Just Enough Fidelity in Student and Expert Modeling for ITS

Making the Practice Practical

Brandt Dargue[1] and Elizabeth Biddle[2]

[1] Boeing Research & Technologies (BR&T), St. Louis, MO, USA
Brandt.W.Dargue@Boeing.com
[2] Boeing Research & Technologies (BR&T), Orlando, FL, USA
Elizabeth.M.Biddle@Boeing.com

Abstract. Intelligent Tutoring Systems (ITSs) are usually comprised of three primary models – an expert model, a domain or system model, and a student model. Many of these models are quite complex to enable just about any learner to get the optimum tailored experience possible. These systems have shown great results, typically at least one standard deviation (a letter grade) better than traditional training (e.g. [1] [2]). This complexity not only ensured that ITSs were successful, it also prohibited their widespread use (e.g. [3]). Results of studies in which the expert and system models were simplified show similar gains in effectiveness (e.g. [4] [5]), suggesting that lower-cost ITSs can be just as effective as those developed at higher costs. This paper compares the results of effectiveness studies in which the ITSs had various levels of fidelity and presents some recommended guidelines in determining the level of fidelity for student, expert, and system models of the ITS.

Keywords: Cognitive Modeling, Perception, Emotion and Interaction, Machine Learning, Neural Networks Techniques for Data Processing: Adaptive User Interfaces, Human performance improvement; Intelligent Tutoring; Adaptive training.

1 Introduction: The Goals of ITS (and This Paper)

The goal of Intelligent Tutoring Systems (ITSs) since the 1970's has been to provide the same level of learning effectiveness as an apprentice in the field with a personal mentor. Thus, the ITS training environments model both the mentor as well as the work environment ("the field" of practice). Research by Bloom [6] showed that one-to-one tutoring provides an improvement of two standard deviations (nearly two letter grades) in effectiveness over group instruction. In a perfectly logical manner, the approach in designing and building ITSs continues to model the expertise and instructional support a mentor provides with the "two sigma" effectiveness improvement goal. The period of rapid ITS growth coincided with the rapid growth of artificial intelligence (AI) software which had the goal of providing a machine that thinks like a person. The Turing Test was exemplified as the goal of AI – to have a

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC 2014, LNAI 8534, pp. 202–211, 2014.

machine/computer/program that has intelligence indistinguishable from that of a human. The Turing Test has also been used to evaluate ITS deployments [7]. This close association between AI and ITS goes back as far as Scholar [8], developed in 1970 by Jaime Carbonell [8]. As early as this system, recognized as the first ITS, a common goal of ITS has been to have the system's artificial mentor be able to hold discussions with the learner in natural language dialogs [9]. Most ITS limited the discussion to the exercise, problem, or topic the learner is currently working. Some ITS strived to enable natural language discussions for any topic in the domain the learner wants to discuss. Some ITSs were built to be able to solve problems posed by the learner (e.g. SOPHIE [10]). Considerable research has also been directed at applying advanced AI techniques to assess and provide detailed mentoring feedback to the learner (e.g. Scholar [8]). These goals have driven the cost of ITS to the point that development and usage never became widespread. Even researchers in 1997 [11] noted that the research in ITS had focused on the science of using AI for dialogs rather than the art of effective pedagogical dialogs. Corbett, Koedinger, and Anderson further stated that "for intelligent tutors to seriously penetrate the educational/training system, the evaluative focus must begin to shift to educational impact and away from artificial intelligence sufficiency [11, p. 850].

> *Fidelity: "1a: the quality or state of being faithful; 1b: accuracy in details: exactness; 2: the degree to which an electronic device (as a record player, radio, or television) accurately reproduces its effect (as sound or picture)." (Merriam-Webster)* [12]*"*

Finding the optimum level of fidelity for the models used with ITS can strike the right balance between effectiveness and cost to make the practice more practical. This paper was born out of the observation that most of the studies performed on measuring the effectiveness of various ITS instances seemed to all result in an effect size of 1 to 1.5 standard deviations when compared to other forms of training. This is true for both very sophisticated ITS that can pass the Turing test and for simple systems that basically only added the capability to select the next problem. The authors set out to perform meta-research testing this observation. The approach was to analyze data from ITS effectiveness studies similar to Durlach & Ray [1] to detect correlations between types of models used for the ITS(s) used in the studies and the effectiveness measured in the study. Inconsistency in available studies and vast differences in the approach used to build the ITS made the comparisons and correlation analysis difficult without harmonization such as that proposed by Pavlik & Roth [13]. As noted by Durlach and Ray: "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." [1, p. v]

2 The Models of ITS

ITS are typically comprised of a user interface which models/simulates the system or problem solving environment; a model of the expert's solution, the expert's cognitive processes, and the rules of the domain; a model of the student/learner's knowledge, skills, attitudes, misunderstandings, and measure of performance; and a model of the instructional approach to mentoring the learner in the pursuit of expertise [14] [11] [15] [16].

