



Fundamentals, Flavors, and Foibles of Adaptive Instructional Systems

Paula J. Durlach^(✉)

U. S. Army DEVCOM Soldier Center, Orlando, FL 32826, USA
Paula.J.Durlach.civ@mail.mil

Abstract. Adaptive instructional interventions have traditionally been provided by a human tutor, mentor, or coach; but, with the development and increasing accessibility of digital technology, technology-based methods of creating adaptive instruction have become more and more prevalent. The challenge is to capture in technology that which makes individualized instruction so effective. This paper will discuss the fundamentals, flavors, and foibles of adaptive instructions systems (AIS). The section on fundamentals covers what all AIS have in common. The section on flavors addresses variations in how different AIS have implemented the fundamentals, and reviews different ways AIS have been described and classified. The final section on foibles discusses whether AIS have met the challenge of improving learning outcomes. There is a tendency among creators and marketers to assume that AIS—by definition—support better learning outcomes than non-adaptive technology-based instructional systems. In fact, the evidence for this is rather sparse. The section will discuss why this might be, and potential methods to increase AIS efficacy.

Keywords: Adaptation · Assessment · Learning

1 Introduction

Park and Lee (2004) defined adaptive instruction as educational interventions aimed at effectively accommodating individual differences in students while helping each student develop the knowledge and skills required to perform a task. Traditionally, this type of customized instruction has been provided by a human tutor, mentor, or coach; but, with the development and increasing accessibility of digital technology, technology-based methods of creating adaptive instruction have become more and more prevalent. Shute and Psotka (1994) point out that availability to computers began occurring at about the same time that educational research was also extolling the benefits of individualized learning; so, it was natural that harnessing computers was seen as a practical way to increase individualized instruction for more students. From the beginning, the challenge has been: how to capture in technology that which makes individualized instruction so effective. This paper will discuss the fundamentals, flavors, and foibles of adaptive instructions systems (AIS). The section on fundamentals covers what all AIS have in common. The section on flavors addresses variations in how different AIS have implemented the fundamentals, and reviews different ways AIS have been described and classified. The final section on foibles discusses whether AIS

have met the challenge of improving learning outcomes. There is a tendency among creators and marketers to assume that AIS—by definition—support better learning outcomes than non-adaptive technology-based instructional systems. In fact, the evidence for this is rather sparse. The section will discuss why this might be, and potential methods to increase AIS efficacy.

2 Fundamentals of AIS

All AIS have three things in common: (1) Automated measurement of the learner's behavior while using the system (2) Analysis of that behavior to create an abstracted representation of the learner's level of competency on the knowledge or skills to be learned, and (3) Use of that representation (the student model) to determine how to adapt the learner's experience while engaged with the system. The first two have to do with assessing learner competency, while the third is concerned with how to ameliorate competency weaknesses. The first two together constitute formative assessment, while the third is the adaptive intervention.

2.1 Formative Assessment in AIS

Formative assessments in AIS can be thought of as experiments about learners' knowledge. The AIS user experience is designed so as to test theories about that knowledge. The theories predict how a learner should behave if they possess such and such knowledge, given the stimuli and behavioral affordances of the learning environment. Consider a multiple choice question. Inferences about knowledge based on the student's selection depend on an implicit theory, though it is rarely explicitly stated: If the student understands the content, they will choose the right answer. But competing theories are possible; e.g., perhaps the student selected the correct answer by chance. Because both theories make the same prediction, it might be necessary to assess the knowledge multiple times and in various ways, until enough evidence has accumulated to make a competency inference with more certainty. Questions can provide more fine-grained information about learner competency by wisely selecting the foils to detect common misconceptions, or by varying the difficulty of the question, based on normative data. This allows a deeper understanding of the student's competency vs. just a binary rating. Other methods besides direct questions, of course, can be used to assess student knowledge. Learners can be asked to create artifacts, or perform activities. "Stealth assessment" refers to assessing student behavior during the performance of an activity, such that assessment occurs without any explicit question-asking (Shute et al. 2016).

The process of designing assessments has been formalized in various ways (Shute et al. 2016). Here I will focus on the Evidence-Centered Design (ECD) framework (e.g., Behrens et al. 2010; Mislevy et al. 2001; Shute et al. 2008). The first step in designing an assessment is to determine what mixture of knowledge, skills, and other factors should be assessed. An explicit representation of this in ECD is the competency model (also referred to as a domain model or an expert model in other contexts). For any particular domain (e.g., troubleshooting a washing machine) the competency

model organizes the things people need to know (e.g., the pump won't work if the lid is open), things they need to do (check there is power to the pump), and the conditions under which they have to perform (e.g., determine why the pump failed to run in the spin cycle). Collecting and organizing the domain information into a competency model may be one of the most challenging steps in creating an AIS, particularly for domains in which there are multiple ways of solving problems, and one best solution might not exist. Traditionally, the domain knowledge must be captured from experts, and experts don't always agree or have explicit access to the knowledge they use. Even for well-defined domains, like algebra, consulting experts is needed in order to capture instructional knowledge (e.g., common mistakes and various teaching methods), which are not represented in text books (Woolf and Cunningham 1987). This knowledge may be used in designing the AIS content and interventions, even if not represented in the competency model.

Making the competency model explicit is not required for assessment, nor for an AIS; but it is recommended, especially for complex knowledge structures and/or stealth assessment. For relatively simple learning, the domain model may remain in the designer's head and only be implicitly represented in the content presented to the student. An AIS designed for learning by rote (e.g., learning the times tables or vocabulary) is an example. An adaptive intervention for these types of AIS is to give each student practice on each association in inverse proportion to their ability to recall it (so for each student, their own pattern of errors determines how frequently specific association probes occur). The domain model may be represented only implicitly by the associations chosen for the student to learn.

Hypotheses about the state of the learner's ability with respect to the elements of the competency model are represented in the student model (or learner model or competency profile in other contexts), by assigning values against each ability. The student model may be a single variable reflecting an overall proficiency, or multiple variables representing finer grains of knowledge, and arranged in a network hierarchy with required prerequisite knowledge also represented. Typically the complexity of the competency model is reflected in the student model. When the competency model is not even explicit, the student model tends to be relatively simple, such as a list of values (e.g., percent correct on each cell of the times tables). For more complex competency models, the student model may be represented computationally, for example as a Bayesian network. A Bayesian network provides the ability to make predictions about how students might respond to a new problem, based on the current student model values. Regardless of the specific model, student model values are updated when new relevant information becomes available. That information comes from the student's performance on tasks specifically designed for that purpose. In ECD, these tasks are the task model. The task model consists of the learning environment the learner experiences, materials presented, actions that the student can take, and products they may produce. It specifies what will be measured and the conditions under which they will be measured. Elements in the task model are linked to elements in the competency model by the evidence model, so as to specify how student actions under different conditions indicate evidence with respect to different competencies. This requires a two-step process. First is evaluating the appropriateness or correctness of a student behavior or product (the evaluation step). This might be something like: correct

vs. incorrect, or low, medium, high. Second is turning these evaluations into student model values. ECD calls this the statistical step. The statistical step turns the evaluations into beliefs (or likelihoods) of student model values. In other words, if the student did x under conditions a and e , the probability is .68 that they understand the relation between the switch activated by lid closure and the pump (in keeping with the washing machine troubleshooting example). The actual probability to assign might be determined using expert input, approximations based on task features, a fitted model, or cognitive task analyses of students at known levels of mastery.

