



*Listening. Learning. Leading.*

# Using Evidence-Centered Design for Learning (ECDL) to Examine the ASSISTments System

Mingyu Feng<sup>1</sup>

Eric G. Hansen<sup>2</sup>

Diego Zapata-Rivera<sup>2</sup>

<sup>1</sup> Worcester Polytechnic Institute

<sup>2</sup> ETS, Princeton, NJ

Paper presented at the annual meeting of the  
American Educational Research Association (AERA) and the  
National Council on Measurement in Education (NCME)

April 13-17, 2009, San Diego, CA.

Unpublished Work Copyright © 2009 by Educational Testing Service. All Rights Reserved. These materials are an unpublished, proprietary work of ETS. Any limited distribution shall not constitute publication. This work may not be reproduced or distributed to third parties without ETS's prior written consent. Submit all requests through [www.ets.org/legal/index.html](http://www.ets.org/legal/index.html).

*Educational Testing Service, ETS, the ETS logo, and Listening. Learning. Leading.* are registered trademarks of Educational Testing Service (ETS).



## **Abstract**

There is a great need for improved methods for designing innovative learning systems. In this paper, we introduce a design methodology – Evidence Centered Design for Learning (ECDL) that supports the design of educational systems in terms of evidentiary arguments. The methodology is then applied to examine an online intelligent tutoring system that blends performance assessment and instructional assistance. We describe how the functionality of the system can be represented using the argument structure of ECDL and present three cases that shows how the ECDL methodology and the system can enhance each other.



## Introduction

Nowadays, there are more and more learning-oriented products, such as intelligent tutoring systems (e.g. Shute & Psozka, 1996; Koedinger, Anderson, Hadley, & Mark, 1997), computer-aided instruction and educational games, and immersive environments (Squire, 2002; Gee, 2003). They include (a) instruction, training or guidance, and (b) assessment for formative and/or summative purposes. For such products, learning effectiveness and learning efficiency are key desirable characteristics. Additionally, the products need to be sufficiently engaging to motivate learners to persist in their use. Yet, since assessments are still an important part of learning-oriented products, concerns such as validity and accessibility are also critical.

Though many learning-oriented products have been developed and delivered, not much effort has been devoted to the research question of “how to enable the design of high-quality learning-oriented products” and thus the quality and effectiveness of these products varies. Consequently, there is a great need for principles or guidelines to lead the design of high-quality learning products. Evidence Centered Design (ECD) (Mislevy, Steinberg, & Almond, 2003)<sup>1</sup> has been useful for the design of high-quality assessments, and it was extended and applied by Hansen, Zapata-Rivera, and Feng as a design methodology for diverse students for learning oriented products (ECDL) (Hansen, in submission; Hansen & Zapata, 2008a Hansen & Zapata, 2008b, Hansen, Zapata-Rivera, & Feng, 2009).

---

<sup>1</sup> Earlier relevant works are those of Cronbach & Meehl (1955), Messick (1994); Kane (1992) and Mislevy (1994).



This paper examines an existing research application, the ASSISTment system in light of ECDL, as part of a more extended effort to pilot test and refine ECDL (Hansen, Zapata-Rivera, Feng, in preparation). The ASSISTment system ([www.assistment.org](http://www.assistment.org); Razzaq et al., 2005; Razzaq et al., 2007) was selected for reasons such as its popularity, the availability of considerable research information, and the fact that it explicitly blends assessment and instructional assistance. Unlike most other assessment systems, ASSISTment offers instructional support to students by introducing a set of scaffolding questions and making available informative hint messages as students work on assessment tasks. Meanwhile, it provides a more detailed evaluation of their abilities to the teachers through a live, online reporting system (Feng & Heffernan, 2007a). It is intended that teachers will be able to use this detailed feedback to tailor their instruction in order to address the particular difficulties identified by the system.

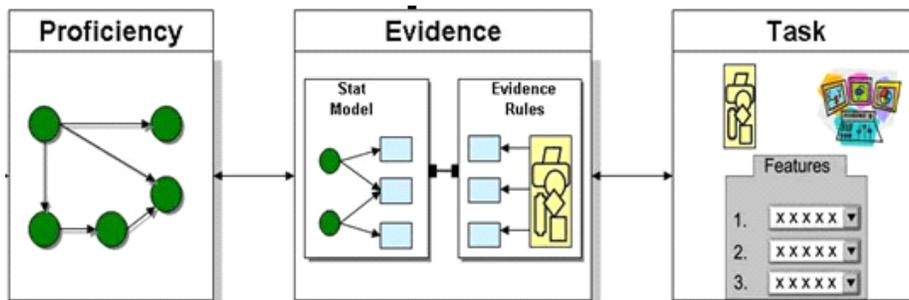
The paper is organized as follows. In the next section, we introduce the theoretical framework of ECD and ECDL. Then, the ASSISTment system is described in section 3. Our methods and evidences will be reported in section 4. Finally, we conclude in section 5.