- The Models of ITS:

- User Interface Model or Problem Solving Environment
- Expert/Cognitive/Domain/System Model
- Student Model
- Pedagogical/Instructional/Tutoring/Mentoring Model

For this paper, we consider the degree to which humans are modeled in an ITS and consequently, the remainder of this paper will focus the expert, student, and instructional models. For a discussion of fidelity of the user interface and the system/domain portion of the expert model, please refer to references such as the US Modeling & Simulation Coordination Office [17] recent and current research on the fidelity of simulations e.g. [18] and training e.g. [19]. Note that the typical models of ITS separate the expert and the tutor whereas an actual human mentor usually has to provide both roles. A human tutor is responsible for both providing the KSA and for adjusting the approach in transferring the KSA. In the models of ITS, it is the expert model that contains what an expert knows about the domain and the Instructional model that determines the best approach for the individual learner to gain that expertise. In an ITS, each of the four models must work closely together. For example, the expert model assesses the learner by observing the user interface model and comparing the learner's actions to the actions predicted by the expert model. The results of this assessment form the primary basis of the student model. The instructional model uses the student model to determine how to best enable the learner to gain a higher level of expertise. The responsibility of providing the specific information that follows that plan of instruction is provided by the expert model. While the nomenclature separates functionality into these models, these tasks do not necessarily need to be authored or implemented as separate software components. However, it is critical that the functionality is well integrated. Before delving into fidelity issues, the following subsections briefly review the components of an ITS.

2.1 Expert Model

The Expert Model contains the expertise of the expert. It models what an expert knows about the domain, what an expert would do in a particular situation, and the reasoning an expert would apply in determining what to do. In other words, it models the knowledge (what), skills (how), and attitudes (why) of an expert. There are many approaches from the most simple documents to systems that determine solutions to problems as mentioned in the introduction section. The overall question is "Does the ITS actually need think like an expert?" We believe it is more effective to describe the fundamental knowledge and thought processes an expert uses to create a solution rather than model the expert's solution for each problem to promote the development of expertise [20].

2.2 Student Model

In its most strict form, the student model is the ITS component that manages the knowledge history and current states of the student [21]. In an expanded form, the student models also manages the current history and state of a wide range of variables that effect learning such as emotion [22], motivation [23] [24], and off-task behavior [25] [26]. The student model acquires data through the student's interaction with the ITS environment. Task performance data is collected by comparing a set of expected actions to the student's actions, such as a keystroke [27], answer selection [28] or specific movement in a virtual simulation [26] or evaluation of metrics that rate task performance [20]. Data for other learning variables is often obtained through comparison of the student actions to actions associated with non-task behaviors such as boredom or anxiety [29] or through physiological or neurophysiological sensor data [30].

2.3 Instructional Model

The instructional model is the component that provides that facilitates the learning activities provided by the ITS. It decides when to intervene with an instructional action, what instructional action to apply, and how the instructional action should be implemented (e.g., hints, feedback, remediation). VanLehn [2] has likened the instructional model functions to the method a human tutor uses, such as working through each step, and sub-step, of a problem.

There exists a wide range of adaptive learning techniques that have been used to develop adaptive training applications. The manner in which adaptive training techniques are applied, or combinations of these techniques are applied, is diverse due to a lack of a common methodology or design principle for implementing adaptive training [1]. Olney and Cade [31] recommend that certain instructional strategies are more or less suitable with different types of adaptive (include ITS) learning technologies.
Durlach [32] proposed the Framework for Instructional Technology (FIT) to provide a structured method for selecting an appropriate adaptation strategy. FIT is based on the assumptions that instructional technology should provide feedback and employ scaffolding and master learning methods [32], and identifies four adaptation decisions to be considered in determining an appropriate adaptive learning strategy to apply in a given learning situation. These decisions are whether to provide: 1) correct feedback, 2) support, 3) micro-sequencing, or 4) macro-sequencing as the adaptation strategy type. These decisions are made based on the student's performance relative to mastery performance criteria, and each strategy type has five levels of adaptation, depending upon the level of intervention the student requires. Similarly, Sampson and Karampiperis [33] describe a detailed process to define adaptive selection and sequence of learning objects in Adaptive Educational Hypermedia Systems (AEHS).

