AIED 2009: 14th International Conference on Artificial Intelligence in Education

Workshops Proceedings

Editors and Co-Chairs:

Scotty D. Craig
University of Memphis, USA

Darina Dicheva
Winston-Salem State University, USA

July 6-7th, 2009
Brighton, UK

Preface

The supplementary proceedings of the workshops held in conjunction with AIED 2009, the fourteen International Conference on Artificial Intelligence in Education, July 6-7, 2009, Brighton, UK, are organized as a set of volumes - a separate one for each workshop.

The set contains the proceedings of the following workshops:

- **Volume 1: The 2nd Workshop on Question Generation**
  Co-chairs: Vasile Rus & James Lester. *University of Memphis, USA & North Carolina State University, USA*.
  http://www.questiongeneration.org/AIED2009/
- **Volume 2: SWEL'09: Ontologies and Social Semantic Web for Intelligent Educational Systems**
  Co-chairs: Niels Pinkwart, Darina Dicheva & Riichiro Mizoguchi. *Clausthal University of Technology, Germany; Winston-Salem State University, USA & University of Osaka, Japan*.
  http://compsci.wssu.edu/iis/swel/SWEL09/index.html
- **Volume 3: Intelligent Educational Games**
  Co-chairs: H. Chad Lane, Amy Ogan & Valerie Shute. *University of Southern California, USA; Carnegie Mellon University, USA & Florida State University, USA*.
  http://projects.ict.usc.edu/aied09-edgames/
- **Volume 4: Scalability Issues in AIED**
  Co-chairs: Lewis Johnson & Kurt VanLehn. *Alelo, Inc., USA & Arizona State University, USA*.
  http://alelo.com/aied2009/workshop.html
- **Volume 5: Closing the Affective Loop in Intelligent Learning Environments**
  Co-chairs: Cristina Conati & Antonija Mitrovic. *University of British Columbia, Canada & University of Canterbury, New Zealand*.
  http://aspire.cosc.canterbury.ac.nz/AffectLoop.html
- **Volume 6: Second Workshop on Culturally-Aware Tutoring Systems (CATS2009): Socio-Cultural Issues in Artificial Intelligence in Education**
  Co-chairs: Emmanuel G. Blanchard, H. Chad Lane & Danièle Allard. *McGill University, Canada; University of Southern California, USA & Dalhousie University, Canada*.
  http://www.iro.umontreal.ca/~blanchae/CATS2009/

- **Volume 7: Enabling Creative Learning Design: How HCI, User Modelling and Human Factors Help**
  Co-chairs: George Magoulas, Diana Laurillard, Kyparisia Papanikolaou & Maria Grigoriadou. *Birkbeck College, University of London, UK; Institute of Education, UK; School of Pedagogical and Technological Education, Athens, Greece & University of Athens, Greece.*
  https://sites.google.com/a/lkl.ac.uk/learning-design-workshop/Home

- **Volume 8: Towards User Modeling and Adaptive Systems for All (TUMAS-A 2009): Modeling and Evaluation of Accessible Intelligent Learning Systems**
  Co-chairs: Jesus G. Boticario, Olga C. Santos and Jorge Couchet, Ramon Fabregat, Silvia Baldiris & German Moreno. *Spanish National University for Distance Education, Spain & Universitat de Girona, Spain*.
  https://adenu.ia.uned.es/web/es/projects/tumas-a/2009

- **Volume 9: Intelligent Support for Exploratory Environments (ISEE'09)**
  Co-chairs: Manolis Mavrikis, Sergio Gutierrez-Santos & Paul Mulholland. *London Knowledge Lab, Institute of Education/Birkbeck College, University of London, UK & Knowledge Media Institute and Centre for Research in Computing, The Open University, UK.*
  http://link.lkl.ac.uk/isee-aied09

- **Volume 10: Natural Language Processing in Support of Learning: Metrics, Feedback and Connectivity**
  Co-chairs: Philippe Dessus, Stefan Trausan-Matu, Peter van Rosmalen & Fridolin Wild. *Grenoble University, France; Politehnica University of Bucharest; Open University of the Netherlands & Vienna University of Economics and Business Administration, Austria.*
  http://webu2.upmf-grenoble.fr/sciedu/nlpsl/

While the main conference program presents an overview of the latest mature work in the field, the AIED2009 workshops are designed to provide an opportunity for in-depth discussion of current and emerging topics of interest to the AIED community. The workshops are intended to provide an informal interactive setting for participants to address current technical and research issues related to the area of Artificial Intelligence in Education and to present, discuss, and explore their new ideas and work in progress.

All workshop papers have been reviewed by committees of leading international researchers. We would like to thank each of the workshop organizers, including the program committees and additional reviewers for their efforts in the preparation and organization of the workshops.

July, 2009
Scotty D. Craig and Darina Dicheva

AIED 2009 Workshops Proceedings
Volume 3

Intelligent Educational Games

Workshop Co-Chairs:

H. Chad Lane
Institute for Creative Technologies,
University of Southern California, USA

Amy Ogan
Human Computer Interaction Institute,
Carnegie Mellon University, USA

Valerie Shute
Educational Psychology and Learning Systems,
Florida State University, USA

http://projects.ict.usc.edu/aied09-edgames/

Preface

There is certainly no shortage of voices promoting the promise of games for learning. And these voices are being heard: from the early days of The Oregon Trail to modern, massively multiplayer 3D immersive games like World of Warcraft, games have caught the attention of educators, researchers, parents, commercial companies, governments, and more. Educational games now exist for almost every K-12 domain and emanate from academic research labs, commercial companies, or non-profit foundations. So what does the AIED community bring to the educational games table? Should AIED researchers think about using games as learning environments?

The call for papers for the Intelligent Educational Games workshop at the 14th International Conference on Artificial Intelligence in Education (AIED 2009) asked authors to consider games from the perspectives of AI and learning (e.g., *motivation, assessment, guidance, affect,* and *classroom use*). We believe that these are the main areas in which AIED research is already making progress and is poised to have an even more profound impact. For example, given the repeated demonstration of the value of intelligent guidance in computer-based learning environments, there is good reason to believe these techniques will also benefit learners in game-based environments. Similarly, AIED researchers focus heavily on motivation, affect, and assessment – these issues converge in interesting and beautiful ways in educational games research.

We are happy to say our authors did not disappoint. In these proceedings, you will find papers that address key issues such as applying educational game techniques in intelligent tutoring systems (McNamara, Jackson, & Graesser), teaching ill-defined domains including ethics (Hodhod, Kudenko, & Cairns; McKenzie & McCalla) and argumentation (Hastings et al.), investigating motivational effects in games (Ogan et al.), teaching K-12 domains including microbiology (Rowe et al.) and reading (Jackson, Boonthum, & McNamara), using advanced AI techniques in a narrative-based system for classroom use (McAlinden et al.), teaching by demonstration (Pareto), and in scenario customization (Niehaus & Riedl), and finally in exploring how games fit into broader issues related to assessment (Shute et al.). In addition to these, six other short papers focus on key issues that delve into other challenging research areas like language learning, multi-player and networked games, and intelligent guidance. In sum, we are proud of this excellent collection of papers and hope that it inspires AIED researchers to continue contributing to educational games research.

The meeting was held at the Thistle Hotel in Brighton, England on Tuesday, July 7, 2009. We received 17 submissions and accepted 11 as full papers and 6 as short papers. The reviews were extraordinary and for that we have our prestigious program committee to thank. One of those committee members, Cristina Conati, ran a similar workshop at AIED 2005, and we thank her and her co-chair, Sowmya Ramachandran, for helping establish a firm place for educational game research in the AIED community. We were also very lucky to have the wonderful support of Scotty Craig and Darina Dicheva, the workshop co-chairs for AIED 2009.

July, 2009
H. Chad Lane, Amy Ogan, & Val Shute

Program Committee

Co-Chair: H. Chad Lane, *University of Southern California, USA* (lane@ict.usc.edu)
Co-Chair: Amy Ogan, *Carnegie Mellon University, USA* (aeo@andrew.cmu.edu)
Co-Chair: Valerie Shute, *Florida State University, USA* (vshute@fsu.edu)

Vincent Aleven, *Carnegie Mellon University*
Jim Belanich, *U.S. Army Research Institute*
Cristina Conati, *University of British Columbia*
Paula Durlach, *U.S. Army Research Institute*
Reva Freedman, *Northern Illinois University*
James Gee, *Arizona State University*
Peter Hastings, *DePaul University*
Randy Hill, *University of Southern California*
Bob Hausmann, *University of Pittsburgh*
Lewis Johnson, *Alelo, Inc.*
James Lester, *University of North Carolina*
Brian Nelson, *Arizona State University*
Mark Riedl, *Georgia Institute of Technology*
Jody Underwood, *Pragmatic Solutions, Inc.*
Mike van Lent, *Soar Technologies*
Diego Zapata, *Educational Testing Services*

Table of Contents

Full Papers

Short Papers

Intelligent Tutoring and Games (ITaG)

Danielle S. McNAMARA[a1], G. Tanner JACKSON[a], and Arthur GRAESSER
[a]*University of Memphis, Psychology/Institute for Intelligent Systems*

Abstract. Intelligent Tutoring Systems (ITSs) have been producing consistent learning gains for decades. One problem consistently faced by ITS researchers is the gap between liking and learning. Students often desire to learn in a passive, effortless manner, while research has shown that the most effective learning requires interaction and active processing. ITSs effectively produce learning gains, but students often dislike interacting with the system (and without return users, establishing a long-term learning goal is near impossible). A potential solution to this problem lies in games. ITS researchers have begun to incorporate game-based elements within learning systems. This paper aims to describe some of those elements, categorize them within functional groups, and provide insight into how elements within each category may affect motivation.

Keywords. Intelligent Tutoring Systems, Games, Serious Games

Introduction

Game research is booming. Indeed, it is likely one of the fastest growing areas of research in the 21st century. There are numerous goals of game research. One area of research has been devoted to examining whether, how, and what students learn from entertainment games (e.g., SimCity, World of Warcraft). A second area of research revolves around the development of serious games and evaluating the degree to which these games engage learners, as well as the effectiveness of those games. A third area of research focuses on identifying the features of games that best promote learning in the context of serious games.

Here, we discuss an additional area of research that has recently garnered attention in mainstream research, namely, incorporating game-based principles into successful ITSs. This area of research should form what we think will be an optimal fusion of principles of learning and motivation to create games that are enhanced with learning principles and ITSs that are enhanced with game principles. We call this *ITaG (Intelligent Tutoring and Games)*. The ultimate goal is to better understand the features of ITSs and games that can be most effectively merged within a learning environment to provide optimal levels of learning, motivation, and engagement.

Intelligent Tutoring Systems (ITSs) afford adaptive, one-on-one tutoring to virtually all students in a classroom. ITSs have demonstrated remarkable success in helping students to learn challenging content and strategies [1]. While ITSs have shown considerable success, a problem arises when the student is required to use the systems over long periods of time. The development of deep knowledge requires time and practice: the mastery of content and learning strategies that will generalize to multiple

[1] Corresponding Author.

contexts and tasks does not happen in hours. However, over time, ITSs can become boring and tedious to students. The novelty of an ITS and the interactive components are quite engaging when they are used for short periods of time (e.g., hours), but can become monotonous and even annoying when a student is required to interact with an ITS for weeks or months. Hence, game-based principles offer a means of enhancing ITSs such that they can be more appealing to students and as a consequence, students will seek to engage with the tutoring systems more seriously and more often. We hypothesize that ITSs can be rendered more engaging to the student, and thus more effective in promoting learning, by incorporating motivational components [2,3]. Thus, our ultimate goal is to examine the benefits of incorporating game-based components within established tutoring systems to improve motivational aspects.

1. Motivation

We use the term motivation in a broad sense that incorporates emotions and social parameters in addition to traditional attributes of motivation. There are several constructs that are related to and intertwined with motivation. These include *self-regulation, self-efficacy, interest,* and *engagement*.

Self-regulation refers to monitoring the success and the weaknesses of the learning process by the learner [4,5]. Theories of self-regulation assume that good students regulate key cognitive, metacognitive, motivational, social, and affective processes. Self-regulation involves actively seeking a coherent understanding of a topic by using strategies and goals, regulating and monitoring cognitive processes, behavior, and motivation, and then modifying behavior to achieve learning goals [4,6].

Self-efficacy regards the learner's sense of success and achievement [7,8]. Lepper and colleagues [9,10] argue that tutoring systems should attempt to build self efficacy and empowerment, particularly when student is struggling or is unsuccessful in the learning task.

Interest refers to the degree to which underlying needs or desires of learners are energized [11]. *Individual interest* is characterized by a desire to develop competence or a personal investment in a particular topic or domain. *Situational interest* is transitory, pertaining to the specific events and contexts [11,12,13]. For the latter, while the nature of the content may play an important role, how the content is framed is more important in driving interest. For example, framing the content within a game can be used as a spring board to capture the interest of the student.

Engagement is another important construct related to both motivation and learning. Assumedly, when students are not engaged, they are bored or inattentive, neither being conducive to learning. For example, Craig et al. (2004)[14] found that increased levels of boredom while students learned about computer literacy topics with an ITS negatively correlated with learning, whereas *flow* (i.e., high engagement; Csikszentmihaly, 1990)[15] was positively correlated with learning. Bored learners are also more likely to strategically bypass the system [16]. Furthermore, boredom may trigger a vicious cycle that prevents students from actively reengaging in constructive learning processes [17].

2. Game-based Features

Serious games are educational games designed with the goal of helping students learn about subject matter content, problem solving strategies, and cognitive or social skills. Instead of learning about history by conventional means, the learner plays a game that successfully integrates the game with instructional curriculum. All games have rules, actions of the player, uncertainty, and feedback on outcomes. Many games have points, rewards, competition, levels, and privileges that are linked to success. And, many games are set in the context of narrative or fantasy. Further, some claim that enjoyment forms the core of the entertainment process, including the experience of games [18].

Aside from the essence of games, there is the question of what renders games successful. Indeed, many researchers have focused on important questions about serious games regarding how the features of games are systematically aligned with pedagogy [19,20,21,22,23]. By contrast, our goal is to identify broad categories of game features that can potentially be used to guide modifications of ITS technologies. Based on the literature, we propose a framework that classifies game-based features into five broad categories, including feedback, incentives, task difficulty, control, and environment. In Table 1, we present example game-based features for each of the categories and the motivational constructs that can be expected to be affected by adding such a feature to an ITS.

2.1. Feedback

Feedback is a feature common and important to both ITS and game technologies. Providing students with accurate, intelligent, and motivating feedback is a critical aspect of learning environments and is a vital component to the learning process [24,25,26]. The literature suggests several aspects of feedback that may influence the learning process, including: timing, content, control, and delivery-method of the feedback. Feedback on performance can be conveyed in various forms, including immediate corrections, delayed corrections, explanations, level of mastery achieved on specific content, paths or actions taken by the system, cumulative points, levels, and skillometers. These forms of feedback can be broadly categorized as information-based, consequence-based, and point-based feedback. Information-based feedback takes the form of providing specific feedback on the accuracy and quality of answers and actions within the ITS or game. Consequence-based feedback occurs when the system reacts to the user's responses or actions by changing the system path. For example, when a student answers incorrectly in an ITS, the student may be redirected into remediation. When a player is killed in a game, the game ends. This is the ultimate form of feedback.

Both information-based and consequence-based feedback are common to most ITSs. By contrast, point-based feedback is a feature more characteristic of games and is less common in ITSs. Point-based feedback can be conveyed in the form of cumulative points, progress bars, and levels. All of these are based directly or indirectly on the accuracy or the quality of responses and actions by the user but they differ in their specificity and detail. A progress bar is a means of conveying performance schematically (e.g., shaded progress bar, *skillometers*), without the specificity of points. Likewise, the use of levels is less specific than points and conveys feedback on performance across a longer time-span.

We expect point-based feedback features to facilitate self-regulation and self-efficacy in computer-based learning environments [24,27,28,29]. Points provide direct feedback to the user on performance, and thus afford regulating and monitoring performance more accurately. If the user/player is successful, such feedback is likely to enhance self-efficacy. Furthermore, feedback can be exploited as a competition. Many games display feedback in the context of others' performance. The player may be told the number of points earned relative to the highest scorer. Or, the player may be compared to other players within a certain area of region, or age range. Many players are highly motivated by competition, and are further driven by having a goal to beat others. Thus, when conveyed as a competition, we can expect feedback to affect many if not all aspects of motivation.

Table 1. Categories of game-based features, their function, and the motivational construct expected to be most influenced by adding the feature to an ITS.

Category	Enhancement Features	Function	Motivational Construct
Feedback	Verbal information, consequences, points, progress bar, skillometer, levels	Information regarding the accuracy or quality of responses is provided to the student	self-regulation, self-efficacy
Feedback	Competition	Information is provided on performance relative to others	self-regulation, self-efficacy, interest, engagement
Incentives	Points, levels, skill bar	Student acquires points or advances in levels by completing tasks successfully	self-regulation, self-efficacy, engagement
Incentives	Mini-games, exchange or modify avatar or environment	Student provided with motivational hooks (e.g., play game, change environment)	self-efficacy, interest, engagement
Task Difficulty	Tasks or materials vary in difficulty, task requirements gradually increase	Task or material is appropriately challenging and scaffolded according to ZPD	self-efficacy, engagement
Task Difficulty	Tasks or materials vary in difficulty	Backsliding: Student is given easier task after failure	self-efficacy
Task Difficulty	Tasks or materials vary in difficulty; feedback varies by performance history	Empowerment: Task is conveyed as difficult, but is below ZPD	self-efficacy
Control	Choosing rewards: mini-game, character, color	Student controls aspects of environment	self-regulation, self-efficacy, interest, engagement
Control	Levels, points, tasks, materials (e.g., texts), choose rewards, change environment	Student sets goals or subgoals to complete	self-regulation, self-efficacy, interest, engagement
Environment	Game-like environment, changeable colors, icons, aesthetically pleasing	ITS is set in a (more) appealing environment	interest, engagement
Environment	Animated agents or avatars	Animated agents improved/incorporated	interest, engagement
Environment	Multi-media: mini-games, graphics, video, simulations,	Simulations and other multimedia improved/incorporated	interest, engagement
Environment	Narrative, immersive environment, fantasy	Situated within a narrative or immersed within an environment (real world or fantasy)	interest, engagement

2.2. Incentives

Incentives are designed to provide motivational hooks that maintain interest and help to prolong student engagement [2,3]. Generally, incentives are contingent on some aspect of performance. In the context of entertainment games, the player may earn points, and those points can be traded for such things as powers, tools, skills, or weapons. Adapting this to the context of an ITS, the student may earn points, and those points can be similarly traded for a variety of objects, choices, or actions. For example, the student may be given options such as changing the avatar, changing color schemas, adding features to an avatar, or playing a mini-game.

Within the context of an ITS, mini-games may provide an optimal form of providing incentives to learners. Mini-games are short 5-15 minute dynamic games. They may be purely for entertainment (e.g., off-road racing or tower defense) or they may present educational content (a mini serious game or puzzle game). The latter option, still provides the student with an incentive to earn points within the ITS, but it also provides a greater variety of learning contexts.

Both the points and the option to trade the points for a utility can be considered as providing separate incentives to the player; however, they would potentially be expected to influence different aspects of motivation. We expect that providing incentives in the form of options that can be chosen by the student will affect self-efficacy, interest, and engagement. We expect self-efficacy to be affected because the student has been afforded the power to make choices and to do something different on the basis of performance. Thus, there is external support for performance. We expect interest in the material or activity to be increased when students engage in educational activities that are more fun, such as the mini-games. By contrast, we would not necessarily expect a student's interest in the domain content to be affected by actions involving modifications of the environment (e.g., changing the avatar, changing the color of the background, etc.). However we might expect that engagement will be influenced through environmental changes because the student has an external, extrinsic reward for performing well, and also because we expect activities such as mini-games to be inherently engaging.

2.3. Task difficulty

Varying the difficulty of the task, and matching the difficulty of the task to the learner is a characteristic of both ITSs and games. Nonetheless, we include this category in the list because the specific manipulations of task difficulty that we discuss have strong associations with games. Indeed, Malone and Lepper (1987)[30] identified *challenge* as a critical feature of successful video games. Many argue that successful games have *optimal levels of challenge*. Game developers have argued that a good game is at the *zone of proximal development* (ZPD) or at the brink of zones of ability, cognition, and emotion [22,30]. A moderately challenging game sustains engagement by providing accomplishment while maintaining effort. Success in the game environment breeds self-efficacy, which is highly correlated with interest in games as well as tutorial environments [31].

We include three subcategories of task difficulty. We view these modifications of task difficulty as means of allowing a system to engage in the ZPD expansion cycle. The first step in this cycle is to use the student's zone of proximal development (ZPD) to help select tasks or material that provides an achievable amount of challenge for the student. The process of overcoming this challenge is expected to provide a boost to self-efficacy, extend the application of student knowledge, and prevent boredom from a task that is too easy. Thus, we expect that maintaining the student at the ZPD should affect both self-efficacy and engagement.

The second step occurs if a student is unable to overcome this new challenge. If the task or material proves to be too difficult for the student, then backsliding can be implemented. Here, the system would provide increased scaffolding or feedback and the requirements of the task would be lowered to an achievable level. Reducing the amount of failure by adapting the task difficulty is expected to help the student to maintain a sufficient level of self-efficacy.

The third step may be implemented shortly after backsliding and adapting to the student's ability level. In this step, the system would attempt to empower the student through feedback strategies and thereby restore the student's sense of self-efficacy. Once a student has regained a stable performance level, this cycle would begin again as the system attempts to expand the student's ZPD.

2.4. Control

The amount of control that a student has in a learning environment such as an ITS is an important issue. For example, an internal locus of control has been found to help students progress through tasks more quickly and accurately [32]. However, while many learners prefer to have some degree of control over their learning environment, less skilled or low ability learners are often unable to make choices that optimize learning [33]. They often lack the self-regulatory skills necessary to make choices regarding learning tasks, and thus need to be provided with sufficient guidance within the learning system. However, game-based features may offer a means of providing a sense of control by the learner, without allowing them to make choices on what and how they learn.

Game designers commonly include multiple features that encourage user personalization and control. Allowing users to control certain aspects of the learning environment provides opportunities for students to become invested in that environment and to identify themselves with some aspect within it. Some research has shown that this investment of "self" may carry over into other effects of the environment (i.e., learning and retention) [34].

In Table 1, we convey two types of control (there could be others of course). The first is affording control over aspects of the environment. For example, the student may be allowed to change the color schemes, the background, or the avatar, or choose a task, such as a mini-game. Such choices may be open to the student at any time. However, we expect that the time to make choices should be controlled because students or players can become focused on changing trivial aspects of the environment (e.g., such as the avatars' clothes).

The second source of control included in Table 1 concerns the student's setting of personal goals or subgoals. For example, if points or levels are included in the system, the student may set the goal of obtaining a certain number of points or reaching the top level in the system. Likewise, the student may have the goal of being able to play a mini-game and perform at a certain level in the game, or they may want to change features of the avatar or other aspects of the environment. Expecting students to naturally set explicit learning goals is somewhat unrealistic (e.g., the goal to learn as much as possible about topic X). However, it seems much more reasonable to expect that students, regardless of ability level, naturally set and achieve explicit, tangible goals, such as obtaining a certain number of points or reaching a certain level. By imposing this game-based feature, the student's overt goal is to achieve points, but the student is induced to learn the target material or skills in order to reach those goals. This is important because, given a goal or subgoal, the student is more likely to monitor performance increasing self-regulation, and have a greater sense of self-efficacy when the choices are earned. Interest is also likely to increase because of the investment in the learning environment.

2.5. Environment

The game environment is what is most apparent to both the players and researchers. Many popular entertainment games are set in colorful, semi-realistic settings. Many games include avatars as well as simulation. Many games are embedded in a story narrative with characters, settings, conflict, competition, action episodes, and outcomes.

We have listed four aspects of the learning environment that are characteristic of games. The first regards ***aesthetics***. A system may be rendered more aesthetically pleasing by improving or varying the colors, using icons rather than words to convey information, increasing graphical quality, or improving the background. These modifications could potentially affect interest and engagement, but probably to a lesser extent than the gameplay features and options. While the aesthetics of a game are most obvious, there is some doubt that they are the most crucial aspects for a learning environment.

The second aspect of the environment regards the use of ***avatars*** or pedagogical agents. Animated agents have become more commonly incorporated in learning technologies over the last decade [35,36,37,38,39]. They appear to captivate students' interest and may affect learning [38,39]. Avatars are useful for many reasons. For example, they can model appropriate learning strategies with high precision and control. Both single agents and ensembles of agents can be carefully choreographed to mimic virtually any activity or social situation: curiosity, inquiry learning, negotiation, interrogation, arguments, empathetic support, helping, and so on. Agents or avatars can interact with students and help them learn by providing hints, clues, feedback or instruction, by modelling good pedagogy, or by holding a conversation. Agents may take on different roles: mentors, tutors, peers, players in multiparty games, or avatars in the virtual worlds [35]. Indeed, several researchers have developed avatars' social skills so that they can convey emotions and attitudes, show sensitivity to the learners' motivational and emotional states, and understand when and how to interact in socially appropriate ways [40].

A third feature that is characteristic of games is the use of ***multimedia***. We have already discussed the potential use of mini-games, which can incorporate multimedia as well as many other features of games. Graphics, video, and simulations can also be used to portray events or concepts that are targeted in the ITS.

Finally, the use of ***narrative,*** often included in both immersive and epistemic environments, is a feature that is uniquely characteristic of games, and is not commonly used in ITSs. Narrative is comprehended quickly and remembered well compared with other genres [41,42]. Unlike text or film, narrative in games has a distinctive status because the story plans can be co-constructed between the player and game system (with or without other players) and because it's possible for a player to experience hundreds of game threads rather than a single episodic sequence [23,43]. Some games are designed to fully immerse the player into a fantasy world [30], similarly, but to a lesser extent, epistemic games require the user to play a specified role and engage in behaviors associated with an expert in a certain field (e.g., you are a doctor and have to examine a fake patient presenting with a specific set of symptoms). Uncertainty in a narrative can also build suspense, one of the prominent features that sustain one's attention [44].

While narratives combined with epistemic and immersive environments may be an important and effective aspect of games, they may also be the most difficult aspects to combine within an existing ITS. That is, ITSs tend to have an obvious learning structure, and thus imposing a narrative on top of an existing ITS structure may be artificial, if possible at all. However, if the learning environment is in the initial development stages, and the developer of an ITaG intends to incorporate both ITS and game principles, then the use of narrative, fantasy, episteme, and immersion are certainly features to consider incorporating.