### **Theoretical framework**

Evidence Centered Design of assessment (ECD) is a design methodology for the design of educational assessments in terms of evidentiary arguments. Figure 1 presents the basic structure of this assessment design approach. Working out these models and their relationships is a way to address concerns posted by Messick (1994, p17):



- “What complex of knowledge, skills, or other attribute should be assessed?” *The proficiency model* (sometimes called student model or competency model) describes characteristics of examinees upon which the inferences are to be based.
- “What behaviors or performances should reveal those constructs, and what tasks or situations should elicit those behaviors?” *The task model* describes features of situations that will be used to elicit performance and how to structure the situations to obtain the kinds of evidence needed for the evidence model.
- “The rational development of construct-based scoring criteria and rubrics.” *The evidence model* expresses how what is observed in a given task constitutes evidence about student model variables. It includes scoring criteria and rubrics and the statistical model of how to update the belief of student knowledge given the performance.



**Figure 1. The three central models of an evidence centered assessment design**

ECDL augments the main ECD models with a pedagogical model that represents how to *foster* growth and learning, given existing evidences and student proficiency level. ECDL also provides a way to consider a wide range of quality characteristics, including learner engagement, learning effectiveness and efficiency, validity, and accessibility, in the



context of a more comprehensive product quality argument. ECDL is intended to be useful not only for the design of new products but also the redesign of existing products.

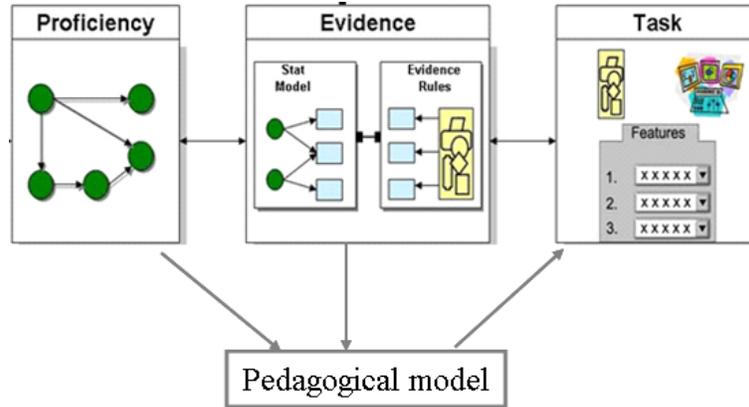
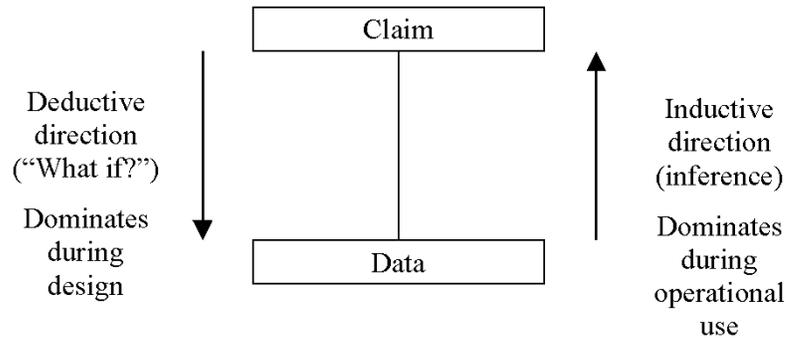