The student model values are made by inferences based on evidence collected, therefore the assessment designer must set rules for how to collect the evidence and when enough evidence has been amassed. In ECD these rules are called the assembly model. The assembly model controls the sequence of tasks and the rules for ending the assessment. When the sole purpose of the assessment is to gauge aptitude (summative assessment), a rule for ending the assessment may be determined by time to complete. A traditional aptitude test allowing a specific amount of time and using paper and pencil is an example of a fixed summative assessment. All the students have the same questions and are given the same amount of time. In contrast, an adaptive summative assessment is possible. It changes the sequence of tasks for each student, depending on the values in the student model, in order to collect evidence as efficiently as possible. Therefore different students may have a different sequence of questions and be allowed to work for different amounts of time. The Graduate Management Admissions Test (GMAT) is an example of an adaptive summative assessment. In AIS, the purpose of assessment is not to rank students, but rather to diagnose student misconceptions and to identify where special attention might be needed. Stopping rules therefore may be based on some sort of threshold (e.g. student model value of .90) or criterion (e.g., correctly diagnosing why the pump won't run in less than three minutes), indicating that mastery of the competency has been achieved, and no further practice is required. Adaptive sequencing rules in AIS are intended to give the student the next "best" learning experience to advance their progress; they are not part of the assessment, but rather one of several possible adaptive interventions.

2.2 Adaptive Interventions in AIS

Once an AIS has an abstracted representation of the learner's competency profile with respect to particular learning objectives, how can an AIS adapt to the individual? To answer this, several researchers have analyzed tutor-learner interactions to extract information about what occurs during one-to-one human tutor-tutee sessions. Much of this research is summarized in VanLehn 2011; see also Lepper and Woolverton 2002). Other sources of adaptive interventions have been educational research (e.g., the mastery learning technique) and psychological research (e.g., spacing). Bloom's research (1976, 1984), as well as Western interpretations of Vgotsky (1978), have had a large influence on instructional interventions in AIS, particularly the use of the mastery learning technique. Both essentially suggested that students should progress at their own rate. Tasks should not be too hard, so as to be frustrating, or too easy, so as to be boring. Learners should be given assistance with new tasks or concepts until they have been mastered. Formative assessments should be used to determine where a learner

needs help and when they are ready to move to a higher level of challenge. Help can be given in a number of ways, and this can also be adapted to the learner. Different learners may need different forms of help, so different methods for remediating weaknesses should be available (e.g., hints, timely feedback, alternative content and media, examples, self-explanation, encouragement, etc.). As far as implementing these recommendations in technology, almost all AIS have incorporated the mastery learning technique, and provide learners with feedback on their performance. Feedback is known to be critical to learning; however, whether it is considered adaptive is kind of a grey area. Positive feedback can be motivating and reduce uncertainty. Sadler (1989) suggested that one critical role of feedback is to support the student in comparing his or her own performance with what good performance looks like, and to enable students to use this information to close that gap. In general, the more fine-grained the feedback, the more likely it is to fit these requirements. Durlach and Spain (2014) suggested that feedback can be at different levels of adaptivity depending on how helpful it is in allowing the learner to self-correct, or how it is used as a motivational prop.

Another adaptive intervention that appears frequently in AIS stems from the psychological literature on the scheduling of practice “trials.” It is a well-established from learning science that retention is better when learning experiences are spaced over time compared to when they are massed (Brown et al. 2014). Repeated recall of information just on the edge of forgetting seems to solidify it better in memory compared to when it is recalled from relatively short term memory (which occurs when repetitions are close together). But people learn and forget at different rates for different items, so some AIS, particularly those aimed at learning associations (like the times tables or foreign language vocabulary), have devised adaptive spacing algorithms such that the sequencing of trials is personalized to the learner based on their accuracy for each association. So if you are learning French vocabulary, once you get the chien-dog association correct several times, you will be tested for it less and less frequently. These AIS are sometimes referred to as electronic flashcards. Although this “drill and practice” method of learning is sometimes looked down on as primitive, Kellman et al. (2010) have argued that it is a good way of learning structural invariance across otherwise variable cases, which is required for expertise. They argue that it can be applied to conceptual materials to increase fluency (greater automaticity and lesser effort), a hallmark of expertise; Kellman and colleagues have demonstrated this for several mathematical concepts (Kellman and Massey 2013). Note that adaptive spacing can be applied within a learning session or across days, weeks, months, etc. Bringing back items from an earlier topic at the right time may be particularly useful when the goal is to discriminate when to apply certain methods of problem solving vs. executing the method itself (Pan 2015; Rohrer 2009).

2.3 PLATO (Programmed Logic for Automatic Teaching Operation)

To illustrate the fundamentals of AIS, the workings of PLATO, one of the first AIS, is summarized (Bitzer et al. 1965). PLATO was developed to explore the possibilities of using computers to support individualized instruction. Several variations were created, but only the simplest of the “tutorial logics” will be outlined here (see Fig. 1). This logic led a student through a fixed sequence of topics, by presenting facts and examples

and then asking questions covering the material presented (the Main Sequence in Fig. 1). PLATO responded with “OK” or “NO” to each answer. Students could re-answer as many times as they liked, but could not move on to new content until they got an “OK.” If required, the student could ask for help by pressing a help button. This took them off to a branch of help material (Help Sequence in Fig. 1), which also had questions and additional branches as needed. After completing each help branch, or short-circuiting the help sequence by selecting an “AHA!” button, the student returned to the original question and had to answer it again. Each question had its own help branch, and in some versions different help branches for different wrong answers. If a student used all the help available, and could still not answer the original question, then PLATO supplied the correct answer. Teachers could create the content and questions on the computer itself, and could also examine stored student records. Records could also be aggregated for different types of analysis. One thing to notice about PLATO is that it is the type of system where the competency model is not explicit. The competency model is only implicit in the content and questions that are included by the teacher/author. Another thing to notice is that the student model is quite simple. It is made up of the correct and incorrect answers chosen by the student. The computer supplies the branching logic, depending on those answers. Although created in the 1960’s PLATO’s outward behavior is quite similar to many AIS in use today.