3. Discussion

Serious games presumably should captivate the attention, motivation, and learning activities of students more so than traditional environments such as ITSs. However, we do not know which components of games are most critical and effective in order to capitalize on the seductive aspects of games for learning. While it is generally assumed that games are more engaging and potentially *could* lead to better or more sustained learning [23,43,45], there is little research comparing the effectiveness of gaming environments to more traditional ITS environments [21]. Further, there is a lack of research that examines the effectiveness of the two approaches to promoting learning. Learning environments that incorporate game-based features are more likely to grab the attention of students and to increase their motivation to participate. On the other side, learning principles, such as those incorporated in current ITSs will increase gains in terms of learning and skill acquisition. As such, ITaGs can be expected to maximize the benefits of both worlds.

Here, we have focused on delineating how features of games can be added to an existing ITS, and how those features can be expected to affect the motivation of the learner. We have discussed motivation in terms of four related constructs: self-regulation, self-efficacy, interest, and engagement. We did so in order to be more precise about what aspects of motivation are likely to be affected by particular game features.

One of our implicit assumptions is that the incorporation of game-based elements within an ITS is unlikely to *directly* influence learning. However, game-based features are directly tied to the motivation of the learner, and the drive to engage with the system. Hence, we are implying within this framework that researchers may not expect to see *direct* benefits on learning from ITaGs. Nonetheless, we do expect to observe *indirect* benefits. When more learners are willing to engage with the system, and for longer periods of time, then we can expect that learning will improve.

We expect that the list of features we have provided may not be exhaustive and we expect that research will provide additional evidence that will inform what is affected by game-based features and how. However, our hope is that the ITaG framework will inspire researchers to vary certain aspects of learning environments and examine which aspects of motivation are enhanced, as well as to examine the potential learning benefits evoked by ITaGs.

Acknowledgements

This research was supported in part by the Institute for Educational Sciences (IES R305G020018-02; R305G040046, R305A080589) and National Science Foundation (NSF REC0241144; IIS-0735682). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES or NSF.

References

[1] B.P. Woolf, *Building intelligent interactive tutors*, Burlington, MA: Morgan Kaufmann Publishers, 2009.

[2] A.C. Graesser, P. Chipman, F. Leeming, & S. Biedenbach, Deep learning and emotion in serious games. In U. Ritterfield, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects,* Mahwah, NJ: Routledge, Taylor and Francis. (2009) 81-100.

[3] R. Moreno, & R.E. Mayer, Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology* **97** (2005), 117–128.

[4] P.R. Pintrich, Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology,* **92** (2000) 544-555.

[5] P.H. Winne, & A.F. Hadwin, Studying as self-regulated learning. In D. J. Hacker & J. Dunlosky (Eds.), *Metacognition in educational theory and practice,* Mahwah, NJ: Erlbaum. (1998) 277–304.

[6] B.J. Zimmerman, & D.H. Schunk, Reflections on theories of self- regulated learning and academic achievement. In B.J. Zimmerman, & D.H. Schunk, (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives,*Mahwah, NJ: Erlbaum (2001) 289-307.

[7] A. Bandura, Self-efficacy: The foundation of agency. In W. Perig & A. Grob (Eds.), *Control of human behavior, mental processes, and consciousness: Essays in honor of the 60th birthday of August Flammer,* Mahwah, NJ: Erlbaum (2000) 17-33.

[8] F. Pajares, Self-efficacy beliefs in academic settings. *Review of Educational Research* **66** (1996) 543-578.

[9] M.R. Lepper, M. Drake, & T. O'Donnell-Johnson, Scaffolding techniques of expert human tutors. K. Hogan & M. Pressley (Eds.). *Scaffolding student learning: Instructional approaches and issues. Advances in learning and teaching,* Cambridge, MA: Brookline Books. (1997) 108-144.

[10]M.R. Lepper, & M. Woolverton, The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.*), Improving academic achievement: Impact of psychological factors on education,* San Diego, CA: Academic Press. (2002) 135-158.

[11]P.A Alexander, P.K. Murphy, B.S. Woods, K.E. Duhon, & D. Parker, College instruction and concomitant changes in students' knowledge, interest, and strategy use: A study of domain learning. *Contemporary Educational Psychology*, **22** (1997) 125–146.

[12]S. Hidi, Interest and its contribution as a mental resource for learning. *Review of Educational Research,* **60** (1990) 549–571.

[13]U. Schiefele, Interest, learning, and motivation. *Educational Psychologist,* **26** (1991) 299-323.

[14]S. Craig, A.C. Graesser, J. Sullins, & B. Gholson, Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media,* **29** (2004) 241-250.

[15]M. Csikszentmihaly, *Flow, the psychology of optimal experience*. New York: Harper Collins. 1990.

[16]M.M.T. Rodrigo, R.S.J.D. Baker, M.C.V. Lagud, S.A.L. Lim, A.F. Macapanpan, S.A.M.S. Pascua, J.Q. Santillano, L.R.S. Sevilla, J.O. Sugay, S. Tep, & N.J.B. Viehland, Affect and usage choices in simulation problem solving environments. In R. Luckin, K.R. Koedinger, & J. Greer, (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, Amsterdam: IOS Press. (2007) 145-152.

[17]S.K. D'Mello, R. Taylor, & A.C. Graesser, Monitoring affective trajectories during complex learning. In D.S. McNamara & J.G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society*. Austin, TX: Cognitive Science Society. (2007) 203-208.

[18]P. Vorderer, C. Klimmt, & U. Ritterfeld, Enjoyment: At the heart of media entertainment. *Communication Theory*, **14** (2004) 388-408.

[19]J.P. Gee, What video games have to teach us about learning and literacy. New York: Palgrave MacMillian, 2003.

[20]M.E. Gredler, Games and simulations and their relationships to learning. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd Ed.) Mahwah, NJ, US: Lawrence Erlbaum Associates, (2004) 571-582.

[21]H.F. O'Neil, R. Wainess, & E.L. Baker, Classification of learning outcomes: Evidence from the computer games literature. *Curriculum Journal,* **16** (2005) 455-474.

[22]L.P. Rieber, Seriously considering play: Designing interactive learning enviornments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development,* **2** (1996) 43-58.

[23]R. Van Eck, Building Intelligent Learning Games. In David Gibson, Clark Aldrich, & Marc Prensky (eds) *Games and Simulations in Online Learning Research & Development Frameworks.* Hershey, PA: Idea Group, 2007.

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

[25]R. Azevedo, & R.M. Bernard, A meta-analysis of the effect of feedback in computer-based instruction. *Journal of Educational Computing Research,* **13** (1995) 109-125.

[26]A.C. Graesser, N.K. Person, & J.P. Magliano, Collaborative dialogue patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology,* **9** (1995) 495-522.

[27]P.W. Foltz, S. Gilliam, & S. Kendall, Supporting content-based feedback in online writing evaluation with LSA. *Interactive Learning Environments,* **8** (2000) 111-129.

[28]G.T. Jackson, & A.C. Graesser, Content Matters: An investigation of feedback categories within an ITS. In R. Luckin, K. Koedinger, & J. Greer (Eds.)*, Artificial Intelligence in Education: Building technology rich learning contexts that work,* Amsterdam, The Netherlands: IOS Press. (2007) 127-134.

[29]V. Shute, *Focus on Formative Feedback*. Unpublished manuscript. Princeton, NJ: Educational Testing Service, 2006.

[30]T. Malone, & M. Lepper, Making learning fun: A taxonomy of intrinsic motivations of learning. In R. E. Snow, & M. J. Farr (Eds.), *Aptitude, learning, and instruction: Vol. 3. Cognition and affective process analyses,* Hillsdale, NJ: Lawrence Erlbaum, (1987) 223-253.

[31]B.J. Zimmerman, & A. Kitsantas, Developmental phases in self-regulation: Shifting from process goals to outcome goals. *Journal of Educational Psychology*, **89** (1997) 29-36.

[32]A.T. Corbett, & J.R. Anderson, Locus of feedback control in computer-based tutoring: impact of learning rate, achievement and attitudes. *Proceedings of ACM CHI-2001 Conference on Human Factors in Computing Systems*, (2001) 245–252.

[33]D.S. McNamara, & A.M. Shapiro, Multimedia and hypermedia Solutions for promoting metacognitive engagement, coherence, and learning. *Journal of Educational Computing Research*, **33** (2005) 1-29.

[34]D.I. Cordova, & M.R. Lepper, Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology,* **88** (1996) 715-730.

[35]A.L. Baylor, & Y. Kim, Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education,* **15** (2005) 95-115.

[36]W.L. Johnson, Pedagogical Agent Research at CARTE, AI Magazine, **22** (2001) 85-94.

[37]J.C. Lester, B.A. Stone, & G.D. Stelling, Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction*, **9** (1999) 1-44.

[38]R. Moreno, & R.E. Mayer, Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology,* **96** (2004) 165-173.

[39]B. Reeves, & C. Nass, *The media equation: How people treat computers, television, and new media like real people and places*, New York: Cambridge University Press, 1996.

[40]W.L. Johnson, S. Kole, E. Shaw, & H. Pain, Socially Intelligent Learner-Agent Interaction Tactics. *Proc. of the 11th Int'l Conference on Artificial Intelligence in Education (AIED)*. Amsterdam: IOS Press. (2003).

[41]A.C. Graesser, & V. Ottati, Why stories? Some evidence, questions, and challenges. In R. S. Jr. Wyer (Ed.), *Knowledge and memory: The real story. Vol. 8, Advances in social cognition,* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. (1995) 121-132.

[42]S.J. Read, & L.C. Miller, Stories are fundamental to meaning and memory: For social creatures, could it be otherwise. In R. S. Wyer, R. P. Abelson, & R. C. Schank (Eds.), *Knowledge and Memory: The Real Story,* Hillsale, NJ: Erlbaum. (1995) 139-152.

[43]J.P. Gee, *What videogames have to teach us about learning and literacy*, New York: Palgrave-MacMillan, 2004.

[44]Y. Cheong, & R.M. Young, A Computational model of Narrative Generation for Suspense. *Proceedings of AAAI 2006 Workshop on Computational Aesthetics.* (2006)

[45]C.A. Steinkuehler, Why game (culture) studies now? *Games and Culture,* **1** (2006) 1-6.

CRYSTAL ISLAND: A Narrative-Centered Learning Environment for Eighth Grade Microbiology

Jonathan P. ROWE[a,1], Bradford W. MOTT[a], Scott W. MCQUIGGAN[b], Jennifer L. ROBISON[a], Sunyoung LEE[a], and James C. LESTER[a]
[a] *Department of Computer Science, North Carolina State University, Raleigh NC 27695*
[b] *Education Practice, SAS Institute, Inc., Cary NC 27513*

Abstract. Narrative-centered learning environments offer significant promise for promoting interactive learning experiences that are both effective and engaging. Models of narrative generation and reasoning can balance the motivational and pedagogical aspects of narrative-centered learning interactions. Affect recognition and affect expression models are useful for shaping students' affective trajectories during narrative-centered learning. Conducting empirical evaluations are critical for determining what factors contribute to the potential pedagogical and motivational benefits of narrative-centered learning environments. This paper presents an overview of progress in these areas by summarizing work on CRYSTAL ISLAND, a narrative-centered learning environment for eighth grade microbiology.

Keywords. Narrative-centered learning environments, game-based learning environments, affect recognition, affect expression, evaluation.

1. Introduction

Narrative-centered learning environments (NLEs) afford significant opportunities for students to participate in motivating story-based educational experiences. By combining commercial game technologies, intelligent tutoring systems, and rich narrative structures, NLEs seek to provide effective, engaging learning experiences that are tailored to individual students. NLEs show promise for encouraging problem solving, strategic and analytical thinking, decision-making, and other 21st century skills [1]. They also serve as a natural platform for adapting learning experiences to individual students. Recent work on narrative-centered learning environments has investigated pedagogical agents with rich models of dialogue [2, 3], affect [4, 5, 6] and social behavior [7], director agents that can manipulate the pedagogical and narrative directions of learning experiences [8], models for detecting students' emotional and motivational states [9], and predictive models of students' goals [10] and other in-game behaviors. NLEs are currently under investigation in a range of domains, including language learning [2, 11], anti-bullying education [5], and science learning [12].

This paper presents work on several experimental versions of CRYSTAL ISLAND, a narrative-centered learning environment developed for middle school students in the

[1] Corresponding Author: Jonathan P. Rowe, Department of Computer Science, North Carolina State University, Raleigh, NC 27695; E-mail: jprowe@ncsu.edu

domain of eighth-grade microbiology. Specifically, it summarizes several lines of investigation that have explored how computational models of narrative and affect can be leveraged in NLEs to create effective, engaging learning experiences. The structure of the paper is as follows. Section 2 provides background and related work on narrative-centered learning environments. Section 3 provides a detailed description of the CRYSTAL ISLAND virtual environment. Following in Section 4 is a summary of work on narrative generation and reasoning, affective recognition and expression, and evaluation in CRYSTAL ISLAND. Section 5 provides concluding remarks and a brief discussion of future work.

2. Background and Related Work

Narrative-centered learning environments offer significant potential for enhancing students' learning experiences. Stories draw audiences into plots and settings, thereby opening perceptual, emotional, and motivational opportunities for learning. Establishing concrete connections between narrative context and pedagogical subject matter has been said to support the assimilation of new ideas in young learners [13]. Narratives can also facilitate students' semantic encoding of new information and making commitments to long-term memory in the form of episodic memories [14]. Furthermore, fantasy contexts in educational games have been shown to provide motivational benefits for learning [15]. Although it is important to remain mindful of potential disadvantages such as seductive details [16], a dynamically generated narrative that draws students into the evolving plot has the potential to be pedagogically compelling.

Narrative-centered learning environments leverage a range of techniques for providing effective, engaging learning experiences. Multi-user virtual environments such as Quest Atlantis [17] and River City [18] use rich narrative settings to contextualize inquiry-based science learning scenarios with strong social and ethical dimensions. Although the systems do not use artificial intelligence to provide tailored narrative or learning experiences, several classroom studies have yielded promising learning results. Other work on narrative-centered learning environments has applied a range of techniques to generating engaging interactive narrative experiences that are pedagogically effective and tailored to individual students. FearNot! uses affectively-driven autonomous agents to generate dramatic, educational vignettes about bullying [5]. In between the non-interactive vignettes, the virtual agent consults the student for advice about prior bullying scenarios, and then uses this feedback to inform its behavior in subsequent vignettes. The Thespian architecture takes a decision-theoretic, multi-agent approach to controlling virtual characters in the Tactical Language and Culture Training System's narrative scenarios [11, 2]. The agents' goal-driven behaviors are trained using a corpus of linear, pre-authored scripts, providing the agents with believable behavior models for conversing with students during language and culture training scenarios [2]. The SASO (Stability and Support Operations) narrative-centered learning environment uses robust, socially intelligent virtual humans as actors in military training scenarios. SASO's virtual humans implement models of multimodal conversational behavior [19, 3], affective reasoning [4], and social behavior [7].

Computational models of affect recognition and affect expression are also important for effective narrative-centered learning environments. The AI in Education community has seen the emergence of work on affective student modeling [20], detecting frustration and stress [21, 22, 23], modeling agents' emotional states [24, 4], devising affectively informed models of social interaction [25, 26, 27], detecting student motivation [28], and diagnosing and adapting to student self-efficacy [29]. All of this work seeks to increase the fidelity with which affective and motivational processes are understood and utilized in intelligent tutoring systems in an effort to increase the effectiveness of tutorial interactions and, ultimately, learning. Recent work has also sought to characterize the affective experiences of learners interacting with intelligent learning environments by considering student affective trajectories during learning [30, 31, 32].

Figure 1. The CRYSTAL ISLAND narrative-centered learning environment.

3. Crystal Island

CRYSTAL ISLAND (Figure 1) is a narrative-centered learning environment built on Valve Software's Source™ engine, the 3D game platform for Half-Life 2. CRYSTAL ISLAND features a science mystery set on a recently discovered volcanic island. The curriculum underlying CRYSTAL ISLAND's science mystery is derived from the North Carolina state standard course of study for eighth-grade microbiology. Students play the role of the protagonist, Alyx, who is attempting to discover the identity and source of an infectious disease plaguing a newly established research station. The story opens by introducing the student to the island and members of the research team for which the protagonist's father serves as the lead scientist. Several of the team's members have fallen gravely ill, including Alyx's father. Tensions have run high on the island, and prior to Alyx's arrival various team members began to accuse one another of having poisoned the sick researchers. It is the student's task to discover the outbreak's cause and source, and determine whether one of the team members is guilty of poisoning.

CRYSTAL ISLAND's setting includes a beach area with docks, an outdoor field laboratory, underground caves, and a research camp. Throughout the mystery, the student is free to explore the world and interact with other characters while forming questions, generating hypotheses, collecting data, and testing hypotheses. The student can pick up and manipulate objects, take notes, view posters, operate lab equipment, and talk with non-player characters to gather clues about the source of the disease. During the course of solving the mystery, the student is minimally guided through a five problem curriculum. The first two problems focus on pathogens, including viruses, bacteria, fungi, and parasites. The student gathers information by interacting with in-game pathogen "experts" and viewing books and posters in the environment. In the third problem, the student is asked to compare and contrast her knowledge of four types of pathogens. In the fourth problem, the student is guided through an inquiry-based hypothesis-test-and-retest problem. In this problem she must complete a "fact sheet" with information pertaining to the disease afflicting members of the research team. Once the "fact sheet" is completed and verified by the camp nurse, the student completes the final problem concerning an appropriate treatment plan for the sickened CRYSTAL ISLAND researchers.

To illustrate the behavior of CRYSTAL ISLAND, consider the following situation. Suppose a student has been interacting with virtual agents in the storyworld and learning about infectious diseases. In the course of having members of the research team become ill, she has learned that an infectious disease is an illness that can be transmitted from one organism to another. As she concludes her introduction to infectious diseases, she learns from the camp nurse that the mystery illness seems to be coming from food items the sick members recently ate. Some of the island's characters are able to help identify food items and symptoms that are relevant to the scenario, while others are able to provide helpful microbiology information. The student discovers through a series of tests that a container of unpasteurized milk in the dining hall is contaminated with bacteria. By combining this information with her knowledge about the characters' symptoms, the student deduces that the disease is *E. coli*. The student reports her findings back to the camp nurse, and they discuss a plan for treatment. The *E. coli* diagnosis ultimately exonerates the members accused of poisoning, and the sick researchers make a speedy recovery.

4. Current Progress

Narrative-centered learning environments are inherently complex systems, and multiple lines of investigation are necessary to develop suites of technologies that can produce adaptive, engaging learning experiences. To date three principal areas of research have been conducted with the CRYSTAL ISLAND virtual environment: narrative generation and reasoning, affect recognition and expression, and empirical evaluation of NLEs.

4.1. Narrative Generation and Reasoning

Because of their interactive nature, narrative-centered learning environments must cope with a wide range of student actions that can be performed in a virtual environment. Providing students with a strong sense of control and agency is important for supporting motivation [33], but agency also introduces opportunities for students to violate or ignore important aspects of an intended narrative experience. This presents a major challenge for narrative-centered learning environments: to simultaneously maintain the coherence and pedagogical effectiveness of a learning experience, but permit significant user agency in the environment. Developing computational models that can reason about students' actions within the narrative, adapt and re-plan narrative events in response to student actions, and promote robust, believable interactions with virtual characters is critical.

Models of narrative generation for CRYSTAL ISLAND have taken a dual planning space approach, with one planning space allocated to tutorial planning and a second allocated to narrative planning [34]. A tutorial planner supports the requirements of inquiry-based learning by formulating tutorial strategies that encourage question formation, hypothesis generation, data collection, and hypothesis testing. A narrative planner is responsible for generating plot elements, sequencing plot elements into coherent and engaging stories, and directing characters' actions and storyworld events to achieve tutorial and narrative goals. However, for the two planners to work in concert, they must effectively coordinate their actions, resulting in a single stream of events occurring in the virtual storyworld. To this end, the tutorial planner posts goals in the tutorial planning space that are achieved by operators in the narrative planning space. This tutorial-driven model seeks to balance plot advancement and tutorial goal achievement seamlessly by the built-in coordination of the two planning spaces via lower-level tutorial constraints and the upper-level narrative goals [34].

The resulting plans are incorporated into a decision-theoretic director agent architecture that manages the narrative-centered learning experience [8]. Decision-theoretic narrative planning offers a unified approach to dynamically guiding narratives in a storytelling environment. The director agent has access to three principal knowledge sources: narrative objectives, storyworld state, and student state. To cope with the uncertainty in narrative planning, the three sets of knowledge sources are integrated into a dynamic decision network (DDN) that the director agent evaluates regularly to select the next narrative action. Once the director agent has fully updated the decision network, it selects the director action that maximizes the expected narrative utility, waits to see what action the student takes (if any) and updates its beliefs as necessary [8].

Student goal recognition is also important for narrative generation in game-based learning environments. Providing narrative planners with the ability to recognize students' goals could enable planners to monitor students to determine if their goals were consistent with the plot, and determine whether sufficient plot progress has been made. Two families of goal recognition models have been investigated in CRYSTAL ISLAND: *n*-gram models (unigrams and bigrams) and Bayesian network models [10]. The models, which exploit knowledge of narrative structure as well as locational information about students' activities in the world, are induced from training data acquired from traces of students' actions in the story environment. Experimental results suggest that probabilistic models can accurately predict students' goals, and that they converge on correct interpretations as observations of a student's activities become available over time [10].

An additional feature of narrative-centered learning environments is their natural support for encouraging rich, believable interactions with embodied pedagogical agents. Character dialogue behavior is crucial for defining and advancing plots, as well as scaffolding learning experiences. Character dialogue generators for interactive narrative environments must meet several requirements. They must generate dialogues that are appropriate for characters' traits, such as personalities, motivations, and preferences; they must consider narrative context and history as they formulate dialogue; and they must be able to robustly handle the large number of possible character-character and character-player interactions that may result in dialogue. CRYSTAL ISLAND takes advantage of a probabilistic unification-based dialogue generation architecture that considers multiple sources of information (character archetypes, narrative context, and communicative goals) to dynamically generate character- and situation-appropriate dialogue [35]. The generated dialogues use preference information encoded within character archetype representations and yield character-specific variations in the dialogue that satisfy the major objectives for conversational interactions during narrative-centered learning.

4.2. Affect Recognition and Expression

Emotion is critical for both narrative and learning. Creating narrative-centered learning environments that are in tune with students' affective experiences can provide support for guiding pedagogical scaffolding and engaging students in virtual story worlds. Affective interactions proceed through a three-stage process termed the *affective loop*: affect recognition, affect understanding, and affect expression (adapted from [36]). Affect recognition is the task of inferring a user's affective state from a sequence of observations of behavior. Affect expression is the task of determining how a system should communicate emotion. Affect understanding is the process of interpreting recognized user emotions, determining what it means for the user to feel the recognized emotion, and then formulating adaptation strategies based on how the user feels. Collectively, affect recognition, expression, and understanding have been the subject of growing attention in the AI in Education community. Work on CRYSTAL ISLAND has primarily focused on data-driven approaches to affect recognition and affect expression, and has yielded promising initial results.

Data-driven models of affect recognition are trained and validated using a rich corpus of student actions, locations, goals, physiological information, and temporal information collected during student interactions with the CRYSTAL ISLAND environment. After students' problem-solving traces have been recorded, affect recognition models are induced using supervised machine learning techniques such as naïve Bayes, decision trees, and support vector machines. This methodology has yielded affect recognition models that are both accurate and efficient, with some models capable of correctly predicting over 95% of students' emotion self-reports [9]. The same methodology has also been successfully applied to induce models of student self-efficacy in the CRYSTAL ISLAND environment [37]. Related work has investigated students' affective transitions in the CRYSTAL ISLAND environment, differentiating typical affective transitions and those stemming from pedagogical agents' empathetic responses to student affect [38]. Also important is the ability to make "early" predictions of student affect. Early detection allows systems adequate time to prepare for particular affective states, opening a window of opportunity for the learning environment to take corrective action. Inductive approaches using a combination of *n*-gram models and decision trees have yielded results with accuracy, precision, and recall exceeding 88%, which was significantly better than baseline comparisons [23].

Computational models of agents' empathetic behavior are an important area of investigation in affect expression. Defined as "the cognitive awareness of another person's internal states, that is, his thoughts, feelings, perceptions, and intentions" [39], empathy enables people to vicariously respond to one another via "psychological processes that make a person have feelings that are more congruent with another's situation than with his own situation" [40]. Initial work on constructing empathetic virtual agents for CRYSTAL ISLAND explored learning empirically grounded models of empathy from observations of human-human social interactions [6]. In this approach, training data is first generated as a by-product of trainers' interactions in a virtual environment, and models of empathy are induced from the resulting datasets. Critically, the training data include only features that can be directly observed in the environment, so that at runtime, the same features can be used by the empathy models to drive the behavior of virtual agents interacting with students. Two complementary lines of evaluation, one investigating predictive accuracy and one investigating perceived accuracy, were conducted on an implemented empathy modeler, and yielded promising results. Follow-up work investigated more detailed models of empathetic agent behavior, namely the use of parallel vs. reactive empathy [41]. In the parallel case, the empathizer mimics the affective state of the target. In the reactive case, empathizers exhibit a higher cognitive awareness of the situation and react with empathetic behaviors that do not necessarily match those of the target's affective state. The results indicated that models of empathy induced from knowledge of the student's situation and the student's affective state can effectively determine which type of empathy is most appropriate for interactions requiring empathetic expression.

4.3. Empirical Evaluation

The third principal line of investigation in CRYSTAL ISLAND has been empirical evaluation of narrative-centered learning environments. NLEs offer myriad opportunities for enhancing learning experiences and motivating students. However, determining appropriate metrics, criteria, and methodologies that can be used to assess NLEs poses a number of challenges. Determining the impact of individual elements of narrative-centered learning environments (setting, plot, game play activities, empathetic pedagogical agents) also introduces a wide range of practical and theoretical challenges. Further, because many narrative-centered learning environments permit activities that are not strictly pedagogical, assessment of players' action traces and problem-solving paths can have implications for environment design as well as models of adaptive scaffolding. The complexity inherent in intelligent narrative-centered learning environments calls for a sophisticated, multi-faceted approach to evaluation. Empirical evaluations of CRYSTAL ISLAND have sought to take initial steps toward this objective.

A controlled, human participant experiment with middle school students investigated the impact of narrative on learning [12]. The study compared two versions of CRYSTAL ISLAND against a more traditional instructional approach, a narrated slideshow that conveyed the same curricular material. The two CRYSTAL ISLAND conditions featured varying levels of narrative content supplementing the curriculum. The results showed that students in the NLE conditions did exhibit learning gains, but that those gains were less than those produced by traditional instructional approaches. However, the motivational benefits of narrative-centered learning, particularly with regard to self-efficacy, presence, interest, and perception of control, were substantial. Students reported the highest levels of presence in the full-narrative condition, a finding that bears important implications for motivation. A recently completed follow-up study with an updated version of the CRYSTAL ISLAND learning environment again found that students in the NLE condition exhibited learning gains, and further, with the updated version, the gains were on par with those of the slideshow condition. Numerically, the learning gains in the CRYSTAL ISLAND condition exceeded those in the slideshow condition, and analyses are underway to investigate these findings.