Figure 2. ECDL models: ECD with an added pedagogical model

It is worth pointing out the importance of assessments in learning-oriented products, in which the assessment may be used to: (a) detect learning, (b) to guide learning (such as by helping identify the next appropriate learning activity or providing task-specific feedback) (Black & William, 1998). Thus, success in the domain of learning-oriented products relies heavily upon high-quality assessments. ECDL addresses the issue of assessment validity in learning-oriented products by establishing arguments for both the *design phase (deductive reasoning)* and the *operational use phase (inductive reasoning)*. Figure 3 shows how a claim is connected to data during the two phases. Reasoning *deductively*, we say, “If the student’s proficiency value is high (or low) *then* high (or low) scores should be obtained on given items.” On the other hand, during operational use of the assessment, we emphasize *inductive* reasoning (i.e., inference), “If high (or low) scores are obtained on these problems, then the student’s proficiency value is high (or low).” While *inductive* reasoning is what we ultimately care about in an assessment, it is



the *deductive* reasoning that takes precedence, both logically and chronologically. In the deductive reasoning we needed to play “what if”—which involves considering what would happen if we knew that the student had a certain level of proficiency and imagining that student interacting with the tasks and then determining what scores would be obtained.



**Figure 3. Basic argument structure for an assessment showing two directions of logical flow**

### **The ASSISTment System**

For this study, we picked the ASSISTments system, an online system, to help identify the high-level quality criteria for learning oriented products. As a relatively popular learning system, ASSISTment system has a significant amount of empirical support, and it has demonstrated its success on improving student learning in mathematics. We will consider various aspects of the ASSISTment system, including the research goal, assessment components, instructional interventions, the domain, the role of users, the usage of the system, research findings, etc.

*Research goal.* As a derivative of the Cognitive Tutors (Anderson et al., 1995; Koedinger et al., 1997), the ASSISTment project started in the year 2004. In many states there are concerns about poor student performance on high-stakes standards-based tests required



by the No Child Left Behind (NCLB) Act. However, limited classroom time compels teachers to choose between time spent *assisting* students' development and time spent *assessing* students' abilities. Noticing these facts, Heffernan at Worcester Polytechnic Institute and his colleagues at Carnegie Mellon University started to build a system ("ASSISTment"<sup>2</sup>) to help resolve this dilemma. Traditionally the two areas of testing (i.e. Psychometrics) and instruction (i.e., math educational research and instructional technology research) have been separate fields of research with their own goals. Therefore, the goal of ASSISTments is to introduce a way whereby students can take an assessment, but at the same time, make sure that students' time is spent primarily on learning.

*Assessment components and instructional interventions.* Figure 4 shows a flow chart of how students interact with the ASSISTment system. This interaction usually consists of a main question and a tutoring session. The *main* question can be treated as an assessment task for which students need to submit an answer. In contrast to a traditional testing environment, students can make a request for tutoring if they don't know the answer, though it is generally thought to be pedagogically more desirable that a student submits a thoughtful answer before accessing the tutoring. The tutoring session consists of a set of *scaffolding* questions that lead the student one-step-at-a-time to the solution. Scaffolds generally provide on-demand, contextual hint messages guiding student in the solution path. After an answer is submitted for a scaffold, the tutor provides specific, constructive feedback (e.g, typically a specific buggy message if the cause for an incorrect response

---

<sup>2</sup> The term "Assistment" was coined by Kenneth Koedinger and blends Assessment and Assisting.



was anticipated by the system authors and, if not, then a generic feedback message). This interactive support provides encouragement for students to learn-by-doing. Meanwhile, the system collects data related to different aspects of student performance such as responses, speed, help-seeking behavior, and efforts as students interact with the system and produces a variety of reports (Feng & Heffernan, 2007a) for teachers to review their students' progress and to inform instruction.

