweller (2004); Triantafillou, et al. (2004); Tseng, et al. (2008); Tsiriga & Virvou (2004); Xu & Wang (2006)	7
Check on learning questions	Corbalan,, et al. (2008); Davidovic,, et al. (2003); Kalyuga & Sweller (2004); Perrin,, et al. (2003); Xu & Wang (2006)	5
Response latency	Metzler-Baddeley and Baddeley (2009)	1
Student input to dialog interactions	Pon-Barry, et al. (2006); Rosé, et al. (2001)	2
Pages visited in hypermedia environment	Triantafillou, et al. (2004); Tsiriga & Virvou (2004)	2
Time spent reviewing content	Xu & Wang (2006)	1
Ability to provide explanations for problem solutions	Aleven & Koedinger (2002); Salden, et al. (2009/10)	2
Pretest on metacognitive skills	Schwonke,, et al. (2006)	1
Student -rated effort	Corbalan,, et al. (2008)	1
Cognitive style	Triantafillou, et al. (2004)	1
Language skills	Tsiriga & Virvou (2004)	1
Conscientiousness	Tsiriga & Virvou (2004)	1
Uncertainty	Forbes-Riley & Litman (2011)	1

Discussion and Conclusions

One obstacle to designing effective adaptive technology-based educational environments (i.e., ones that result in superior learning outcomes compared to nonadaptive ones), is that guidance with respect to which techniques are most effective is lacking. In this review we attempted to analyze the empirical evidence regarding different adaptive approaches; but, we found the evidence to be relatively undiscriminating. Many of the experiments producing learning benefits used multiple adaptive techniques, making assignment of responsibility for the observed benefits problematic. Although the few experiments that used a single technique are suggestive, they still fail to inform us about precise implementation in computer software, which might generalize across domains. For example, the technique of fading worked examples has been a popular topic for research, and the data essentially support the idea that that adaptively transitioning from worked examples to problem solving is more effective than a fixed mixture of examples and problems. Including worked examples is not a novel procedure in traditional learning. It is employed in many text books, such as when a new mathematical principle is applied in a worked-out example, often with the rationale for each step provided. In the text book, the fading is student-determined: the student is to read through the examples provided before attempting to solve related problems. The adaptive technology-based version of this intends to provide a customized amount of worked out examples, to ensure that the student does not attempt a full problem solution until they understand the logic behind the examples. The question is, how does one determine when the student is ready? What are the precise rules by which the fading should occur? How much evidence of mastery is required before moving on to the next more challenging level? Determining these specifics is needed for implementation of adaptive training techniques. Use of one of the adaptive procedures called out below is no guarantee of enhanced learning outcomes, because there are multiple ways a procedure could be implemented. A procedure implemented poorly may fail to obtain the desired effect, and the precise rules or algorithms employed may require iterative refinement. One way to tune parameters of adaptation is through analysis of past student performance data using data mining techniques (cf. Arroyo, Mehranian, & Woolf, in press; Cen, Koedinger, & Junker, 2007).

Below we offer the following as the mostly likely adaptive techniques to provide learning payoffs; but preface this recommended list by a caveat. The caveat is, these techniques cannot yet (based on scientific results) be specified precisely enough to turn directly into software code. Instructional design experts are required to make both qualitative and quantitative decisions with respect to implementation, and some iterative testing and revision may be required. Thus, each technique is accompanied by some comments about implementation.

Error-sensitive Feedback

Feedback about student performance should not only inform the student about whether they were correct or incorrect, but also should aim to repair errors. The easiest way to do this is to point the student back to the original relevant learning content; but whether this is effective or not depends on why the student erred. This approach would be expected to be useful only if the error were due to mere forgetting. On the other hand, if the student failed to comprehend the content initially, merely re-presenting it is likely to be ineffective, and some other form of remediation may be required. Thus, care must be taken as to how the "repair process" is implemented. There is no real consensus, based on empirical data, about the best ways to provide feedback (e.g., timing, content). Potentially, the way feedback is provided itself needs to be adaptive (model-based). One method (e.g., immediate, detailed error correction) may be best for novices, while another (delayed, abstract reminders) may be best for students with greater degrees of mastery. Some attention to past error history in deciding how to handle an error may be beneficial as well (e.g., Was this the first time this type of problem was encountered or has this same error been made multiple times? Has the student responded correctly numerous times before on analogous problems?). At least one study has shown providing feedback based on student certainty, as well as accuracy, can also be beneficial (Forbes-Riley & Litman, 2011).