However, the only mechanism used to sequence to didactic content for instruction and remediation in the adaptive training used in independent studies by Perrin [4] and Durlach [34] [5] was student scores on assessments. The studies by Perrin, Banks and Dargue [4] directly compared instructor-led classroom training with three types of computer -based training (linear, mastery learning, and adaptive remediation) with as identical as possible content. The software for both the mastery learning (ML) and the adaptive remediation (AR) conditions assessed the learner and remediated to specific content not mastered in the assessment. The key difference for the adaptive remediation condition is that each error the learner made could be associated with more than on content module. For both of these groups, learning performance exceeded self-directed or classroom instruction by 1.2 standard deviations or more (see Figure 1). Perrin further compared the results with predictions of performance gains from additional time. The actual learning performance scores of these groups was more than 1.8 standard deviations greater than learning scores predicted from student allocated study time as shown in Figure 2.

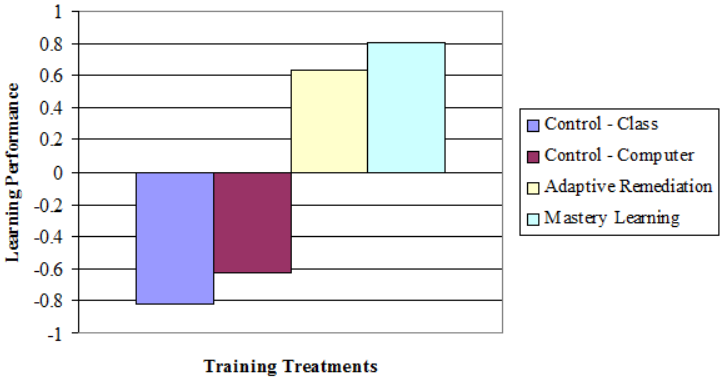

Fig. 1. Mean Learning Score by Group (Perrin [4])

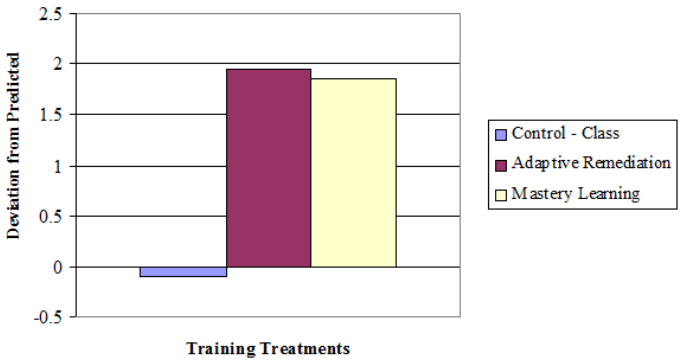

Fig. 2. Difference between Predicted and Actual Learning by Group (Perrin [4])

3 ITS Model Fidelity Framework

In order for the ITS to integrate the contributions of the Expert, Student, and Instructional Models into a cohesive instructional experience, it is necessary for these components to be highly synchronized. Given this necessity, it is vital that these individual functions or components are implemented with the same degree of fidelity. First, we will examine the concerns and issues regarding ITS fidelity as a whole. Then, we will present a framework that can be used to further examine and provide guidance in determining ITS fidelity requirements.

The development of ITSs has traditionally been a time and resource intensive effort due to the focus of modeling the expert, and subsequently the student's, knowledge at the cognitive level. In the case of student models, there have been efforts over the past decade focused on the design and development of student models that are still effective in accurately representing the student's learning state, while reducing the development burden. Aleven, McLaren and Sewall [35] have been investigating the development of a large-scale ITS through the use of an open-access web site. VanLehn [36] contends that of the granularity of student assessment, "the degree of aggregation over the task domain's knowledge" (p. 248) should be appropriate for the type of task or skill being evaluated. This notion that the methods and level of detail that student knowledge is modeled is dependent upon the needs of the particular learning task or domain is also supported by Woolf [37].

