Published on July 7, 2015
Olney, A. M., & Cade, W. L. (2015). Authoring Intelligent Tutoring Systems Using Human Computation: Designing for Intrinsic Motivation. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), Foundations of Augmented Cognition (Vol. 9183, pp. 628–639). Springer International Publishing.

Authoring Intelligent Tutoring Systems Using Human Computation: Designing for Intrinsic Motivation

Andrew M. Olney and Whitney L. Cade

Institute for Intelligent Systems
University of Memphis, Memphis TN 38152, USA,
aolney@memphis.edu,
WWW home page: http://andrewmolney.name

Abstract. This paper proposes a methodology for authoring of intelligent tutoring systems using human computation. The methodology embeds authoring tasks in existing educational tasks to avoid the need for monetary authoring incentives. Because not all educational tasks are equally motivating, there is a tension between designing the human computation task to be optimally efficient in the short term and optimally motivating to foster participation in the long term. In order to enhance intrinsic motivation for participation, the methodology proposes designing the interaction to promote user autonomy, competence, and relatedness as defined by Self-Determination Theory. This design has implications for learning during authoring.

Keywords: authoring, intelligent tutoring system, human computation, motivation

1 Introduction

It is commonly believed that it takes several hundred hours of authoring effort to create one hour of instruction for an intelligent tutoring system [3, 11]. What is less commonly considered is that those are "expert hours," namely the time spent by highly trained knowledge engineers, instructional designers, and subject matter experts. Typical authoring tools for ITS are intended to reduce this ratio. However these authoring tools do not address the shortage of experts needed to use the tools.

In our current work, we are trying a radically different approach to address this shortage of experts. We address expertise by letting novices do the authoring but then let other novices check the work to ensure quality. We address motivation by disguising the authoring task as another task that novices are already engaged in. We call this system BrainTrust.

The idea is that as students read online, they work with a virtual student on a variety of educational tasks related to the reading. These educational tasks are designed to both improve reading comprehension and contribute to the creation of an intelligent tutoring system based on the material read. After the human students read a passage, they work with the virtual student to summarize, generate concept maps, reflect on the reading, and predict what will happen next. The tasks and interaction are inspired by reciprocal teaching [30], a well known method of teaching reading comprehension strategies.

The virtual student's performance on these tasks is a mixture of previous student answers and answers dynamically generated using AI and natural language processing techniques. As the human teaches and corrects the virtual student, they in effect improve the answers from previous sessions and author a domain model for the underlying intelligent tutoring system. It should be pointed out that while this process is domain-independent, the domain model that results is specifically designed for a conversational, conceptual style of tutoring, described in detail below.

In developing BrainTrust, several interaction designs were created and evaluated. The early designs were rigidly aligned with intelligent tutoring system authoring tasks. Although early designs were efficient from an authoring standpoint, they were perceived as boring in our focus groups, leading to concerns about the motivation of students to participate. After iterating through many storyboards, we adopted the principles of Self-Determination Theory [14] in order to enhance the intrinsic motivation of users. Designing for intrinsic motivation increases the amount of time users spend in non-authoring activities, which is at odds with the goal of efficient authoring. However, we argue that the same design choices have positive implications for the user's reading comprehension and learning.

2 Background & Motivation

2.1 Tutoring by Humans and Computers

It is well established that human tutoring is a highly effective form of instruction that yields better outcomes that typical classroom instruction. An early meta-analysis of tutoring studies found that even novice human tutors enhanced learning with a medium effect size ($d = .4$) compared to classroom and comparable control conditions, an improvement of approximately half a letter grade [9], and an early study of expert tutors reported a very large effect size ($d = 2$) for mathematics skill training, an improvement of approximately two letter grades [5]. Early studies like these were influential in driving the emerging field of intelligent tutoring systems [36], or ITS.

During the last 30 years, researchers have made some important progress in developing ITSs that have the potential to seriously increase learning gains at deeper levels of comprehension and mastery [18]. The ITSs implement systematic strategies for promoting learning, such as error identification and correction, building on prerequisites, frontier learning (expanding on what the learner already knows), building on the zone of proximal development, student modeling (inferring what the student knows and having that guide tutoring), modeling-scaffolding-fading, and building coherent explanations [37, 38]. The defining characteristic of an ITS is that it tracks knowledge and adaptively responds to the learner [43], using computational modeling techniques like production rules, graphical models, and vector spaces. Recent meta-analyses have found that ITS learning gains are indistinguishable from human tutor controls [21, 40], suggesting that ITS research has sufficiently matured to make it broadly applicable to K-16 education.