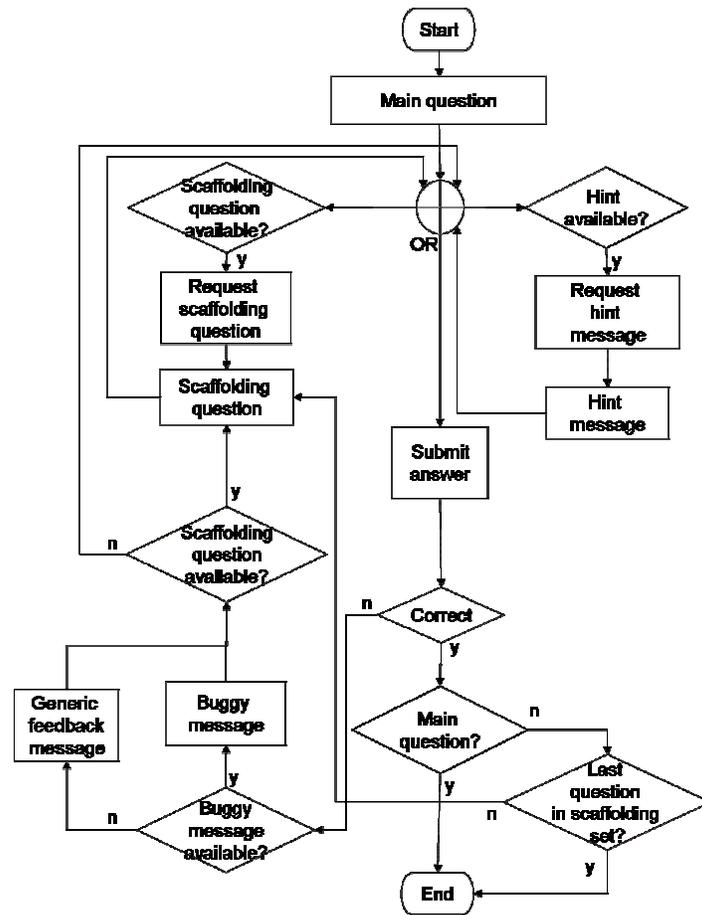


Figure 4. Flowchart showing interaction between an ASSISTment tutor and a learner.



*Research findings.* As a hybrid approach, ASSISTments demonstrated its success on both assessment and learning achievement. Administering scaffolding questions that focus on one step gives ASSISTments a better chance to assess student cognitive skills at a fine grained level (Feng, Heffernan, Mani, & Heffernan, 2006; Feng & Heffernan, 2007b; Pardos et al., 2007; Feng, Heffernan, Heffernan & Mani, accepted). The information collected during student-system interaction such as help-seeking behavior and problem-solving speed were used to obtain a more accurate estimate of student proficiency level (Feng, Heffernan & Koedinger, 2006; Feng, Heffernan, & Koedinger, in press). In terms of learning effectiveness, analysis of data from within ASSISTment usage shows ASSISTment effectively teaches (Razzaq et al., 2005; Feng, Heffernan, Beck & Koedinger, 2008). Also, several randomized controlled experiments have been carried out to compare the effectiveness of different tutoring strategies (Razzaq & Heffernan, 2006; Razzaq, Heffernan & Lindeman, 2007).

*The domain and usage.* ASSISTment has evolved to the third generation, changing from a java-based application, to a web-based e-learning and e-assessment system with rich user interface. It is targeted at helping middle school and high school students to be more prepared for the standardized test at the end of a school year. Over 3000 middle school and high school students from Worcester and Pittsburgh used the system as a part of their normal math class during the school year of 2006-2007. Teachers and students show positive attitudes towards the ASSISTment system in terms of effectiveness on fostering learning and help achieving instruction objectives.



## Methods and Evidences

While reviewing the system in light of an ECDL-based effort to develop a product argument, we tried to represent the major function of ASSISTment in an ECDL argument structure. We were also looking for ways to improve ECDL based upon the practice of ASSISTments, and how the system can be improved through this review process. In this section, we first illustrates an argument structure that represents the functionality of an ASSISTment tutor, followed by three cases, each showing a way in which ECDL applies to the ASSISTments system and/or what design principles in ASSISTments can be incorporated to enhance ECDL.

*An ECDL argument structure for an ASSISTment tutor* - ECDL provides a formal method, the Knowledge Skill and other Attribute (KSA) Value Matrix method (Hansen, in submission) that represents a wide range of considerations during the product design in a relatively simple way. The method essentially treats the design as an argument, wherein claims are made (e.g., claims about specific quality criteria), and then evidence is marshaled to support those claims. The table in Figure 5 demonstrates an argument structure as a KSA matrix for an ASSISTment tutor that addresses two 8<sup>th</sup> grade math skills: Congruence and Perimeter. This argument portrays a student who starts the tutor without being able to satisfy the requirement for “know math concept - congruence” and “know math concept – perimeter” but who acquires that through use of the system, specifically during the tutoring process of scaffolding questions.