Mastery Learning

Applying the mastery learning technique has proven effective in traditional educational settings and should be considered an essential technique to improve effectiveness of technology-based instructional environments. It has a basis in several theories of learning, in particular cognitive load theory (Sweller, 1988). It can be used to control both the sequencing and content of learning materials, when the domain can be organized according to various dimensions, such as difficulty and/or complexity. Clearly, it is appropriate for domains where learning one capability (e.g., solving simultaneous equations) depends on prior learning (e.g., solving single algebraic equations). Pretesting level of existing knowledge or skill can be used to allow students to “test out” of content they have already acquired, or in setting the difficulty or complexity of a practical exercise. Despite its proven effectiveness, application of the mastery technique to any specific instructional environment may require some fine-tuning, and will depend heavily on the quality and nature of the upfront domain analysis conducted, as well as the ability to create valid and diagnostic performance measures. This is particularly important when initially there is little knowledge about what is more or less difficult for students, as is often the case in less than well defined domains (e.g., influencing skills). Moreover, by analogy to the comments made above with respect to error-sensitive feedback, attention must also be paid to how remediation for slower students is provided. Recycling them through the same content again may not be adequate, and provision for multiple ways of presenting content may be required.

Adaptive Spacing and Repetition for Drill-and-Practice Items

Many findings from cognitive science experimentation have been collected in situations where learners are presented with repeated, short learning opportunities, and much is understood about how people learn in these kinds of relatively simple situations. These findings can be readily incorporated into adaptive training for “drill-and-practice” content, as demonstrated in the experiment by Metzler-Baddeley and Baddeley (2009). Their experiment used a form of paired-associate learning (English-Spanish vocabulary); but the technique is also likely applicable to cases of discrimination learning or categorization, such as learning to identify different types of vehicles or learning to tell the difference between images with and without tumors, or threatening vs. nonthreatening facial expressions, for example. Indeed, any form of perceptual learning (Manfred & Poggio; 2002) would seem amenable to application of adaptive spacing and scheduling of learning items on the basis of item difficulty. Some preliminary data collection would likely be required in order to fine tune the spacing and repetition algorithms used, and in determining the optimal training stimuli to include in the case of perceptual learning.

Fading of Worked Examples for Problem Solving Situations, or Fading of Demonstrations for Behavioral Tasks (such as in scenario-based simulations)

When applying this technique the precise parameters of fading need to be decided, and this may require iterative testing and evaluation, varying the parameters of adaptive fading. The evidence suggests incorporating two types of fading. First, students should be given rationales with the examples. Next, students should be required to provide the rationale once shown a solution, and finally, students should be required to provide both the solution and the rationale. The performance criteria required to trigger fading from one phase to the next is an open question requiring further study.

Metacognitive Prompting, Both Domain Relevant and Domain Independent

The role of metacognition and self-regulation in deliberate learning cannot be understated. It is a characteristic that separates good and poor learners. Good learners self-explain, self-evaluate, self-correct, and paraphrase. Poor learners fail to engage in these behaviors, or engage in them erroneously (such as the common mistake of incorrectly judging oneself as having understood material sufficiently). One function a human tutor serves is to compensate for poor metacognitive skills by requiring the elaboration, reasoning, and evaluation that learners fail to perform adequately for themselves. To the extent possible, technology-based instructional environments should also compensate for students lacking good metacognitive skills. Just like for the previous items, however, the most effective techniques for doing this have not adequately been established.