2.2 ITS Authoring with Natural Language

Many ITS have been developed for mathematically well-formed topics, including algebra, geometry, programming languages [33], and physics [41]. Unfortunately, developing mathematically oriented ITS is problematic in terms of development costs, which can be as high as 100 hours of development time for 1 hour of instruction, even with special authoring tools [3]; see [25] for a review of emerging authoring methods. However, a number of ITS have been built over the last decade that tackle knowledge domains with a natural language foundation as opposed to mathematics and subject matters that require precise analytical reasoning [23]. The learning gains on these natural language ITS are consistent with large effects found in ITS meta-analyses [40]. and the development cost for these natural language ITS tends to be very low, the lowest reported time being two hours of development time for one hour of instruction [17].

These conversational ITS based in natural language share two defining attributes (see [27, 29] for a review). First, they are based on naturalistic observations and computational modeling of human tutoring strategies embedded in tutorial dialogue. A common strategy is the so-called five-step dialogue frame:

1. Tutor asks a deep reasoning question,
2. Student gives an answer,
3. Tutor gives immediate feedback or pumps the student,
4. Tutor and student collaboratively elaborate an answer, and
5. Tutor assesses the student's understanding.

The five-step dialogue frame illustrates the other defining attribute of natural language ITS, which is their interactive and collaborative nature: the tutor and student are co-constructing an explanation together. According to theories of learning, the interactive and collaborative nature of tutoring is what makes it more effective than activities like individual problem solving [7].

From a computational perspective, the goal of a natural language ITS is to help the student construct an explanation to a given problem. A full explanation is has multiple points, which are commonly called *expectations* because they are the expected parts of the correct answer. A natural language ITS manages the tutoring session by keeping track of the expectations and directing the student's attention to expectations that have not been covered. The ITS directs the student's attention by asking questions, ranging from relatively vague *pumps* like "What else can you say?" to *hints* like "What can you say about the force of gravity?" to very specific *prompts* like "The direction of gravity is?" Authoring a natural language ITS consists of constructing a paragraph length correct answer, generating questions for each expectation in the paragraph, and using text similarity measures like latent semantic analysis [20] to judge the difference between the student's answers and the expectations. The correct answer, expectations, questions, and vector space for latent semantic analysis are collectively referred to as the domain model of the ITS – the key components that must be authored every time a new topic must be covered. Even so, one of the reasons that natural language ITS are relatively easy to author is because the authoring is done in natural language.

2.3 Conversational ITS Authoring and Reading Comprehension

Recent attempts have been made to fully automate the authoring of natural language ITS using natural language processing technologies like semantic parsing, coreference resolution, automated inference, and ontology extraction [26, 28, 29]. The core aspects of automation were keyword identification, concept map generation, and question generation using manually generated summaries (equivalent to correct answers and expectations) as resources. After the keywords, concept maps, and questions were automatically generated, they were checked manually and corrected for errors. The BrainTrust approach extends this work by using human computation to correct errors.

BrainTrust maps authoring to human computation tasks using the key insight that keyword identification, summarization, concept map extraction, and question generation are not just authoring tasks but also reading comprehension strategies. Several meta-analyses have concluded that strategies like these should be taught explicitly to maximize reading comprehension, particularly for low-achieving students who lack the knowledge and skill to effectively comprehend reading at their grade level [15, 22]. A specific program of multiple-strategy instruction is *reciprocal teaching* [30]. In this program, as instructors read the text, they think aloud to model their comprehension process to the student including their reasoning for when to use each strategy. In a classic modeling-scaffolding-fading paradigm, the instructor and student take turns as the student gradually learns the strategies and practices them while the instructor provides feedback. More specifically, students read paragraph by paragraph and generate questions, summarize, clarify terms and concepts, and make predictions about what is coming up in the text. This practice becomes a dialogue as the instructor comments on and contributes to the student's questions, summaries, and other activities, or as other students make similar contributions in small group sessions.