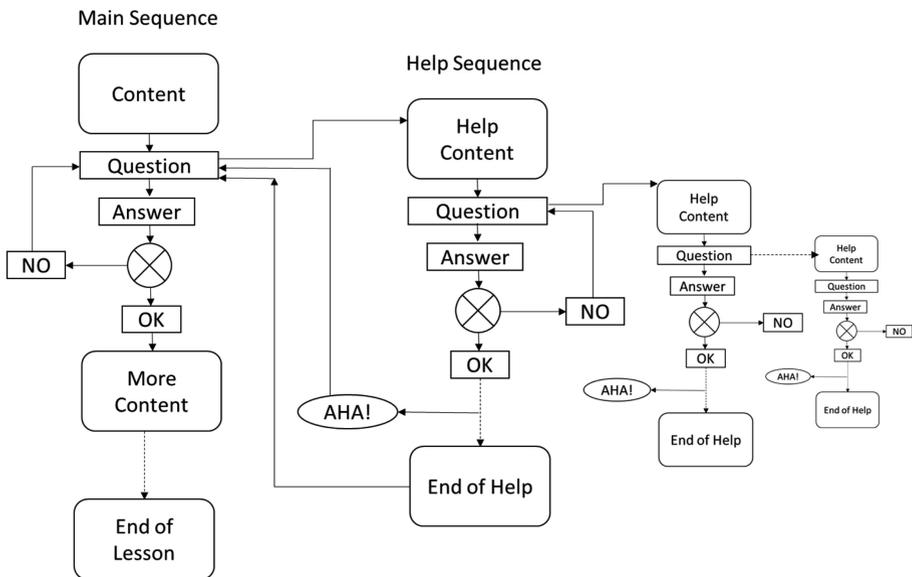


Fig. 1. Illustration of sequencing in PLATO. Students could request help for each question. The depth of help branches was determined by the content author.

Advances in computing power have allowed today's AIS to be more sophisticated, by using explicit competency models and artificial intelligence; however, it is not entirely clear whether AIS in 2019 are any more effective.

3 Flavors of AIS

Like ice cream, AIS come in many different flavors and qualities. You can get basic store-brand vanilla. You can get artisanal Salt & Straw's Apple Brandy & Pecan Pie;[©] or you can get up-market varieties in stores (e.g., Chunky Monkey[®]). PLATO is kind of like a store brand. It provides different paths through material and problems, with each decision based on only an isolated student action. The equipment is the same for each domain; only the ingredients (the content) differ. More complicated machinery (artificial intelligence) or customization (hand crafting) are required for more up-market brands. The analogy for "Artisanal" consist of AIS from academic researchers. These typically have been hand-crafted, use computational models, and are experimented with and perfected over years. Such AIS are typically referred to as Intelligent Tutoring Systems (ITS). ITS tend to provide coached practice on well-defined tasks requiring multiple steps to reach a solution (such as solving linear equations), although one particular type—constraint-based reasoning ITS (e.g., Mitrovic 2012) have had success with more open-ended domains such as database design. ITS are domain-specific, and cannot be repurposed easily for a different domain. Defining what makes an ITS "intelligent" has never been clearly agreed upon. Some researchers suggest it is only the appearance of intelligence, regardless of the underlying engineering, while others require explicit competency, student, and/or learner models (Shute and Psotka 1994). Most AIS use mastery learning and provide feedback; but other ingredients can be added: on-demand help, dashboards showing progress, interactive dialogue, simulations, and gamification, to name a few. Some AIS also use student models that include student traits (e.g., cognitive style) or states (frustration, boredom), in addition to the student's competency profile. The states or traits are used to influence adaptive interventions such as altering feedback or content (e.g., D'Mello et al. 2014; Tseng et al. 2008). Durlach and Ray (2011) reviewed the research literature on AIS and identified several adaptive interventions for which one or more well-controlled experiments showed benefits compared to a non-adaptive parallel version. These were (1) Mastery learning, which has already been explained; (2) Adaptive spacing and repetition, which also has already been explained; (3) Error-sensitive feedback—feedback based on the error committed; (4) Self-correction, in which students find errors in their own or provided example problems; (5) Fading of worked examples (here students are first shown how to solve a problem or conduct a procedure; they are not asked to actually solve steps until they can explain the steps in the examples); (6) Metacognitive prompting—prompting students to self-explain and paraphrase. Because there are many ways that AIS can differ, different ways to classify the various types have been suggested.

4 Macro and Micro Adaptation

Shute (1993) and Park and Lee (2004) describe adaptive interventions as being either macro-adaptive or micro-adaptive. A single AIS can employ both. Macro-adaptation uses pre-task measures or historical data (e.g., grades) to adapt content before the instructional experience begins. Two types of macro-adaptive approaches are Adaptation-as-Preference, and Mastery. With Adaptation-as-Preference, learner preferences are collected before training and this information is used to provide personalized training content. For example, it may determine whether a student watches a video or reads; or whether examples are given with surface features about sports, business, or the military. With Mastery macro-adaptation, a pretest determines the starting point of instruction, and subsequently already-mastered material, as determined by the pretest, may be skipped. For both these macro-adaptations, the customization is made at the beginning of a learning session or learning topic. In contrast, micro-adaptive interventions respond to learners in a dynamic fashion while they interact with the AIS. These systems perform on-going adaptations during the learning experience, based upon the performance of the learner or other behavioral assessment (e.g., frustration, boredom). They may use a pattern of response errors, response latencies, and/or emotional state to identify student problems or misconceptions, and make an intervention in real-time. Micro-adaptive interventions may be aimed at correcting specific errors or aimed at providing support, such as giving hints or encouragement.

4.1 Regulatory Loops

VanLehn (2006) popularized the idea of characterizing ITS behaviors in terms of embedded loops. He originally suggested that ITS can be described as consisting of an inner loop and an outer loop. The inner-loop is responsible for providing within-problem guidance or feedback, based on the most recently collected input. The outer loop is responsible for deciding what task a student should do next, once a problem has been completed. In ITS, data collected during inner-loop student-system interactions are used to update the student model. Comparisons of the student model with the competency model drive the outer loop decision process, tailoring selection (or recommendation) of the next problem. In order to do this, the available problems must be linked to the particular competencies their solution draws upon. Zhang and VanLehn (2017) point out that different particular AIS use different specific algorithms for making this selection or recommendation. The complexity of the algorithm used tends to depend on how the competency model is organized, e.g., whether knowledge is arranged in prerequisite relationships, difficulty levels, and/or whether there are one-to-one or a many-to-one mappings between underlying competencies and problems. There is no one algorithm, nor even a set of algorithms proven to be most effective.

VanLehn (2011) introduced the interaction granularity hypothesis. According to this hypothesis, problem solving is viewed as a series of steps, and a system can be categorized by the smallest step upon which the student can obtain feedback or support. The hypothesis is that the smaller the step, the more effective the AIS should be. This is because smaller step sizes make it easier for the student (and the system) to identify and address gaps in knowledge or errors in reasoning. The granularity hypothesis is in

accord with educational research on the effects of “immediacy” (Swan 2003). Combining the notion of loops and of granularity, it seems natural to extend the original conception of inner and outer loops to a series of embedded loops in which the inner loop is the finest level of granularity. Loops can extend out beyond the next task or problem to modules, chapters, and courses. Different types of adaptations can be made to help regulate cognitive, meta-cognitive, social, and/or motivational states at different loop levels (VanLehn 2016).