Additional Considerations

This review has been concerned with the comparison of adaptive to nonadaptive technology-based learning environments, asking, is there evidence for the benefits of adaptation when all other factors are held constant? Our conclusion is that there is evidence; however, we do not yet have sufficient information about technique implementation to enable the mass production of effective adaptive learning environments. We do not have a tried and true recipe that will guarantee superior learning outcomes in the absence of iterative system evaluation and refinement. The techniques that were addressed in this paper were those for which we could find some empirical evidence. There may be other fruitful techniques for which data are currently lacking, at least according to our inclusion criteria. For example, instructional interventions which take into account student psychophysiologcal or affective state (e.g., confusion, attention, arousal, boredom, etc.) may have promise. The majority of the work in this area to date has been in developing methods to measure and infer these states, and less attention has been devoted to interventions intended to do something about them to optimize affect for learning (e.g., D’Mello, Picard, & Graesser, 2007; Carroll et al., 2010).

This review focused on learning outcome benefits of automated adaptive training techniques; however, there are other potential benefits besides posttest measures of learning gain. For some of the experiments we examined, which failed to find learning outcome differences, there were time benefits in the adaptive conditions (e.g., Kalyuga, 2006; Salden, et al., 2004). This can be seen as a benefit when there is limited time to devote to learning. Another potential benefit is student attitude toward the instructional system. If students prefer to learn in an adaptive environment compared to a nonadaptive one, they may spend more time on task, be more engaged, and develop a more positive attitude about the domain.

As discussed in the report, adaptive instructional technology can make adaptive decisions using local data about the student (what the student just did), or using model-based decisions (a collection of data amassed over the time), or both. Some of the adaptive techniques recommended require model-based decisions (e.g., adaptive spacing and repetition), while others are agnostic as to the required data (e.g., error-sensitive feedback). In general, model-based decisions ought to be superior; although we were unable to provide any concrete evidence as to this, because many of the systems examined used both. Brusilovsky (2003) elegantly builds a case for the need for meta-adaptation in hypermedia environments; i.e., that the methods of providing guidance themselves need to adapt as the student changes over the learning experience. Specifically, novices seem to benefit most from restrictive techniques that limit their options (e.g., link hiding), whereas more knowledgeable students benefit more from somewhat more freedom (e.g., link annotation). Essentially, his conclusion implies that more data about the student needs to be taken into account to determine the most effective adaptive interventions; not just local data, but also data about the learning trajectory itself. Adaptive techniques effective at one point in the learning trajectory may be different from those most effective at a different point (Kalyuga, 2007). Domain novices seem to need more structure, spoon-feeding, and guidance; but as mastery advances, structures need to be loosened, learners need to start thinking for themselves, and to take a more active role in creating their own learning path. Thus, the need for meta-adaptation: adaptive techniques that themselves adapt over the course of student learning.

Implications for Future Development of Technology-based Training for the Army

Current Army procurement of technology-based training and education does not take into account the range of adaptive techniques that could be applied, like those examined in this report. A common specification in the current procurement process is interactive multimedia (IMI) level. IMI levels address the degree of passivity vs. activity on the part of the student. It is fairly well agreed that interactivity does support learning; but only to the extent that it focuses cognitive processing on the central concepts and principles to be learned (Chi, 2009; Renkl & Atkinson, 2007). Future specifications for procurement of technology-based training and education should include requirements for adaptive techniques like those listed here – Adaptive Multimedia Interventions (AMI), perhaps.

As discussed above, however, a particular adaptive technique could be implemented in multiple ways, and any specific implementation may or may not produce superior learning outcomes, compared with nonadaptive training. One reason is that, due to time or other constraints, the instructional design may have to be implemented without thorough analysis of the student learning process. Ideally, the designer would have the opportunity to iteratively test, norm, and refine instructional materials and assessment methods; but the time and resources required are not always available. One practice, which could greatly facilitate instructional design in the future, would be to start saving student data now. Current IMI offerings typically include some form of learning assessment; but student responses on individual assessment items are difficult to access, or are not saved at all. Collection and analysis of these data would reveal which of the assessment items were sensitive, discriminating, and predictive of student mastery (or lack thereof). Good assessment items could be retained, while poor ones could be replaced. These data could also contribute to improvement of the training itself. By providing insights into the average relative difficulty of different learning objectives, training sequence could be optimized. By providing insights into common student misconceptions, training could be revised to recognize those common misconceptions and provide appropriate error-sensitive feedback and targeted practice situations to students.