A different series of studies examined the impact of virtual characters' empathetic behavior on student presence in narrative-centered learning environments [42]. In a study with middle school students comparing non-empathetic and empathetic characters, it was found that empathetic characters in narrative-centered learning environments had a significant effect on measurements of students' overall presence (total PQ), involvement and control, naturalism of the experience, and resolution. When the study was replicated with high school students, the same effects were found. In short, it appears that empathetic interactions with characters in narrative-centered learning environments can contribute to increased student presence. The results are encouraging for the motivational potential of narrative-centered learning environments, and they extend other results illustrating the relationship between narrative and presence. The work points to a need for continued examination of the impact of narrative content in virtual environments, as well as social and emotional interactions that take place during those experiences [42].

Empirical evaluations of CRYSTAL ISLAND have also focused on the nature of student behavior in narrative-centered learning environments. Work examining students' note-taking is one example. CRYSTAL ISLAND provides students with a note-taking feature so that students can document useful information encountered during learning interactions. A corpus of student notes was collected from a study involving 116 middle school students [43]. A team of judges annotated the corpus by classifying individual notes into one of several categories including narrative, curricular, and hypothesis notes. An analysis of the tagged corpus revealed that students who took hypothesis notes performed better on posttests, confirming inquiry-based learning findings suggesting the importance of scaffolding students' hypothesis generation activities. Individual differences were also able to suggest which students are likely to take notes. Results illustrated significant gender effects on note-taking, where females took significantly more notes than males. Goal orientation and efficacy for self-regulated learning also exhibited significant correlations with note-taking behavior in narrative-centered learning environments.

5. Conclusions and Future Work

Narrative is the subject of increasing attention in the AI in Education community as a powerful medium for contextualizing learning. Narrative-centered learning environments present a range of opportunities for investigating how different computational models can be leveraged to create effective, engaging learning experiences. Work on the CRYSTAL ISLAND environment has begun to illustrate how models of narrative generation and reasoning, as well as affect recognition and expression, can bear on game-based learning environments. A series of empirical evaluations has begun to demonstrate the motivational and pedagogical potential of narrative-centered learning.

Results to date suggest several promising directions for future work on narrative generation and reasoning, affect recognition and expression, and empirical evaluation in narrative-centered learning environments. Currently under investigation are adaptive models for scaffolding students' narrative and pedagogical progress through a learning environment. Devising models that can integrate knowledge about the state of a virtual environment, sets of intended narrative objectives, and individual student qualities, poses serious challenges. Additionally, exploring computational models of affect understanding will become increasingly important for closing the "affective loop" that exists in interactive learning environments. Finally, a critical step in this research agenda will be conducting extensive empirical investigations to explore the relationships between narrative, affect, character behavior, and their collective impact on learning gains, self-regulated learning, and motivation.

Acknowledgements

The authors wish to thank members of the IntelliMedia lab for their assistance, Omer Sturlovich and Pavel Turzo for use of their 3D model libraries, and Valve Software for access to the Source™ engine and SDK. This research was supported by the National Science Foundation under Grants REC-0632450, IIS-0757535, DRL-0822200 and IIS-0812291. Portions of this material are based upon work supported under a National Science Foundation Graduate Research Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Partnership for 21st Century Skills. A Policymaker's Guide to 21st century Skills. Retrieved April 1, 2006, from http://www.21stcenturyskills.org/images/stories/otherdocs/P21_Policy_Paper.pdf

[2] M. Si, S. Marsella, D. Pynadath, Thespian: Modeling socially normative behavior in a decision-theoretic framework, in *Proceedings of the 6th International Conference on Intelligent Virtual Agents* (2006), 369–382.

[3] D. Traum, S. Marsella, J. Gratch, J. Lee and A. Hartholt, Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents, in *Proceedings of the 8th International Conference on Intelligent Virtual Agents* (2008), 117–130.

[4] J. Gratch and S. Marsella, A domain-independent framework for modeling emotion, *Journal of Cognitive Systems Research*, 5(4) (2004), 269-306.

[5] R. Aylett, S. Louchart, J. Dias, A. Paiva and M. Vala, FearNot! an experiment in emergent narrative, in *Proceedings of the 5th International Conference on Intelligent Virtual Agents* (2005), 305-316.

[6] S. McQuiggan and J. Lester, Modeling and evaluating empathy in embodied companion agents, *International Journal of Human-Computer Studies* 65(4) (2007), 348-360.

[7] J. Gratch, N. Wang, J. Gerten, E. Fast and R. Duffy, Creating rapport with virtual agents, in *Proceedings of the 7th International Conference on Intelligent Virtual Agents* (2007), 223–236.

[8] B. Mott and J. Lester, U-Director: A Decision-theoretic narrative planning architecture for storytelling environments, in *Proceedings of the 5th International Conference on Autonomous Agents and Multiagent Systems* (2006), 977-984.

[9] S. Lee, S. McQuiggan and J. Lester, Inducing user affect recognition models for task-oriented environments, in *Proceedings of the 11th International Conference on User Modeling* (2007), 380-384.

[10] B. Mott, S. Lee and J. Lester, Probabilistic goal recognition in interactive narrative environments, in *Proceedings of the 21st National Conference on Artificial Intelligence* (2006), 187-192.

[11] W. L. Johnson, Serious use of a serious game for language learning, in *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (2007), 67-74.

[12] S. McQuiggan, J. Rowe, S. Lee and J. Lester, Story-based learning: the impact of narrative on learning experiences and outcomes, in *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (2008), 530-539.

[13] C. Wells, *The meaning makers: children learning language and using language to learn*, Heinemann, Portsmouth, NH, 1986.

[14] J. Ormrod, *Human Learning*. 4th Edition. Pearson Education, Upper Saddle River, NJ, 2004.

[15] L. Parker and M. Lepper, Effects of fantasy contexts on children's learning and motivation: Making learning more fun, *Journal of Personality and Social Psychology* 62(4) (1992), 625 – 633.

[16] S. Harp and R. Mayer, How seductive details do their damage: a theory of cognitive interest in science learning, *Journal of Educational Psychology* 90(3) (1998), 414-434.

[17] S. Barab, T. Dodge, H. Tuzun, K. Job-Sluder, C. Jackson, A. Arici, L. Job-Sluder, R. Carteaux, J. Gilbertson Jr. and C. Heiselt, The Quest Atlantis Project: A socially-responsive play space for learning, in B. E. Shelton and D. Wiley (Eds.) *The Educational Design and Use of Simulation Computer Games,* (pp. 159-186), Sense Publishers, Rotterdam, The Netherlands, 2007.

[18] D. J. Ketelhut, C. Dede, J. Clarke, B. Nelson and C. Bowman, Studying situated learning in a multi-user virtual environment, in E. Baker, J. Dickieson, W. Wulfeck and H. O'Neil (Eds.) *Assessment of problem solving using simulations*, (pp. 37-58), Lawrence Erlbaum, Mahwah, NJ, 2007.

[19] J. Lee and S. Marsella, Nonverbal behavior generator for embodied conversational agents, in *Proceedings of the 6th International Conference on Intelligent Virtual Agents* (2006), 243-255.
[20] C. Conati, and H. Mclaren, Data-driven refinement of a probabilistic model of user affect, in *Proceedings of the 10th International Conference on User Modeling* (2005), 40-49.
[21] H. Prendinger, and M. Ishizuka, The empathic companion: a character-based interface that addresses users' affective states, *Applied Artificial Intelligence* 19(2005), 267-285.
[22] W. Burleson, Affective learning companions: strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance, PhD thesis, Massachusetts Institute of Technology (2006).
[23] S. McQuiggan, S. Lee, and J. Lester, Early prediction of student frustration, in *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction* (2007), 380-384.
[24] E. André and M. Mueller, Learning affective behavior, in *Proceedings of the 10th International Conference on Human-Computer Interaction* (2003), 512-516.
[25] W. L. Johnson and P. Rizzo, Politeness in tutoring dialogs: "run the factory, that's what I'd do", in *Proceedings of the 7th International Conference on Intelligent Tutoring Systems* (2004), 67-76.
[26] K. Porayska-Pomsta and H. Pain, Providing cognitive and affective scaffolding through teaching strategies, in *Proceedings of the 7th International Conference on Intelligent Tutoring Systems* (2004), 77-86.
[27] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall and C. Zoll, Learning by feeling: evoking empathy with synthetic characters, *Applied Artificial Intelligence* 19(2005), 235-266.
[28] A. de Vicente and H. Pain, Informing the detection of the students' motivational state: an empirical study, in *Proceedings of the 6th International Conference on Intelligent Tutoring Systems* (2002), 933-943
[29] C. Beal and H. Lee, Creating a pedagogical model that uses student self reports of motivation and mood to adapt ITS instruction, in *Workshop on Motivation and Affect in Educational Software, in conjunction with the 12th International Conference on Artificial Intelligence in Education* (2005).
[30] S. Craig, A. Graesser, J. Sullins and B. Gholson, Affect and learning: an exploratory look into the role of affect in learning with AutoTutor, *Journal of Educational Media* 29(3) (2004), 241-250.
[31] R. Baker, M. Rodrigo and U. Xolocotzin, The dynamics of affective transitions in simulation problem-solving environments, in *Proceedings the 2nd International Conference on Affective Computing and Intelligent Interactions* (2007), 666-677.
[32] S. D'Mello, R. S. Taylor and A. Graesser, Monitoring affective trajectories during complex learning, in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (2007), 203-208.
[33] T. Malone and M. Lepper, Making learning fun: a taxonomy of intrinsic motivations for learning, in R. Snow and M. Farr (Eds.) *Aptitude, learning, and instruction: III. Conative and affective process analyses,* (pp. 223-253), Erlbaum, Hillsdale, NJ, 1987.
[34] B. Mott and J. Lester, Narrative-Centered Tutorial Planning for Inquiry-Based Learning Environments, in *Proceedings of the Eighth International Conference on Intelligent Tutoring Systems* (2006), 675-684.
[35] J. Rowe, E. Ha and J. Lester, Archetype-Driven Character Dialogue Generation for Interactive Narrative, in *Proceedings of the Eighth International Conference on Intelligent Virtual Agents (2008)*, 45-58.
[36] P. Sundstrom, A. Stahl and K. Hook, In situ informants exploring an emotional mobile messaging system in their everyday practice*, International Journal of Human-Computer Studies* 65(4) (2007), 388-403.
[37] S. McQuiggan, B. Mott and James Lester. Modeling Self-Efficacy in Intelligent Tutoring Systems: An In*ductive Approach*. *User Modeling and User-Adapted Interaction 18(1-2) (2008), 81-12*3.
[38] S. McQuiggan, J. Robison and J. Lester, Affective Transitions in Narrative-Centered Learning Environments, in *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems* (2008), 490-499.
[39] W. Ickes, *Empathic Accuracy*, Guilford Press, New York, NY, 1997.
[40] M. Hoffman, *Empathy and Moral Development: Implications for Caring and Justice*, Cambridge University Press, Cambridge, UK, 2000.
[41] S. McQuiggan, J. Robison, R. Phillips and J. Lester, Modeling Parallel and Reactive Empathy in Virtual Agents: An Inductive Approach, in *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems* (2008), 167-174.
[42] S. McQuiggan, J. Rowe and J. Lester, The effects of empathetic virtual characters on presence in narrative-centered learning environments, in *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2008), 1511-1520.
[43] S. McQuiggan, J. Goth, E. Ha, J. Rowe and James Lester. Student Note-Taking in Narrative-Centered Learning Environments: Individual Differences and Learning Outcomes, in *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems* (2008), 510-519.

Designing a Game for Teaching Argumentation Skills

Peter HASTINGS [a] [1] Anne BRITT [b] Brad SAGARIN [b] Amanda DURIK [b]
Kris KOPP [b]

[a] *College of Computing and Digital Media, DePaul University*
[b] *Department of Psychology, Northern Illinois University*

Abstract. This paper describes the design of a game which is aimed at teaching argumentation skills to college students. Ability to understand and generate arguments is critical for STEM and a broad range of other fields, but it is sorely lacking in students today. Building on our prior experience creating intelligent tutoring systems for teaching argumentation skills, we have designed a game so that we can compare the effectiveness of the two approaches. This paper describes the design of the game from the viewpoint of cognitive principles and frameworks for serious game design.

Keywords. Game design, game design principles, game design frameworks

Introduction

Argumentation is a central component of social and political decision making as well as a fundamental skill required by many class assignments and by entrance exams to post-secondary education (e.g., SAT and GRE). It is also an ability that we expect our educational system to impart to students during their schooling [1]. Sadly, many students leave high school unable to identify and write arguments [2].

In previous work, we have created intelligent tutoring systems for argumentation and other areas [3,4,5,6, for example]. In the current project, we are attempting to develop a computer game which will teach the same types of information about argumentation as the tutoring systems did. Our design for the game includes the high-level conceptual knowledge that we have gleaned from our experiments with the tutoring system and some of the underlying AI technology for evaluating generated arguments. We did not, however, try to import any of the modes of tutorial interaction into the game.

Our game is called, "Advisor to the King" (hereafter AttK). In AttK, the student/player is put in the role of a new employee in the King's bureaucracy. At each level, the player's "job" is to evaluate petitions from the King's subjects using a different argumentation skill. The player submits her "work" to her boss. If the boss approves it, she can get a promotion. If she fails, she'll be sent back to do it right. At present, we have created the initial design for the game, and most of the materials. We still need to do playtesting, iterative design, and learning testing.

This paper analyzes our game design from the standpoints of different principles of educational game design, namely those of Gee [7] and Clark and Mayer [8]. When we started this project, we were relatively ignorant of any theoretical or empirical approaches to game design. Instead, our game design process was primarily driven by our knowledge of learning sciences and our intuitions about effective game design. Thus, this paper presents a *post hoc* analysis according to the frameworks mentioned, and is meant to indicate:

- how our intuitions fit with recent frameworks for serious game design,
- where our game could use some improvement,
- more generally, the benefits of using these principles and frameworks to guide serious game design.

1. Gee's Principles of Learning in Games

Based on his own experience in playing video games with his son, the semiotician James Gee published an influential text which described 31 principles that non-educational video games follow to support learning [9]. In 2005, he published a condensed set of 13 principles in three different conceptual areas [7]. In this section, we describe how these principles apply to the AttK game.

1.1. Empowered Learners

1.1.1. Co-Design

According to Gee, this principle holds that good games give players the feeling that they are actively creating part of their experience, having an effect on the virtual world they're inhabiting, and influencing their playing experience. As Gee explains, this can be rather trivially true; in almost every game, each player has a somewhat different experience. It is most effective when a player feels like her choices have a significant impact on the tasks she is attempting and how she approaches them.

In AttK, this principle is in evidence, but not as strongly as it might be. The player has some choices in taking on some side tasks which were primarily intended to support transfer by giving the player practice in different settings within the game. For example, the player may visit a pub where an argument soon breaks out between two other patrons. The player has the opportunity to use their argumentation skills to help settle the argument before it gets out of hand, or they can choose not to act and leave the other patrons to work it out for themselves.

AttK could provide more opportunities for the players to control their own experiences by providing a wider range of practice tasks. In this first stage of development, however, we are concentrating on demonstrating the effectiveness of the general concept, and with our limited budget, we must be satisfied with having one type of task in each level of the game.

1.1.2. Customize

This principle addresses a more fundamental way that the player can influence gameplay: by tailoring it to their favored style of learning. Whether due to gender [10], multiple intelligences [11], or simply individual preference, different game players prefer different types of games. Gee gives examples of various games that allow players to engage in very different modes of gameplay. AttK does not do this. As mentioned above, the initial version of the game is relatively narrowly defined. Its current game mechanics are designed to provide the player with ample time for reflection on the concepts. A more fast-paced version could be created, and it might well be more attractive to a certain type of player, but it would probably also decrease learning by minimizing time for reflection.

1.1.3. Identity

Teachers know how motivating it can be for students to feel ownership of and invested in a learning task. Conversely, work which must be done "just because the teacher said so," is highly demotivating. Games can provide an exceptionally strong method of fostering this type of investment by immersing players in an alternate reality where they take on a different identity. When a player takes on the role of Wizard, or a Tank, or a Healer in a game which they play hour after hour, the player is highly motivated to learn how to help that character succeed.

In AttK, the player starts out in the role of an entry-level civil service worker. It doesn't pay well, and the boss is a jerk, but if the character/player succeeds (by processing argumentative texts), she can advance in the game, possibly displacing the boss. This is a central conceit of the game situation, and the central hope for its success. If the player develops a sense of unified identity with the character in the game, they should be motivated to learn the argumentation skills to help that character succeed.

1.1.4. Manipulation and Distributed Knowledge

This principle deals with the connection between perception and action. In learning to do any task, the learner must come to recognize the situations in which action must be taken, and also learn what actions are possible. Game play provides an excellent environment for focusing the learner on this task with significant advantages over traditional schooling. Immersing the learner in a (simulated) environment provides a much richer setting than a worksheet or other homework assignment could. The player's character acts as a tool with which the player manipulates objects in the game world. As Gee says, this activity "causes humans to feel as if their bodies and minds have stretched into a new space" [7, p. 8].

This touches on the rather controversial issue of endogenous vs exogenous games [12,13,14]. In exogenous games, the learning content is often added into a general game framework like a quiz show or a shooter game. One well-known example is Math Blaster. In endogenous games, the content material is intimately tied in with the gameplay. Oregon Trail is commonly held up as a good example of an early endogenous game, though its critics contend that many students are more engaged in developing their shooting skills (by shooting animals from trains), than they are in learning about the more conceptual learning objectives of the game. Commercial developers of educational games like exogenous games because they're much less expensive to create [15]. Researchers prefer endogenous games because of their theoretical advantage in learning effectiveness [14].

Although we weren't familiar with the term at the time, we designed AttK as an endogenous game. We created a story in which the player is put into a position where she must use the skills that we are trying to reinforce. Inability to perform the skills will bring feedback and extra practice. Mastery of the skills will bring success and progress within the game. In this case, the skills are more conceptual than perceptual, but the game provides interaction with just the type of conceptual materials that we want students to learn.

If the player's avatar in a game is a tool which allows them to influence the game world, it is a "smart tool." Game characters generally free the player from having to worry about the very low-level aspects of their actions. Pacmen already know how to move and eat, shooters know how to shoot various weapons, and Mario knows how to drive a kart. The player just needs to tell the characters when to act and point them in the right direction.

For educational games, designers have the opportunity to provide avatars with knowledge equivalent to the player's prior knowledge — at least within the narrow context of the game world. Then the learning of the content material within the game can fall neatly in the learner's zone of proximal development [16].

AttK's interface is primarily linguistic, using audio and on-screen text. There is no avatar moving around in a 3D world. But there are assumptions about what the character (and by extension, the player) can and can't do. We assume they can read and make some judgments on texts. We don't assume that they are good at processing argumentative texts. If they are, they should progress quickly to a challenging level within the game. If they are not, the game will help them improve. This should provide learners with minimal barriers to getting started in the game and allow them to learn at their own pace.

1.2. Problem Solving

1.2.1. Pleasantly Frustrating

One general difficulty in education is that students tend to overestimate their knowledge of a subject. This is especially true in dealing with argumentative texts. In a previous study, we showed that more-skilled readers precisely remember the predicate of an argument, while less-skilled readers only retain the general theme [6]. Because these less-skilled readers believe they know what the text is about [17], we expect that they will initially have difficulty completing the tasks which require precise recall of text components. In the context of the game, however, they will get immediate, non-threatening feedback that indicates they are missing something important. If they want, they can also ask for instruction about what is missing within the context of trying to solve a goal. Thus, the game design provides a challenging situation for the player, but also provides mechanisms for solving the challenge, and opportunity to practice the solution.

1.2.2. Information On-Demand and Just-In-Time

In traditional tutoring situations, it is either assumed that new content material was previously taught in class, or the tutor begins a session by didactically presenting the new content. Games take a radically different approach. Although most games come with some sort of manual, few players ever read them. Instead, players rely on the subtle hints that good games give them about what to do. Just-in-time information is incredibly useful for learning because it is directly connected to the player's current goals. The player can see immediately how applying that new knowledge will help them advance in the game.

In AttK, we made an explicit design decision to provide minimal instruction about how to do each task. The player gets feedback during play and can request additional help if stuck. This supports learning in connecting actions with solving goals.

1.2.3. Well-ordered Problems

Good games "teach" players to play by starting with simple challenges and then requiring players to use what was learned to solve increasingly complex problems. There are two negative outcomes associated with presenting players with overly complex tasks. Players may simply give up because it's too frustrating. Alternatively, because humans can be very clever in finding *some* solution to a problem, they may invent solutions which do not apply to related tasks — they don't transfer well [18,19].

AttK's levels are based on theoretical and empirical work on argumentation. [20,21,6, for example]. We have created levels for claim identification, predicate identification, and argument classification. We have plans to create two additional levels for argument evaluation and policy creation from arguments.

1.2.4. Cycles of Expertise; Skills as Strategies

As Gee summarizes from [22],

> Expertise is formed in any area by repeated cycles of learners practicing skills until they are nearly automatic, then having those skills fail in ways that cause the learners to have to think again and learn anew.[7, p. 10]

As mentioned above, in AttK, each new level brings a new challenge that builds on previously-learned skills. Students advance between levels when a certain level of proficiency is reached. Then they continue to practice those skills in the service of higher level goals. Practice helps the student automatize the new knowledge and feel pride in their growing expertise.

At higher levels in the game, students are using the more basic skills in varied ways, applying them to different tasks, learning when to focus on one skill or another [23]. As skills become automated, they serve as components in the higher level strategies that the students learn.

1.2.5. Fish Tanks; Sandboxes

As Gee defines the term, *fish tanks* are "stripped-down versions of the game,"[7, p. 12] where complexity is reduced, allowing the player to avoid getting lost and instead focus on acquiring important preliminary skills. *Sandboxes* may have the full complexity of the game, but major negative consequences like "dying" or losing are removed. Sandboxes give players free rein to explore the range of choices in a game environment without feeling pressure to perform optimally or choose too quickly.

The "full" AttK game is much less complex than most video games. It more closely resembles what Gee calls a fish tank. With the exception of the game story elements which are designed to motivate the student to learn the argumentation skills, the content of the game consists entirely of argumentative texts.

1.3. Understanding

1.3.1. System Thinking

Learning of isolated facts and knowledge is not useful learning. Learning is only useful if it comes with understanding — understanding of associations, applications, conditions, causes and effects [24]. Typical classroom worksheets have students practice skills in isolation. On the other hand, the real world is a complex system where actions are based on goals, and have meaningful consequences. Good games mirror the real world in this way. They present a complex system to the player. The player learns most effectively when she understands her role within the system and can use that knowledge to set goals and determine actions [7, p. 14]. Ironically, exogenous educational games do not provide this type of support.

The AttK story was created precisely to give students this type of system within which to learn and practice argumentation skills. While the system is not nearly as complex as in most commercial video games, it was designed to provide the player with the conceptual connections required for learning with understanding. The side tasks mentioned above supply an extra level of complexity, encouraging the player to understand that argumentation skills are not only applicable within the (simulated) job context.

1.3.2. Meaning as Action Image

As a semiotician, Gee's central concern is with meaning making. By linking meaning with action in this principle, Gee formulates his ideas about knowledge and learning in a way that fits very nicely with cognitive architectures like Soar and ACT-R [25,26]. In these architectures, thinking is all about selecting the next action which will make progress towards the agent's goal, and learning is all about making that selection process more efficient. Gee describes the same process in terms of the game player's experiences and of strengthening conceptual learning by linking perception and action.

As we have stressed throughout this paper, the AttK design is all about situating the learning and use of argumentation skills within a rich context that enables the player learn with deep understanding. AttK was designed to help the student learn not just the actions that are required, but also the perceptual conditions in which they apply.

2. Clark and Mayer's e-Learning and the Science of Instruction

In 2003, Ruth Clark and Richard Mayer, a professional training developer and an educational psychologist, published a collection of principles for creating effective e-learning systems. The principles were based on cognitive science theories and backed with empirical studies. In 2008, they released a second edition of the book [8] which included specific recommendations for educational games. Although they admit that significant research to fill out our understanding of what makes educational games effective, they offer several general principles which provide a useful alternative view on game design.

2.1. Match Game Types to Learning Goals

Clark and Mayer are not very specific about which types of games to use for which purposes. In fact, the only game type they mention is Jeopardy-style games which, they say (citing Van Eck [27]) are best suited for what Bloom classified simply as knowledge, e.g. facts, labels, and dates [23]. This principle does fit in well with the general comments made above about endogenous and exogenous games.

2.2. Make Learning Essential to Progress

This principle makes clear a fundamental difference between games and education. In most educational settings, a student is only required to repeat an assignment, class, or grade if that student's performance is dismal, and that repetition is seen as punitive by both teachers and students. In games, on the other hand, players are not allowed to advance to the next level unless they demonstrate a high level of performance, often by battling a level's boss. To defeat the boss, players must apply what they've learned in that level in novel ways (*a la* Gee's "Skills as Strategies"). AttK adheres to this principle by requiring the student to correctly process a high percentage of the texts that they receive in each level before they can "get promoted." If they fail, students receive instructional feedback and more practice items. Within the context of the game, this feedback is not face-threatening and should motivate students to learn to perform the skills correctly.

2.3. Build in Guidance

Clark and Mayer suggest a variety of ways to build guidance into an educational game. These are listed below, along with explanations if necessary, and examples from AttK.

Incorporate explanations In AttK, the boss character and a co-worker both can provide explanations to the player, before they take on a task and if they make incorrect judgments.

Encourage reflection Self-explanation is not currently part of the game design, but when a player fails to get a promotion within the game, she is implicitly encouraged to figure out why. Furthermore because the game is not timed, players can take a break from the game to reflect and plan new approaches.

Optimize interface fidelity Learning can be hindered if the game interface is too complex — too many distractions — or if it is so simple that players don't get the appropriate environmental cues. We have tried to include only the content materials and enough story to motivate the player. Play testing is required, however, to see if we achieved our goal.

Provide instructional support At the beginning of each level in AttK, the boss character demonstrates how the task is done. We do not currently provide memory support and visualization support as Clark and Mayer suggest, but these could be incorporated in future versions.

2.4. Promote Reflection on Correct Answers

This principle is nicely orthogonal to Gee's. It is based on empirical studies which showed that students who reflected on correct answers learned better than those who reflected on their own answers, some of which were wrong. It was not in the original design of AttK, but we have decided to include it, due to the cited research.