4.2 Levels of Adaptation

Durlach and Spain (2012, 2014) and Durlach (2014) proposed a Framework for Instructional Technology (FIT), which lays out various ways of implementing mastery learning, corrective feedback, and support using digital technology. Mastery learning and feedback have already been discussed. Support is anything that enables a learner to solve a problem or activity that would be beyond his or her unassisted efforts. In FIT, mastery learning is broken down into two separate components, micro-sequencing and macro-sequencing. Micro-sequencing applies to situations in which a given mastery criterion has yet to be met, and a system must determine what learning activity should come next to promote mastery. It can roughly be equated with remediation. Macro-sequencing applies to situations in which a mastery criterion has just been reached and a system must determine the new mastery goal—what learning activity to provide next. It can be equated with progression to a new topic or deeper level of understanding. For each of the four system behaviors (micro-sequencing, macro-sequencing, corrective feedback, and support), FIT outlines five different methods of potential implementation. These are summarized in Table 1. Except for macro-sequencing, the five methods of implementation fall along a continuum of adaptation. At the lowest level (Level 0), there is no adaptation – all students are treated the same. Each successive level is increasingly sophisticated with respect to the information used to trigger a system’s adaptive behavior. The type of intervention (macro-sequencing, micro-sequencing, feedback, and support) crossed with the level of adaptation (0 to IV) do not fit into a nice neat four x five matrix, however. This is because the adaption levels advance differently for the different types, based on differences in the pedagogical functions of the types. Macro-sequencing and micro-sequencing are about choosing the next learning activity. So, e.g., the adaptive levels of micro-sequencing have to do with how personalized the remedial content is, which is in turn, based on the granularity of the student model. The more granular the student model, the more personalized the remedial content can be (though this is not explicitly acknowledged in the framework). Analogously, the macro-sequencing levels depend on the student information used to determine the content of the next topic (e.g., job role). Clougherty and Popova (2015) also proposed levels of adaptivity for content sequencing, but they were organized as seven levels along a single dimension. The levels progress with increasingly flexible methods of “remediation” and “advancement,” where remediation is analogous to FIT’s micro-sequencing, and “advancement” is analogous to FIT’s macro-sequencing. At their first level, all students receive the same content and the only flexibility is that they can proceed at their own pace. The next level is quite similar to the sequencing outlined in Fig. 1 for Plato. Clougherty and Popova’s top level, which they call an

adaptive curriculum, exceeds what is specified by FIT. The adaptive curriculum allows students, assisted by an AIS, to pursue personalized, interdisciplinary learning.

Table 1. Framework for Instructional Technology (FIT)

Corrective feedback	
Level	Description
0	No item level feedback—summary score only
I	Minimal item feedback—item accuracy information
II	Correct answer or explanation of correct answer
III	Error-sensitive feedback—compares and contrasts correct answer to specific incorrect answer
IV	Context-aware feedback—knowledge of past performance used to influence feedback message
Support	
Level	Description
0	None
I	Fixed sources of information (e.g. glossary); problem-determined hints; student initiates access
II	Locally adaptive hints (answer determined) a. on request b. triggered
III	Context-aware hints, prompts, or pumps (faded scaffolding) a. on request b. triggered
IV	Level III + mixed initiative interactive dialogue
Micro-sequencing	
Level	Description
0	Recycling (repeat already seen content)
I	Supplemental Remediation (one remedial content set)
II	Supplemental Remediation levels (multiple remedial content sets)
III	Adaptive Content (remedial content composed for the individual)
IV	Real Time Adaptation (remediation via within—task adaptation)
Macro-sequencing	
Level	Description
0	Fixed sequence—one version only
I	Student choice or hybrid choice/fixe
II	Test-out (placement)
III	Adapted ahead (multiple versions)
IV	Adapted During (multiple versions)

In contrast to sequencing, the functions of feedback and support are to guide attention and to assist recall and self-correction for the current learning activity. The advancing FIT levels of adaptation for these are based on how much of the learning

context is taken into account in deciding how feedback and support are provided. Support will be used as an example. For Level 0 there is no support. Level I support is the same for all learners, and accessed on learner initiative. This includes support like glossaries and hyperlinks to additional explanatory information. Level I also includes problem-determined hints, where students are given the same advice for fixing an error no matter what error was made. Answer-based hints are at Level II. Answer-based hints are different, depending on the type of error made. FIT refers to this type of support as locally adaptive because it depends on the specific error the student made at a specific (local) point in time. Nothing needs to be accessed from the student model in order to supply answer-based support. In contrast, context-aware support needs to draw upon information from the student model. Context-aware support provides different levels of support as the student gains mastery; this is often referred to as fading. For example, at Level III the availability of hints may change depending on a student's inferred competency. Finally, Level IV adds more naturalistic interactive dialogue to Level III. It should be noted, however, that interactive dialog can be implemented at Level II; it does not necessarily require access to a student model that persists across problems (Level III). Autotutor (D'Mello and Graesser 2012) is an example of a Level II AIS with interactive language. Autotutor supports mixed-initiative dialogues on problems requiring reasoning and explanation in subjects such as physics and computer science; but the interaction is not influenced by information about student mastery from one problem to the next.

One intention behind FIT was to provide procurers of instructional technology with a vocabulary to describe the characteristics they desire in to-be-purchased applications. Terms like “adaptive” and “ITS” are not precise enough to ensure that the customer and the developer have exactly the same idea of the features that will be engineered into the delivered product. However, in an attempt to be precise, the resulting complexity of the FIT model may have undermined its ability to meet the intention. And despite this complexity, there are multiple factors that FIT does not deal with at all. For example, it does not make distinctions according to different kinds of competency, student, or domain models. Nor does it address factors like the quality of the instructional content, instructional design, nor user interface considerations.

Tyton Parners' white paper (2013) presented another framework for instructional technology, using six attributes: (1) Learner profile, (2) Unit of adaptivity, (3) Instruction coverage, (4) Assessment, (5) Content Model, and (6) Bloom's Coverage. An AIS can be located at a point along each attribute's continuum (e.g., an AIS can be high on one attribute and low on another). Explanation of the attributes and each's continuum are shown in Table 2. The white paper suggests that, along with maturity (defined by eight attributes), the taxonomy can be a useful guide to selecting instructional technology. They also suggest consideration of instructor resources (e.g., dashboards and data analytics) and evidence on learning outcomes. FIT and Tyton's frameworks are largely complementary. The Tyton taxonomy covers attributes not addressed by FIT (Assessment frequency, Instruction Coverage, Content Model, Bloom's coverage), whereas FIT provides finer-grained detail on Tyton's Unit of adaptivity and Learner profile attributes.