A key element of the method is defining the targeted proficiency to be measured or improved by the learning-oriented product.<sup>3</sup> In Figure 5, this is done by identifying a set of KSAs (KSA column) and defining them as either “focal” (shown with a number in the “Focal value” column) or “nonfocal”. For instance, students need the ability to see (row 1) or know (row 2) certain math vocabulary to be able to correctly answer the questions in the ASSISTment, yet these are not abilities that the tutor intends to address. Therefore, these KSAs are defined as “nonfocal” and their focal values as “n/a”. Although the nonfocal KSAs are not the constructs to be assessed or to be taught, they are essential elements to be considered for the design of an assessment or a learning system. A system that imposes a higher level of requirements on the nonfocal KSAs than the initial profile values of the users may be confronted with issues on the validity of the assessments and the instructional efficacy of the system. (Such issues are critical not only for students with disabilities and English language learners, but also for students, who for any reason, e.g., lack of opportunity to learn; poor understanding of prior instruction, cannot satisfy the requirement for a nonfocal KSA.) On the contrary, the tutor is intended to teach a student the math concepts of “congruence” and “perimeter” if it finds evidence that the student has not mastered these constructs. Thus, the two KSAs in row 6 and 7 are defined as focal KSAs with focal value being 4. (Higher number indicates a higher level of capability.) Each row in the table also shows the requirement value for each KSA (i.e., the level of cognitive or other demand imposed by the task situation in order to perform well). The post-activity profile indicates the student’s status after the activity.

---

<sup>3</sup> Here, we assume the purpose of main item is to *measure* a targeted proficiency (construct), while the purpose of scaffolds is to *foster* the targeted proficiency (learning objective)



Particularly, rows 6 and 7 show that (a) before starting the tutor, the student's profile value is 2 (column 3); (b) the desired proficiency level of "congruence" is 4 (column 4); (c) since the main item in the ASSISTment serves as an assessment task, the intended growth outcome is not applicable (column 5), thus the student's profile value remains the same as 2 after the student answers the main question (column 6); (d) because the student's profile level on "congruence" is lower than the required level, he will answer the main item incorrectly, which will invoke the tutoring session (scaffolds), starting with scaffolding #1 whose requirement value for "congruence" is 3 (column 7). Additionally, scaffolding question #1 intends to improve students' mastery of math concept "congruence" to a level of 3 (column 8) as well. (e) After finishing the first scaffolding question, the second one will show up, focusing on the other skill "Perimeter" (column 10-12). And finally after two scaffolds, the student's profile value on both concepts increase to level 3. Notice that in this example, although both of the two concepts are required by the main item, each scaffolding question focuses on only one concept. Therefore, the requirement value and intended growth outcome are "n/a" for concept "congruence" (row 6, column 10-11) in scaffolding #2, and for concept "perimeter" in scaffolding #1 (row 7, column 7-8). This approach is consistent with a pedagogical approach of letting the student focus on one concept at time in the scaffolds.



	1	2	3	4	5	6	7	8	9	10	11	12
	Both main item and scaffolds	Main item	Main item	Main item	Main item	Scaffold #1	Scaffold #1	Scaffold #1	Scaffold #2	Scaffold #2	Scaffold #2	Scaffold #2
KSA	Focal value	Pre-activity profile value	Requirement value	Intended growth outcome	Post-activity profile value	Requirement value	Intended growth outcome	Post-activity profile value	Intended growth outcome	Requirement value	Post-activity profile value	Post-activity profile value
1 See	n/a	3	3	n/a	3	3	n/a	3	n/a	3	3	3
2 Hear	n/a	3	n/a	n/a	3	n/a	n/a	3	n/a	n/a	n/a	3
3 Have computer skills	n/a	3	2	n/a	3	2	n/a	3	n/a	2	3	3
4 Know non-math vocabulary	n/a	3	3	n/a	3	3	n/a	3	n/a	3	3	3
5 Know math vocabulary	n/a	3	3	n/a	3	3	n/a	3	n/a	3	3	3
6 Know math concept - Congruence	4	2	4	n/a	2	3	3	3	n/a	n/a	3	3
7 Know math concept - Perimeter	4	2	4	n/a	2	n/a	n/a	2	3	3	3	3

**Figure 5. The KSA matrix for an ASSISTment tutor with one main question and two scaffolding questions**

*Case A: Potential improvements on ASSISTments.* The design approach for ASSISTments is arguably more focused on tasks (development of main questions and scaffolding materials starts from released state exam items) than is ECD/ECDL, which makes explicit the argument structure between tasks and unobservable skills or other latent variables. But like ECD, ASSISTments also directly attribute individual differences to unobservable skills or other latent variables. In the development of ASSISTments, educational researchers and domain experts conducted cognitive task analysis on the released exam items to identify the fine-grained skills. For 8<sup>th</sup> grade mathematics, they built a cognitive model consisting of 106 skills (Feng, Heffernan, Heffernan & Mani, accepted), which from the perspective of ECDL can be considered as the proficiency model. During a coding session, content authors are asked to tag their questions (including both main questions and scaffolding questions) with one or more skills in the



cognitive model. To estimate student knowledge of a particular skill, the system considers questions that are tagged with the skill as assessment tasks (the task model) and treats student performance on the questions as evidence of their knowledge level of the particular skill to be assessed (the evidence model).