2.5. Manage Complexity

This principle is generally in line with Gee's "Well-Ordered Problems" principle, but Clark and Mayer provide some alternative perspectives.

Simple to complex goals This is closest to "Well-Ordered Problems."

Minimize interface complexity As mentioned above, too much complexity in the interface can hinder learning by providing distractions.

Training wheels This is similar to Gee's notions of Fish Tanks and Sandboxes

Use faded worked examples As mentioned above, we do provide a worked example at the beginning of each level. AttK does not include fading, however, because the different skills are practiced and perfected individually before they are combined into more complex skills.

Manage Pace Games which are too fast hinder learning by requiring "twitch speed" responses and leaving no time for reflection. Games which are too slow can reduce player motivation. ("This is so boring!") Because AttK is a turn-taking game, the pace is primarily determined by the player. Only play testing will show if we need to speed it up or slow it down in other ways.

3. Conclusions

This paper has analyzed the design of the AttK game using two different sets of principle. One set was intended to demonstrate how good commercial video games address an interesting paradox: although learning is hard work, and these games require lots of learning, players are highly motivated to continue playing them until they achieve mastery. The other set of principles was based on cognitive theories and empirical studies of learning. These principles are intended to provide a set of "best practices" to help create educational games which are effective for learning. Despite their different origins and intentions, the different frameworks largely agree. We feel that they validate many of the design choices that we made. There is still two huge questions that remain though:

- Will the game be fun?
- Will it help students learn?

These questions can only be answered empirically. We plan on addressing them in our future research this year.

References

[1] National Research Council. National science education standards. Technical report, National Academy Press, Washington, DC, 1996.

[2] D. Salahu-Din, H. Persky, and J Miller. National assessment of educational progress. the nation's report card: Writing 2007. Technical Report NCES 2008-468, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2008.

[3] E. Arnott, P. Hastings, and D. Allbritton. Research methods tutor: Evaluation of a dialogue-based tutoring system in the classroom. *Behavior Research Methods*, 40(3):694–672, 2008.

[4] A. Britt, P. Wiemer-Hastings, A. Larson, and C. Perfetti. Using intelligent feedback to improve sourcing and integration in students' essays. *International Journal of Artificial Intelligence in Education*, 14:359–374, 2004.

[5] M. A. Britt and A. Larson. Construction of argument representations during on-line reading. *Journal of Memory and Language*, 48(4):749–810, 2003.

[6] M. A. Britt, C. Kurby, S. Dandotkar, and C. Wolfe. I agreed with what? Memory for simple argument claims. *Discourse Processes*, 45(1):52–84, 2008.

[7] J. Gee. Learning by design: Good video games as learning machines. *e-Learning*, 2005.

[8] R. Clark and R. Mayer. *e-Learning and the Science of Instruction*, chapter Simulations and Games in e-Learning. Pfeiffer, 2008.

[9] J. Gee. *What video games have to teach us about learning and literacy.* Palgrave Macmillan, 2003.

[10] C. Heeter, C. Chu, A. Maniar, B. Winn, P. Mishra, R. Egidio, and L. Portwood-Stacer. Comparing 14 plus 2 forms of fun (and learning and gender issues) in commercial versus educational space exploration digital games. Presented at the International Digital Games Research conference, 4-6 November 2003, University of Utrecht, The Netherlands, accessed from http://gel.msu.edu/carrie/publications/forms_of_fun.pdf on May 1, 2009, 2003.

[11] H. Gardner. *Intelligence reframed: Multiple intelligences for the 21st century.* Basic Books, New York, 1999.

[12] T. Malone and M. Lepper. Making learning fun: A taxonomy of intrinsic motivations for learning. In R. Snow and M. Farr, editors, *Aptitude Learning and instruction*, volume 3, pages 223–253. Erlbaum, Hillsdale, NJ, 1986.

[13] M. Habgood, S. Ainsworth, and S. Benford. Endogenous fantasy and learning in digital games. *Simulation and Gaming*, 36(4):483–498, 2005.

[14] B. Winn. The design, play, and experience framework. In *Handbook of Research on Effective Electronic Gaming in Education*, volume III, pages 1010–1024. Information Science Reference, 2009.

[15] M. Prensky. *Digital game-based learning.* McGraw-Hill, 2001.

[16] L.S. Vygotsky. *Mind in society.* Harvard University Press, Cambridge, MA, 1978.

[17] A. Koriat and R. Bjork. Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory and Cognition*, 34(5):959–972, 2006.

[18] R. Koster. *A Theory of Fun*. Paraglyph Press, Scottsdale, AZ, 2005.
[19] A. Luchins and E. Luchins. Einstellung effects. *Science*, 238(4827):598, 1987.
[20] S. E. Toulmin. *The Uses of Argument*. Cambridge University Press, Cambridge, MA, 1958.
[21] M. A. Britt, P. Wiemer-Hastings, A. Larson, and C. Perfetti. Using intelligent feedback to improve sourcing and integration in students' essays. *International Journal of Artificial Intelligence in Education*, 14:359–374, 2004.
[22] C. Bereiter and M. Scardamalia. *Surpassing Ourselves: an inquiry into the nature and implications of expertise*. Open Court, Chicago, 1993.
[23] B.S. Bloom, editor. *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain*. Longmans, Green, New York, Toronto, 1956.
[24] J. Bransford, A. Brown, and R. Cocking, editors. *How People Learn: Brain, Mind, Experience, and School*. National Academic Press, expanded edition, 2000.
[25] J. R. Anderson. *Rules of the Mind*. Erlbaum, Hillsdale, NJ, 1993.
[26] A. Newell. *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA, 1990.
[27] R. Van Eck. Digital game-based learning. *Educause*, 41(2):17–30, 2006.

Teachable Agents that Learn by Observing Game Playing Behavior

Lena PARETO[1]
[a]*Media production and Informatics Departments, University West, Sweden*

Abstract. An educational game in mathematics targets cognitive skills oincluding pattern finding and generalization from example by teaching an agent. These can be difficult for young children, let alone teaching about them. Therefore, we have designed a new way of teaching an agent, the teach-by-guiding model. We describe the model and its basic implementation. Because this is a work-in-progress, we end by raising questions of how such an agent is perceived, and in what respect it may differ from other pedagogical agents.

Keywords. Teachable agent, Educational game, CSCL, Intelligent tutor

Introduction

Computer games as learning tools have considerable potential for contextual learning and boosting affect and motivation [7]. Yet, not all computer games produce the same effects, and findings of effects are often contradictory or promotional [24]. Creating an effective educational game entails more than creating (or using) an engaging game and building in age-appropriate educational content [6],[14]. Educational content must be represented in appropriate form; content must be closely intertwined with game play; and, feedback and hint structures must be provided to scaffold children into challenging content [6]. According to [9], students may be successful game players by learning superficial heuristics rather than reasoning about the underlying domain knowledge. In particular young children, who can be so engaged in their own problem solving and game playing that they do not reflect on what they are doing. As a result, the transfer of knowledge outside the game environment is not apparent. The enthusiasm for educational games is high, but it is not evident how to design game environments so they foster deep domain understanding particularly of difficult symbolic content [12].

Providing intelligent computer based support to educational games is both valuable and challenging [9]. There is evidence that even simple educational games can be highly motivating, but little evidence of their learning effect unless they are integrated with supporting activities. The role of guidance and reflection for promoting understanding in agent-based games was examined in [12], and the conclusion was that inexperienced learners often need structural guidance in combination with reflection techniques, to help them achieve deep understanding. Providing this support is challenging because it requires careful trade-offs between fostering learning and maintaining positive engagement [9]. Often, the type of knowledge gained in game

[1] Corresponding Author.

playing is tacit, contextually bound knowledge (especially for younger children), so it is important to find ways to foster ways to develop articulated, explicit knowledge.

Teachable Agents (TAs) [1][2][19], are pedagogical computer agents that build on the wisdom that people learn by teaching. In contrast to most instructional technology where the computer teaches the students, with TAs students teach the agent. Artificial intelligence techniques guide the agent's behavior based on what it was taught and students can revise their agents' knowledge (and their own) based on such behavior. For a teachable agent, directly teaching the agent in a pre-defined format is the most common model [1][2][5][20]. For example in [1], students teach the agent Betty by explicitly creating a network of entities and their relations, like a concept map, and then they can ask her questions to check her knowledge. This model of teaching presumes that the teacher is able to articulate and model knowledge in the required format.

We have designed teachable agents for an educational game, and the idea is that children can teach an agent to play, and then let the agent play the game instead of playing themselves. The game targets conceptual understanding in arithmetic through a game model where the semantics of the content is built into the game – discovering the rules of the game is the same as learning the mathematics. Observation indicates that children become strategic using patterns in the game; children as young as 6 years old can play the game well. Yet, they often have great difficulty explaining why they play as they do. So identifying patterns (playing well), and being able to explain the game playing behavior, are two different abilities. There is a gap between tacit game playing knowledge and the ability to articulate formal knowledge. Teachable agents may bridge this gap, but it requires a model of teaching that does not depend on spontaneous, explicit formulation of strategies. This leads us to our research question: *How can a teachable agent learn by other means than direct instruction?*

This research is work-in-progress, and is part of an ongoing project. The aim is to provide a model of an engaging educational game that not only is fun to play but which have proven to be effective for learning conceptual understanding and general cognitive skills such as strategic thinking and reflection (in our case for arithmetic). We are currently conveying learning and motivational effect studies in two countries, in five different schools and with 9 classes ranging from 1^{st} to 7^{th} grade. Of today, over 2000 games are played and 300 agents are taught. The paper describes preliminary findings of the teachable agent implemented in this game.

The paper is organized as follows: we start by describing our educational game and shortly review teachable agents before we go into our version of teaching the agent: the teach-by-guiding model. Then we turn to more technical section, where the implementation of the agent is described. We summarize by a few conclusions and many intriguing questions related to this type of teachable agent.

1. The Educational Game – Graphical Arithmetic Model (GAM)

Our educational game [16] is based on a metaphor for arithmetic with numbers as graphical objects and arithmetic operations as animated actions on these objects. Digits are represented as a group of one-colored squares, and a number as a sequence of such digit-groups. For example, addition is represented by adding squares onto the game board (the middle panel in figure 1), and subtraction by removing squares from the board. The model represents all arithmetic operations, positive and negative numbers. Rules and properties are built into the model, insuring valid mathematical computations.

This means for instance that if the right-most blue unit on the game board already contains red 9 squares, then it is full. If a 10^{th} square is added it will be placed in the arrow above the game board, the packing arrow, and all 10 squares are packed into an orange 10-box, which falls down into the compartment to the left (the tens). This way, the model ensures that only valid numbers can be put on the board game. Moreover, children can discover that any compartment (as well as any 10-base unit) obeys such behavior. Thus, the purpose of the model is *to discover and explore* computational behavior of arithmetic, rather than to perform computations which is done automatically. Concepts, structure and behavioral properties are the aimed for learning content.

To motivate usage and exploration of the model, it is used in several 2-player board games, intertwining game play with learning content [6]. Each player acts an operation and receives a hand of number cards visible to both players. The game proceeds by taking turns laying a card on the common game board. In figure 1, both players act addition and the player to the left has added a card 44, yielding the result 0+44=44 on the game board (4 orange tens, 4 red ones). The player to the right has the choice of cards 84, 56, 35, and 78. In this game, the goal is to get as many carry-overs (packings) as possible. The choice 44+84 will yield one carry-over (the tens), 56 and 78 two carry-overs (ones and tens) and 35 no carry-over. The strategy comes from choosing the best card, which in this case is 44+56=100 since it gives maximum points (2 carry-overs) and also leaves the opponent with the worse possible situation (no 2-digit number will yield carry-over when added to 100). Feedback is given immediately on successful choices, by receiving stars (points).

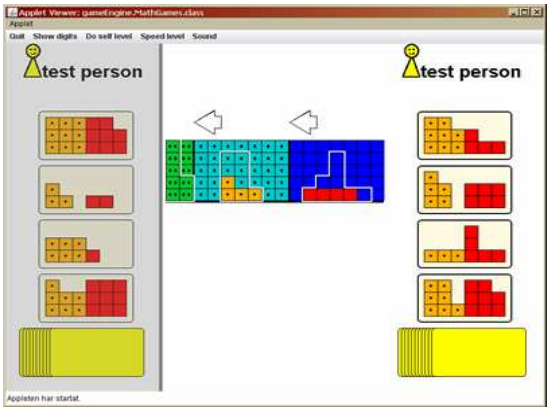


Figure 1 – Game play example of addition game

The overall purpose of game play is to acquire understanding of the graphical model behavior, and the games are designed to be playable and practice aspects of arithmetic. For instance, in the example above, the idea is that by giving points for carry-overs, this should attract attention to the fact that when a unit exceeds 9 in the base-10 system, something happens. Another game with positive and negative numbers has as goal to get as many zeros as possible in the result. The purpose of this is to attract attention to that positive and negative squares “outnumber” each other, and that a department can be emptied by making squares add up to exactly 10 as well.

Strategic game play – inventing effective computation strategies

To think strategically and to invent computational strategies are part of playing the game well, and is good for learning. Most mathematics instruction tells students a symbolic procedure, and then they practice repeatedly on a set of problems In multiple studies [3][22][23] that compared tell-and-practice instruction with a pedagogy of student invention, students who invented their own solutions and then heard about standard methods learned in a deeper and more lasting way. Also, they were able to transfer their understanding and learnt future concepts and methods better. In GAM, students always work to invent their own strategies for choosing cards.

Even for the simpler games as the one in the example above, strategic thinking is present. In order to play well the player needs to choose a good, preferable the best possible, card. To judge the goodness of a card, the player needs to understand how the model behaves, and beforehand estimate or calculate the effect of choosing a card (since once a card is chosen, the operation is performed by the system). Numerous observations show that children become good (strategic) game players quickly, and are able to see and identify behavioral patterns of the model, such as “units can only contain 9 objects and yield a carry-over if exceeded” or “adding positive numbers have the same effect as removing negative numbers”.

The player can and will sometimes choose based on guessing, but to perform well they need to count and understand the basic structures and rules of numbers and operations. The success rate of each turn (i.e. the chance of getting points) is rather high, ranging from about 50% (easier games) to 15% for more difficult ones, to maintain interest and encourage exploration of the model’s behavior. Games can be played competitively or cooperatively. The same kind of thinking is involved: a player should either maximize possibilities for the co-player to score, or minimize the chances for the opponent, respectively.

Even though the graphical model can help children’s intuition for arithmetic, they need eventually to transfer this knowledge to the ordinary world of mathematics with numbers and symbols. To enhance the game playing further, and to support transfer from tacit game playing knowledge to articulated knowledge, we have implemented teachable agents (TA) in the game. Before describing how TAs work in our game, we will review the concept of teachable agents.

2. Teachable Agents

Teachable Agents (TAs) [1][2][19], are pedagogical computer agents that capitalize on the wisdom that people learn by teaching. In contrast to most instructional technology where the agent teaches students, with TAs students teach agents. Artificial intelligence techniques guide the agent’s behavior based on what it was taught and students can revise their agents’ knowledge (and their own) based on such behavior.

Teaching typically involves three phases: To prepare, to teach, and to observe the effect of teaching (often by judging/assessing the learner) [14]. The preparation phase involves reflecting of one’s own knowledge and articulating such knowledge in an explanatory form (i.e. self explanation). Self explanation has gained much interest as a meta cognitive skill that results in improved problem solving and better learning [11]. By explaining to themselves (or others), students integrate new knowledge with existing knowledge. Although many students do not spontaneously self-explain, they do so when prompted and can learn to do it effectively [11]. Thus, as a teacher to the agent, the student is encouraged to reflect and self-explain their own knowledge.

The effect phase involves observation and judgment of the effort, which for game playing means watching and judging the agent play the game. Observing performance actively gives extremely valuable feedback and improves learning compared to just “doing it oneself” [14]. Attachment to the agent comes from being its teacher.

Students readily adopt the fiction of teaching an agent. They find it motivating, and it helps them to organize otherwise complex human-computer interactions and learning tasks. Benefits of using a teachable agent include 1) making (a mirror of) the children's own understanding explicit and observable, 2) creating an affective relation which motivates engagement of the agent's performance, and 3) detachment of performance to self, i.e., that the agent performs and not the child, which can effect prior negative attitudes, and be easier to handle and discuss.

According to [1], the four most important principles of a teachable agent are to 1) use explicit and well-structured visual representations, 2) enable the agent to take independent actions, 3) model productive learner behaviors, and 4) include environments that support teaching interactions. To model learner behaviour can for example mean that the agent can occasionally start to check her own knowledge, and make remarks (right or wrong) that her derived answer does not seem to make sense. The independent performance of TAs offers excellent assessment opportunities. A hidden expert map can for example determine the correct answer, and other agents' answers can thus be checked. This means that agents can take a test and the test be scored automatically. Since the agent's performance on tests are highly correlated with students' own performance on paper and pencil tests [20], the agent can take the test *instead of* the child. Moreover, it is possible to automatically score student homework.

3. Teach-by-guiding model

We have clear indications from earlier studies of the game [16] that children quickly learn to use their game knowledge effectively (i.e., perform well in the game), but how should children teach the agent this knowledge?

We have chosen to try a novel teaching model for the TA, resembling the way masters teach apprentices in craft professions. This means that the child can teach the agent in two ways; 1) by showing how to play (using their own tacit or aware game playing knowledge), and 2) by letting the agent try to play and give corrective feedback, i.e., accepting or rejecting the choice of the agent. If the child rejects the agent's choice, she needs to provide an alternative and thereby showing the agent a more appropriate action

This way the child can, so to speak, teach her tacit game playing knowledge to the agent. Since we have judged self-articulated explanations of general game patterns to be too difficult for our age group, we have included explanations in a different way, to benefit from the reflective injection of explanations. According to [12], students achieve better transfer scores; give better explanations and more correct answers when provided by explanatory feedback compared to corrective feedback, so explanations are important. When the child shows the agent which choice she would make, the agent reacts by posing questions related to that particular choice and its consequences. The questions are multiple-choice explanations of the current situation (see figure 2), false, irrelevant or true. It is easier to recognize an explanation than to articulate it, and

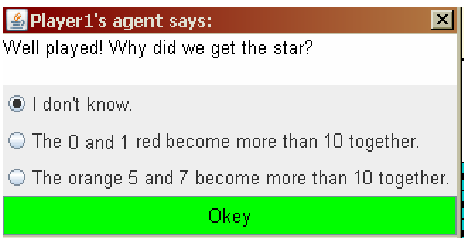


Figure 2 - The agent prompts the child with explanatory choices (stars refer to points)

children are encouraged to select the correct answer since the agent's learning depend on their answer. Our TA combines the reflection and guidance into an explanatory inquiry, i.e., the learner is asked to select correct explanation of their own actions.

The TA is trying to mirror the child's knowledge, but can only do so by observing her game play and analyzing the explanatory inquiry. The TA estimates the likelihood of the action being a conscious (explicit, aware knowledge) or unconscious (implicit, tacit knowledge) act of successful behavior. We will never know if the choice is a guess, but over time the likelihood decreases that successful choices are made by chance. Correct explanations indicate conscious and appropriate knowledge. The child's accumulative behavior is analyzed and reflected in a barometer (for each basic game rule) of the estimated level of her current knowledge. Successful choices give positive indications whereas missed, better alternatives give negative indications. This idea relates to the "guess-parameter" in [9], except that they only judge potential mistakes of the user. Here, we try to estimate the awareness of performed behavior.

When the child teaches in try-and-correct mood, the agent uses its current knowledge, which means that if the agent has reached a threshold of "knowing a rule well enough", then the rule become "visible" to the agent (i.e., it can be applied). We have set several threshold levels which monitor the agent behavior (for instance, the agent asks more elaborative questions when the level is higher). If the child accepts the agent's choice, its corresponding knowledge is reinforced. If the child rejects it and chooses an alternative, the agent asks the child to explain why (see figure 3). The agent's learning depend on the corrected choice and its relation to the rejected choice, as well as on the explanation selected by the child.

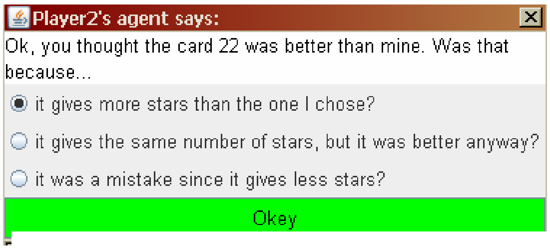


Figure 3 - Agent asks for explanation of its rejected choice (stars refer to points in the game)

Pedagogical Rationale

There is little agreement among researchers if guided or unguided exploration is the best for learning [13][10][8]. The role of guidance and reflection for promoting understanding in agent-based games was examined in [12], and the conclusion was that inexperienced learners often need structural guidance in combination with reflection techniques, to help them achieve deep understanding. Providing this support is challenging because it requires careful trade-offs between fostering learning and maintaining positive engagement [9]. By embedding adaptive, informative feedback and providing collaborative learning situations for discussions between peers, an educational game's effect may be enhanced [7]. Our game provides both: guided exploration with corrective and explanatory feedback by means of the TA, and unguided exploration and corrective feedback (received stars) during ordinary game-play.

4. Implementation of the Teach-by-Guiding Agent

A teachable agent needs to be able to do three things: to represent knowledge, to learn, and to act according to its knowledge. The knowledge we are trying to convey in our game is not arithmetic calculations (i.e., $a+b=c$), but structural pattern rules of arithmetic concepts (the base-10 system, for instance) and strategic rules of how to play the game well. The pattern rules correspond to patterns that are more or less easy to identify in the graphical model, but since it is isomorphic to arithmetic behavior, this is also rules of arithmetic (but not formalized in any traditional way). For example, recall the previously described game where the goal is to get as many carry-overs as possible. Most children rather quickly identify the pattern that if the red squares on the game board and the red squares on a card together become 10 or more, then these will be packed and yield a point. The same pattern applies to orange 10-boxes. However, there is an additional pattern for this game, which also generates a point: if the red squares carry-over to an orange box and the orange boxes are exactly 9, then they will get two carry-overs – one for the red and one for the orange squares. Note that this pattern involves both red and orange squares, corresponding to the fact that units affect each other under arithmetic operations (i.e., carrying for addition or borrowing for subtraction algorithms). Game knowledge is therefore represented as a set of production rules, where each rule will generate a point if it matches. For our simple game example, the three production rules generating points look like this:

rule*(addition, red squares, ≥10)*
rule*(addition, orange squares, ≥10)*
rule*(addition, orange squares, one carry-over from red squares, =9)*

To generalize the idea that a unit may be affected by any adjacent unit, we have defined the notion of *effects.* Any transaction between two units result in one incoming effect and one outgoing effect, for example a carry-over yields a *take-ten effect* for the lower unit and an *add-one effect* for the higher. Correspondingly, a borrowing yields a *take-one effect* from the higher unit and an *adds-ten effect* to the lower. This way, we can define pattern production rules for all games in the form:

rule*(operation ,unit, incomingEffect, goalOfUnit)*

For each choice the child or the agent makes, the corresponding pattern rules of the game is matched against the particular situation (maximum one match per unit). So, for instance, adding *47* to *55,* will match the first rule for the ones ($5+7\geq 10$)*,* and yield a carry-over, and the tens ($4+5=9$) thus match the second rule. With this generalization, the rules look exactly the same for all units except for the ones, which is one of the key insights of the base-10 system (i.e., that the hundred-unit behaves just like the ten-unit, and so forth). Such transfer of knowledge from one unit to another is something we have observed in children's game playing behavior.

Beside the production rules, there are strategic rules related to each game. These depend on the particular situation: for instance how good a card is related to the alternatives. A basic game playing strategy is to maximize your own points in a turn, choosing the best among the alternatives that generate the maximum. What is best vary in different games and depends on if the players collaborate or compete. In our example game, a good choice is a card that leaves as few squares in the result as possible, or as a more advanced strategy actually looks at the opponents/collaborators exact choices. Currently, only the simpler variant is implemented as an evaluation function on the result, but we have already seen that we need to make the agent able to learn more advanced strategies. We have heard children formulate conscious strategies, such as "*if the game board is empty you can never get any points, so it is better not to start this game*", and they expect the agent to be able to reason accordingly.

The way the agent learns is by observing the child's behavior during teaching, and this behavior is represented as a *trace of cumulative behavior.* In each situation where a pattern rule occur (both in the chosen card and in the alternatives) it leaves a trace. These traces are generalized descriptions of the situation, such as *Chosen, Chosen correct reason,* or *Chosen wrong reason*. These traces correspond to when a child shows the agent how to play, and the two latter refer to a situation when the agent prompted the child to explain and the child selected a correct or wrong reason, respectively. The trace *Chosen* is used when the agent has no information of the reason behind the choice (i.e., no question asked or the "don't know" alternative chosen). For the other teaching mood, when the agent tries to play and the child gives corrective feedback, the situation becomes more complicated, especially when the child rejects the agent's card and chooses an alternative. In this situation, the agent learns from both cards (the dismissed and the corrected to cards) as well as their relation to each other. The corresponding traces are *Dismissed for better* or *Dismissed for worse* for the dismissed card when the alternative is better or worse, respectively. Similarly, the traces *Corrected to is better* or *Corrected to is worse* reflect a situation when the corrected card is better or worse, respectively. Cards neither chosen by the agent nor the child may cause the trace *Missed in better* when a rule was missed in a better choice, and *Occurred in worse* for rules that occur in worse cards. This way, all production rules are associated with a trace list of cumulative behavior:

rule*(operation ,unit, incomingEffect, goalOfUnit)* $\rightarrow [trace_1,\ldots,trace_n]$

Each trace gives raise to either a positive or a negative indication, i.e., an estimation of whether such behavior indicates that the child knows the rule or not.

Table 1 – Some of the rule traces and their indicators

Rule Trace	Indicate	Used in	Description of pattern rule counter
Chosen	++	all	
Chosen correct reason	+++	Agent show	Child's choice. Verified with correct explanation
Chosen wrong reason	+	Agent show	Child's choice. Selected incorrect explanation
Acknowledged	+	Agent try	Agent's choice. Acknowledged by child.
Dismissed for better	++	Agent try	Agent's choice. Dismissed for a better choice of child
Dismissed for worse	–	Agent try	Agent's choice. Dismissed for a worse choice of child
Corrected to is better	++	Agent try	Child's corrected choice. Better than dismissed
Corrected to is worse	–	Agent try	Child's corrected choice. Worse than dismissed

The indicators in the table denote strong, medium or weak positive indication (+++,++, or +) and medium or weak negative indication, respectively.

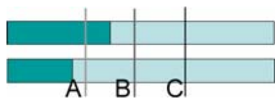


Figure 4 - rule barometer

The trace lists are used to estimate a knowledge level for each rule, like a barometer going up and down. Figure 2 illustrate the barometer for two different rules where A, B, and C denote three barriers to reach, for example A=aware of rule, B=knows rule, and C=knows rule very well. The first barometer has passed the "aware of rule"-level, and the agent can take actions accordingly (e.g., starting to ask questions about the rule), whereas the bottom barometer has not reached any level yet and the rule remains "invisible" to the agent.