Consequently, the issue of student model fidelity is complex, and various methods for categorizing and comparing student models have been proposed. Brusilovskiy [21] recommended that the design and development of a student model needs to consider knowledge representation and diagnostics. Murray [38] defined a set of design trade-offs for developing a student model that considered Power and Flexibility in terms of breadth and depth, Usability in terms of learnability and productivity, Fidelity, and Cost. Similarly, Ohlsson and Mitrovic [39] proposed that the method or technique used to representation knowledge consider two dimensions: cognitive fidelity and cost.

Woolf [16, p. 51] suggested a two-dimensional mapping of knowledge domain complexity based on Lynch's recommendations on defining ill-defined domains [40] as shown in Figure 3.

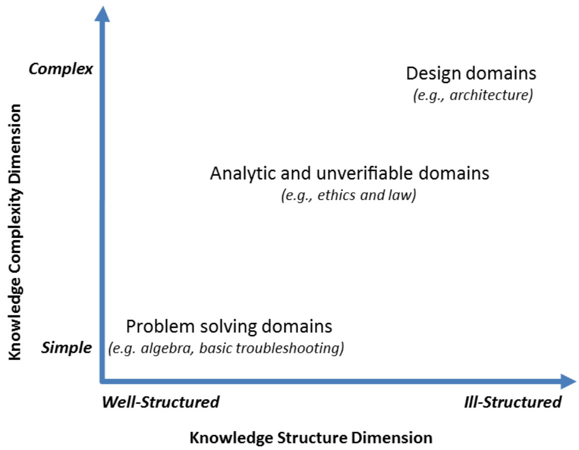

Fig. 3. Domain Knowledge Complexity (Based on Woolf [16] & Lynch [40])

Because the answers to problems in the lower-left are explicit, assessment of learners and instructional feedback are also somewhat simple. In the center of the graph, there is no clear-cut right and wrong. In the design domains, the goal of applying the knowledge is creativity and innovation. The studies by Perrin [4] and Durlach [34] [5] were in the center of this figure. Both taught content that includes declarative or background knowledge, as well as the use of this information in the exercise of judgment. The former was an export control ethics class and the later was on tactical command and control of small unmanned aerial systems.

Consider Bloom's taxonomy and three domains of learning: cognitive, affective and psychomotor. The cognitive domain has been studied quite extensively and in 2001, the Bloom's Cognitive Taxonomy was revised [41], and is in current use by various educational institutions (e.g., Iowa State, Penn State). Propose to evaluate the fidelity of representation required for cognitive, affective and psychomotor models in ITSs. The revised Bloom Cognitive Taxonomy is shown in Figure 4 plotted in two dimensions. These taxonomies can be used to assess where in the taxonomy to which the ITS application is focused to guide the determination of fidelity level, with the assumption that the lower the knowledge dimension (e.g., factual knowledge) and lower the cognitive process dimension (e.g., remember) the more granular the cognitive assessment required.

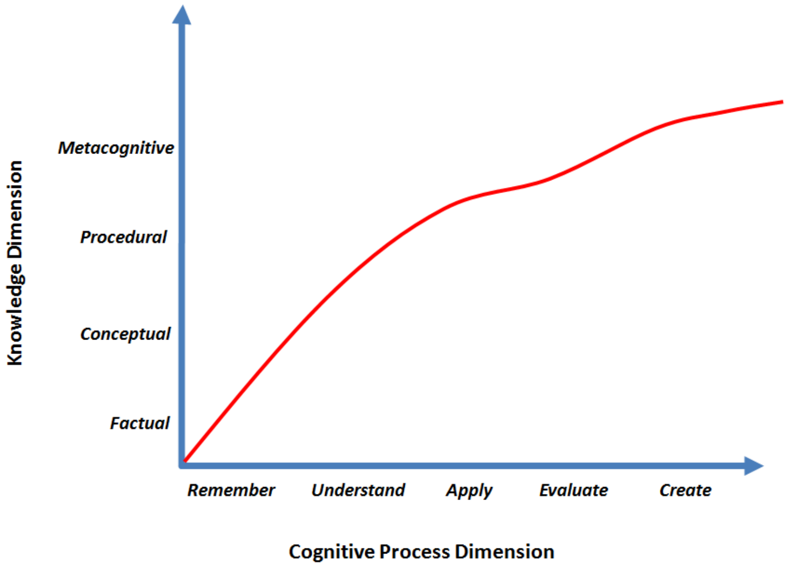

Fig. 4. Revised Bloom Cognitive Taxonomy (adapted from Krathwohl [41])

4 Conclusions

When building adaptive learning systems, the goal should not be to create the best tutor, the focus should be to produce proficient learners. While the system can model the techniques used by expert tutors, the system should leverage technology to go beyond the capabilities of a human tutor. Adaptive training does not need to be intelligent to be effective.