As used in this review, adaptive instructional environments are ones that alter their behavior with the intention of supporting learning; but what is to be learned is fixed. There is another sense in which instructional environments could be adaptive, however. The learning objectives themselves might adapt to the needs of the learner. Such just-in-time training has also been referred to as mission-based (Johnson, Friedland, Schrider, Valente, & Sheridan, 2011), and is especially desirable when time for learning is limited. For example, a Soldier being deployed to a position where manning traffic check points will be a principle part of his work could be given language and cultural training with practice scenarios situated at traffic check points. Another Soldier, going to the same area, but to train host nation forces, could receive language and cultural training with practice scenarios situated in the context of advising a host nation counterpart. To accomplish such tailoring, the instructional system would need to contain a variety of learning modules, some basic and perhaps used by all students, and some more specific, allowing the learner to practice use of new knowledge in the context in which they are likely to need it (cf. Johnson, et al., 2011). It would also need a way to recommend modules for students based on information known by the systems or provided by the student.

Yet another way the term adaptive is relevant to Army instructional systems is with respect to modifiability. Instructors or trainers should be able to modify instructional content or practice exercises, without having to go to a programmer or systems developer. This ability is desirable either for purposes of tailoring (as per previous paragraph), or to keep content up-to-date in rapidly changing domains. These systems are sometimes referred to as authorable, or at least editable. In the long run, it would be desirable to have learning environments that embody all three aspects: adaptive, mission-based, and authorable.

References

Atkinson R. (1972) Optimizing the learning of a second language vocabulary. *Journal of Experimental Psychology, 96*, 124-129.

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26(2),* 147-179.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, 228, 456-462.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*(2), 167-207.

Arroyo, I., Mehranian, H., Woolf, B. P. (in press). A data mining approach to intelligent tutoring. *Journal of Educational Data Mining*.

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

Bausell, R. B., Moody, W. B., & Walzl, F. N. (1972). Factorial study of tutoring versus classroom instruction. *American Educational Research Journal*, *9*(4), 591-597.

Billings, D. R. & Durlach, P. J. (2010). How input device characteristics contribute to performance during training to operate a simulated micro-unmanned aerial vehicle (Technical Report 1273). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Bloom, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*, 4-16. Retrieved from http://en.wikipedia.org/wiki/Bloom%27s_2_Sigma_Problem

Brusilovsky, P. (2003). Adaptive navigation support in educational hypermedia: the role of student knowledge level and the case for meta-adaptation. *British Journal of Educational Technology*, *34*, 487 – 497.

Brusilovsky, P., Karagiannidis, C., & Sampson, D. (2004). Layered evaluation of adaptive learning systems. *International Journal of Continuing Engineering Education and Lifelong Learning*, *14*, 402 – 421.

Carroll, M., Fuchs, S., Carpenter, A., Hale, K., & Abbott, R.G. (2010). Development of an auto-diagnostic adaptive precision trainer for decision making (ADAPT-DM). *International Test and Evaluation Association Journal*, *31*, 247–263.

Cen, H., Koedinger, K.R., & Junker, B. (2007). Is over practice necessary? – improving learning efficiency with the Cognitive Tutor through educational data mining. In R. Luckin, K.R.

Koedinger, & J. Greer (Eds.). *Proceedings of the13th International Conference on Artificial Intelligence in Education* (pp. 511-518). Amsterdam: IOS Press.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73-105.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145-182.

Chi, M. T., H., de Leeuw, N., Chui, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.

Chien, T. C., Yunus, A. S., Ali, W. C. W., & Bakar, R. (2008). The effect of an intelligent tutoring system (ITS) on student achievement in algebraic expression. *International Journal of Instruction*, *1*, 25-38.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*, New Jersey: Lawrence Erlbaum.