Table 2. Taxonomy and adaptivity continua proposed by Tyton Partners (2013).

Attribute	Low ← Adaptivity Level Continuum → High		
<u>Learner Profile</u> is a structured repository of information about the learner used to inform and personalize the learning experience	Student data informs initial placement	Student data drives adaptivity during a learning sequence	Student data is dynamic following each adaptive experience
<u>Unit of Adaptivity</u> refers to the structure of the instructional content and the scale at which that content is modified for specific learner needs	Course prerequisite level	Unit/lesson	Learning object level
<u>Instruction Coverage</u> refers to the pedagogical flexibility of a product to deliver an adaptive learning experience and the scope/scale of that experience within the context of a course	Targeted study aid	Supplemental instruction	Whole course
<u>Assessment</u> is the frequency, format, and conditions under which learners are evaluated	Infrequent/Benchmark	Formative	Adaptive/Continuous
<u>Content Model</u> describes the accessibility of the product's authoring environment to instructors or other users and their ability to add and/or manipulate instructional content in the system	Closed, with some configurability	Authoring capability offered as a service	Open, authoring platform
<u>Bloom's Coverage</u> highlights to what extent a product can support the learning objectives within the Cognitive Domain of Bloom's Taxonomy	Understanding/ Remembering	Analyzing/Applying	Creating/Evaluating

5 Foibles of AIS

AIS have been seen as a way to scale up individualized instruction. While the aim is to produce better learning outcomes than can be achieved with less adaptive methods, head-to-head comparisons of AIS to equivalent-in-content non-adaptive instructional technology has been rare (Durlach and Ray 2011; Murray and Perez 2015; Yarnall et al. 2016). More typically, AIS have been assessed as a supplement to traditional classroom instruction (e.g., Koedinger et al. 1997), and/or by looking at pre-AIS vs. post-AIS assessment outcomes. In that context, it has been suggested that an effect size of one be set as a benchmark for AIS learning gains; i.e., learning gains should raise test scores by around one standard deviation, on average (Slavin 1987; VanLehn 2011), to be on par with human tutoring (or, if compared with a non-adaptive control intervention, an effect size of about .75 or greater). While some AIS have been shown to meet or exceed this benchmark (e.g., Fletcher and Morrison 2014; VanLehn 2011); this is not always the case (Durlach and Ray 2011; Kulik and Fletcher 2016; Pane et al. 2010; Vanderwaetere et al. 2011; Yarnall et al. 2016). Consequently, the inclusion of some kind of adaptive intervention in instructional technology does not guarantee better learning outcomes. A generalized “secret sauce” – the right combination of instructional design, content selection, and individualized feedback and support that will boost learning outcomes for any given set of learning objectives (and for any learners in any context)—has yet to be identified. Evaluation studies conducting systematic manipulation of AIS features and behaviors to determine those which are required for superior learning outcomes have rarely been conducted (Brusilovsky et al. 2004; Durlach and Ray 2011); but, perhaps this is not surprising due to the myriad decisions that must be made during AIS design (thus the large number of variables that could be systematically examined), and the challenge of conducting educational research in naturalistic settings (Yarnall et al. 2016). This section will review some of the factors that might help improve AIS efficacy.

5.1 AIS Design

There is still much “art” in AIS design. Many decisions are made by intelligent guesswork (Koedinger et al. 2013; Zhang and VanLehn 2017). What should the mastery criteria be? What algorithms should be used to update the student model or to select the next learning task? There are no standard solutions to these questions. Even if mirroring approaches that have proven successful for others (e.g. using a Bayesian model), there are still many parametric decisions required; and what if the designed model does not align with learners’ actual knowledge structures? There is little learning science available to guide these decisions. Some researchers have looked to improve student models by re-engineering them based on whether the data produced by students using their AIS fit power-law curves (e.g., Koedinger et al. 2013; Martin and Mitrovic 2005). Poor fits suggest a misalignment. More recently, machine learning approaches are being applied to optimize pedagogical interventions (e.g., Chi et al. 2011; Lin et al. 2015; Mitchell et al. 2013; Rowe et al. 2016). Whether these data-driven approaches lead to better learning outcomes remains to be seen.

Much AIS design has been inspired by what human tutors do; but, AIS pedagogical interventions have focused more on knowledge and skill acquisition compared with motivational interventions. Tutoring behaviors concerned with building curiosity, motivation, and self-efficacy are known to be important (Clark 2005; Lepper and Woolverton 2002). Therefore AIS learning outcomes may benefit from greater incorporation of tactics aimed at bolstering these. Many AIS do not incorporate “instruction;” but, rather focus on practice. They therefore do not necessarily abide by instructional design best practices. Merrill (2002) described five principles of instruction and multiple corollaries; but, it is not clear that these have had much impact on the design of AIS content. Instructional design advises on what content to provide and when during the learning process (e.g., ordering of concrete vs. abstract knowledge). “Content is king” and may outweigh the effects of adaptive intervention. The technology used to provide content is less relevant than the content itself (Clark 1983). Finally, with respect to content, it is possible that greater effort at including “pedagogical content knowledge” in AIS intervention tactics may improve AIS effectiveness. Pedagogical content knowledge is knowledge about teaching a specific domain (Hill et al. 2008). This includes what makes learning a specific topic difficult, common preconceptions, misconceptions, and mistakes. An AIS that can recognize these may be better able to help students overcome them. Only a few AIS have included explicit representation of pedagogical content knowledge, often referred to as bug libraries (Mitrovic et al. 2003).

Greater attention to heuristics for good human-computer interaction might also help AIS effectiveness. Certainly, interfaces that incur a high cognitive load will leave students with fewer cognitive resources for learning. Beyond good usability design, it has been suggested that the structure of interactions can affect knowledge integration and the mental models formed by students (Swan 2003). Woolf and Cunningham (1987) suggest that AIS environments should be intuitive, make use of key tools for attaining expertise in the knowledge domain, and as far as possible resemble the real world in which the tasks need to be conducted. Learners should have the ability to take multiple actions so as to permit diagnosis of plausible errors.