This representation allows for a flexible mapping of a knowledge level estimation algorithm, since both indications on traces and algorithms to compute a level from the traces can be altered. Moreover, it is possible to assign different behavioral properties to an agent, for instance a fast "error-prone" learner or a more conservative, slower learner.

5. Conclusion

We could answer our question, *how can a teachable agent learn by other means than direct instruction*, by suggesting the apprentice-agent that learn from observing the game play and receiving corrective feedback from the game player, i.e., the teach-by-guiding model. However, this is quiet a shallow answer, since many questions arise from this model of teaching, which we will discuss below.

6. Discussion

We have already gained much insight of additional features desired by the students, such as more direct feedback of agent's current knowledge levels (and good suggestions how to do this!), or features that need to be developed further such as more advanced strategic rules in the agents knowledge system (they want the agent to be sensitive to learn the same strategies they have discovered themselves, and do get a bit frustrated if the agent is not "smart enough" to follow their own level of strategic play.

A question that we have discussed is how to make the agents' knowledge system transparent to the players in a good way. Recall the principles of teachable agents, the first being "use explicit and well-structured visual representations" for the knowledge representation, which is not the case here. And it is not apparent how to show the rule-system and the barometer in a comprehensible way. We even got the very insightful question from 5th grades students: "*Do agents learn better from observing my game play and ask questions or from trying to play and get corrective feedback*"?

The teach-by-guiding model also raises interesting questions of how students perceive the agents' behavior, well illustrated by a 3rd grade student's comment during the first session with the agent: "*I think the agent just pretends not to know*". The role of the agent has also been discussed among the older students with math learning problems. And this is a relevant question, because who is teaching who? Is the agent an apprentice, a mentor, or a coach? Or all? With our approach, the clear teacher-student role of most other TAs becomes fuzzy. The agent can be made to be more or less suggestive, changing the balance between who is actually teaching who. In a sense our agent behaves more like a supervisor that "teach" by hiding her knowledge and injecting questions to invoke reflection and push the learner to higher insights. Will this make the children feel deceived when they are told that they should teach the agent?

The last question concern the very attractive idea of automatic scoring, and assessing children by assessing their agent instead. When the child is not explicitly articulating their knowledge in direct instruction, but the agent actually has "complete" knowledge that it successively uncovered when the estimation algorithm judge the child's behavior, it seems more risky to use this information since we only interpret an observed behavior, which may be over interpreted since we know what we want to see.

7. Acknowledgements

This work is currently financed by Wallenberg Global Learning Network. I would also like to thank my collaborating partners in the project for fruitful collaboration and for sharing the same interests.

References

[1] K Blair, DL Schwartz, G Biswas, K Leelawong. Pedagogical Agents for Learning by Teaching: Teachable Agents. *Educational Technology*, Saddle Brook Then Englewood Cliffs NJ, (2007)

[2] G. Biswas, T. Katzlberger, J. Brandford, D. Schwartz & TAG-V. Extending intelligent learning environments with teachable agents to enhance learning. In *Artificial Intelligence in Education*, J.D. Moore et al. (Eds.) IOS Press, (2001). 389–397.

[3] J. D. Bransford. & D. L. Schwartz (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad and P. D. Pearson (Eds.), Review of Research in Education, 24, 61-100. Washington, D.C.: American Educational Research Association.

[4] R. Case, Y. Okamoto, S. Griffin, A. McKeough, C. Bleiker, B. Henderson,, K. M. Stephenson, R. S. Siegler & D. P. Keating, (1997). The role of central conceptual structures in the development of children's thought. Monographs of the Society for Research in Child Development, 61, 1-295.

[5] J. Davis, K. Leelawong, K. Belynne, G. Biswas, N. Vye, R. Bodenheimer and J. Bransford. (2003). Intelligent user interface design for teachable agent systems. In *Proceedings of the 8th international Conference on intelligent User interfaces* (Miami, Florida, USA, January 12 - 15, 2003)

[6] S. M. Fisch. Making educational computer games "educational". In *Proceedings of the 2005 Conference on interaction Design and Children* (Colorado 2005). IDC '05. ACM, New York, NY, 56-61

[7] F. Ke. Alternative goal structures for computer game-based learning. *International Journal of Computer-Supported Collaborative Learning* (2008) 3:429–445

[8] P. A Kirschner, J. Sweller & R. E Clark,. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75–86.

[9] M Manske & C. Conati. Modelling Learning in Educational Games. *Proceedings of the International Conference on Artificial Intelligence in Education*, (2005). 411-418.

[10] R. E. Mayer (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. Educational Psychologist, 59, 14–19.

[11] A. Mitrovic. The Effect of Explaining on Learning: a Case Study with a Data Normalization Tutor. *Proceedings of the International Conference on Artificial Intelligence in Education*, (2005), 499-506

[12] R. Moreno and R. E. Mayer. Role of guidance, reflection and interactivity in an agent-based multimedia game. Journal of Educational Psychology, (2005). 97(1), 117-128.

[13] B. Nelson and D. Ketelhut, Exploring embedded guidance and self-efficacy in educational multi-user virtual environments. *Int Journal of CSCL,* Vol. 3, No. 4. (2008), pp. 413-427.

[14] S.Y Okita, and D. L Schwartz. When observation beats doing: learning by teaching. In *Proceedings of the 7th international Conference on Learning Sciences*, Indiana, (2006), pp 509-515.

[15] L. Pareto (2007), Utility Games – gaming as a design strategy to achieve utility effects. In proceedings of LearnIT Conference Game In Action, Gothenburg University 13-15 June 2007.

[16] L. Pareto, D.L. Schwartz & L. Svensson (2009). Learning by Guiding a Teachable Agent to Play an Educational Game. To appear in Proceedings of the International Conference on Artificial Intelligence in Education, (2009).

[17] L. Pareto. The Squares Family: A Game and Story based Microworld for Understanding Arithmetic Concepts designed to attract girls. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Vol. 2004, Issue 1, (2004), pp. 1567-1574.

[18] C. P Rosé, D. Bhembe, S. Siler, R. Srivastava, and K. Van-Lehn. Exploring the Effectiveness of Knowledge Construction Dialogues. In *Proceedings of AI in Education* (2003).

[19] D. L. Schwartz, K.P. Blair, G. Biswas, K. Leelawong & J Davis. Animations of thought: Interactivity in the teachable agents paradigm. In R. Lowe & W. Schnotz (Eds). Learning with Animation: Research and Implications for Design. (2007). UK: Cambridge University Press

[20] D. L Schwartz, C. Chase, J. Wagster, S. Okita, R. Roscoe, D. Chin & G. Biswas. (in press). Interactive metacognition: Monitoring and regulating a teachable agent. In D. J. Hacker, J. Dunlosky, and A. C. Graesser (Eds.), Handbook of Metacognition in Education.

[21] G. Stahl. (2002) "Contributions to a theoretical framework for CSCL" in G. Stahl (Ed.), (2002) Computer support for collaborative learning: foundations for a CSCL community, (Cscl 2002 Proceedings), Mahwah, NJ: Lawrence Erlbaum Associates.

[22] D. L Schwartz,. & J. D Bransford,. (1998). A time for telling. Cognition & Instruction, 16, 475-522.

[23] D. L Schwartz & T. Martin (2004). Inventing to prepare for learning: The hidden efficiency of original student production in statistics instruction. Cognition & Instruction, 22, 129-184

[24] J. F. Vogel, D. S. Vogel, J. Cannon-Bowers, C.A. Bowers, K. Muse & M. Wright. Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3) (2006), 229–243.

UrbanSim: A Game-based Simulation for Counterinsurgency and Stability-focused Operations

Ryan MCALINDEN, Andrew S. GORDON,
H. Chad LANE[1] and David PYNADATH
Institute for Creative Technologies, University of Southern California, USA

Abstract. The UrbanSim Learning Package is a simulation-based training application designed for the U.S. Army to develop commanders' skills for conducting counterinsurgency operations. UrbanSim incorporates multiple artificial intelligence (AI) technologies in order to provide an effective training experience, three of which are described in this paper. First, UrbanSim simulates the mental attitudes and actions of groups and individuals in an urban environment using the PsychSim reasoning engine. Second, UrbanSim interjects narrative elements into the training experience using a case-based story engine, driven by non-fiction stories told by experienced commanders. Third, UrbanSim provides intelligent tutoring using a simulation-based method for eliciting and evaluating learner decisions. UrbanSim represents a confluence of AI techniques that seek to bridge the gap between basic research and deployed AI systems.

Keywords. simulation-based training, serious games, social simulation, narrative-based learning environments, intelligent tutoring

Introduction

The last decade has seen enormous changes in the way that simulation technology is being used in military training applications. Fuelled by advances in the computer game industry, contemporary military training simulations are the product of equal parts pedagogical design and computer game technologies. The adoption of computer game technologies, in particular, has opened up new opportunities for the innovative application of artificial intelligence (AI) research, but often not in the most obvious ways. While there continues to be enormous interest and progress in the creation of autonomous virtual characters for virtual training environments [7], many of the recent success stories in the application of AI technologies for military training involve "under-the-hood" software components that, on their own, represent only a small portion of current AI research. Still, progress over the last decade has led us further than the simple path-planning AI of the early game-based military simulations that were deployed (e.g. [11]). For example, the recently deployed *Tactical Iraqi* training application [6] incorporates sophisticated speech recognition techniques to allow users to converse with virtual characters in a foreign environment, while the recently

[1] Corresponding Author: H. Chad Lane, USC Institute for Creative Technologies, 13274 Fiji Way, Marina del Rey, CA 90292 USA, email: lane@ict.usc.edu

deployed *BiLat* training application [9] integrates a generative character animation model to allow trainees to better learn about intercultural communication in the context of bilateral negotiations.

A common factor in each of these previous cases was that the AI technology was well suited to support the specific pedagogical goals of the training application, one of several practical considerations to be made when introducing technology into military training environments [4]. Although these considerations impose certain constraints on the appropriateness of certain AI technologies, the design space of simulation technologies that successfully incorporate AI remains large and largely unexplored. In order to make progress on the innovative application of AI for training purposes, novel combinations of learning objectives, simulation environments, and AI technologies should be explored and evaluated.

In this paper, we explore a research effort that combines learning and AI to support the U.S. Army in developing the skills of military commanders in the conduct of counterinsurgency operations. We describe the *UrbanSim Learning Package*, a practice environment for counterinsurgency operations based roughly on the design of commercial games that allow players to manage cities (e.g. Electronic Arts' *SimCity*). UrbanSim incorporates several "under-the-hood" AI technologies, three of which are described below after a brief overview of this training application.

1. UrbanSim Overview

Today's military leaders face extremely stressful and demanding situations that are, in many cases, not covered by standard tactics and doctrine. These operations, which combine both lethal and non-lethal aspects of warfare, have been referred to as "armed social work," in which military forces attempt to "redress basic social and political problems while being shot at" [8]. The overarching challenge is to develop leaders who possess adaptive expertise and function effectively in complex environments. The goal is to prepare leaders for novel situations unlike any they may have experienced in the past [1,17]. To address this challenge, we developed an instructional software suite for military commanders and their staffs to practice directing and coordinating operations with a "stability-focused" component. The *UrbanSim Learning Package* (or *UrbanSim* for short) focuses predominantly, but not exclusively, on military operations in support of the local citizenry and government that take place after primary offensive and defensive efforts have concluded. UrbanSim has adopted a formal instructional design approach to guide the development of the underlying simulation-based experience. This design approach has proven successful in the development of similar pedagogical training aids, such as the BiLAT system [9], a precursor to UrbanSim that focuses on bilateral negotiations to improve cultural awareness. Building on the lessons learned from BiLAT, and applying the principles from Guided Experiential Learning (GEL) [2], UrbanSim was designed, developed, and deployed with a strong pedagogical focus. However, the resulting learning objectives called for a complex, dynamic, yet highly realistic simulated environment. This brought about the need to employ agent-based research technologies and transition them to software that would eventually be used in the classroom.

Consisting of two separate but tightly interwoven applications, UrbanSim adheres to the GEL model by providing learners with a complete cognitive foundation required to conduct these types of complex, dynamic operations. The first application, the

UrbanSim Primer, provides the requisite conceptual and task knowledge required for the learner to lead a full-scale stability operation, from analyzing background information to coordinating the actions that are carried out in support of reaching a desired end state. Taking the form of an interactive tutorial, the UrbanSim Primer is broken into eight lessons, each of which contain a narrative, interview segments from former Commanders, and assorted practice exercises as a means of demonstrating specific tasks to the learner. Taking approximately one to two hours to complete, the self-paced Primer prepares the learner for the second application, the more complex UrbanSim Practice Environment.

The UrbanSim Practice Environment is a game-based social simulation that allows a learner to plan, prepare, execute, and assess a full stability operation. Similar to the mechanisms employed for a turn-based strategy game (such as *Civilization* or *Age of Empires*), the learner directs subordinate units to take action with and against agents in a virtual environment (shown in Figure 1), and attempts to successfully complete a mission using the products/strategies learned in the UrbanSim Primer. These actions are taken against key individuals, groups, and structures in an area of operation in the attempt to reach a desired end state, which is often associated with support of the local populace for the local government. Each turn-cycle in the game represents one day in simulation time, though actions can take multiple turns (i.e., days), and can be interrupted if conditions in the world do not allow the action to complete (e.g., money runs out to construct a school). Upon completion of a scenario, the learner is brought to a debrief phase where a summary of the mission is presented for the learner to evaluate their progress.

The UrbanSim Practice Environment is unique in that it incorporates several AI technologies in order to support the training objectives of the application. The AI technologies that were selected each address specific challenges in simulating the complexities of urban environments and in guiding users in learning the skills of counterinsurgency operations. First, we utilized the PsychSim social simulation tool to

Figure 1. UrbanSim practice environment

model the goals, interrelationships, and beliefs of the population of the urban environment, and to select the behaviors of these agents using a decision-theoretic framework. PsychSim was particularly appropriate for this application, as it simplified scenario authoring by automatically generating behaviors from a reusable library of entities that could be composed in different combinations for different scenarios. Second, we developed a case-based Story Engine to interject narrative elements into the simulation environment, where events are drawn from a case library of the real-world experiences of military commanders. The use of a data-driven Story Engine allows UrbanSim to generate events in the user experience that would be difficult to model in rule-based simulations, and provides a framework for quickly modifying UrbanSim content to reflect changes in the contemporary operating environment of the U.S. Army. Third, we developed a new intelligent tutoring system to deliver informative feedback on learner actions, facilitated by a look-ahead procedure provided by the PsychSim system. This technology was particularly appropriate for UrbanSim due to the nature of the skill set of counterinsurgency operations, where the lack of strict doctrine and procedural rules prohibit the development of an expert model of the sort typically employed in contemporary intelligent tutoring systems.

The integration of these AI technologies is facilitated by UrbanSim's system architecture, which follows a data-driven distribution model where these AI components to work together in a synchronous cycle. Each cycle begins when a learner specifies a set of actions to be executed by subordinate units for the given turn. These actions are then sent to the intelligent tutoring system for evaluation, which may initiate a question-answer tutoring dialogue with the learner. Once this dialogue is complete, the learner commits the actions and the simulation cycle is executed. The simulation cycle is comprised of two separate components that alter the current state of the world: the PsychSim social simulation tool and the Story Engine. Each component is run one step (a simulated day) to produce a set of effects based on the actions designated by the learner as well as those taken by the non-player characters in the environment. These effects are then aggregated and presented back to the learner in the form of various display mechanisms, such as spatial views and social network diagrams, and the cycle repeats.

Each of these three AI technologies are described further in the following sections.

2. Population modeling with PsychSim

We simulate the population in UrbanSim (including relevant people, groups, and environment) using the PsychSim social simulation tool [15]. PsychSim can model an entire social scenario, where a diverse set of entities, either groups or individuals, interact and communicate among themselves. Each entity has its own goals, relationships with other entities (e.g., friendship, hostility, authority), private beliefs, and mental models about other entities. PsychSim generates the behavior for these entities and provides explanations of the result in terms of each entity's goals and beliefs.

2.1. Authoring a Population Model

An author creates the population model for a given scenario by instantiating the desired agents from a library of generic PsychSim models. These generic models (e.g., “Mayor”, “Neighborhood”, “Insurgents”, “Utility”) specify the default state feature values (e.g., a “Mayor” has “political power”), actions (e.g., “Insurgents” may “detonate an IED”), goals (e.g., a “Neighborhood” wants to maximize its “physical security”), action dynamics (e.g., “detonate an IED” decreases the “physical security” in the target “Neighborhood”), and relationships (e.g., a “Mayor” may be “politically aligned” with “Insurgents”). Having selected the instances out of these generic models, the author then specializes them, giving them appropriate names and overriding the default values as necessary (e.g., one neighborhood may have a higher “physical security” value than default, or the specific Mayor may *not* be aligned with any of the insurgent groups in the given scenario). This instantiation process generates a set of operational PsychSim agents corresponding to the relevant entities in the scenario.

2.2. Behavior Generation

During each execution cycle, these agents generate their behavior in turn through a bounded look-ahead procedure that seeks to maximize expected reward by simulating the behavior of the other agents and the dynamics of the world in response to their selected actions. Each agent computes a quantitative value of each possible action by using the authored dynamics to project the effects of each action, weighing the effects against its goals. Thus, the agent is seeking to maximize the expected reward of its behavior (as in a Markov Decision Problem). However, PsychSim's agents exhibit only bounded rationality, constrained by the finite horizon of their look-ahead procedure.

2.3. Behavior Explanation

We also exploit PsychSim's quantitative agent models to provide explanations of the agents' behavior that serve as input to UrbanSim's intelligent tutoring. The agent's behavior generation process constitutes a piecewise linear function that we can invert for a fast sensitivity analysis [14]. We can thus examine each agent's decision and immediately compute the conditions under which it would have chosen a different, preferable action (e.g., the insurgents would not have detonated the IED if the neighborhood had been more secure). PsychSim can also examine the dynamics of the player's actions to identify those moves that may bring about the desired state change (e.g., the neighborhood would have been more secure if there had been a patrol present). The complete set of explanations and suggestions provides a raw input to UrbanSim's Intelligent Tutoring System.

3. Story-driven simulation

There is a rich tradition of pedagogical storytelling in military organizations, where "war stories" are swapped in both formal and casual settings to illustrate points and debate tactics and strategy. This tradition naturally transfers to computer-based learning environments for military trainees. For example, the *Air Campaign Planning Advisor* [5] guided trainees through the process of military air campaign planning through the presentation of approximately one thousand video clips from 12 experts, who told stories of their experiences in the Gulf War, Bosnia, Somalia, and Haiti. More recently, there has been interest in story-based learning environments [3], where the real-world stories of experts are used to design analogous fictional situations that relay the points of these stories. Often realized using computer game technologies, story-based learning environments move the content of expert stories closer to the experience of the trainees, albeit mediated through the designers who craft the fictional situations.

For UrbanSim, we sought to take this characteristic of story-based learning environments one step further in the direction of story-*driven* learning environments, where the events in the fictional training simulation are more directly generated from a corpus of real-world stories. For this purpose, we developed a story-driven simulation engine that works in parallel with PsychSim and created a new story repository based on the real-world experiences of commanders of counterinsurgency operations during the Second Gulf War.

3.1. Story Collection

In the fall of 2007, we conducted five hours of story-collection interviews with five U.S. Army Lieutenant Colonels, each having just completed service as a battalion commander during Operation Iraqi Freedom. We collected sixty-four stories related to counterinsurgency operations using story-collection techniques that had been successful in the development of previous story-based learning environments [3].

As a representative example, one commander told us the story of a suicide-bombing attack resulting in numerous casualties, carried out by a woman and child at a location where the U.S. Army was recruiting police officers. In response, the commander launched a campaign to sway public opinion against the insurgents. He solicited the families of children that were killed in similar attacks for cute photographs of these children, then printed and distributed pamphlets with these pictures around the city with the explanation that these were the children that were being killed by insurgent bombers. The campaign effectively turned public opinion against the insurgents, and fewer suicide bombings subsequently occurred.

3.2. The UrbanSim Story Engine

To use these stories to generate events in UrbanSim, these stories were encoded as sequences of events, each with preconditions and effects, using an XML formalism in the language of the UrbanSim world state model. For example, the story above was represented as a sequence of two events (the insurgent bombing and the pamphlet campaign). Event preconditions were encoded in two forms. First, *world state preconditions* specify characteristics of the fictional simulation environment that would have to be true in order for a similar event to occur. For example, suicide bombing attacks can occur when the military power of the insurgency exceeds a threshold. Second, *user action preconditions* specify actions that the trainee must have previously taken for a similar event to occur. For example, a bombing during the recruiting of police officers can only occur if the learner has directed a company of soldiers to recruit in a city neighborhood. These two preconditions are represented as follows:

```
<WORLD_STATE_PRECONDITION CLASS="Insurgents"
  ATTRIBUTE="Military Power"
  OPERATOR="GREATERTHAN" VALUE="0.1" />
<USER_ACTION_PRECONDITION SUBJECT="Company"
  VERB1="Recruit Police"
  OBJECT="Neighborhood" />
```

During runtime, these preconditions are matched against the current world model, where high-level class variables in the preconditions (e.g. *Neighborhood*) are bound to simulation instances (e.g. the *Northwest Sector*) using a class hierarchy. When all preconditions of a story event unify with the simulation state, the story event executes. Information about the story event is presented to learners as pre-authored situation reports (text), and the underlying simulation state is modified to reflect the effects of the event. For example, the first event of the story described above causes reductions in the physical security of the targeted neighborhood, the support for the U.S. forces, and the military power of the Iraqi police force. These effects are encoded as follows:

```
<RELATIVE_EFFECT SUBJECT="Neighborhood"
  FEATURE="Physical Security"
  VALUE="-0.2" />
<RELATIVE_EFFECT SUBJECT="Neighborhood"
  FEATURE="Coalition Support"
  VALUE="-0.2" />
<RELATIVE_EFFECT SUBJECT="Police"
  FEATURE="Military Power"
  VALUE="-0.1" />
```

The stories of counterinsurgency operations that we collected and encoded into UrbanSim provide an initial proof-of-concept of our story-driven simulation approach. However, we believe that the success of this approach will depend most heavily on the size and relevance of the story corpus to particular training objectives. Toward this end, our current research focus is to develop authoring tools that enable military training developers, instructors, and learners to contribute their own stories to UrbanSim without the assistance of our research group.

4. Intelligent tutoring in UrbanSim

The UrbanSim Practice Environment presents learners with a huge problem space with many solution paths of varying degrees of quality. This complexity is a strength of the application in that it attempts to provide a realistic practice environment through modeling of human behavior. However, it also presents challenges in terms of learning. Specifically, large open learning environments that rely on discovery learning can be problematic for novices [10]. The need for guidance is a reoccurring and established principle of instructional design [12] and is delivered in two key forms: through the Urban Primer (discussed earlier) and by an intelligent tutoring system (ITS) that provides feedback in the UrbanSim Practice Environment.

Expert human tutors and the best ITSs deliver *formative* feedback – that is, “information communicated to the learner that is intended to modify his or her thinking or behavior to improve learning” [16]. Explicit feedback can be used for a variety of reasons, such as to verify the correctness of an action, explain correct answers, remediate misconceptions, reveal goal structure, and more. Feedback can be delivered immediately after an action, or after some delay. The best choices for feedback content and timing depend on many things, including task domain, nature of the skill being learned, the aptitude of the learner, whether the learner has a performance orientation, and more [16]. The UrbanSim ITS supports both immediate and delayed feedback.

One of the goals of UrbanSim is to teach about the broader and unintended effects of actions taken in stability-focused operations (specifically, directing action to anticipate 2nd and 3rd order effects). Understanding the role of non-player characters (NPC) in the PsychSim models is part of this. Each NPC agent acts to achieve its goals and makes decisions based on the state of the world. Although the learner cannot directly order non-U.S. Army NPCs to take (or not take) certain actions, she or he can certainly affect the world state. A key goal for the ITS is to help the learner understand this idea, and to take actions that 1) limit the ability of NPCs to take harmful actions, and 2) enable NPCs to take helpful actions. In other words, *the learner should be thinking about how their actions influence the actions of others* – the ITS frames its feedback in this light and attempts to reveal the reasoning behind NPC's actions: why they made the decisions they made, what consequences (seen or unseen) were most relevant, and under what circumstances different decisions would have been made.

To support learning of unintended consequences, we have implemented an *anticipate-wait-relate* tutoring strategy, which is similar to the "Pause, Predict, Ponder" strategy for cultural learning described in [13]. After the learner has proposed an action and the ITS has decided to apply the strategy, three steps are taken. (1) Elicit the anticipated effects of that action. That is, ask the learner to assess that choice by indicating, via drop down menus, how she or he expects that action to affect the world state. (2) After this input, allow the game to proceed for some number of turns (which is only one at the time of this writing, but longer delays are possible). (3) Finally, the ITS presents the learner with the *actual* results for comparison.

The system is able to provide feedback before or after the initial action proposal, or later (step 3), along with the comparison between anticipated and actual outcomes. Our focus thus far has been on this delayed form of feedback. Application of this strategy requires answers to at least two questions. First, what is used to trigger the strategy? In other words, when should the learner be prompted to anticipate the effects of an action? Second, how should the ITS support reflection on the results of the comparison? Of course, the learner could easily be asked to anticipate outcomes to every action, but this would quickly become a distraction. Also, learning could potentially occur by simply allowing the learner to inspect the predicted versus actual outcomes and learn from them. This also is unappealing, especially given the rich PsychSim models that drive NPC behavior. In fact, our approach leverages these models and the reasoning capabilities of PsychSim.