In ASSISTments, the inference of student proficiency level is rather simple. Students get full credit for a skill when they correctly answer the questions tagged with the skill, while in the case of a wrong answer to a question tagged with multiple skills, the system relies upon responses to scaffolding questions (typically tagged with only one skill per scaffold) to determine which skill "to blame" (i.e., attribute the cause of the wrong answer to the main question). Thus, the connection between proficiencies and tasks is relatively loose and informal. Also student proficiency level is not dynamically updated and thus assessment tasks are not assigned correspondingly during the tutoring process, which may impose a validity issue for assessment. Based on ECD, we argue for a more formal integration—perhaps by a more elaborated evidence model - between proficiencies and tasks, especially in the case of multi-tagged questions. This may improve the validity of the assessment. In ECD, the evidence model gives special attention to the role of probability-based reasoning in accumulating evidence across task performances, in terms of belief about unobservable variables that characterize the KSAs of students.

As the most substantive enhancement in ECDL relative to ECD, is the key function of the pedagogical model that describes “how” to move a learner from one state of proficiency to a higher state. The pedagogical model serves as a basis for determining what feedback to provide to the learner or what next learning activity to prescribe. The pedagogical model is *semi-dynamic* in the ASSISTment system in the sense that although the in-



problem feedback and scaffolding questions are presented based on a student's response, the set of scaffolds is fixed as is the order of scaffolds within the set rather than being determined based on the updated belief of student knowledge. Because the student knowledge level is constantly changing in a learning environment, administering a fixed set of tasks may be pedagogically suboptimal, thereby adversely impacting learning effectiveness. An ECDL-based analysis might suggest determining the next activity based on information such as the specific instructional intent (e.g., "focal value"), an estimate of student ability in a particular skill ("profile value"), and cognitive or other demands ("requirement value") imposed by a task situation, etc. It should be noted that such improvements to the pedagogical model of the ASSISTment system may entail greater effort in authoring than is currently employed.

*Case B: Reflection of ECDL principles in the ASSISTment system.* In contrast to ECD for assessment, the ECDL framework gives special emphasis to providing rationales for strategies to enable learning effectiveness and learning efficiency. For example, consider the strategy wherein if a student has a hard time answering a hard problem with a high level of requirement, then consider lowering the requirement for focal KSAs (i.e. the constructs to be measured or to be taught) to allow students to gradually "warm-up". A rationale for this strategy is that doing so will reduce the likelihood of cognitive overload that might impede student learning (Mayer, 2008). One can also employ another strategy of gradually increasing the complexity of instructional tasks to avoid cognitive overload while building skill (Van Merriënboer, Kirschner, and Kester, 2003). The scaffolding strategy in the ASSISTments can be considered as a good example of avoiding cognitive overload. By decomposing a multi-step problem into sub-steps, and



presenting them sequentially, ASSISTments reduces the amount of new information novices must process at a time, and enhances learning effectiveness.

The ECDL methodology suggests that designers document the rationale of what activities might be *efficient* in fostering learning. Specifically, it points out that lowering requirements for nonfocal KSAs (in contrast to focal KSAs) can potentially lead to improved learning efficiency, which is supported by an extensive body of research on cognitive load theory regarding how learning efficiency can result from reducing or minimizing “extraneous cognitive load” (Kirschner, Sweller, & Clark, 2006). Several practices in ASSISTments reflect this idea. For instance, because knowledge in vocabulary and math formulas is nonfocal (not the construct to be measured or to be improved) in ASSISTments, it keeps low the requirements for this knowledge by providing students with a reference sheet and the definition of the words (See Figure 6) to help them understand the question text. As another practice of reducing extraneous cognitive load in ASSISTments, whenever appropriate, the authors try to present the instructional support using visual diagrams and animated gif images to leverage the “modality effect” (Clark et al., 2006), and to use cues, signs to draw attention of the learners to avoid the “split-attention effect” (Sweller, 1998) (see Figure 7). These efforts potentially leave a larger working memory available for learners to assimilate instructional materials, i.e., to learn the targeted skills.