Finally, a shift from focusing on improving performance during the use of an AIS to improving retention and transfer may enable AIS to provide better learning outcomes. The performance measured during or immediately after an AIS experience may fail to translate into retention and transfer. Interventions aimed at improving performance during practice may actually have the inverse effect on retention and transfer (Soderstrom and Bjork 2015). An example is massed vs. spaced practice: performance during practice may appear better when practice opportunities are massed in time; however, retention over the long term is better when it is spaced in time. Research has shown that the optimal spacing depends on encoding strength, and so may be different for different items and different across learners. Therefore an inferred measure of encoding strength can be used to adapt when a repetition will likely enhance that strength further. As already mentioned, some AIS do use adaptive spacing algorithms to schedule tasks so as to improve retention. Most of these are applied in the context of simple associative learning, and use measures such as accuracy, reaction time, and/or student judgement of certainty to infer encoding strength. This tactic has found less application in AIS concerning more complex types of learning (Kellman and Massey

2013). Therefore more consideration of when tasks should be retested might improve AIS impacts on retention. Methods of practice that tend to impair performance during initial learning but facilitate retention and transfer have been referred to as “desirable difficulties” (Bjork and Bjork 2011). These desirable difficulties include making the conditions of practice less predictable, including variations in format, contexts, timing, and topics. Formative assessment using recall (vs. recognition) is also desirable. Not all sources of difficulty are desirable (e.g., unintuitive user interfaces, or content too advanced for the learner). Desirable difficulties are ones that reinforce memory storage and de-contextualize retrieval. Just as spacing effects can be made adaptive, it is possible that implementation of some of the other desirable difficulties can be done adaptively (e.g., personalizing when to vary the format of a task based on prior performance). Except for spacing, the potential benefits of implementing these desirable difficulties adaptively has received little attention.

5.2 AIS Deployment

The impact of AIS on learning outcomes is not only dependent on AIS design, but also on the deployment environment. The quality and context of AIS implementation is important. The same piece of courseware can be used in different ways, and the specifics of usage can affect learning outcomes (Yarnall et al. 2016). First-order barriers to successful technology integration include limitations in hardware capability, hardware availability, time, systems interoperability, technical support, and instructor training (Ertmer and Ottenbreit-Leftwich 2013; Hsu and Kuan 2013). But AIS are socio-technical systems, so their impact also depends on human activities, attitudes, beliefs, knowledge, and skills (Geels 2004). Educational technology integration is facilitated when instructors believe that first-order barriers will not impede its use, have had adequate training, and perceive institutional support (Ertmer and Ottenbreit-Leftwich 2013; Hsu and Kuan 2013). Additionally, the instructor’s ability to co-plan lessons and/or content with the technology can have an impact on technology integration (Ertmer and Ottenbreit-Leftwich 2013) and learning outcomes (Yarnall et al. 2016). Much research has examined the factors that affect technology integration in education in general (e.g., Knezek and Chritensen 2016); and much advice can be found on how to facilitate technology integration (e.g., <https://www.edutopia.org/technology-integration-guide-description> and <https://www.iste.org/standards/essential-conditions>); but, there is rather less available on how to assess it objectively or on its relation to learning outcomes. The literature on human aspects of technology integration in education has tended to focus almost exclusively on the teacher and has devoted far less attention to other factors including social variables like class size, student demographics, or student attitudes.

With respect to implementing AIS specifically, Yarnall et al. (2016) suggest that AIS providers work with their institutional partners and researchers to articulate and validate implementation guidelines. The technology needs to fit the environment in which it will be used. AIS providers need to fully explore the boundary conditions of recommended use. If the provider recommends one hour per day but the program of instruction can only accommodate three hours per week, what are the implications? Instructors deviating from recommended guidelines may undermine AIS effectiveness;

but it is not entirely clear that guidelines have been validated through empirical testing for multiple types of learning environments and learners. Beyond following guidelines, Tyton Partners (2016) recommend establishing a shared understanding of the impact of AIS on the teaching process among all stakeholders, and involving faculty in the selection and implementation processes throughout. They also recommend educating faculty on AIS in general, including the concepts, models and techniques used, and the learner data that can be accessed.

6 Conclusions

AIS have three fundamental characteristics: (1) Automated measurement of the learner's behavior while using the system (2) Analysis of that behavior to create an abstracted representation of the learner's level of competency on the knowledge or skills to be learned, and (3) Use of that representation (the student model) to determine how to adapt the learner's experience while engaged with the system. Different flavors of AIS do this in different ways, however, and there is still much art in their design. AIS can provide learning benefits including reduction in time to learn and increasing passing rates; but, this is not a guaranteed outcome. AIS is not a homogeneous category that can be established as "effective" or "ineffective" (Yamall, Means, & Wetzell). It is likely an AIS will not be any less effective than a non-adaptive analog; however, an AIS will typically require more upfront resources, and potentially more radical changes to an existing program of instruction. This is because AIS require more content and under-the-hood engineering, and allow for different learners within a class to be at substantially different places in the curriculum. The decision to use AIS also introduces new issues of privacy and security concerning the usage of the learner data generated (Johanes and Lagerstrom 2017). Some have suggested that a great value of AIS may be in the revelation of those data to the instructor and perhaps the learners themselves (e.g., Brusilovsky et al. 2015; Kay et al. 2007; Martinez-Maldonado et al. 2015). If the data can be provided in a form that is meaningful for instructors and learners, that may enable beneficial instructional adaptation on the part of an instructor, or self-regulation on the part of the learner, without necessarily requiring especially effective use of the data for adaptive intervention by the system itself. Given the current state of the art of artificial intelligence applied to AIS, one of AIS's greatest potential benefits may be to empower the "more knowledgeable other" (words attributed to Vygotsky) to guide learning. Further breakthroughs may provide AIS to become more reliably effective in and of themselves.

References

- Behrens, J., Mislevy, R., DiCerbo, K., Levy, R.: An evidence centered design for learning and assessment in the digital world. CRESST Report 778. University of California, Los Angeles (2010)
- Bitzer, D., Lyman, E., Easley, J.: The used of Plato: a computer controlled-teaching system. Report R-268. University of Illinois, Urbana Illinois (1965)