Our initial approach to answering the question of when to ask the learner to anticipate is to use PsychSim's look-ahead functionality. If progress toward the desired end-state of the world is about to decrease from a learner action, the ITS will ask the learner to anticipate the action's effect on features that are used to assess this progress. We currently focus on potentially damaging actions the learner can take, but also ask for predictions on positive actions to prevent gaming of the system. Regarding feedback, we have implemented an approach based on *causal chains* of the reasoning behind the NPCs. These causal chains reveal the state changes that occur based on learner actions, allowing the learner to see the connections between their actions, the world state, and the ensuing NPC actions. Figure 2 shows two examples of how causal chains show the effect of actions on the world state and their subsequent impact on lines of effort. Additionally, the ITS also queries PsychSim to reveal what conditions would have led to different NPC actions. For example, an NPC's ability to have taken one very bad action may have been impossible had the learner taken a different action

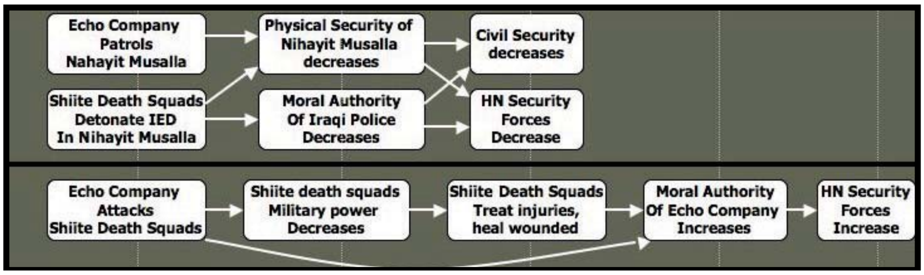


Figure 2. Causal chains in UrbanSim. Actions (on the left) induce changes to the world state, which subsequently lead to changes in *lines of effort* (e.g., Civil Security).

at that turn. The ITS's aim here is to support the learner's reflection in imagining what other actions could have produced such world states, and scaffold the identification of alternative courses of action that would have produced different outcomes.

5. Conclusion

Contemporary simulation-based training systems, which utilize advances in computer game technology to address specific training needs, offer many new opportunities for the innovative application of AI technologies. However, the design space of possible combinations of learning objectives and technologies is large and largely unexplored. In this paper, we have presented our implementation of a design solution for a particularly novel combination of learning objectives and technologies. The skill set related to the conduct of military counterinsurgency operations was well suited to the design of turn-based strategy games and city management simulations built by the entertainment industry. However, to support the acquisition of skills in this environment, several "under-the-hood" AI technologies were required. To effectively model the behavior and complex interrelationships among stakeholders in urban environments, we used the PsychSim social simulation tool. To interject narrative elements that reflect the real-world experiences of commanders in counterinsurgency operations, we developed a story-driven simulation engine. To deliver informative feedback to learners about the counterinsurgency skills that they are practicing, we developed a new intelligent tutoring system that uses an anticipate-wait-relate tutoring strategy enabled by simulation-based look-ahead.

We believe that this design solution generalizes well to a wide range of other learning environments. The skill set related to effective counterinsurgency operations is analogous to a variety of other skills that involve managing and intervening in complex social environments, e.g. reducing gang violence in an urban environment through law enforcement. In addition to turn-based strategy games, the AI technologies that we employed in UrbanSim could be adapted for use in real-time strategy games and god-games. The exploration and evaluation of systems in this broader design space is an important direction for future work.

Acknowledgements

The project or effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

[1] Army Training and Leader Development Panel Washington DC, (2003) *The Army Training and Leader Development Panel Officer Study Report to the Army*. http://handle.dtic.mil/100.2/ ADA415810
[2] Clark, R. (2004) *Design Document For A Guided Experiential Learning Course*. Final report on contract DAAD 19-99-D-0046-0004.
[3] Gordon, A. (2009) Story-Based Learning Environments. In Nicholson, D., Schmorrow, D., & Cohn, J. (eds.) *The PSI Handbook of Virtual Environments for Training and Education, Vol 2*. Westport, CT: Praeger Security International.
[4] Hart, J., Wansbury, T., & Pike, W. (2009) So you want to us a game: Practical considerations in implementing a game based trainer. In J. Cohn, D. Nicholson, and D. Schmorrow (eds.) *The PSI Handbook of Virtual Environments for Training and Education, Vol. 3*. Westport, CT: Praeger Security International.
[5] Johnson, C., Birnbaum, L., Bareiss, R., and Hinrichs, T. (2000) War Stories: Harnessing Organizational Memories to Support Task Performance. *Intelligence 11*(1):16-31.
[6] Johnson, W. L. (2007) Serious use of a serious game for lan*guage learning. Proceedings of the International Conference on Artificial in Intelligence*. IOS Press, Amsterdam, 2007.
[7] Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S., & Piepol, D. (2007) Building Interactive Virtual Humans for Training Environment. *Proceedings of 2007 Interservice/Industry Training, Simulation & Education Conference, Orlando, FL, November 2007*.
[8] Kilcullen, D. (2006) Twenty-Eight Articles: Fundamentals of Company-level Counterinsurgency, *Military Review*, May-June 2006, 103-108.
[9] Kim, J., Hill, R.W., Durlach, P., Lane, H.C., Forbell, E., Core, C., Marsella, S. Pynadeth, D. & Hart, J. (*in press*) BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education*.
[10] Kirschner, P., Sweller, J., & Clark, R. (2006) Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86.
[11] Korris, J. (2004) Full Spectrum Warrior: How the Institute for Creative Technologies Built a Cognitive Training Tool. Proc*eedings of the 24th Army Science Conference, Orlando, Fl, December 2004*.
[12] Merrill, M. (2002) Pebble-in-the-pond model for instructional development. *Performance Measurement 41,* 39-44.
[13] Ogan, A., Aleven, V., & Jones, C. (2008). Pause, predict, and ponder: Use of narrative videos to improve cultural discussion and learning. *Proceedings of the 26th annual SIGCHI conference on Human factors in computing systems,* 155-162.
[14] Pynadath, D. & Marsella, S. (2004) Fitting and Compilation of Multiagent Models through Piecewise Linear Functions. *Proceedings of the 2004 International Conference on Autonomous Agents and MultiAgent Systems*, 1197-1204.
[15] Pynadath, D. & Marsella, S. (2005) PsychSim: Modeling Theory of Mind with Decision-Theoretic Agents. *Proceedings of the 2005 International Joint Conference on Artificial Intelligence*, pp. 1181-1186.
[16] Shute, V. (2008) Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.
[17] Wong, L. (2004) *Developing Adaptive Leaders: The Crucible Experience of Operation Iraqi Freedom*. Carlisle, PA: Strategic Studies Institute of the U.S. Army War College.

Explicit social goals and learning: enhancing a negotiation game with virtual characters

Amy OGAN[a,1], Julia KIM[b], Vincent ALEVEN[a], and Christopher JONES[a]
[a] *Carnegie Mellon University*
[b] *University of Southern California*

Abstract. Games are increasingly being adapted for use as educational tools. One relatively new use of games is to facilitate learning social or interpersonal skills such as conflict resolution by simulating human behavior with virtual characters. This work investigates students' social goals to understand how they help motivate students to acquire cultural understanding in BiLAT, one such system designed to teach cross-cultural negotiation skills. We hypothesized that students who were given explicit social goals (e.g. "Come to understand your partner's point of view") would be more successful learning from the game than students who were given task-only goals. We ran a randomized controlled experiment in which 59 students played BiLAT. 30 students played the game as designed, with negotiation task goals. 29 were additionally given a social goal. The results did not confirm our hypothesis – the group without the social goal learned more according to most measures. However, on further investigation, students who reported having social goals in a manipulation check, regardless of condition, seemed to learn the most. These results suggest that social goals and interactions are important in learning cultural negotiation, but that setting explicit social goals may not be the right scaffold.

Keywords. Virtual environments, motivation

Introduction

Simulation-based instructional systems are increasingly being used to facilitate learning social or interpersonal skills such as conflict resolution by simulating human behavior with virtual characters [14]. These skills are currently taught through methods like role-playing exercises and tutoring, both of which are very resource-intensive teaching methods [9]. Computer-based simulations, which are growing more realistic, offer a major advantage for social learning by providing a cheaper solution to a much larger number of students. Examples of existing game-based instructional systems with social learning components include FearNot [6], PeaceMaker [2], and BiLAT [7]. BiLAT, the game in which we situate this work, is a virtual environment that supports cross-cultural interactions in the context of a negotiation task.

While results on learning from such simulations are preliminary, they are purported to be highly motivating (e.g., 8, 11, 12). Motivation is important in learning contexts because it can lead students to make greater effort, seek greater challenges, and have higher achievement (see 15). However, there has been little research on how

[1] Corresponding Author: Amy Ogan; Carnegie Mellon University, SCS, HCII; 5000 Forbes Avenue; Pittsburgh, PA 15213

aspects of student motivation affect learning in social simulations, and even less on how they affect learning of intercultural competence in such environments. Since motivation is a very multi-faceted notion, it is important from a practical and a theoretical perspective to better understand what aspects of student motivation are particularly conducive (or not) to student learning in this context.

One standard framework of student motivation that is relevant to games which has been studied extensively within the realm of educational psychology is goal orientation. Goal orientation researchers tend to focus on mastery and performance orientation (e.g., [1, 5]). However, in a domain like culture that focuses on social interactions, social factors might have an even greater influence on learning than these academic orientations [17]. Cultural interaction is an inherently social process between people with different cultural identities. While culture may be largely unconscious, it becomes more salient when interacting with someone of another culture who is then categorized as a member of an "outgroup" [13]. Such cross-cultural contact can exacerbate ingroup-outgroup biases and lead to social goals like the desire to be seen as distinct from and positively compared to the outgroup [16]. These motives may be detrimental to learning about a new culture. On the other hand, social goals such as a need for affiliation or the desire to conform to social rules, may promote learning. It is an open question how such goals can be promoted in a way that improves learning of intercultural competence in environments (such as BiLAT) in which learners interact with virtual characters from a different culture.

In this paper we describe an empirical study that investigates student motivation in an educational game for learning to negotiate in a cross-cultural context. We attempt to answer the following question: when learners are given a social goal as a part of a game-based experience, what is the impact on their learning? We present results from a first study in which students played a game with or without the addition of a social goal, and a discussion of the emergent subgroup of the population that expressed these types of goals. This work contributes to understanding whether students who approach virtual characters as culturally distinct social beings increase their learning over those with a task focus, and will develop improved ways to support intercultural learning through technology.

Game Context

The context we use for our investigation is BiLAT [7], a game-based simulation for practicing bilateral engagements in a cross-cultural context.[2] BiLAT was designed to address learning objectives related to negotiation generally as well as the specific cultural knowledge and skills [see 3 for a description of cultural learning objectives] that support more effective negotiations in that particular culture. A primary learning objective is considering the counterpart's interests in order to achieve "win-win" results. Scenarios presented to the student drive the game experience. The initial scenarios are set in an Iraqi town, and the student is put into the role of a U.S. Army officer tasked with meeting with members of the town in order to accomplish specific tasks. The student is given these concrete, negotiation task-related goals related to each meeting such as "Learn why the market is not being used."

[2] The BiLAT architecture is built on Unreal Engine 2.5 and integrates research technologies such as virtual human characters and intelligent tutoring support.

To play, the student begins by preparing for a meeting in the “prep room.” Here, the student learns about the character from a variety of different sources of varying degrees of trustworthiness. The student then moves into the meeting (see Fig. 1). The student communicates with each BiLAT character by selecting from a menu of hand-authored communicative actions such as “Discuss [Hassan’s] favorite authors and literature”, or “Complement your host on his generosity”. Characters’ responses depend on a number of factors, including the current meeting phase, the trust level, and a virtual dice roll. Each action entails a possible change to the trust variable of the character. The dice roll is intended to simulate uncertainty in human behavior – cognitive and emotional modeling techniques can be used to simulate these reactions in more principled ways [7]. The character responds to the actions in both text and synthesized speech as well as non-verbal behaviors such as gestures.

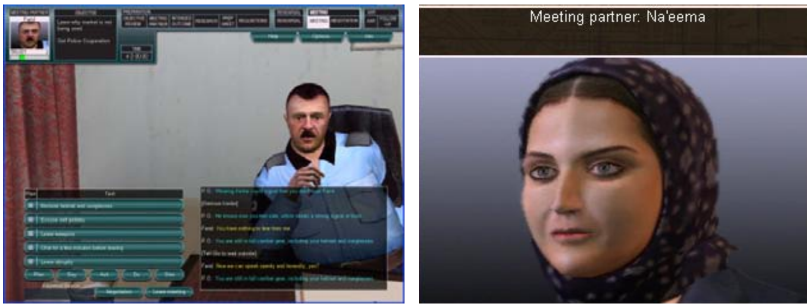


Figure 1. At left, a meeting in BiLAT with police captain Farid, with the goal of solving a problem with a market in an Iraqi town. At right, meeting partner Na’eema, a doctor.

Experimental Study

We ran a randomized controlled study to test the effect of giving learners a social goal as a part of a game-based experience. Our hypotheses were: 1) the social goal group would perform better than the control on overall learning measures, 2) the control group would perform better on task-related items while the social goal group would perform better on social items, and 3) the social goal group would perform better on cultural transfer measures.

1. Method

Participants were 59 students, mostly undergraduates, recruited from two institutions. They were compensated $40 for a two and a half hour session. Each student began by watching an introductory video about the concepts and skills related to the cultural learning objectives. The student then took a number of pre-tests (described below). Next, they entered the game and met with two different characters. A set of measures relating to the information learned in the “prep room” (described below) was administered. The student then met with the character until an agreement on the negotiation was reached. If the student did not come to an agreement when the allotted time ran out, the experimenter moved the student along. After the student met with both characters, he or she closed the game and took the post-tests.

Students were randomly placed into one of two conditions. The control condition played BiLAT with its standard task-related goals, e.g. "Learn who is responsible for enforcing the tax." Success in achieving these goals was clearly defined, and they were clearly related to progress through the task. The social goal condition received an additional objective in each meeting which was labeled "Come to understand your partner's point of view," which was intended to encourage them to focus on the interpersonal aspects of the interaction. Only the social goal condition received this explicit social goal in the game.

2. Measures

Assessment of intercultural competence is not a trivial task. It is an ill-defined domain; there is not always a clear distinction between right and wrong answers, and even experts at times may disagree. A number of measures have been developed that may be used in different situations. We assessed the cultural and negotiation learning objectives with a Situational Judgment Test. This assessment asks students to rate the appropriateness of various actions based on situations related to the learning objectives [10]. This test has been used previously to collect data with students playing BiLAT. As a measure of transfer, we introduced a selection of questions from the Cultural Assimilator [4]. In this assessment, students read a scenario about people experiencing a foreign culture and chose the best of four possible cultural explanations for the events in the scenario.

Additionally, the students were administered a learning assessment related to the specific scenario and characters. Prior to and after meeting a new character, we asked students to rate the truth of various items relating to the task or to the character (e.g., "Farid could be described as a family-oriented man"). These items were taken from the information students received in the "prep room". Students evaluated the items as true, false, or "I don't know". The goal of this measure was to assess a student's ability to develop an accurate model of the character and the scenario. A successful student would be able to elicit information while meeting with the character and integrate this knowledge with the information from the sources in the prep room. We called this measure *information integration*.

Finally, we wanted to determine whether our manipulation had the desired effect on students' goals in the game. We therefore asked them to list their goals in free text after meeting with a character.

Results

While 59 participants completed the study, we dropped 3 students from the analyses due to computer error and 2 students who demonstrated a complete lack of attention or engagement. In our final analyses, we compared 25 experimental students who received the additional social objective presented at the beginning of each meeting to 29 control students playing with the standard game objectives. To evaluate the success of our conditions, we coded students' responses to the manipulation check into 2 categories, "no social goals" and "social goals". Students who reported no social goals tended to write out the task goals they were given verbatim, such as "Learn why the market is not being used". Two independent coders rated anything that focused on social interaction with the virtual character in the "social goals" category, such as "Learn more about my partner" or "Better understand my partner." Table 1 has the results of this coding. A chi squared test showed that the number of students with social goals was significantly influenced by condition ($\chi^2(1, N = 54) = 5.868, p = .015$). However, because almost a third of the students did not report their goals as expected, we examined all of the learning results in this light.

Table 1: Number of students by condition and reported goals

Given Condition	Reported goals		
	No social goals	Social goals	Total
Control	20	9	29
Social Goals	9	16	25
Total	29	25	54

To analyze the data we conducted repeated measures ANCOVAs on students' pre and post scores with condition as the between-subjects variable and reported social goals as the covariate. On the Situational Judgment Test, we observed no significant differences between pretest and posttest for either condition. It may be that watching the introductory video was sufficient to hit ceiling on this test; mean correlations with experts were around .8 on both the pretest and posttest. Next we looked at the information integration items. On these items, students were given a point for each item that matched a subject matter expert's rating of the information. An ANCOVA showed that condition significantly influenced learning ($F(1,49) = 5.307, p = .026$); the control condition learned more than the social goals condition (see Figure 3 for learning gain means).

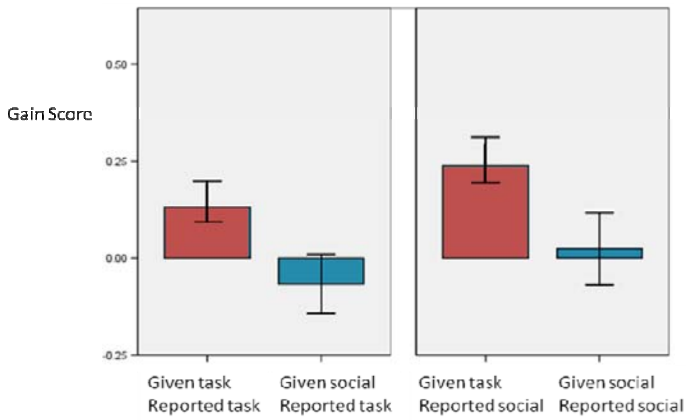


Figure 3: Mean information gain scores by condition and reported social goals

We then divided the information integration items into task-related items, such as "The U.S. must set up checkpoints to increase security in the market," and social-related items, such as "Farid is a family man and is good with children." On the social items, an ANCOVA showed that reported social goals ($F(1,49) = 3.979, p = .052$) and condition ($F(1,49) = 3.285$, $p = .076$) fell just short of statistical significance; the control condition learned more than the social goals condition and reported social goals learned more than no reported social goals (see Figure 4). There was significant overall learning from pre to post ($F(1,49) = 9.213$, $p = .004$). The ANCOVA on the task items showed no difference between conditions, but there was significant overall learning from pre to post ($F(1,49) = 13.552$, $p = .001$). See Figure 5 for gain means.

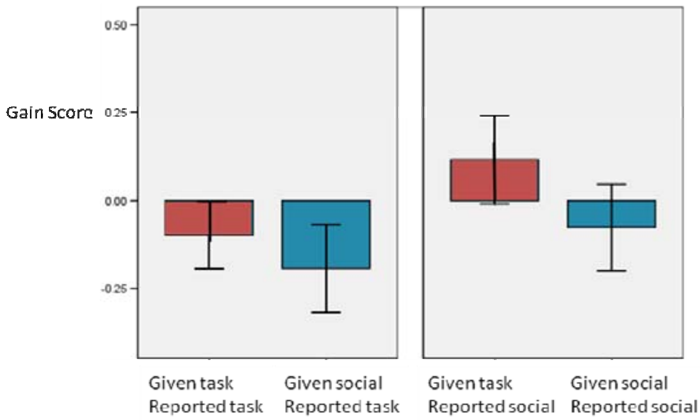


Figure 4: Mean social information gain scores by condition and reported social goals

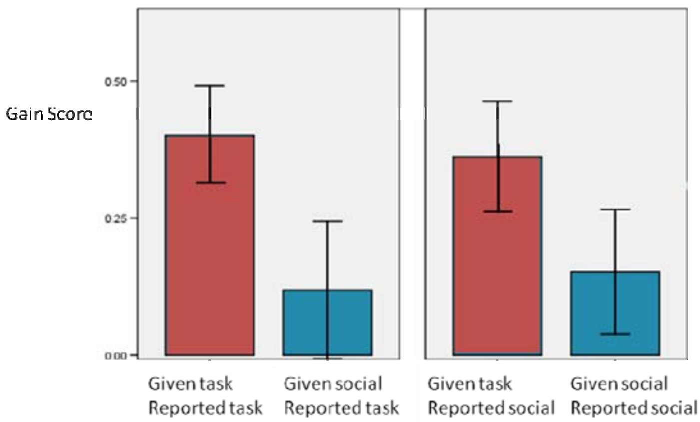


Figure 5: Mean task information gain scores by condition and reported social goals

Finally, an ANCOVA model of the culture assimilator transfer test showed that condition significantly influenced student learning ($F(1,47) = 11.873$, $p = .001$), as did reported social goals ($F(1,47) = 8.314$, $p = .006$). The control condition outperformed the experimental and reported social goals outperformed no reported social goals (see Figure 6 for a comparison of gain score means). Also, there was significant overall learning from pre to post ($F(1,47) = 4.582$, $p = .038$).

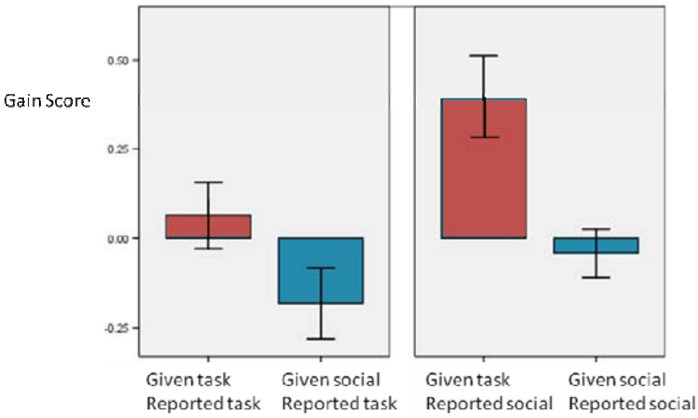


Figure 6: Mean assimilator gain scores by condition and reported social goals

Discussion and Conclusions

The results did not directly confirm our hypothesis – in fact, the group given only task goals performed significantly better on most measures. However, students with self-reported social goals, regardless of condition, learned more. This provides suggestive evidence that social goals and interactions are critical in learning cultural negotiation, although the study manipulation was an ineffective means of integrating social considerations into the learning experience for students who do not have them spontaneously. Explicit presentation of such goals may even be detrimental. There are several viable hypotheses about why these results occurred. An additional goal may have caused more cognitive load which divided students' focus, students may not have understood how to achieve the goal, or perhaps the addition of extrinsic motivation may have reduced intrinsic motivation to be social.

Next steps in this work include a program of research to understand the role of various types of social goals (*integrative* and *self-assertive* goals) in learning cultural negotiation, and how to promote them. We know that some students arrive at the intervention already holding social goals. We will develop a model of how social goals are influenced by and interact with learner characteristics such as social intelligence and personality traits. Then, we will investigate how social goals can be promoted without interfering with their pre-existence. We will design a scaffold for a cultural discussion embedded in the game that implicitly primes students to focus on social goals. This will allow us to manipulate students' goals in BiLAT to empirically examine how and when social goals are beneficial or detrimental to learning intercultural competence.

Recent research has been investigating how people interact with virtual humans. This work will provide a better understanding of how people interact socially with virtual humans in a cultural learning context, and will develop improved ways to support intercultural learning through technology. It will increase understanding of how social goals influence learning in the context of a cross-cultural negotiation task, and how they can be promoted in a way that is beneficial to learning. This work contributes to understanding whether students who approach virtual characters as culturally distinct social beings increase their learning over those with a task focus.

Acknowledgments

Thanks to Eric Forbell for his great effort at support and implementation. Part of this work has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred. The research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B040063 to CMU.

References

[1] Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84,* 261-271.

[2] Burak, A., Keylor, E., & Sweeney, T. (2005). PeaceMaker: A Video Game to Teach Peace. In *Intelligent Technologies for Interactive Entertainment* (pp. 307-310).

[3] Byram, M. (1997). Teaching and Assessing Intercultural Communicative Competence. Clevedon: Multilingual Matters.

[4] Cushner, K., Brislin, R, (1995). Intercultural Interactions: A *Practical* Guide. Second Edition. Thousands Oaks, CA: Sage Publications, Inc

[5] Dweck, C.S., & Leggett E.L. (1988). A social-cognitive approach to motivation and personality. Psychological Review, 95 256-273.

[6] Hall, L., Vala, M., Hall, M., Webster, M., Woods, S., Gordon, A., & Aylett, R. (2006) FearNot's appearance: Reflecting Children's Expectations and Perspectives. *In 6th International Conference, IVA 2006,* Springer, LNAI 4133, pp 407-419.

[7] Hill, R.W., Belanich, J., Lane, H.C., Core, M.G., Dixon, M., Forbell, E., Kim, J., & Hart, J. (2006). Pedagogically Structured Game-based Training: Development of the ELECT BiLAT Simulation, *in the Proceedings of the 25th Army Science Conference (ASC 2006),* Orlando, FL.

[8] Klein, J. D., & Freitag, E. (1991). Effects of using an instructional game on motivation and performance. *Journal of Educational Research, 84(5),* 303-308.

[9] Landis, D., Bennett, J., & Bennett, M. J. (2003). *Handbook of Intercultural Training* (3rd ed.). Sage Publications, Inc.

[10] Legree, P. & Psotka, J. (2006). Refining situational judgment test methods. In Proceedings of the 25th Army Science Conference. Orlando, FL.

[11] Lepper, M. R., & Malone, Th. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. *Aptitude, learning, and instruction: Vol. 3,* 255-286.

[12] Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology*, *62*, 625-633.

[13] Prentice, D. A., & Miller, D. T. (1999). The psychology of cultural contact. In D. A. Prentice & D. T. Miller (Eds.), *Cultural divides: Understanding and overcoming group conflict* (pp. 1-19). New York: Russell Sage Foundation.

[14] Raybourn, E. M., & Waern, A. (2004). *Social Learning Through Gaming.* In Extended Abstracts of CHI Proceedings 2004. ACM Press.

[15] Schunk, D. H., Pintrich, P. R., & Meece, J. (2007). *Motivation in Education: Theory, Research, and Applications* (3rd ed.). Prentice Hall.

[16] Taylor, D. M., & Moghaddam, F. M. (1994). Theories of intergroup relations: International social psychological perspectives. 2nd. Ed. New York: Praeger

[17] Urdan, T., Maehr, M. (1995) Beyond a Two-Goal Theory of Motivation and Achievement: A Case for Social Goals. Review of Educational Research, Vol. 65, No. 3, 213-243, 1995.

iSTART-ME: Situating extended learning within a game-based environment

G. Tanner JACKSON[a,1], Chutima BOONTHUM[b], and Danielle S. McNAMARA[a]
[a]*University of Memphis, Psychology/Institute for Intelligent Systems*
[b]*Hampton University, Computer Science*

Abstract. iSTART is an intelligent tutoring system designed to improve students' reading comprehension by teaching helpful and effective comprehension strategies. Students using iSTART have consistently demonstrated significant improvement in reading comprehension; however, these benefits improve with extended practice (taking place over a span of weeks and months rather than the typical experimental intervention of a few minutes or hours). Due to the long-term nature of this interaction, the extended practice module of iSTART is currently being situated within a game-based environment called iSTART-ME (motivationally enhanced). This game-based environment provides students a chance to interact with texts, earn points, advance through levels, purchase in-game rewards, personalize a character, or even play an educational mini-game (designed to use the same strategies as in practice).