2.4 Human Computation for Knowledge and Language

Recently a new subfield of computer science has emerged, known as *human computation*, that studies how to represent computationally difficult tasks so that humans will be motivated to work on them [1, 31, 35]. Human computation can be extremely powerful. In a recent example, a human computation game called Foldit was used to find the lowest energy form of a protein causing AIDS, a long-standing problem that had defied solution for nearly 15 years [10, 19]. Foldit makes use of humans' spatial reasoning abilities and motivates them to work by presenting the task as a game. However, this simple description belies the complexity involved in representing human computation tasks and executing them to produce a desired result. In essence, a human computation is a step in a larger algorithm that distributes tasks, checks their quality, and aggregates them into a solution. Much of the advantages and challenges of human computation stem from the issue that a human is "in the loop," because while humans are capable of solving complex and difficult problems, they are also autonomous beings with their own motivations and physical limits.

Several human computation games have been proposed to create knowledge representations and language data. FACTory is a human computation game designed to validate the truth of propositions in the Cyc Knowledge Base [12]. Users vote on the correctness of a proposition's natural language interpretation, for example, "Conjunctivitis is a symptom of earache," until enough users agree that FACTory stops asking for confirmation. Verbosity is a human computation game that presents itself as a two-player guessing game where each player has a secret word and a set of sentence template *cards* and chooses the card that will best allow the other player to guess the word, e.g. the word may be "cat" and the played card may be "tiger is a kind of ___" [1]. A related game, 1001 Paraphrases, uses a similar template providing strategy, except that its goal is to generate alternative phrasings of statements rather than facts [8]. Human computation systems like these often present previously proposed solutions to new users to improve upon, a process called *iterative improvement* in the human computation literature. Because even simple tasks, such as determining if an image includes the sky, can have non-agreeing "schools of thought" that systematically respond in opposing ways [39], it is preferable to use Bayesian models of agreement jointly to determine the ability of the user (and their trustworthiness as teachers) as well as the difficulty of the items they correct [32].

Currently underway is a human computation project called Duolingo, whose stated purpose is to help people learn a language while simultaneously translating the Web [2, 35]. Thus Duolingo appears to make use of two human computation motivators previously described [31]: altruism and implicit work. Altruism stems from helping others by translating the Web. Implicit work means the work achieved as a side-effect of the main task; in this instance the implicit work of learning the language is translation. In Duolingo, users translate sentences from a foreign language into their own language, with some computer support that provides dictionary translations of individual words. As users proceed, Duolingo increases the complexity of the task. A recent review of Duolingo praised its use of hints and feedback in guiding the translation process but also questioned the use of a translation-based approach to learning a language, an educational approach that fell out of fashion about 50 years ago [16].

3 Designing BrainTrust

3.1 Motivating Human Computation with Virtual Students

Designs for human computation include user motivation to participate, typically focusing on pay, enjoyment, altruism, reputation, and implicit work [31]. Psychological theories of motivation contrast intrinsic and extrinsic motivation, such that of the typical motivators used in human computation, only altruism (without recognition) and enjoyment would qualify as intrinsic motivators [14]. This is an important distinction because numerous experiments have found that introducing extrinsic motivators, like pay, can actually diminish intrinsic motivation for an activity [13]. Therefore, if the goal of BrainTrust is to increase learning and the desire for learning, it is important to design for intrinsic motivation rather than extrinsic motivation.

Self-Determination Theory identifies three factors influencing intrinsic motivation: competence, autonomy, and relatedness [34]. Competence is enhanced by maintaining optimal challenge so that participants achieve success and positive feedback. Autonomy interacts with competence, enhancing motivation when the participant feels in control. In contrast, when the participant feels controlled, pressured, or manipulated, autonomy and motivation decrease. Relatedness occurs when the participant is socially connected to others who pay attention to or even care about what the participant is doing. By supporting competence, autonomy, and relatedness, a human computation design should maximize intrinsic motivation.

The BrainTrust approach to maximizing intrinsic motivation is to present the human computation tasks through a virtual student, sometimes called a teachable agent [4], as shown in Figure 1. The virtual student's performance on these tasks is a mixture of previous student answers and answers dynamically generated using AI and natural language processing techniques. As the human teaches and corrects the virtual student, they in effect improve the answers from previous sessions and author a domain model for the underlying intelligent tutoring system. From the perspective of Self-Determination Theory, users may demonstrate competence if teaching the student presents an optimal level of challenge, experience autonomy if their interaction with the student is loosely directed, and feel relatedness because the student is presented as a animated conversational character. Although previous research has not directly assessed the effects of virtual students on intrinsic motivation, studies have shown that students spend more time with virtual students, attribute mental states to them, and are more likely to acknowledge their own errors [6].