Conati C.& VanLehn K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation . *International Journal of Artificial Intelligence in Education*, *11*, 389-415.

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2008). Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology*, *33*, 733-756.

D'Mello, S.K., Picard, R., & Graesser, A.C. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, *22*, 53 - 61.

Dallal, G. E. (2007). The little handbook of statistical practice. Retrieved from http://www.tufts.edu/~gdallal/LHSP.HTM.

Davidovic, A., Warren, J., & Trichina, E. (2003). Learning benefits of structural example-based adaptive tutoring systems. *IEEE Transactions on Education*, *46*(2), 241-251.

Forbes-Riley, K. and Litman, D. (2011). Designing and evaluating a wizarded uncertainty-adaptive spoken dialogue tutoring system. *Computer Speech and Language*, *25*, 105-126.

Goetzfried, L., & Hannafin, M. J. (1985). The effect of the locus of CAI control strategies on the learning of mathematics rules. *American Educational Research Journal*, *22*, 273-278.

Gouli, E., Gogoulou, A. ,Papanikolaou, K. A., & Grigoriadou, M. (2006). An adaptive feedback framework to support reflection, guiding and tutoring. In G. D. Magoulas & S. Y. Chen (Eds.), *Advances in Web-Based Education: Personalized Learning Environments* (pp. 178 – 202). Hershey, PA: Information Sciences Publishing.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 180-192.

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

Johnson, C. I. & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior*, *26*, 1246 – 1252.

Johnson, W. L., Friedland, L. E., Schrider, P. J., Valente, A., & Sheridan, S. (2011). The Virtual Cultural Awareness Trainer (VCAT): Joint Knowledge Online's (JKO's) solution to the individual operational culture and language training gap. *ITEC 2011*, 10 - 12 May 2011, Cologne, Germany.

Kalyuga, S. (2006). Assessment of learners' organized knowledge structures in adaptive learning environments. *Applied Cognitive Psychology*, 20, 333-342.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19, 509–539.

Kalyuga, S. & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, *96*, 558-568.

Kavcic, A. (2004). Fuzzy user modeling for adaptation in educational hypermedia. *IEEE Transactions on Systems, Man, and Cybernetics*, *34*(4), 439-449.

Koedinger, K. R. & Aleven, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review, 19* (3): 239-264

Koedinger, K. R., & Anderson, J. R. (1998). Illustrating principled design: The early evolution of a cognitive tutor for algebra symbolization. *Interactive Learning Environments*, *5*, 161-180.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30-43.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in younger and older adults. *Psychology and Aging*, *25*, 498-503.

Kulik, C., Kulik, J., & Bangert-Drowns, R. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*(2), 265-306.

Landsberg, C R., Van Buskirk, W. L., Astwood Jr., R. A., Mercado, A. D., & Aakre, A. J. (2010). Adaptive training considerations for use in simulation-based systems (Special Report 2010-001). Orlando, FL: Naval Air Warfare Center Training Systems Division.

Lane, H. C., & VanLehn, K. (2005). Teaching the tacit knowledge of programming to novices with natural language tutoring. *Computer Science Education, 15*(3), 183-201.

Manfred. F. & Poggio, T. (2002). *Perceptual Learning*. Cambridge, MA: MIT Press.

Mathan, S. A. & Koedinger, K. R. (2005) Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist*. *40*, 257-265.

McKendree, J. (1990). Effective feedback content for tutoring complex skills. *Human-Computer Interaction*, *5*, 381 – 413.

Metzler-Baddeley, C. & Baddeley, R. J. (2009). Does adaptive training work? *Applied Cognitive Psychology*, *23*, 254-266.

Moundridou, M. & Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning*, *18*, 253 – 261.