The fidelity of the models in an ITS should be driven by the learning analysis and resulting learning objectives. The Expert/Domain model needs to be able to impart the same level of cognitive KSA required expected of the trained person. The Student model needs to have the correct level of fidelity to measure whether the appropriate KSA has been achieved. The Instructional model needs to have the fidelity required to properly guide the learner through the Domain model. The Revised Bloom Cognitive Taxonomy can be used as a guide to determine the appropriate level.

The more complex the models are, the more expensive they are to develop, validate, and maintain. Evidence shows that simple approaches are effective but direct study comparisons of complexity/fidelity are scarce. As the saying goes, “Keep it simple, stupid” and follow the definition of fidelity to remain faithful in the endeavor.

References

1. Durlach, P.J., Ray, J.M.: Designing adaptive instructional environments: Insights from empirical evidence. Technical Report 1297, Army Research Institute, Arlington, VA (2011)
2. VanLehn, K.: The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems. Educational Psychologist 46(4), 197–221 (2011)
3. Brawner, K., Holden, H., Goldberg, B., Sottilare, R.: Recommendations for Modern Tools to Author Tutoring Systems. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL (2012)
4. Dargue, B.W., Banks, F.Z., Perrin, B.M.: Student vs. Software Pacing of Instruction: An Empirical Comparison of Effectiveness. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL (2004)
5. Durlach, P.J., Dargue, B.W.: An Adaptive Training Prototype for Small Unmanned Aerial System Employment. In: 23rd Florida Artificial Intelligence Research Society Conference (FLAIRS-23), Daytona Beach, FL (2010)
6. Bloom, B.S.: The Two Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. Educational Researcher 13(4-6) (1984)
7. Person, N.K., Graesser, A.C.: Human or computer?: AutoTutor in a bystander Turing test. In: Cerri, S.A., Gouardéres, G., Paraguaçu, F. (eds.) ITS 2002. LNCS, vol. 2363, pp. 821–830. Springer, Heidelberg (2002)
8. Carbonell, J.: Mixed-Initiative Man-Computer Instructional Dialogue. MIT, Cambridge (1970)
9. Stevens, A.L., Collins, A.: The goal structure of a Socratic tutor. In: Proceedings of the ACM Conference, New York (1977)
10. Brown, J.S., Burton, R.R., Zdybel, F.: A Model-Driven Question-Answering System for Mixed-Initiative Computer-Assisted Construction. IEEE Transactions on Systems, Man and Cybernetics SMC-3(3), 248–257 (1973)
11. Corbett, A.T., Koedinger, K.R., Anderson, J.R.: Intelligent Tutoring Systems, Chapter 37. In: Handbook of Human-Computer Interaction, 2nd Completely Revised edn., pp. 849–874. Elsevier, Amsterdam (1997)