3.2 Designing Motivating Interactions

We do not claim that simply adding a virtual student makes the design intrinsically motivating. If the reading comprehension tasks themselves do not reinforce competence, autonomy, and relatedness, then the design will fail in this regard.

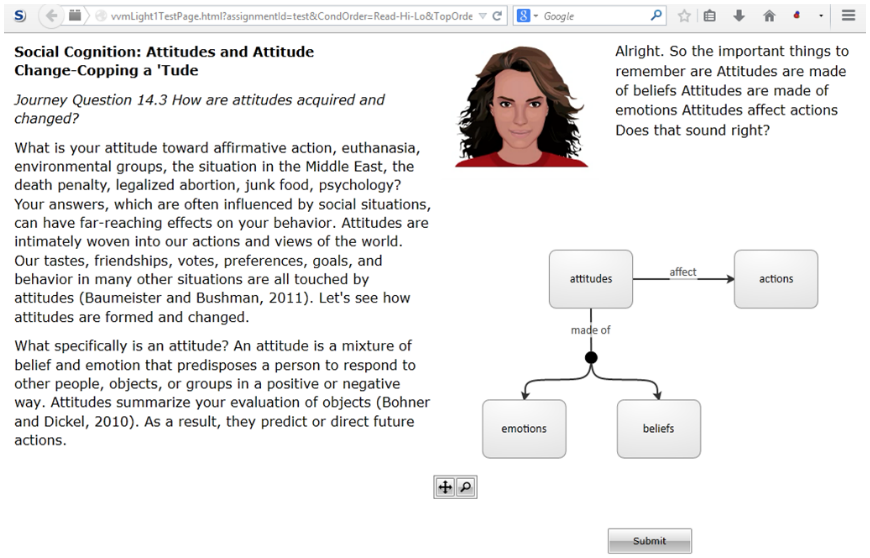

Fig. 1. BrainTrust during a concept mapping activity

And it is these reading comprehension activities that individually represent specific human computation tasks.

As we developed BrainTrust's human computation tasks, we iterated through six different interaction designs before settling on one that best supports intrinsic motivation. Because of space limitations, we will only describe the first and the last designs, as their differences best illustrate how competence, autonomy, and relatedness can be enhanced. The earliest design was rigidly aligned with intelligent tutoring system authoring tasks. The original storyboard proceeded as follows:

1. Virtual student reads the selected paragraph aloud.
2. Virtual student summarizes the material by selecting key sentences.
3. Human corrects the summary.
4. Virtual student generates questions and answers on important facts.
5. Human corrects or adds questions and answers.
6. Virtual student clarifies by identifying key concepts and linking them in a concept map.
7. Human corrects or adds key concepts and links.
8. Next paragraph is selected and process is repeated.

Although this earliest design was efficient from an ITS authoring standpoint, it was perceived as boring in our focus groups, leading to concerns about the intrinsic motivation of students to participate. Using Self-Determination Theory as a lens, we can identify several design weaknesses in this storyboard. First, the interaction is very mechanical, with the virtual student controlling as much of the interaction as possible, leading to low autonomy. For example, summaries involve sentence selection rather than free-response. Likewise questions are generated complete with answers, so again little opportunity for free-response. Second, user competence is diminished because the virtual student is essentially asking the user to do the same kinds of tasks repeatedly: the questions are just rephrased pieces of the summary, and the keywords/concept maps are just rephrased pieces of the questions. Finally, relatedness is reduced because the virtual student gives the user very little opportunity to inject their own ideas, and as a consequence creates fewer opportunities to learn from teaching the virtual student [7]. The first design, perhaps counterintuitively, has low intrinsic motivation precisely because it is closer in spirit to typical human computation tasks like template filling [1, 8, 12] but without providing a game-like metaphor to make the task more enjoyable.