Nückles, M., Hübner, S., & Renkl, A. (2008). Short-term versus long-term effects of cognitive and metacognitive prompts in writing-to-learn. In G. Kanselaar, V. Jonker, P. A. Kirschner, & F. J. Prins (Eds.), *Proceedings of the 8th International Conference of the Learning Sciences 2008 . Utrecht, NL: ICLS.*

Park, O., & Lee, J. (2004). Adaptive instructional systems. In D.H. Jonassen (ed.), *Handbook of Research on Educational Communications and Technology 2nd edition* (pp. 651-684). Mahwah, NJ: Lawrence Erlbaum.

Park, O., & Tennyson, R. D. (1986). Computer-based response-sensitive design strategies for selecting presentation form and sequence of examples in learning in coordinate concepts. *Journal of Educational Psychology, 78*(2), 153-158.

Pashler, H., McDaniel, M., Rohrer, D. & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, *9*, 105 – 119.

Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1051 – 1057.

Perrin, B. M., Dargue, B. W., & Banks, F. Z. (2003). Dynamically adapting content delivery: An effectiveness study and lessons learned*Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Session E-10, 1180-1190.*

Pon-Barry, H., Schultz, K., Bratt, E. O., Clark, B., & Peters, S. (2006). Responding to student uncertainty in spoken tutorial dialogue systems. *International Journal of Artificial Intelligence in Education, 16*, 171 – 194.

Renkl, A. & Atkinson, R. K. (2007). Interactive learning environment: Contemporary issues and trends. An introduction to the Special Issue. *Educational Psychology Review*, *19*, 235 – 238.

Rosé , C. P., Jordan, P., Ringenberg, S. S., VanLehn, K., & Weinstein, A. (2001). Interactive conceptual tutoring in Atlas-Andes. *Proceedings of Artificial Intelligence in Education, 2001*, 256 -266.

Salden, R., Aleven, V., Renkl, A., & Schwonke, R. (2009). Worked examples and tutored problem solving: Redundant or synergistic forms of support? *Topics in Cognitive Science, 1*, 203-213.

Salden, R., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, *38*, 289-307.

Salden, R. J.C.M., Paas, F., Broers, N. J., van Merriënboer, J. J.G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science*, *32,* 153-172.

Schwonke, R., Hauser, S., Nückles, M., & Renkl, A. (2006). Enhancing computer-supported writing of learning protocols by adaptive prompts. *Computers in Human Behavior, 22,* 77-92.

Shute, V. J., & Zapata-Rivera, D. (2008). Adaptive technologies. In J. M. Spector, M. D. Merril, J. J. G. van Merriënboer, & M. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 277–294). New York, NY: Taylor and Francis.

Suraweera, P. & Mitrovic, A. (2004). An intelligent tutoring system for entity relationship modeling. *International Journal of Artificial Intelligence in Education*, *14*, 375 – 417.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257 –285.

Sweller, J., & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2,* 59–89.

Triantafillou, E., Pomportsis, A., Demetriadis, S., & Georgiadou, E. (2004). The value of adaptivity based on cognitive style: An empirical study. *British Journal of Educational Technology, 35*(1), 95-106.

Tseng, J. C. R., Chu, H., Hwang, G., & Tsai, C. (2008). Development of an adaptive learning system with two sources of personalization information. *Computers & Education, 51*, 776-786.

Tsiriga, V., & Virvou, M. (2004). Evaluating the intelligent features of a web-based intelligent computer assisted language learning system. *International Journal on Artificial Intelligence Tools, 13*(2), 411-425.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, *16*, 227-265.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31,* 3-62.

VanLehn, K., & Jones, R. M. (1993). What mediates the self-explanation effect? Knowledge gaps, schemas or analogies? In M. Polson (Ed.), *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 1034-1039). Hillsdale, NJ: Erlbaum.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., & Wintersgill, M. (2005). The Andes physics tutoring systems: Lessons learned. *International Journal of Artificial Intelligence in Education, 15*, 147-204.

Vanderwaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*, 118-130.

Winne, P. H. & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, .J. Dunlosky, & A. C. Graesser (Eds),. *Handbook of Metacognition in Education* (pp. 260 -277). Mahwah, NJ: Erlbaum.

Wood, H. & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers and Education*, *33*, 153 – 169.