Triangles ABC and DEF are congruent. The perimeter of triangle ABC is 23 inches.

What is triangle?

A geometric figure with 3 sides and three vertices.

Figure 6. A pop-up window shows the definition of a word in a math problem.

Hint 1:

$$\frac{2}{3} = \frac{x}{15}$$

The first step in cross multiplication is to multiply the numerator of one fraction by the denominator of the other.

Hint 2:

$$\frac{2}{3} = \frac{x}{15}$$

Do this for both numerators and set the products equal to each other.

Figure 7. Scaffolding and hints use animated gif and signals to help students in ASSISTments.

*Case C: enhancing ECDL based upon empirical practice of ASSISTments.* Anderson et al. (1995) address the importance of practical classroom deployment and coordination with teachers for Cognitive Tutors. They pay attention to the curriculum that educators wanted to teach, what happens to students after they passed through the cognitive tutors, coordination with teachers and issues of the deployment of the tutors. However, designers of learning oriented products often ignore these factors. For instance, the ASSISTments system gives special attention to getting teachers involved in the process of content creation, class management, assessment and instruction. The content administered in ASSISTments is closely connected to school curricula. It provides user friendly web-



based authoring tools so that typical teachers can create tutors themselves without any programming. It allows teachers to share their problems with others, organize problems into problem sets and to assign the problem sets to their classes through a teacher toolkit. It presents live, online reports to help teachers to analyze the items, evaluate progress of their students and to find out the skills and steps that students have difficulty on so that teachers can adjust their instruction accordingly. All the factors, together with other effort in designing valid assessments and effective interventions, enable the success of the ASSISTments system. ECDL might be enhanced to better address these teacher-related variables.

Analyzing the ASSISTment system brings up additional issues and opportunities such as suggestions for improving the ECDL process as well as the identification of areas where the products or their supporting documentation might be improved. While it is easier to represent pure instruction or pure assessments than it is to represent assessment-instruction mixtures like ASSISTments, it can be done, as shown above, by subdividing the complex applications into sub-arguments, many of which might be considered either pure instruction or pure assessment. The articulation of the argument structure could be improved to reflect the role of teachers, and the intended outcomes of reports and other forms of feedback. If, for example, improvements in students' knowledge or abilities, including meta-cognition (self-awareness of one's abilities and areas for improvement) are essential outcomes of the use of an application, then these outcomes need to be articulated and means provided for validating that these outcomes are occurring. Our examination of ASSISTments also points out the need to improve the scalability of the procedure. For example, there is a need to develop principled ways of scaling the



procedure up to better accommodate a large scale system like ASSISTment with 100+ skills and 1400+ questions. Ideally, one would analyze many profiles, for example, for each application and then develop ways of summarizing the results and implications.

### **Conclusions**

In this paper, we illustrate how a successful learning oriented product can be described by ECDL and how ECDL's argument-based approach can help identify areas for improvement. Also, we illustrate how applying ECDL on existing research applications can help improving ECDL to make it more generally useful. As a part of the future work, we are working on applying ECDL to improve the accessibility of the ASSISTment system to accommodate for students with disabilities.