The final design enhances intrinsic motivation without gamifying the task by rethinking autonomy, competence, and relatedness. To make the task more motivating and useful to the user, we accepted that some of the activities the user performs will have low utility to the end goal of creating an intelligent tutoring system; however those same activities will have high utility to the goals of enhancing intrinsic motivation and helping the user comprehend the text they are reading. The final design is inspired by the methodology of reciprocal teaching [30], which provides a natural interaction paradigm in which these reading comprehension activities can be learned and practiced. The final storyboard proceeds as follows:

1. Human reads the selected paragraph, and, if desired, activates the virtual student.
2. Virtual student voices the gist, or topic, of the paragraph.
3. Human corrects the gist as free-response.
4. Virtual student generates open ended, authentic questions.
5. Human provides their answers as free-response
6. Virtual student clarifies by identifying key concepts and linking them in a concept map.
7. Human corrects or adds key concepts and links.
8. Virtual student predicts the topic of the next paragraph.
9. Human corrects as free-response.
10. Next paragraph is selected and process is repeated.

In this design, autonomy is increased because all tasks are open-ended and are answered using free-response. Open-ended tasks like gists, authentic questions, and predictions allow for interpretations and personalized responses, and free-response options for these tasks further allow for autonomy. Competence is strengthened because the tasks are now decoupled from each other, and the tasks themselves are both more challenging because they require free-response as well as less evaluative because they allow for personalized responses. Relatedness improves through open-ended, authentic questions [24], such as “What are your beliefs about gun control?” which invite the user to contribute their own ideas and interpretations of the text. Both of these features make the tasks less about the *facts* of the text and more about the global meaning of the text – a key aspect of reading comprehension.

Clearly, making the tasks open-ended and free-response makes the corresponding answers more difficult to use for ITS authoring. The tasks of gist, authentic questions, and prediction, do not clearly correspond to ITS authoring tasks like summarization and question generation. Indeed, the core piece of ITS authoring is now largely encapsulated in the concept map. This is not a problem for ITS authoring as concept maps can be used to generate the questions, summaries, and other materials needed for a natural language ITS [26, 28, 29]. On the other hand, the shift from the earliest interaction design to the final design does illustrate the tension between making the human computation for authoring efficient in the short term and keeping the process viable in the long term by enhancing intrinsic motivation. Similar trade-offs occur when human computation is embedded in actual games with extraneous game play [42].

However, BrainTrust seems to differ from previous human computation systems in the sense that the tasks users engage in have three side effects: the tasks improve the users' understanding of the texts they wish to read, the process of correcting the virtual student improves reading comprehension skills, and the tasks create knowledge representations and content for an intelligent tutoring system. In other words, the tasks help students understand what they read, improve their reading skills for the future, and use their efforts to help other students in the same way.

4 Conclusion

This paper presented a methodology, called BrainTrust, for authoring of intelligent tutoring systems (ITS) using human computation. In BrainTrust, as users read online, they work with a virtual student on reading comprehension tasks that are aligned with authoring tasks in natural language ITS. This approach circumvents the shortage of experts who are typically needed to create ITS by leveraging novice users who are already engaged in reading a text.

Although our earliest design included superficially motivating components like a virtual student and tasks inspired by reciprocal teaching, it did not carefully address the intrinsic motivation of users. Using Self-Determination Theory, we presented an analysis of the earliest design and our final design in light of the core principles of autonomy, competence, and relatedness defined by that theory. To include these principles, the final design included more open-ended tasks with free-response options, many of which are not directly applicable to the task of authoring a natural language ITS. However, the open-ended tasks bring the final design closer to collaborative dialogue that various studies suggest is optimal for learning [7, 24, 30]. Thus designing BrainTrust for intrinsic motivation may also optimize for student learning, not only by increasing participation in authoring, but by making participation itself a beneficial learning experience.

Acknowledgements. This research was supported by the National Science Foundation (1235958 & 1352207), Institute of Education Sciences (R305C120001 &

R305A130030), Office of Naval Research (N00014-12-C-0643), and Army Research Laboratory (W911NF-12-2-0030). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the author and do not represent the views of these organizations.