Woziak, P. A., & Gorzelanczyk, E. J. (1994). Optimization of repetition spacing in the practice of learning. *Acta Neurobiologiae Experimentalis*, *54*, 59–62.

Xu, D., & Wang, H. (2006). Intelligent agent supported personalization for virtual learning environments. *Decision Support Systems, 42,* 825-843.

Zhou, Y., Freedman, R., Glass, M., Michael, J. A., Rovick, A. A., & Evens, M.W. (1999). Delivering hints in a dialogue-based intelligent tutoring system. *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, Orlando, FL, July.

Appendix A

Explanation of Effect Sizes

When effect sizes were presented in papers, the measures presented were reported in Tables 1 and 2. Otherwise, effect sizes were computed, if enough of the required data allowed it. Effect size can be interpreted as the improvement in outcomes in units of standard deviation. Thus, an effect size of 1 indicates that the manipulation improved performance, on average, by one standard deviation (sd). The "gold standard" for educational interventions is 2 sds, set by Bloom's work (Bloom, 1984), on human tutoring; however, this size is seldom achieved by educational interventions. According to Cohen (1988), .2 - .5 is considered a small effect size, .5 - .8 a medium effect size, and .8 and above a large effect size.

Cohen's *d*

Cohen's *d* is defined as the difference between two means divided by the sd for the data

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

Cohen's f^2 and *f*

Cohen's f^2 is an appropriate effect size measure to use in the context of an F-test for ANOVA or multiple regression.

In a balanced design (equivalent sample sizes across groups) of ANOVA, the corresponding population parameter of f^2 is

$$\frac{SS(\mu_1, \mu_2, \ldots, \mu_K)}{K \times \sigma^2},$$

wherein μ_j denotes the population mean within the j^{th} group of the total *K* groups, and σ the equivalent population sd's within each group. *SS* is the sum of squares manipulation in ANOVA. Often reported as Cohen's f, which is simply the square root of Cohen's f^2

Cohen's f can be calculated "backwards" from an ANOVA as

$$\hat{f}_{\text{effect}} = \sqrt{(df_{\text{effect}}/N)(F_{\text{effect}} - 1)}.$$

(from http://en.wikipedia.org/wiki/Effect_size#Cohen.27s_.C6.922)

Hedges' *g**

Hedges' *g*, is based on a standardized difference.

$$g = \frac{\bar{x}_1 - \bar{x}_2}{s^*}$$

where s^* is computed as

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

As an estimator for the population effect size, it is biased. However, this bias can be corrected for by multiplication

$$g^* = J(n_1 + n_2 - 2)g \approx \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right) g$$

where J represents a gamma function.
(From http://en.wikipedia.org/wiki/Effect_size#Cohen.27s_.C6.922)

Proportion of Variance Accounted for
When researchers reported proportion of variance accounted for by the manipulation, this has also been reported in Tables 1 and 2. It should be noted that proportion of variance accounted for can range between 0 and 1, and is not on the same scale as effect size. Thus, proportion of variance accounted for should not be compared with effect size.

Eta squared η^2 and Partial Eta squared, η_p^2
Eta squared is the proportion of the total variance that is attributed to an effect. It is calculated as the ratio of the effect variance (SS_{effect}) to the total variance (SS_{total}).
$\eta^2 = SS_{effect} / SS_{total}$
The partial Eta squared is the proportion of the effect + error variance that is attributable to the effect. The formula differs from the Eta squared formula in that the denominator includes the SS_{effect} plus the SS_{error} rather than the SS_{total}.
$\eta_p^2 = SS_{effect} / (SS_{effect} + SS_{error})$
(from http://www.uccs.edu/~faculty/lbecker/SPSS/glm_effectsize.htm#Eta%20squared%20(h2))

Appendix B

Acronyms

ALC2015	Army Learning Concept 2015
AMI	Adaptive Multimedia Interventions
dL	Distributed Learning
DTIC	Defense Technical Information Center
IMI	Interactive Multimedia Instruction
ITS	Intelligent Tutoring Systems
TRADOC	Training and Doctrine Command