Keywords. Intelligent Tutoring Systems, educational games, game-based learning

Introduction

Games and game-based environments constitute an area of rapid growth in private, public, and research sectors. In 2007, while industries such as music and movies saw either negative or stagnant growth (-10.0% and +1.8% respectively), the gaming industry reported dramatic growth (+28.4%) [1]. Researchers of Intelligent Tutoring Systems (ITSs) have begun to leverage the engagement and appeal of games by incorporating game-like features within learning environments [2].

While it is intuitively clear that games are engaging and can often sustain interest over extended amounts of time, it is still relatively unclear how this process occurs and which specific features are essential to the essence of games. Previous research has attempted to identify and investigate specific gaming components such as challenge, fantasy, complexity, control, rules, strategy, goals, competition, cooperation, and chance [3,4,5]. However, these components have been primarily observed within the context of entertainment games. Only recently have these components been implemented and observed (and sometimes tested) in the context of learning environments [6,7]. Establishing the effects of game components on learning and motivation is important for those who are interested in developing systems that maximize learning benefits in computer-based systems. ITS developers and researchers often struggle to create just the right balance between implementing effective learning practices, while at the same time enhancing motivational aspects of the learning

[1] Corresponding Author.

environment [8,9]. The principal goal of ITS technologies is most often to produce significant learning gains (e.g., learn a new skill or understand concepts within a specific domain). However, these systems, though often effective at producing learning gains, are sometimes uninspiring to those who use them. Focusing on maximizing learning benefits can suffice for experimental purposes, but it creates a problem for systems that are used repetitively and over long periods of time. Additionally, improving motivational aspects of learning environments is likely to produce indirect gains in learning, particularly if the modifications result in heightened engagement on the part of the learner [10].

The intersection of these two fields (games and ITSs) provides a fertile ground to develop effective learning environments that maximize learning while at the same time fully engaging the user and instilling a desire to interact with the system. The remainder of this paper describes a work-in-progress to develop a learning system that borrows effective design elements from both games and ITS technologies.

1. iSTART

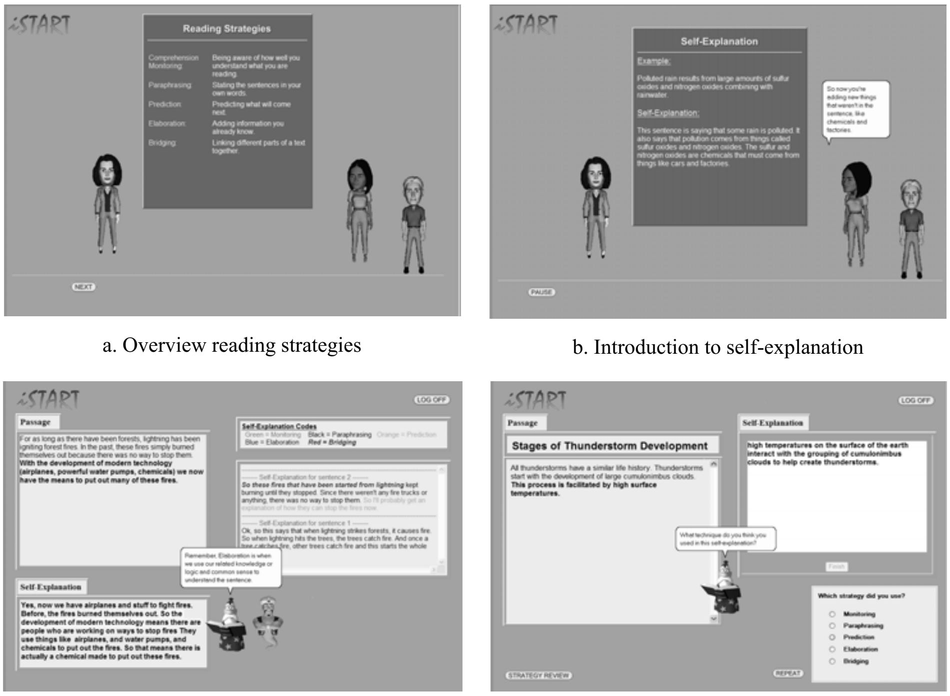
Interactive Strategy Training for Active Reading and Thinking (iSTART) is a tutoring system designed to improve students' reading comprehension [11,12]. This web-based system implements a modeling-scaffolding-fading approach to teach students self-explanation strategies. iSTART originated as a human-based intervention called Self-Explanation Reading Training, or SERT [13,14,15]. Similar to the original SERT program, the automated iSTART program has produced significant learning gains [15,16,17]. Unlike SERT, iSTART is completely automated and can potentially train any student at any school with internet capable computers. Furthermore, it has been designed to work with students on an individual level and can provide self-paced instruction. iSTART also maintains a log of student performance and can use this information to provide appropriate feedback and instruction for each individual student. Lastly, iSTART incorporates pedagogical agents and an automated linguistic analysis that can engage students in an interactive dialog and creates an active learning environment [10,18,19,20].

1.1. iSTART Modules

iSTART incorporates pedagogical agents that use dialogue to engage users with the system and to tutor them on how to apply effective reading strategies. The agents introduce students to self-explanation as a general comprehension strategy, and also demonstrate specific reading strategies to improve their self-explanations and reading comprehension. iSTART is composed of three modules that instantiate the pedagogical principle of modeling-scaffolding-fading: introduction, demonstration, and practice.

During the introduction module, three agents (a teacher and two student agents) engage in a classroom-like discussion to present the relevant reading strategies used within iSTART. These three agents interact with each other, provide students with information, pose questions to each other, and provide examples of self-explanations to illustrate appropriate strategy use (including counterexamples). These interactions use the agents to model the active processing that students should use when creating their own self-explanations. After each strategy is introduced and the agents have concluded their interaction, the students complete a set of multiple-choice questions that assess the basic understanding of the recently covered concepts.

After all of the iSTART strategies have been introduced, students progress to the demonstration module. The demonstration module incorporates two new animated characters (Merlin & Genie) that interact and scaffold the students as they are asked to analyze example explanations, provided by Genie. Within the demonstration module, Genie acts as the student agent by reading text aloud and providing a self-explanation for each sentence. After Genie self-explains a sentence, Merlin instructs the student to identify the strategies used within the example. Merlin provides feedback to Genie on his explanations and also scaffolds the students in identifying strategies. For example, Merlin may tell Genie that his explanation is too short and then ask him to add information, or Merlin might applaud when the student correctly identifies appropriate strategies. Merlin's feedback to Genie is similar to the feedback that Merlin gives students when they complete that section and move on to the practice module.



a. Overview reading strategies

b. Introduction to self-explanation

c. Demonstration module

d. Practice module

Figure 1. iSTART Trainer: screen shots

The practice module in iSTART provides an opportunity for students to apply the recently learned strategies. This module fades out most direct instruction and uses feedback to guide the interaction. Merlin is the only agent in this module and his role is to serve as a self-explanation coach by providing feedback on student explanations and prompting them to use the newly acquired repertoire of strategies. The main focus of this module is to provide students with an opportunity to apply the reading strategies to new texts and to integrate knowledge from different sources in order to understand a challenging text. A good student explanation may include information from previous text, or come from world knowledge and domain knowledge. Merlin provides feedback each time the student generates an explanation. For example, Merlin may prompt them to expand the explanation, incorporate more information, or suggest that they link to other parts of the text. Merlin sometimes asks students to identify which strategies they used and where they were used within their explanation. All of Merlin's responses are adapted to the quality of each student's explanation. For example, when students provide longer and more relevant explanations Merlin gives more enthusiastic expressions, while short and irrelevant student explanations prompt Merlin to provide scaffolding to work the student toward better self-explanations.

Within iSTART there are two types of practice modules. The first practice module is situated within the context of the initial 2-hour training (core iSTART modules). That is, initially, a student typically goes through the introduction, demonstration, and practice modules in about 2 hours. The initial practice module includes practice with only two texts, on which the iSTART feedback algorithm was trained [21]. The second phase of practice, called extended practice, begins immediately after the regular practice module. Extended practice is most often used in situations where a classroom, teacher, or student has committed to using the system over a longer period of time, such as over the course of a year. During this extended practice phase, a teacher can upload and assign specific texts for students to read. Teachers can select texts already available within the system or they can upload their own text into the system *on the fly*. Because of the need to incorporate texts with no preprocessing, the iSTART feedback algorithms must be able to provide appropriate and accurate assessments, not only for the texts during initial practice (for which the iSTART algorithms are highly tuned), but also for any newly entered text. Recent analyses have demonstrated the ability to provide accurate assessments on untrained texts [22] and have provided support for the current enhancements to iSTART.

2. iSTART-ME

iSTART-ME (i.e., iSTART-Motivationally Enhanced) is an extension of the previous tutoring system. In this version of iSTART, the three main modules (introduction, demonstration, and regular practice) remain relatively unchanged (only minor modifications). By contrast, the extended practice module is modified to incorporate various game elements and principles.

The main thrust of the iSTART-ME project is to implement several game-based principles and features that are expected to support effective learning, increase motivation, and sustain engagement throughout a long-term interaction with an established ITS. Previous research has indicated that increasing self-efficacy, interest, engagement, and self-regulation should positively impact learning [23,24,25,26,27]. The iSTART-ME project attempts to manipulate these motivational constructs via game-based features that map onto one of the following five categories: feedback, incentives, task difficulty, control, and environment. These categories are discussed in detail in [2]; so only a brief description is provided here.

Feedback provides an assessment of performance, a measure of system progress, and is used to scaffold student self-regulation and increase self-efficacy. *Incentives* are designed to provide motivational hooks that maintain interest and help to prolong student engagement. Adapting the level of *task difficulty* allows a system to engage the student within their zone of proximal development (select a task that is challenging but still achievable). Allowing users to make choices and exhibit *control* within the system increases interest, engagement, and locus of control (which increases positive attributions). The *environment* is the most outwardly apparent feature and can serve as an overall representation of the system. This organizational scheme creates a multiple-to-multiple mapping between individual game features and these motivational categories. In essence, multiple game features (e.g., points, levels, progress bars) can contribute to each one of these categories (e.g., incentives), and likewise, a single game feature (e.g., points) may map onto multiple categories (e.g., incentives, feedback, environment).

Each of the game features implemented within iSTART-ME corresponds to at least one of these five categories. The game features implemented within iSTART-ME include: earning points, advancing through levels, and purchasing rewards such as creating a personalized avatar, changing environment attributes, or playing an educational mini-game.

2.1. Points

Earning points has become a ubiquitous feature within game design and it plays a central role for iSTART-ME. Within iSTART-ME, students can earn points as they interact with texts and provide their own self-explanations. Each time that a student submits a self-explanation, it is assessed by the iSTART algorithm and points are awarded based on a scoring rubric. The rubric has been designed to reward consistently good performance. So, students earn more points if they repeatedly provide good self-explanations, but earn fewer points if they fluctuate between good and poor performance.

These points provide a quantifiable assessment that complements the qualitative verbal feedback provided by the tutor. For example, a student can discern that a score of 50 is clearly better than a score of 30, whereas the verbal feedback may not indicate such a measurable difference (by contrast, its purpose it to provide information on how to improve the explanation that received only 30 points). Points can provide a turn-by-turn indicator of self-explanation quality, but they can also provide a global indicator of how well the student has been self-explaining over an extended period of time (thus providing both local and global feedback).

In addition to providing another form of feedback, earning points within iSTART-ME serves two main purposes: advancing through levels and purchasing rewards.

2.2. Levels

As students accumulate more points, they advance through a series of levels. Each subsequent level requires an increasing number of points, requiring students to put in slightly more effort with each subsequent advancement. The levels are labeled to help increase interest (e.g., “ultimate bookworm”, “serious strategizer”, etc.), and also help to serve as global indicators of progress across texts. Including the levels along with points allows students to set goals and subgoals based on external, extrinsic indicators [2]. For example, students may set a goal to reach the next level or a goal to earn enough points to obtain a reward. These types of goals can be induced by the structure of the system without explicitly prompting the student. The hope is that this approach can afford a more smooth and enjoyable interaction that more naturally builds interest in interacting with the system.

2.3. Purchasing Rewards

As previously mentioned, points can also be used to “purchase” rewards within the system. Being able to spend points means that the new extended practice module must maintain two point counters: total points earned and points available to spend. The total point score is used to progress the student through the iSTART-ME levels and as global feedback. The available points counter keeps track of the points that are currently available to spend. When a student uses points to obtain a reward, the point value associated with that reward is subtracted from their current available points (the total points score is not affected). Allowing students to choose their own reward provides them a great deal of control. After every completed text, the students are taken to the main selection menu for iSTART extended practice (Figure 2). This menu displays information for each user and it allows students to earn more points by self-explaining a new text, cash in points to change the interface, swap out the tutoring agent, edit their own avatar/character, or spend points to play a mini-game.

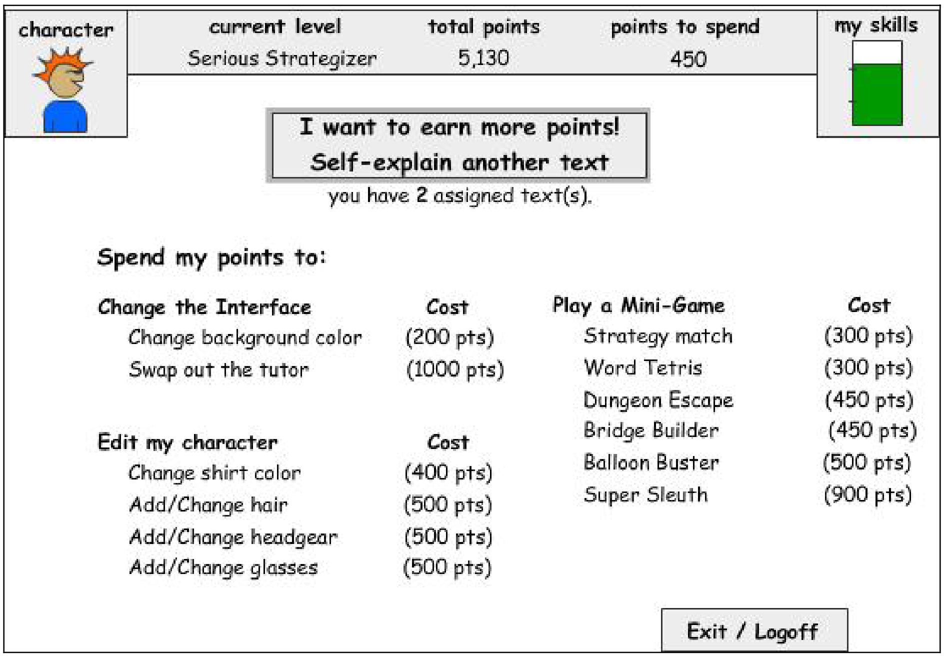


Figure 2. Example prototype screen shot of the iSTART-ME selection screen

2.3.1. Interface and Tutor

One of the options available as a reward is for students to change aspects of the learning environment. They can spend some of their points to choose a new tutor agent or to change the interface to a new color scheme. If a student decides to spend points to obtain a new tutor agent, that new agent remains as their agent until they spend points to change it again (or they discard it for the default tutor). Similarly, if a student uses their points to convert the interface to a new color scheme, that new scheme remains active until they spend points to change it again. These rewards are designed as purchasable replacements, rather than always available options, to help reduce off-task behaviors (such as switching back and forth between agents just to see what they all look like).

2.3.2. Personal Avatars

The extended practice module also includes a personal avatar for students. There is a graphical avatar that represents the students throughout their extended practice sessions (upper-left corner of Figure 2). Points earned while self-explaining texts can be spent to make changes to the avatars. Students can spend points to add accessories, to their character (hat, sunglasses, hair, etc), or change the color of existing accessories. This avatar remains on screen during the interaction and for each student it serves as their iSTART-ME identity. This feature should provide students with a substantial amount of control and personalization. These character customizations are also designed as purchasable replacements to reduce the amount of time spent "playing" with their character. Once a change has been purchased, it is automatically applied, and remains until another change is purchased.

2.3.3. Mini-Games

Several educational mini-games have been designed to be incorporated within the iSTART-ME extended practice module. The mini-games have been designed to use the reading strategies presented within iSTART. Some games require identification of strategy use, while others may require students to generate their own self-explanations. Most of the games incorporate an internal point system; however, those points are not carried over into the iSTART-ME point system. Students can only earn spendable iSTART-ME points by self-explaining texts within extended practice.

The mini-games are designed to incorporate increasing difficulty changes. By this we mean that as students progress within each of the mini-games, the task difficulty increases accordingly. This change is represented as starting a new level within the game and is achieved by having the students interact with increasingly difficult texts, identify an increasing number of strategies within a self-explanation, or complete an increasing number of tasks to reach the next level or to finish the game. The majority of the games are designed to incorporate the same leveling structure and can be completed within 10-20 minutes. After completion of a mini-game, students are directed back to the main iSTART-ME selection screen.

3. Conclusions, Expectations, and Future Directions

The overarching goal of the iSTART-ME project is to further our understanding of the benefits of adding game-based elements to ITSs (coined ITaG, Intelligent Tutoring and Games, by McNamara et al., 2009). The ultimate goal of the iSTART-ME project is to assess the benefits of motivational aspects of tutoring systems in terms of both learning and enjoyment, particularly in sustained learning situations. One approach to this issue is to examine how to incorporate learning principles such as those instantiated in successful ITSs within games. Indeed, the incorporation of learning principles into games is likely to significantly improve their effectiveness. However, we are exploring the other side of the coin. The approach we describe here is the integration of game components with an established ITS, in this case iSTART.

The development of iSTART-ME will allow us to examine the effectiveness of a combined ITaG system, as well as to more systematically evaluate the effects of game components in the context of an ITS. The current system has been designed with distinct and separable features so that multiple combinations can be tested across a variety of experiments. Thus, many questions can be investigated. For example, does the presence of a personalized avatar increase interest, engagement, or investment within the environment (or does it detract from the interaction as some form of "seductive detail")[28]? What types of rewards are sought most often by students (rewards that provide control over the environment, that customize their identity, or that provide entertainment)? Do the educational mini-games contribute to any changes in learning, interest, and motivation or do they simply serve as a distraction and a change of tasks? Addressing and answering such questions using systems such as iSTART-ME will further advance our understanding of the effects of game-based features as well as the effects of motivation during learning.

A further consideration regards the effects of game components in combination versus individually. We suspect that many game components may result in interactive effects, such that they produce one pattern in isolation and another in combination with other features. For example, points may need to be paired with levels or rewards in order to provide more meaningful game-play. Along these lines, research has shown that adding game-like features in isolation (without context) provides no benefits to either learning or motivation [9]. Simply adding arbitrary points or levels to an ITS will likely not result in any significant learning changes, nor will it improve motivation. In order to be effective, these game features must be fully integrated within a coherent system, and must serve some purpose. This "purpose" could be as simple as using points to compare performance between students (competition), or it could require a combination of multiple features, such as redeeming points to modify a personal avatar. With this in mind, the iSTART-ME project has been designed to incorporate multiple game-based features in such a manner that each feature can be contextualized and provide some specific gaming purpose.

A notable omission from the iSTART-ME game elements is narrative. This feature was specifically omitted from its design for two principal reasons: orthogonality of game elements and longevity of interaction. Situating the extended practice within a narrative would have made integration and separation of other elements difficult, and it would inherently limit the duration of interaction (whenever the plot/storyline ends). Excluding the narrative element maintains orthogonality between game features, still allows for natural subgoals (advancing through levels), and provides an open ended interaction with almost limitless dynamic game-play. Perhaps more important is the challenge of incorporating narrative into an existing ITS. Whereas some game-based features seem to combine relatively seamlessly with an existing ITS structure (e.g., points, rewards, avatars, personalization), narrative by contrast seems to call for a different structure altogether (requires an additional method of conveying the plot and integrating/scripting the content within a storyline). As such, the use of narrative may arise more naturally when ITaGs are developed from the ground up (rather than post-development integration as we describe here). Nonetheless, narrative is a considerable feature of many games and thus we hope that other researchers explore the benefits of such game components to ITSs and other computer-based training systems.

In closing, the iSTART-ME project is clearly a work-in-progress. We expect that many modifications and improvements will arise as we more fully develop and test the system. Our future research agenda entails examining the benefits and effects of iSTART-ME, as well as systematically examining the effects of the various game-based features. Ultimately, we expect hybrid ITaG learning environments to dramatically impact, if not revolutionize, the effectiveness of computer-based training as well as further our understanding of the complex motivational aspects of learning environments and their interplay with learning. In sum, developing systems that exploit the best practices of games and ITSs is expected to significantly enhance learning systems, as well as our understanding of learning environments. iSTART-ME is a first step toward that goal.

Acknowledgements

This research was supported in part by the Institute for Educational Sciences (IES R305G020018-02; R305G040046, R305A080589) and National Science Foundation (NSF REC0241144; IIS-0735682). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES or NSF.

References

[1] S. Combs. Current Trends in the Media Industry: File Production, Television Commercials and Video Games, *The Current and Potential Economic and Fiscal Impacts of Texas' Moving Media Industry*, Texas Controller of Public Account report (2008), 4-7.

[2] D.S. McNamara, G.T. Jackson, & A.C. Graesser. Intelligent tutoring and games (ITaG). In Y. Baek, Ed. *Gaming for Classroom-Based Learning: Digital Role Playing as a Motivator of Study*. (submitted). IGE Global Press.

[3] D. Crookall, R.L. Oxford, & D. Saunders. Towards a reconceptualization of simulation: From representation to reality, *Simulation/Games for Learning* **17** (1987), 147-171.

[4] Garris, R., Ahlers, R. & Driskell, J.E. Games, motivation, and learning: A research and practice model, *Simulation and Gaming* **33** (2002), 441-467.

[5] T.W. Malone. Towards a theory of intrinsically motivating instruction, *Cognitive Science* **4** (1981), 333-369.

[6] S. Barab, M. Thomas, T. Dodge, R. Carteaux, & H. Tuzun, Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development,* **53** (2005), 86-107.

[7] W.L. Johnson, S. Kole, E. Shaw, & H. Pain, Socially intelligent learner-agent interaction tactics. *Proc. of the 11th Int'l Conference on Artificial Intelligence in Education (AIED)*. (2003) Amsterdam: IOS Press.

[8] K.E. Boyer, R. Phillips, M. Wallis, M. Vouk, & J. Lester, Balancing cognitive and motivational scaffolding in tutorial dialogue, *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems*, Montreal, Canada, (2008), 239-249.

[9] G.T. Jackson, & A.C. Graesser, Content matters: An investigation of feedback categories within an ITS. In: Luckin, R., Koedinger, K.R., Greer, J. (eds.) *Proceedings of AIED 2007 conference* **158** (2007), 127-134.

[10] A.C. Graesser, X. Hu, & D.S. McNamara, Computerized Learning Environments that Incorporate Research in Discourse Psychology, Cognitive Science, and Computation Linguistics (in press).

[11] D.S. McNamara, I.B. Levinstein, & C. Boonthum, iSTART: Interactive strategy trainer for active reading and thinking, *Behavior Research Methods, Instruments, and Computers* **36** (2004), 222-233.

[12] I.B. Levinstein, C. Boonthum, S.P. Pillarisetti, C. Bell, C., & D.S. McNamara, iSTART 2: Improvements for Efficiency and Effectiveness, *Behavior Research Methods Instruments and Computers* **39** (2007), 224-232.

[13] D.S. McNamara, SERT: Self-explanation reading training, *Discourse Processes* **38** (2004), 1-30.

[14] D.S. McNamara & J.L. Scott, Working memory capacity and strategy use, *Memory & Cognition* **29** (2001), 10-17.
[15] T. O'Reilly, R. Best, & D.S. McNamara, Self-explanation reading training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, Mahwah, NJ: Erlbaum, (2004), 1053-1058.
[16] J.P. Magliano, S. Todaro, K.K. Millis, K. Wiemer-Hastings, H.J. Kim, & D.S. McNamara, Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research* **32** (2005), 185-208.
[17] T. O'Reilly, G.P. Sinclair, & D.S. McNamara, Reading strategy training: Automated versus live, *Proceedings of the Annual Meeting of the Cognitive Science Society*, Austin, TX, (2004), 1059-1064.
[18] J.D. Bransford, A.L. Brown, & R.R. Cocking, *How people learn.* Washington, D.C.: National Academy Press (2000).
[19] A.C. Graesser, X. Hu, & N.K. Person, Teaching with the Help of Talking Heads, *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, Los Alamitos, CA: IEEE Computer Society (2001), 460-461.
[20] M.M. Louwerse, A.C. Graesser, A. Olney, & TRG, Good computational manners: Mixed-initiative dialog in conversational agents, *Proceedings of the AAAI Fall Symposium on Computer Etiquette* (2002).
[21] D.S. McNamara, C. Boonthum, I.B. Levinstein, & K.K. Millis, Evaluating Self-Explanation in iSTART: Comparing Word-based and LSA Systems, In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch eds, *Handbook of Latent Semantic Analysis*, Lawrence Erlbaum, Mahwah, NJ (2007), 227-241.
[22] G.T. Jackson, R.H. Guess, & D.S. McNamara. Assessing cognitively complex strategy use in an untrained domain. *To appear in the Proceedings of the annual meeting of the Cognitive Science Society.* (in press).
[23] P.A. Alexander, P.K. Murphy, B.S. Woods, K.E. Duhon, & D. Parker, College instruction and concomitant changes in students' knowledge, interest, and strategy use: A study of domain learning, *Contemporary Educational Psychology* **22** (1997), 125–146.
[24] A. Bandura, Self-efficacy: The foundation of agency, in W. Perig & A. Grob (Eds.), *Control of human behavior, mental processes, and consciousness: Essays in honor of the 60th birthday of August Flammer* , Mahwah, NJ: Erlbaum (2000), 17-33.
[25] F. Pajares, Self-efficacy beliefs in academic settings, *Review of Educational Research* **66** (1996), 543-578.
[26] P.R. Pintrich, Multiple goals, multiple pathways: The role of goal orientation in learning and achievement, *Journal of Educational Psychology* **92** (2000), 544-555.
[27] B.J. Zimmerman, & D.H. Schunk, Reflections on theories of self- regulated learning and academic achievement, in B.J. Zimmerman, & D.H. Schunk, (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives,* Mahwah, NJ: Erlbaum (2001), 289-307.
[28] R. Garner, M. Gillingham, & C. White, Effects of seductive details on macroprocessing and microprocessing in adults and children. *Cognition and Instruction* **6** (1989), 41-57.