References

1. von Ahn, L.: Human Computation. Doctoral thesis, Carnegie Mellon University (2005), uMI Order Number: AAI3205378
2. von Ahn, L.: Three human computation projects. In: Proceedings of the 42nd ACM technical symposium on Computer science education. pp. 691–692. SIGCSE '11, ACM, New York, NY (2011), http://doi.acm.org/10.1145/1953163.1953354
3. Aleven, V., Mclaren, B.M., Sewall, J., Koedinger, K.R.: A new paradigm for intelligent tutoring systems: Example-tracing tutors. International Journal of Artificial Intelligence in Education 19(2), 105–154 (2009)
4. Biswas, G., Schwartz, D., Leelawong, K., Vye, N.: Learning by teaching: A new agent paradigm for educational software. Applied Artificial Intelligence 19, 363–392 (March 2005)
5. Bloom, B.S.: The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher 13(6), 4–16 (Jun 1984)
6. Chase, C.C.: Motivating Persistence in the Face of Failure: The Impact of an Ego-protective Buffer on Learning Choices and Outcomes in a Computer-based Educational Game. Ph.D. thesis (2011), http://books.google.com/books?id=uHsdh6mKHuYC
7. Chi, M.T.H.: Active-constructive-interactive: A conceptual framework for differentiating learning activities. Topics in Cognitive Science 1(1), 73–105 (2009), http://dx.doi.org/10.1111/j.1756-8765.2008.01005.x
8. Chklovski, T.: Collecting paraphrase corpora from volunteer contributors. In: Proceedings of the 3rd international conference on Knowledge capture. pp. 115–120. K-CAP '05, ACM, New York, NY (2005), http://doi.acm.org/10.1145/1088622.1088644
9. Cohen, P.A., Kulik, J.A., Kulik, C.L.C.: Educational outcomes of tutoring: a meta analysis of findings. American Educational Research Journal 19, 237–248 (1982)
10. Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popovic, Z., players, F.: Predicting protein structures with a multiplayer online game. Nature 466(7307), 756–760 (2010), http://dx.doi.org/10.1038/nature09304
11. Corbett, A.T.: Cognitive tutor algebra I: Adaptive student modeling in widespread classroom use. In: Technology and Assessment: Thinking Ahead. Proceedings from a Workshop. pp. 50–62. National Academy Press, Washington, DC (November 2002)
12. Cycorp: FACTory. http://game.cyc.com/ (2005), http://game.cyc.com/, accessed: 7/23/12
13. Deci, E.L., Koestner, R., Ryan, R.M.: A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin 125(6), 627–668 (1999)
14. Deci, E.L., Ryan, R.M.: Handbook of Self-Determination Research. University of Rochester Press (2004), http://books.google.com/books?id=DcAe2b7L-RgC

15. Edmonds, M.S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K.K., Schnakenberg, J.W.: A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. Review of Educational Research 79(1), 262–300 (2009), `http://rer.sagepub.com/content/79/1/262.abstract`
16. Garcia, I.: Learning a language for free while translating the web. does duolingo work? International Journal of English Linguistics 3(1), 19–25 (2013)
17. Graesser, A., Rus, V., D'Mello, S., Jackson, G.T.: Learning through natural language dialogue that adapts to the cognitive and affective states of the learner. In: Robinson, D.H., Schraw, G. (eds.) Recent Innovations in Educational Technology That Facilitate Student Learning, pp. 95–125. Current perspectives on cognition, learning, and instruction, Information Age Publishing (2008), `http://books.google.com/books?id=66sJmAycE3cC`
18. Graesser, A.C., Conley, M.W., Olney, A.: Intelligent tutoring systems. In: Harris, K.R., Graham, S., Urdan, T., Bus, A.G., Major, S., Swanson, H.L. (eds.) APA educational psychology handbook, Vol 3: Application to teaching and learning, pp. 451–473. American Psychological Association, Washington, DC, US (2011)
19. Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popovi, Z., Jaskolski, M., Baker, D.: Crystal structure of a monomeric retroviral protease solved by protein folding game players. Nat Struct Mol Biol 18(10), 1175–1177 (Oct 2011), `http://dx.doi.org/10.1038/nsmb.2119`
20. Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W. (eds.): Handbook of Latent Semantic Analysis. Lawrence Erlbaum, Mahwah, New Jersey (2007)
21. Ma, W., Adesope, O.O., Nesbit, J.C., Liu, Q.: Intelligent tutoring systems and learning outcomes: A meta-analysis. Journal of Educational Psychology 106(4), 901–918 (2014)
22. National Institute of Child Health and Human Development: Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. NIH Publication No. 00-4769, U.S. Government Printing Office, Washington, DC (2000)
23. Nye, B.D., Graesser, A.C., Hu, X.: Autotutor and family: A review of 17 years of natural language tutoring. International Journal of Artificial Intelligence in Education 24(4), 427–469 (2014), `http://dx.doi.org/10.1007/s40593-014-0029-5`
24. Nystrand, M. (ed.): Opening Dialogue: Understanding the Dynamics of Language and Learning in the English Classroom. Language and Literacy Series, Teachers College Press, New York (1997)
25. Olney, A.M., Brawner, K., Pavlik, P., Keodinger, K.: Emerging trends in automated authoring. In: Sottilare, R., Graesser, A., Hu, X., Brawner, K. (eds.) Design Recommendations for Intelligent Tutoring Systems, Adaptive Tutoring, vol. 3. U.S. Army Research Laboratory, Orlando, FL (2015)
26. Olney, A.M., Cade, W., Williams, C.: Generating concept map exercises from textbooks. In: Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications. pp. 111–119. Association for Computational Linguistics, Portland, Oregon (June 2011), `http://www.aclweb.org/anthology/W11-1414`
27. Olney, A.M., Graesser, A.C., Person, N.K.: Tutorial dialog in natural language. In: Nkambou, R., Bourdeau, J., Mizoguchi, R. (eds.) Advances in Intelligent Tutoring Systems, Studies in Computational Intelligence, vol. 308, pp. 181–206. Springer-Verlag, Berlin (2010)