## References

1. 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.
2. Black P. J., & Wiliam D. (1998). Assessment and classroom learning. *Assessment in Education* 5, 7-74.
3. Clark, R., Nguyen, F., and Sweller, J. (2006). *Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load*. San Francisco: Pfeiffer.
4. Feng, M., Heffernan, N.T., & Koedinger, K.R. (2006). Addressing the Testing Challenge with a Web-Based E-Assessment System that Tutors as it Assesses. *Proceedings of the Fifteenth International World Wide Web Conference*. pp. 307-316. New York, NY: ACM Press. 2006.
5. Feng, M. & Heffernan, N. (2007a). Towards Live Informing and Automatic Analyzing of Student Learning: Reporting in ASSISTment System. *Journal of Interactive Learning Research*. 18 (2), pp. 207-230. Chesapeake, VA: AACE.
6. Feng, M. & Heffernan, N. (2007b). Assessing Students' Performance: Item Difficulty Parameter vs. Skill Learning Tracking. Paper presented at the National Council on Educational Measurement 2007 Annual Conference, Chicago.
7. Feng, M., Beck, J., Heffernan, N. & Koedinger, K. (2008). Can an Intelligent Tutoring System Predict Math Proficiency as Well as a Standardized Test? In Beck & Baker (Eds.). *Proceedings of the 1st International Conference on Education Data Mining*. Montreal, 2008.
8. Feng, M., Heffernan, N., Beck, J. & Koedinger, K. (2008). Can we predict which groups of questions students will learn from? In Beck & Baker (Eds.). *Proceedings of the 1st International Conference on Education Data Mining*. Montreal, 2008.
9. Feng, M., Hansen, E. G. & Zapata-Rivera, D. (in preparation). Using evidence centered design for learning to examine an Intelligent Tutoring System that assesses at it tutors.
10. Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.
11. Hansen, E. G., & Mislevy, R. J. (2005). Accessibility of computer-based testing for individuals with disabilities and English language learners within a validity framework. In M. Hricko & S. Howell (Eds.), *Online assessment and measurement: Foundation, challenges, and issues*. Hershey, PA: Idea Group Publishing, Inc.
12. Hansen, E. G. (in submission). Towards an evidence centered methodology for designing games for learning. Submitted to ETS Research Report Series.
13. Hansen, E. G., & Zapata-Rivera, D. (2008a). *Extending Evidence Centered Design for Learning (ECDL) for diverse educational applications*. Final report for the ECD for Learning-oriented Products Pilot Project.



14. Hansen, E. G., & Zapata-Rivera, D. (2008b). *Evidence Centered Design for Learning: A method for identifying educational potential in games*. Final report for the Games and Learning Project.
15. Hansen, E. G., Zapata-Rivera, D., & Feng, M. (2009, April). Beyond accessibility: evidence centered design for improving the efficiency of learning and assessment. Presentation at the annual meeting of the National Council on Measurement in Education, San Diego, California. April 16, 2009.
16. Koedinger, K., Anderson, J., Hadley, W., & Mark, M., (1997). Intelligent tutoring goes to school in the big city. Pittsburgh, PA: Human-Computer Interaction Institute, *Carnegie Mellon University. International Journal of Artificial Intelligence in Education*, 8, 30-43.
17. Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Education Researcher*, 23(2), 13-23.
18. Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-67.
19. Pardos, Z. A., Heffernan, N. T., Anderson, B. & Heffernan, C. (2007). The effect of model granularity on student performance prediction using Bayesian networks. The International User Modeling Conference 2007.
20. Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N.T., Koedinger, K. R., Junker, B., Ritter, S., Knight, A., Aniszczyk, C., Choksey, S., Livak, T., Mercado, E., Turner, T.E., Upalekar, R., Walonoski, J.A., Macasek, M.A., Rasmussen, K.P. (2005). The Assistment Project: Blending Assessment and Assisting. In C.K. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.) *Proceedings of the 12th International Conference on Artificial Intelligence In Education*, 555-562. Amsterdam: ISO Press.
21. Razzaq, L., Heffernan, N.T. (2006). Scaffolding vs. hints in the Assistment System. In Ikeda, Ashley & Chan (Eds.). *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*. Springer-Verlag: Berlin. pp. 635-644. 2006.
22. Razzaq, L., Heffernan, N.T., Lindeman, R.W. (2007). What level of tutor interaction is best? In Luckin & Koedinger (Eds.). *Proceedings of the 13th Conference on Artificial Intelligence in Education*. Amsterdam, Netherlands: IOS Press.
23. Razzaq, Feng, Heffernan, Koedinger, Nuzzo-Jones, Junker, Macasek, Rasmussen, Turner & Walonoski (2007). Blending Assessment and Instructional Assistance. In Nadia Nedjah, Luiza deMacedo Mourelle, Mario Neto Borges and Nival Nunesde Almeida (Eds). *Intelligent Educational Machines within the Intelligent Systems Engineering Book Series*. pp.23-49. (see <http://www.isebis.eng.uerj.br/>). Springer Berlin / Heidelberg.
24. Shute, V. J., & Psofka, J. (1996). Intelligent tutoring systems: Past, present, and future. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 570-600). New York: Macmillan Library Reference USA.



25. Squire, K. (2002, July). Cultural framing of computer/video games. *Game Studies*, 2(1). Retrieved from <http://www.gamestudies.org/0102/squire/> on September 28, 2007.
26. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257-285.



## Notes

1



**Appendix A**

**Appendix Title**



**Appendix B**

**Appendix Title**