12. Fidelity, http://www.merriam-webster.com/dictionary/fidelity (accessed February 17, 2014)
13. Pavlik Jr., P., Toth, J.: How to Build Bridges between Intelligent Tutoring System Subfields of Research. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010, Part II. LNCS, vol. 6095, pp. 103–112. Springer, Heidelberg (2010)
14. Wikipedia: Intelligent tutoring system (December 3, 2013), http://en.wikipedia.org/wiki/Intelligent_tutoring_system#Structure (accessed February 16, 2014)
15. Nkambou, R., Bourdeau, J., Mizoguchi, R.: Introduction: What Are Intelligent Tutoring Systems, and Why This Book? In: Nkambou, R., Bourdeau, J., Mizoguchi, R. (eds.) Advances in Intelligent Tutoring Systems. SCI, vol. 308, pp. 1–12. Springer, Heidelberg (2010)
16. Woolf, B.P.: Building Intelligent Interactive Tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann Publishers/Elsevier, Amsterdam (2009)
17. Fidelity, http://www.msco.mil/VVA_RPG.html, http://vva.msco.mil/Special_Topics/Fidelity/default.htm (accessed February 20, 2014)
18. Gross, D.C.: Report from the Fidelity Definition and Metrics Implementation Study Group (FDM-ISG). Simulation Interoperability Standards Organization (SISO), Orlando (1999)
19. Stacy, W., Walwanis, M.M., Wiggins, S., Bolton, A.: Layered Fidelity: An Approach to Characterizing Training Environments. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL (2013)
20. Perrin, B.: Intelligent tutoring systems: Facilitating learning while holding to standard practice. In: Proceedings to the 2009 ITEC, Brussels (2009)
21. Brusilovskiy, P.L.: The construction and application of student models in intelligent tutoring systems. Journal of Computer and Systems Sciences International 32(1) (1994)
22. McCombs, B.L., McDaniel, M.A.: On the design of adaptive treatments for individualized instructional systems. Educational Psychologist 16(1), 11–22 (1981)
23. del Soldato, T., du Boulay, B.: Implementations of motivational tactics in tutoring systems. Journal of Artificial Intelligence in Education 6(4), 337–378 (1995)
24. Tong, A., Agelides, M.: Formalising tutoring strategy selection in multimedia tutoring systems. Information and Software Technology 41, 67–90 (1999)
25. Baker, R.S.: Modeling and understanding students' off-task behavior in intelligent tutoring systems. In: Computer/Human Interaction (CHI) 2007, San Jose (2007)
26. Rickel, J., Johnson, W.L.: Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. Applied Artificial Intelligence 13, 343–382 (1999)
27. Anderson, J.: Rules of the Mind. L. Erlbaum Associates, Hillsdale (1993)
28. Lane, H.C., Core, M.G., Gomboc, D., Karnavat, A., Rosenberg, M.: Intelligent tutoring for interpersonal and intercultural skills. In: Proceedings to the Interservice/Industry, Training, Simulation & Education Conference (I/ITSEC), Orlando (2007)
29. D'Mello, S.K., Graesser, A.: Language and discourse are powerful signals of student emotions during tutoring. IEEE Transaction on Learning Technology 5(4), 304–317 (2012)
30. Sottilare, R.A., Goldberg, B.S., Brawner, K.W., Holden, H.K.: A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In: Proceedings to the Interservice/Industry, Training, Simulation & Education Conference (I/ITSEC), Orlando (2012)
31. Olney, A.C.W.: Matching learner models to instructional strategies. In: Design Recommendations for Intelligent Tutoring Systems. Learner Modeling, vol. 1, pp. 23–28

32. Durlach, P.J.: Vanilla, Chocolate, or Chunky Monkey: Flavors of Adaptation in Instructional Technology. In: iFest 2012, Orland, FL (2012)
33. Sampson, D.G., Karampiperis, P.: Decision Models In The Design Of Adaptive Educational Hypermedia Systems. In: Graf, S., Lin, F., Kinshuk, McGreal, R. (eds.) Intelligent and Adaptive Learning Systems, Hershey, PA, pp. 1–18. IGI Global (2011)
34. Durlach, P.J., Dargue, B.W.: Adaptive and Non-Adaptive Training Technology for Small Unmanned Aerial System Employment. Poster Paper for The 27th Army Science Conference (ASC), Orlando, FL (2010)
35. Aleven, V., McLaren, B., Sewall, J.: Scaling up programming by demonstration for intelligent tutoring systems development: an open-access web site for middle school mathematics learning. IEEE Transactions on Learning Technologies 2(2) (April-June 2009)
36. VanLehn, K.: The behavior of tutoring systems. International Journal of Artificial Intelligence in Education 16(3), 227–265 (2006)
37. Woolf, B.P.: Student modeling. In: Nkambou, R., Bourdeau, J., Mizoguchi, R. (eds.) Advances in Intelligent Tutoring Systems. SCI, vol. 308, pp. 267–279. Springer, Heidelberg (2010)
38. Murray, T.: A overview of the intelligent tutoring system authoring tools: Updated analysis of the state of the art. In: Murray, T., Blessing, S., Ainsworth, S. (eds.) Authoring Tools for Advanced Technological Learning Environments, The Netherlands, pp. 493–546. Kluwer Academic Publishers (2003)
39. Ohlsson, S., Mitrovic, A.: Fidelity and Efficiency of Knowledge Representations for Intelligent Tutoring Systems. Technology, Instruction, Cognition and Learning (TICL) 5(2-3-4), 101–132 (2007)
40. Lynch, C., Ashley, K.D., Aleven, V., Pinkwart, N.: Defining 'Ill-Defined Domains ': A Literature Survey. In: Eighth International Conference on Intelligent Tutoring Systems Workshop on Intelligent Tutoring Systems for Ill-Defined Domains, Hongli, Taiwan (2006)
41. Krathwohl, D.R.: A revision of Bloom's taxonomy: An overview. Theory into Practice 41(4) (2002)