28. Olney, A.M., Graesser, A.C., Person, N.K.: Question generation from concept maps. Dialogue and Discourse 3(2), 75–99 (2012)
29. Olney, A.M., Person, N.K., Graesser, A.C.: Guru: Designing a conversational expert intelligent tutoring system. In: McCarthy, P., Boonthum-Denecke, C., Lamkin, T. (eds.) Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches, pp. 156–171. IGI Global, Hershey, PA (2012)
30. Palincsar, A.S., Brown, A.L.: Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction 1, 117–175 (1984)
31. Quinn, A.J., Bederson, B.B.: Human computation: a survey and taxonomy of a growing field. In: Proceedings of the 2011 annual conference on Human factors in computing systems. pp. 1403–1412. CHI '11, ACM, New York, NY (2011), `http://doi.acm.org/10.1145/1978942.1979148`
32. Raykar, V.C., Yu, S., Zhao, L.H., Valadez, G.H., Florin, C., Bogoni, L., Moy, L.: Learning from crowds. The Journal of Machine Learning Research 99, 1297–1322 (2010)
33. Ritter, S., Anderson, J.R., Koedinger, K.R., Corbett, A.: Cognitive tutor: Applied research in mathematics education. Psychonomic bulletin & review 14(2), 249–255 (2007)
34. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 55(1), 68–78 (2000)
35. Savage, N.: Gaining wisdom from crowds. Commun. ACM 55(3), 13–15 (Mar 2012), `http://doi.acm.org/10.1145/2093548.2093553`
36. Sleeman, D., Brown, J.S. (eds.): Intelligent tutoring systems. Academic Press, New York (1982)
37. Sottilare, R., Graesser, A., Hu, X., Goldberg, B. (eds.): Design Recommendations for Intelligent Tutoring Systems: Instructional Management, Adaptive Tutoring, vol. 2. U.S. Army Research Laboratory, Orlando, FL, isbn 978-0-9893923-3-4 edn. (2014), available at: https://gifttutoring.org/documents/
38. Sottilare, R., Graesser, A., Hu, X., Holden, H. (eds.): Design Recommendations for Intelligent Tutoring Systems, Adaptive Tutoring, vol. 1. U.S. Army Research Laboratory, Orlando, FL, isbn 978-0-9893923-0-3 edn. (2013), available at: https://gifttutoring.org/documents/42
39. Tian, Y., Zhu, J.: Learning from crowds in the presence of schools of thought. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 226–234. KDD '12, ACM, New York, NY, USA (2012), `http://doi.acm.org/10.1145/2339530.2339571`
40. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist 46(4), 197–221 (2011)
41. VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., Rose, C.: When are tutorial dialogues more effective than reading? Cognitive Science 31, 3–62 (2007)
42. Vannella, D., Jurgens, D., Scarfini, D., Toscani, D., Navigli, R.: Validating and extending semantic knowledge bases using video games with a purpose. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1294–1304. Association for Computational Linguistics (2014), `http://aclweb.org/anthology/P14-1122`
43. Woolf, B.P.: Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning. Morgan Kaufmann Publishers/Elsevier (2008), `http://books.google.com/books?id=MnrUj3J\_VuEC`