- Bjork, E., Bjork, R.: Making things hard on yourself, but in a good way: creating desirable difficulties to enhance learning. In: Gernsbacher, M., Pew, R., Hough, L., Pomerantz, J. (eds.) *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, pp. 56–64. Worth Publishers, New York (2011)
- Bloom, B.: *Human Characteristics and School Learning*. McGraw-Hill, New York (1976)
- Bloom, B.: The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educ. Res.* **13**, 4–16 (1984)
- Brown, P., Roediger, H., McDaniel, M.: *Make It Stick*. The Belknap Press of Harvard University, Cambridge (2014). <https://doi.org/10.1021/ed5006135>
- Brusilovsky, P., Karagiannidis, C., Sampson, D.: Layered evaluation of adaptive learning systems. *Int. J. Continuing Eng. Educ. Lifelong Learn.* **14**, 402–421 (2004). <https://doi.org/10.1504/IJCEELL.2004.005729>
- Brusilovsky, P., Somyurek, S., Guerra, J., Hosseini, R., Zadorozhny, V., Durlach, P.: Open social student modeling for personalized learning. *IEEE Trans. Emerg. Top. Comput.* **4**(3), 450–461 (2015). <https://doi.org/10.1109/etec.2015.2501243>
- Chi, M., VanLehn, K., Litman, D., Jordon, P.: An evaluation of pedagogical tutorial tactics for a natural language tutoring system: a reinforcement learning approach. *Int. J. Artif. Intell. Educ.* **21**, 83–113 (2011). <https://doi.org/10.3233/JAI-2011-014>
- Clark, R.: Reconsidering research on learning from media. *Rev. Educ. Res.* **53**(4), 445–459 (1983). <https://doi.org/10.3102/00346543053004445>
- Clark, R.: What works in distance learning: motivation strategies. In: O’Neil, H. (ed.) *What Works in Distance Learning: Guidelines*, pp. 89–110. Information Age Publishers, Greenwich (2005)
- Clougherty, R., Popova, V.: How adaptive is adaptive learning: seven models of adaptivity or finding Cinderella’s shoe size. *Int. J. Assess. Eval.* **22**(2), 13–22 (2015)
- D’Mello, S., Blanchard, N., Baker, R., Ocumpaugh, J., Brawner, K.: I feel your pain: a selective review of affect-sensitive instructional strategies. In: Sottolare, R., Graesser, A., Hu, X., Goldberg, B. (eds.) *Design Recommendations for Adaptive Intelligent Tutoring Systems: Adaptive Instructional Strategies*, vol. 2, pp. 35–48. U.S. Army Research Laboratory, Orlando, FL (2014)
- D’Mello, S., Graesser, A.: AutoTutor and affective AutoTutor: learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Trans. Interact. Intell. Syst.* **2**, 23–39 (2012). <https://doi.org/10.1145/2395123.2395128>
- Durlach, P.: Support in a framework for instructional technology. In: Sottolare, R., Graesser, A., Hu, X., Goldberg, B. (eds.) *Design Recommendations for Adaptive Intelligent Tutoring Systems: Adaptive Instructional Strategies*, vol. 2, pp 297–310. Army Research Laboratory, Orlando (2014)
- Durlach, P., Spain, R.: Framework for instructional technology. In: Duffy, V. (ed.) *Advances in Applied Human Modeling and Simulation*, pp. 222–223. CRC Press, Boca Raton (2012)
- Durlach, P., Spain, R.: Framework for instructional technology: methods of implementing adaptive training and education. Technical report 1335. U.S. Army Research Institute for the Behavioral and Social Sciences (2014). www.dtic.mil/docs/citations/ADA597411. Accessed 26 Dec 2018
- Durlach, P., Ray, J.: Designing adaptive instructional environments: insights from empirical evidence. Technical report 1297. U.S. Army Research Institute for the Behavioral Social Sciences (2011)
- Ertmer, P., Ottenbreit-Leftwich, A.: Removing obstacles to the pedagogical changes required by Jonassen’s vision of authentic technology-enabled learning. *Comput. Educ.* **64**, 175–182 (2013). <https://doi.org/10.1016/j.compedu.2012.10.008>

- Fletcher, J., Morrison, J.: Accelerating development of expertise: a digital tutor for navy technical training. Institute for Defense Analysis Document D-5358. Institute for Defense Analysis, Alexandria, VA (2014)
- Geels, F.: From sectoral systems of innovation to social-technical systems: insights about dynamics of change from sociology and institutional history. *Res. Policy* **33**, 897–920 (2004). <https://doi.org/10.1016/j.respol.2004.01.015>
- Hill, H., Ball, D., Schilling, S.: Unpacking pedagogical content knowledge: conceptualizing and measuring teachers' topic-specific knowledge of students. *J. Res. Math. Educ.* **39**(4), 372–400 (2008)
- Hsu, S., Kuan, P.: The impact of multilevel factors on technology integration: the case of Taiwanese grade 1–9 teachers and schools. *Educ. Technol. Res. Dev.* **61**(1), 25–50 (2013)
- Johanes, P., Lagerstrom, L.: Adaptive learning: the premise, promise, and pitfalls. *Am. Soc. Eng. Educ.* (2017). <https://peer.asee.org/adaptive-learning-the-premise-promise-and-pitfalls>. Accessed 08 Jan 2019
- Kay, J., Reimann, P., Yacef, K.: Mirroring of group activity to support learning as participation. In: Luckin, R., Koedinger, K., Greer, J. (eds.) *Artificial Intelligence in Education*, pp. 584–586. IOS Press, Amsterdam (2007)
- Kellman, P., Massey, C.: Perceptual learning, cognition, and expertise. In: Ross, B. (ed.) *The Psychology of Learning and Motivation*, vol. 58, pp. 117–165. Academic Press, Elsevier (2013). <https://doi.org/10.1016/b978-0-12-407237-4.00004-9>
- Kellman, P., Massey, C., Son, J.: Perceptual learning modules in mathematics: enhancing students' pattern recognition, structure extraction, and fluency. *Top. Cogn. Sci.* **2**(2), 285–305 (2010). Special Issue on Perceptual Learning
- Knezek, G., Christensen, R.: Extending the will, skill, tool model of technology integration: adding pedagogy as a new model construct. *J. Comput. High. Educ.* **28**, 307–325 (2016). <https://eric.ed.gov/?id=ED562193>
- Koedinger, K., Anderson, J., Hadley, W., Mark, M.: Intelligent tutoring goes to school in the big city. *Int. J. Artif. Intell. Educ.* **8**, 30–43 (1997)
- Koedinger, K., Booth, J., Klahr, D.: Education research. Instructional complexity and the science to constrain it. *Science* **342**, 935–937 (2013). <https://doi.org/10.1126/science.1238056>
- Koedinger, Kenneth R., Stamper, John C., McLaughlin, Elizabeth A., Nixon, Tristan: Using data-driven discovery of better student models to improve student learning. In: Lane, H., Chad, Yacef, Kalina, Mostow, Jack, Pavlik, Philip (eds.) *AIED 2013. LNCS (LNAI)*, vol. 7926, pp. 421–430. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_43
- Kulik, J., Fletcher, J.: Effectiveness of intelligent tutoring systems: a meta-analytic review. *Rev. Educ. Res.* **86**(1), 42–78 (2016). [https://doi.org/10.1016/0004-3702\(90\)90093-F](https://doi.org/10.1016/0004-3702(90)90093-F)
- Lawless, K., Pellegrino, J.: Professional development in integrating technology into teaching and learning: knowns, unknowns, and ways to pursue better questions and answers. *Rev. Educ. Res.* **77**(4), 575–614 (2007). <https://doi.org/10.3102/0034654307309921>
- Lepper, M., Woolverton, M.: The wisdom of practice: lessons learned from the study of highly effective tutors. In: Aronson, J. (ed.) *Improving Academic Achievement: Impact of Psychological Factors on Education*, pp. 135–158. Academic Press, San Diego (2002)
- Lin, H., Lee, P., Hsiao, T.: Online pedagogical tutorial tactics optimization using genetic-based reinforcement learning. *Sci. World J.* **2015** (2015). Article ID 352895. <https://www.hindawi.com/journals/tswj/2015/352895/cta/>. Accessed 04 Jan 2019
- Martin, Brent, Mitrovic, Antonija: Using learning curves to mine student models. In: Ardissono, Liliana, Brna, Paul, Mitrovic, Antonija (eds.) *UM 2005. LNCS (LNAI)*, vol. 3538, pp. 79–88. Springer, Heidelberg (2005). https://doi.org/10.1007/11527886_12

- Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., Clayphan, A.: The LATUX workflow: designing and deploying awareness tools in technology-enabled learning settings. In: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, pp 1–10, ACM, New York (2015). <https://doi.org/10.1145/2723576.2723583>
- Merrill, M.: First principles of instruction. *Educ. Technol. Res. Dev.* **50**(3), 43–59 (2002)
- Mislevy, R., Steinberg, L., Almond, R.: On the structure of educational assessments. *Measur. Interdisc. Res. Perspect.* **1**(1), 3–62 (2001)
- Mitchell, Christopher M., Boyer, Kristy Elizabeth, Lester, James C.: A Markov decision process model of tutorial intervention in task-oriented dialogue. In: Lane, H.Chad, Yacef, Kalina, Mostow, Jack, Pavlik, Philip (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 828–831. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_123
- Mitrovic, A.: Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User Model. User Adap. Interact.* **22**(1–2), 39–72 (2012)
- Mitrovic, Antonija, Koedinger, Kenneth R., Martin, Brent: A comparative analysis of cognitive tutoring and constraint-based modeling. In: Brusilovsky, Peter, Corbett, Albert, de Rosiis, Fiorella (eds.) UM 2003. LNCS (LNAI), vol. 2702, pp. 313–322. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-44963-9_42
- Murray, M., Pérez, J.: Informing and performing: a study comparing adaptive learning to traditional learning. *Inf. Sc. Int. J. Emerg. Transdiscipline* **18**, 111–125 (2015). <https://doi.org/10.1080/13636820000200120>, <http://www.inform.nu/Articles/Vol18/ISJv18p111-125Murray1572.pdf>. Accessed 03 Jan 2019
- Pan, S.: The interleaving effect: mixing it up boosts learning. *Sci. Am.* (2015). <https://doi.org/10.1177/1529100612453266>, <https://www.scientificamerican.com/article/the-interleaving-effect-mixing-it-up-boosts-learning/>. Accessed 26 Dec 2018
- Pane, J., McCaffrey, D., Slaughter, M., Steele, J., Ikemoto, G.: An experiment to evaluate the efficacy of cognitive tutor geometry. *J. Res. Educ. Effectiveness* **3**(3), 254–281 (2010). <https://doi.org/10.1080/19345741003681189>
- Park, O., Lee, J.: Adaptive instructional systems. In: Jonassen, D.H. (ed.) *Handbook of Research on Educational Communications and Technology*, pp. 651–684. Lawrence Erlbaum Associates Publishers, Mahwah (2004)
- Rohrer, D.: The effects of spacing and mixing practice problems. *J. Res. Math.* **40**(1), 4–17 (2009)
- Rowe, J., et al.: Extending GIFT with a reinforcement learning-based framework for generalized tutorial planning. In: Sottolare, R., Ososky, S. (eds.) *Proceedings of the 4th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym4)*, pp. 87–97. U.S. Army Research Laboratory, Orlando, FL (2016)
- Sadler, D.: Formative assessment and the design of instructional systems. *Instr. Sci.* **18**, 119–144 (1989)
- Shute, V.: A comparison of learning environments: all that glitters... In: Lajoie, S., Derry, S. (eds.) *Computers as Cognitive Tools*, pp. 47–74. Lawrence Erlbaum Associates, Hillsdale (1993)
- Shute, V., Hansen, E., Almond, R.: You can't fatten a hog be weighing it – Or can you? evaluating as assessment for learning system called ACED. *Int. J. Artif. Intell. Educ.* **18**(4), 289–316 (2008)
- Shute, V., Leighton, J., Jang, E., Chu, M.: Advances in the science of assessment. *Educ. Assess.* **21**(1), 34–39 (2016)
- Shute, V., Psotka, J.: Intelligent tutoring systems: past, present, and future. In: Jonassen, D. (ed.) *Handbook of Research for Educational Communications and Technology*, pp. 570–600. Simon & Schuster Macmillan, New York (1994)

- Slavin, R.: Mastery learning reconsidered. *Rev. Educ. Res.* **57**(2), 175–213 (1987). <https://doi.org/10.3102/00346543057002175>
- Soderstrom, N., Bjork, R.: Learning versus performance: an integrative review. *Perspect. Psychol. Sci.* **10**(2), 176–199 (2015). <https://doi.org/10.1177/1745691615569000>
- Swan, K.: Learning effectiveness online: what the research tells us. In: Bourne, J., Moore, J. (eds.) *Elements of Quality Online Education, Practice and Direction*. Sloan Center for Online Education, Needham, MA, pp. 13–45 (2003)
- Tseng, J., Chu, H., Hwang, G., Tsai, C.: Development of an adaptive learning system with two sources of personalization information. *Comput. Educ.* **51**, 776–786 (2008)
- Tyton partners: learning to adapt: understanding the adaptive learning supplier landscape (2013). http://tytonpartners.com/tyton-wp/wp-content/uploads/2015/01/Learning-to-Adapt_Supplier-Landscape.pdf. Accessed 03 Jan 2019
- Tyton Partners: learning to adapt 2.0: the evolution of adaptive learning in higher education (2016). <http://tytonpartners.com/tyton-wp/wp-content/uploads/2016/04/tyton-Partners-Learning-to-Adapt-2.0-FINAL.pdf>. Accessed 09 Jan 2019
- Vanderwaetere, M., Desmet, P., Clarebout, G.: The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Comput. Hum. Behav.* **27**, 118–130 (2011). <https://doi.org/10.1016/j.chb.2010.07.038>
- VanLehn, K.: The behavior of tutoring systems. *Int. J. Artif. Intell. Educ.* **16**, 227–265 (2006). <https://doi.org/10.1145/332148.332153>
- VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educ. Psychol.* **46**(4), 197–221 (2011)
- VanLehn, K.: Regulative loops, step loops, and task loops. *Int. J. Artif. Intell. Educ.* **26**, 107–112 (2016). <https://doi.org/10.1007/s40593-015-0056-x>
- Vygotsky, L.: *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge (1978)
- Woolf, B., Cunningham, P.: Multiple knowledge sources in intelligent teaching systems. *IEEE Expert* **2**(2), 41–54 (1987). <https://doi.org/10.1109/MEX.1987.4307063>
- Yarnall, L., Means, B., Wetzel, T.: *Lessons learned from early implementations of adaptive courseware*. SRI Education, SRI International. Menlo Park, CA. (2016). https://www.sri.com/sites/default/files/brochures/almap_final_report.pdf. Accessed 19 Sep 2018
- Zhang, L., VanLehn, K.: Adaptively selecting biology questions generated from a semantic network. *Interact. Learn. Environ.* **25**(7), 828–846 (2017). <https://doi.org/10.1080/10494820.2016.1190939>