Serious Games for Professional Ethics: An Architecture to Support Personalization

Adam MCKENZIE and Gord MCCALLA
ARIES Laboratory
Department of Computer Science
University of Saskatchewan
176 Thorvaldson Building
110 Science Place
Saskatoon, Saskatchewan, S7N 5C9
Canada
akm220@mail.usask.ca, mccalla@cs.usask.ca

Abstract. Ethics is difficult to teach because it is an ill-defined domain. There are no right or wrong answers and it is expensive and difficult to give learners realistic experience in making ethical decisions. In this paper we will describe an architecture for a serious game, Conundrum, that will give learners experience in making ethical decisions in realistic scenarios. Students will be able to take on various roles in each scenario and make decisions that lead to consequences, both positive and negative, anticipated and not. Through game play they will gain a much deeper appreciation of the ethics issues that can arise in the construction and deployment of information technology. To further student reflection, models of student behavior will be made available (opened) to the student during and after game play, and can be used by the system to personalize its interactions with each student. We also hope to provide tools to the students that will allow them to build their own scenarios, thus further deepening their ability to understand ethical situations. The game is being designed to fit into the regular curriculum, as part of a senior undergraduate Computer Science ethics course, and will be tested in this course when it is next offered.

Keywords. Ethics, Serious Games, Open Learner Modeling

Introduction

Ethics is taught to Computer Science students during their undergraduate degree at most post secondary institutions, either in its own course or as a part of the rest of their courses. Teaching ethics is difficult because it is an ill defined domain [8] and has no explicit right or wrong answers. Different ethical decisions can be defended based upon different criteria and it often depends on the context of the situation. Currently there are three major ways to help learners appreciate the necessity of understanding the ethical dilemmas they may encounter in their career or the social context their solutions will be embedded in.

The first is to do a retrospective case study of a particular ethical dilemma that occurred [1]. Usually this is a high profile case in the media such as Therac-25 case [11]. Case studies are great at allowing students to develop their skills in identifying stakeholders and analyzing the ethical dimensions of a particular context but do not give them experience making ethical decisions in a realistic context with imperfect information.

Learners can also be taught ethics through the learning and application of ethical frameworks such as Kantianism or Utilitarianism [15]. This gives them a tool set that they can use in making ethical decisions and analyzing ethical issues from different perspectives. However, it doesn't allow learners to make ethical decisions in a realistic and practical context.

Role playing is the last major approach for teaching ethics to Computer Science students. The instructor creates a stage for an ethical drama and has roles that she or the learners can take on. The different participants in the drama interact to explore the different perspectives the stakeholders have in the ethical context. Unfortunately, all of the participants need to be committed to taking the perspective required of their role. If they aren't committed they won't explore the deeper issues of the context or understand what it means to have that role in a real situation. Role playing is most effective when the instructor facilitates it so that it stays on track exploring the issues, which usually means taking up considerable class time.

None of these approaches is perfect. A new approach should be able to provide students with the ability to make ethical decisions in a realistic context so that they will be prepared for their own ethical dilemmas. It shouldn't require the instructor to facilitate but it should allow the instructor to have feedback on the learner's progress. It would ideally cause the students to reflect on what they would actually do in a particular ethical situation. Finally it should be personalized for each learner as they each have differing skills and need to experience different aspects of the ethical situations. This suggests a fourth approach: a "serious" game that students could play that would give them experience in making ethical decisions in realistic simulated situations.

A serious game should be highly motivating for students [7] and would be an ideal complement to case studies, ethical frameworks and role playing. The game play would involve students making ethical decisions where the consequences would be dependent on those decisions. The game play would take place in a realistic social context created in a virtual environment supported by the game. Ideally, the game would allow multi-players. As the students play the game they won't need the instructor to facilitate their minute to minute decision making; instead, the game could keep track of each student's progress in a learner model and report it to the instructor who could intervene as needed. The learner models could also be opened to the students to encourage reflection and could support personalization to appropriately support or challenge each particular learner. Finally, a set of tools that make it easy to build new scenarios in the game could be used by either the instructor to enrich the game play, or even by the students to build their own scenarios, thus furthering the depth of their understanding of ethical issues.

This is not the first research done into building realistic virtual worlds to support learning. In the ethics domain Winter & McCalla [18] built a game for exploring ethical decisions for the ethics class at the University of Saskatchewan. It was text based and was limited to students answering yes and no questions to affect the outcome of an ethical scenario. There was also only one possible outcome for each choice.

There has been an interactive narrative designed to teach ethics [8]. In an interactive narrative a director monitors a scenario where the learner can make certain pivotal decisions that will affect the outcome of the scenario. As the scenario progresses the director alters the scenario path to allow for the most pedagogically effective experience for the learner. In this case, the learner is presented an ethical situation and they engage in a Socratic discussion with the system.

Beyond ethics, the Tactical Language Training System [10] has been designed for US Army personnel to practice making decisions in social situations they might face in Iraq and other battle zones. TLTS uses realistic scenarios in which are embedded difficult decisions the player must make. The player has to interact with simulated agents that are intelligent in how they represent their culture and can carry out real time speech processing. TLTS demonstrates that AIED and game design techniques can be blended together.

Another educational game is based on *goal based scenarios* [16]. This game has been used to help urban planners make difficult subjective decisions. The learners move towards a goal that is designed to expose them to micro skills that they will use later when actually carrying out urban planning in a real situation.

1. Conundrum

Conundrum is a serious game that allows learners to experience ethical decision making in realistic scenarios. We hope to capture enough complexity in the ethical scenarios to challenge a learner's preconceived notions about the consequences of their ethical choices. We also want to have both ethical and unethical paths with both positive and negative consequences so that we can choose consequences based on pedagogical purposes. There should also be no explicit "winning" or "losing" scenarios, only different stakeholders' perspectives. We want the learner to reflect on what they have chosen and the consequences of those choices.

Conundrum represents these ethical scenarios in an XML format as shown in Figure 1 to support the flexibility required to develop new scenarios and to make it easy to communicate between the game's components. The structure of the XML draws from a metaphor of drama productions. Each stakeholder encountered in the game is an actor. Some of the actors have roles that the player can take while others are controlled by the game. Both game and player controlled actors have choices available to them. The choices available to the player are there to allow them to explore realistic ethical decision making. The choices available to the game are for guiding the player down pedagogically relevant paths.

Each ethical situation that the learner plays is encoded as an *act*, a sequence of *scenes* that lead to a conclusion. Each scene represents one of: the start of the act, a consequence from a student choice, or the end of an act. The player moves from scene to scene making important ethical choices. In each scene there are individual pieces of *dialog* that the player can use to gather further information on the ethical context and to make ethical choices. An *intention* is the XML entity for what appears in the thought bubbles for the characters. These represent possible choices available to the player that they can click on to choose for their actor to say. The *speech* is what the actors in the scene actually say aloud to one another after choosing one of the intentions. The *result* specifies the resulting dialog or scene after an *intention* is chosen for an actor and the *speech* is said aloud.

```
<act id="A">
<scenes firstscene="A1">
```

```
<scene id="A8">
<scenedialog firstdialog="A8A">
<choices>
    <choice id="C2">
        <intention>Ask customer for more time</intention>
        <speech>Let's ask the customer for more time.</speech>
        <result type="scene" id="A9">
            <ethicalprinciple type="can't avoid choices" amount="0.7"/>
        </result>
    </choice>
    <choice id="C2">
        <intention>Ask customer to omit the security tests</intention>
        <speech>I think it is best if we ask the customer to skip the security tests to meet the
deadline.</speech>
        <result type="scene" id="A10">
            <ethicalprinciple type="can't avoid choices" amount="0.7"/>
        </result>
    </choice>
    <choice id="C4">
        <intention>
        <ethicalprinciple type="can't avoid choices" minimumthreshold="0.1">
            Don't tell the customer, it will work out fine.
        </intention>
        <speech>I don't think we need to tell the customer, I think we can meet the deadline.</speech>
        <result type="scene" id="A11">
            <ethicalprinciple type="can't avoid choices" amount="0.1"/>
        </result>
    </choice>
</choices>
<dialogs>
<dialog id="A8B">
    <actor role="Project Manager" name="Steven"/>
    <choices>
        <choice id="C2">
        <choice id="C3">
        <choice id="C4">
    </choices>
</dialog>
…
```

Figure1: Conundrum Example XML

Conundrum is designed as a basic client server architecture as shown in Figure 2. The server will act as a centralized repository for the acts, learner models and messages. This will allow learners to access acts as they are altered by either the instructor or fellow students. The server will also allow their clients to access the other player's learner models and allow the instructor to see their progress through the game. Finally the server will route messages sent between actors in the game.

The client has several components including an *act processor*, *director*, *learner model*, *open learner model*, and *act editor*. The first subsystem of the client required for Conundrum is the act processor that takes the act in XML form and converts it into a graph that can be crawled as the player makes their ethical choices.

After the act has been processed and the player starts an act, the *director* prunes dialog choices on a scene by scene basis, drawing on the learner model of the player to help individualize the set of choices for a particular student playing the game. For example, if the player has been previously identified to know that they cannot avoid making ethical decisions then the *director* could prune dialog options and responses that are designed to demonstrate this. This guides the player to try other dialog choices that lead to the unexamined aspects of their learner model.

The learner model keeps track of the progress of the learners through the game. This includes all the dialog choices they have made previously to allow instructors to review the learner's play sessions. The learner model also has the decisions the learners have made that could affect how other acts might proceed. For example, if the player chooses to whistle blow about their company, an employer in subsequent acts might not trust the player as much given his history. The other dimensions of the model are based on the Impact CS report [9]. The Impact CS report was created by a steering committee of ethicists, philosophers, sociologists and computer scientists from universities in the United States. They determined the ethical skills, social skills, ethical principles and social principles that a computer scientist should develop during their undergraduate program as shown in Table 1.

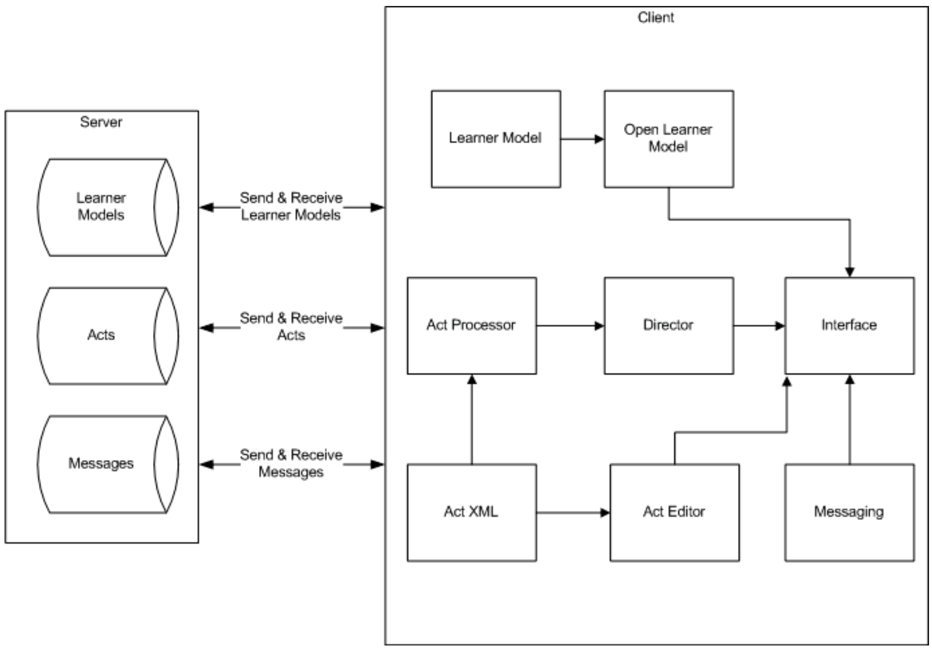


Figure 2: Conundrum Architecture

Each of these skills and principles is represented by a value normalized between 0 and 1 in the learner model. When a learner has 0 in a skill or principle it means they don't appear to know the principle or to be able to apply the skill. If they have a 1 in the learner model, they appear to have mastered the principle or skill. The dimensions are initialized to 0.5 to represent that we don't know initially whether they know the principles and skills. Each ethical decision is encoded with the skills and principles it represents with a value between 0 and 1. This is combined with the existing value in the learner model using an aging function to allow new evidence to alter the model but allow it to retain a memory of previous evidence. Skills and principles that are unrelated to the current decision are not affected in the model. Choices can also be encoded with thresholds that the *director* can enforce against the student's learner model. If the player's learner model doesn't meet the threshold they are not allowed to see the choice or choose it.

Table 1: Ethical Skills, Ethical Principles, Social Skills and Social Principles of Impact CS

Ethical		**Social**	
Skills	**Principles**	**Skills**	**Principles**
Using ethical skills in concrete situations.	Every ethical situation can be discussed rationally.	Determining the social context that the software will be used in.	The context changes the design and technology used.
Using examples, analogies or counter-examples to argue a point.	Ethical choices can't be avoided and are embedded in design decisions in a social context.	Finding assumptions and tradeoffs in design.	Power relations exist in all relationships.
Identify and evaluate possible courses of action.	Easy solutions are often questionable ones that will lead to problems.	Using empirical evidence to decide the balance of the tradeoffs.	The context changes the design and technology used.
Identify stakeholders in the situation and the ethical principles that apply.	It is possible to defend an ethical position with reasons.		Designers make tradeoffs.
			Populations are diverse, use empirical data to decide the tradeoffs.

As the player proceeds through the acts they are able to click on any actor's avatar and view an open learner model representing that actor. The open learner model appears like game play statistics from video games such as Left 4 Dead [17], character sheets from RPGs such as World of Warcraft [5] or a profile page for a social networking site (basing the model on such familiar systems should help the learner understand the model and how to navigate it). The model shows the actor's picture, name, position, work history, skills and principles exhibited so far. By exposing the internal representation of the learner's knowledge we hope to encourage reflection [12]. The open learner model is a subset of the learner model that only displays their progress through the acts through the work history and their estimated progress on the Impact CS ethical skills, social skills, ethical principles and social principles.

The interface component allows the learner to have conversations with other actors making ethical choices for their avatars, messaging other learners, viewing the open learner models and creating new acts, scenes and dialogs.

The act editor allows instructors or learners to create new alternatives, scenes, and acts. Every act is created by creating the first scene with its dialog options and moving progressively down each possible path the learners will be able to follow. Each scene is independent of other scenes so that acts can be created incrementally and as new options are required, for example because of learner input. Current scenes can be edited and new scenes added without affecting the act as a whole.

2. Sample Act

To ground the architecture design, consider an example act where the learner has to balance the security of a software project with a strict deadline. The player in this act takes on the role of a new project manager working at a software design firm on a project that is nearly complete. The act begins by him conversing with his manager about the new project. The conversation is controlled by the player through choosing conversation options. The player in this case asks the manager to illuminate why the deadline is strict. He responds that the client is currently licensing the software they are using from another software firm. This software is mission critical to their organization and their license expires with the other software firm soon after the deadline leaving only enough time for a proper deployment of the new system. If the player's company cannot deliver on its deadline, the client may choose to stay with the other software firm.

The player then asks the manager what the current status of the project is. The manager replies that the project is in the testing phase with a quality assurance period followed by a rigorous testing protocol for security. The project team is concerned that the testing will not be completed by the deadline. Inquiring further, the player discovers from the manager that even if they start working overtime, the project won't finish on time but he is willing to risk this if that is what the player advises. The player then asks the manager if they omit the security testing will the project get back on schedule. The manager replies that it most certainly would. However, this could leave the client open to a security vulnerability that if exploited would hurt their reputation with the client. It also might be unethical to advise a client to omit their security testing when the original requirements for the project indicated it as necessary. The player asks if it would be possible to ask the client for more time. The manager says that he will support the player's decision but it might mean the client will return to their original provider. He also reminds the player that this would eliminate the ability of the player to tell the client that it will finish on time. At this juncture the manager asks the player the question in Figure 4, "How do you want to proceed?" The player at this point needs to choose one of the three options in the thought bubbles by clicking on one.

The current player has a problem with trying to avoid ethical decisions. In previous acts he continually chose to assume everything will work out if he ignores ethical dilemmas and his learner model reflects this. He has not ignored making ethical decisions enough times for the *director* to remove the option of him telling the customer that the project will be finished on time. However, he has chosen similar decisions enough times that giving him negative consequences for choosing to ignore the problem yet again is pedagogically useful and it might help him consider another option next time. The player once again decides to ignore the problem by not telling the customer and assuming the project will finish on time. The learner model is updated to show that he has once again chosen to avoid facing the ethical dilemma.

The *director* at this point has two possible scenes for the player to experience. In the first scene the player has succeeded in completing the project on time and the client is pleased that there were no complications. In the second scene the player must argue with the client as to why he didn't give any indication that the project wouldn't finish on time. This ultimately culminates with the client choosing the software firm they are already licensing software from and suing the player's company for not completing the contract. Looking at the player's learner model the *director* chooses to have the project not complete on time which will hopefully make the player reconsider his options next time.

The player is surprised by the negative consequences, so he opens his open learner model by clicking on his actor and sees the representation of his learner model. He notices that, just as his learner model indicates, he doesn't make ethical choices and instead tries to avoid them. This causes him to reflect on how he has been making his ethical decisions and he agrees that he has been avoiding making difficult ethical choices. The player decides to retry the act to see what other options he had at that pivotal moment. Since the player's learner model indicates that he has avoided making ethical choices, the *director* decides he no longer has the option to tell the client that everything will be completed on time. The player notices this and decides to ask the customer for more time. The *director,* seeing that the player has not been making difficult ethical decisions previously, rewards the player for choosing something different by having the customer agree to more time. The player's learner model is updated to reflect his improved status at applying the ethical principle that ethical choices cannot be avoided.

Figure 4: Conundrum Example Game Play

3. Status of the Project

The current status in the implementation of the architecture shown in Figure 2 is that a single act has been fully created in XML. The act parser has been completed allowing act XML to be parsed into a traversable graph structure usable by the game. The dimensions of the open learner model have been identified so that new acts can be created around their skills and principles. The design of the interface and open learner model for the final proof of concept prototype has been completed. The game interface has been developed but will need more assets to be created including more actors, backgrounds and props for other acts. One of the scenes from the developed act in the interface can be seen in Figure 4.

What remains to be implemented is the open learner model and a messaging system so that learners can message each other in the game. A server needs to be developed to store the necessary data structures. The *director* needs to be implemented.

The learner model needs to be implemented including tweaking parameters for the aging functions to give accurate feedback. Currently the acts are created by writing raw XML so authoring tools will need to be created.

The proof of concept prototype will be evaluated in the ethics class at the University of Saskatchewan in spring of 2010. The metrics we hope to gather from the evaluation is how immersive and motivating the game is to the students and whether it has changed their ethical thinking process. We also plan to have the students that are evaluating Conundrum create their own acts as a way of learning by authoring, a variant of learning by teaching [13] [14]. This will allow the students to extrapolate from what they have learned in the course and from playing Conundrum to create new acts that their fellow students could learn from.

For this evaluation there will be some limitations so that the project will be tractable. The authoring tools don't have to be as robust as authoring tools for other domains because they will be used by either computer science instructors or upper year computer science students. The game will have simple 2D isometric graphics instead of current state of the art 3D graphics. We don't expect students to have acquired the skills needed to navigate 3D spaces virtually. We have based the gameplay and interface on previous games with ethical dilemmas such as Fallout [4], Baldur's Gate [2], Mass Effect [3] or the Witcher [6] which rely on a dialog system to discuss the ethical dilemmas.

The game at evaluation time will also be single player. Each player that is added to an act increases the number of scenes exponentially as each player has several different ethical decisions to make that combine to make new scenes. After several acts have been developed and tested we will revisit developing multiplayer acts which are supported by the architecture. The advantage to developing a multiplayer version is that the players will be able to experience the perspective of different roles and then switch those roles. For example, assume one player is the manager and the other player is an employee in an act. The manager player will make decisions that will impact the decisions the employee can make. Having them switch roles after playing through an act allows them to see the act from both perspectives. This would require the students to be playing the game at the same time, or at least wait for one another to make decisions. The *director* could take the role of any missing players and make their ethical choices estimating what they would have decided based on their student model. This would allow students to drop in and out of acts whenever it was convenient for them while their fellow players could continue playing the act, oblivious to whether the other roles in the act are played by the *director* or other players.

4. Conclusion

Conundrum is exploring supporting learners in exploring ethics, a domain that has not been researched extensively in AIED. It is a novel approach that blends open learner modeling and game design. It demonstrates that the two disciplines can be combined to deliver a motivational experience with feedback and reflection. Learner modeling is useful in serious games in that it allows personalization of the game play experience for individual learners. Learner modeling in serious games can be extended by using open learner modeling as feedback to the player. Learning by authoring is a promising capstone activity after learners have played the game. The architecture of Conundrum is robust enough to support incremental design of interesting new elements such as new ethical choices, scenes and acts.

5. Acknowledgements

The authors wish to thank the Natural Sciences and Engineering Research Council of Canada for their funding of this research project. We would also like to thank Scott Bateman for aiding the paper by reviewing it and Christopher Brooks for contributing the learning by authoring concept.

6. References

[1] Baase, S. (2008). *A Gift of Fire: Social, Legal, and Ethical Issues for Computing and the Internet.* Third Edition: Prentice-Hall.

[2] Bioware. (n.d.). *Baldur's Gate Official Site*. Retrieved May 1, 2009, from Bioware Website: http://www.bioware.com/games/baldurs_gate/

[3] Bioware. (n.d.). *Mass Effect Official Site*. Retrieved May 1, 2009, from Bioware Website: http://masseffect.bioware.com/

[4] Black Isle. (n.d.). *Fallout Official Site*. Retrieved May 1, 2009, from Fallout: A Post Nuclear Role Playing Game: http://www.nma-fallout.com/fallout1/official_site/

[5] Blizzard Entertainment. (n.d.). *Leeroy Jenkins Character Sheet*. Retrieved May 1, 2009, from The World of Warcraft Armory: http://www.wowarmory.com/character-sheet.xml?r=Laughing+Skull&n=Leeroy

[6] CD Projekt (n.d.). Retrieved May 1, 2009, from The Witcher Official Website: http://www.thewitcher.com

[7] Gee, P. (2003). *What Video Games Have to Teach Us About Learning and Literacy.* New York: Palgrave MacMillan.

[8] Hodhod, R., & Kudenko, D. (2008). Interactive Narrative and Intelligent Tutoring for Ethics Domain. *Workshop Proceedings of Intelligent Tutoring Systems, 20th International Conference*, (pp. 13-21). Montreal, Canada.

[9] Huff, C., & Martin, C. D. (1995). Consequences of computing: A framework for teaching ethical computing (First Report of the Impact CS Steering Committee). *Communications of the ACM*, 75-84.

[10] Johnson, W. L., Vilhjalmsson, H., & Marsella, S. (2005). Serious Games for Language Learning: How Much Game, How Much AI? *12th International Conference on Artificial Intelligence in Education.* Amsterdam, The Netherlands: IOS Press.

[11] Leveson, N., & Turner, C. S. (1993). An Investigation of the Therac-25 Accidents. *IEEE Computer, Vol. 26, No. 7*.

[12] Mitrovic, A., & Martin, B. (2002). Evaluating the Effects of Open Student Models on Learning. *Second International Conference of Adaptive Hypermedia and Adaptive Web-Based Systems* . Málaga, Spain: Springer Berlin / Heidelberg.

[13] Nichols, D. (1993) Intelligent student systems: Learning by teaching. In P.Brna, S. Ohlsson and H. Pain (Eds.) *Artificial Intelligence and Education: Proceedings of AIED'93*(pp. 576), AACE:Charlottesville, VA.

[14] Palthepu, S., Greer, J., McCalla G. (1991) Learning by Teaching. *In: Proceedings of the International Conference on Learning Sciences*, Illinois, USA.

[15] Quinn, M. J. (2008). *Ethics for the Information Age 3rd Edition.* Boston, MA: Pearson Education Inc.

[16] Schank, R. (1996). Goal-Based Scenerios: Case-Based Reasoning Meets Learning by Doing. In D. B. Leake, *Case-Based Reasoning: Experiences, Lessons & Future Directions* (pp. 295-347). AAAI Press / MIT Press.

[17] Valve Corporation. (n.d.). *Lord Perekles Left 4 Dead Stats*. Retrieved May 1, 2009, from Steam Community: http://steamcommunity.com/id/LordPerekles/stats/L4D

[18] Winter, M., & McCalla, G. (1999). The emergence of student models from an analysis of ethical decision making in a scenario-based learning environment. *Seventh International Conference on User Modeling* (pp. 265-274). Banff, Canada: Springer-Verlag New York, Inc.

AEINS: Adaptive Educational Interactive Narrative System to Teach Ethics.

Rania HODHOD [a,b] Daniel KUDENKO [a] and Paul CAIRNS [a]

[a] *Computer Science Department, University of York*
Heslington, York, YO10 5DD, UK; E-mail: rania.hodhod@cs.york.ac.uk.

[b] *Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt*

Abstract. Student modeling has proven its importance over the years. It allows intelligent tutoring systems to offer adaptive teaching which has proven its usefulness in the educational process. Ill-defined problems are characterized by the ambiguity in classifying a problem's answer as correct or incorrect, in addition to the lack of defined rules that helps in transferring the solution path from the initial step to the final step. These properties make it very difficult to use intelligent tutoring to teach in an ill-defined domain. In this paper we present a model that uses an intelligent tutor in an educational narrative environment to assist teaching in the citizenship and ethics domain. The tutor monitors and analyzes students' actions in order to provide an individualized story-path and a personalized learning process. The paper also discusses preliminary evaluation results.

Keywords. Intelligent tutoring, interactive narrative, ill-defined domains, ethics and citizenship

1. Introduction

Problems around us can be classified as either well-defined or ill-defined. Well-defined problems are characterized by having a clearly defined domain structure that can be modeled and well-defined operators that transfer the solver from the initial state of the problem answer to the goal state. The well accepted models of these domains make it easy to unambiguously classify problem solutions as correct or incorrect. On the other hand, we consider problems that lack these characteristics, ill defined; there is no single right or wrong answer to cases nominally of the same type. Problems can involve more than one aspect that inter correlate to different degrees with each other. Learning how to solve ill-defined problems can not be achieved through acquiring knowledge only but it tends more to be problem based learning or inquiry based learning. This potential difficulty in ill-defined domains such as argumentation, history, and ethics makes them of great interest to computer-supported learning.

Ethics and citizenship is one of the important ill-defined domains; the development of skills of participation and responsible action is a fundamental part of the citizenship curriculum. Teaching citizenship and ethics at schools has used discussions and brainstorming possible solutions to moral dilemmas [10] and classroom games [5]. Allowing students to be involved in moral dilemmas helps them to express their character through the kind of choices they make. This kind of problem solving and decision making allows children to learn about basic human values including honesty and kindness. In addition they help in the student's cognitive development. Such development is not mainly the result of gaining more knowledge, but rather consists of a sequence of qualitative changes in the way an individual thinks [14].

For some children the classroom environment is not private enough to talk freely and express themselves, so they may feel shy to present certain unethical actions because this not what is expected from them. Children also have different personalities, different children need different ways of learning and getting feedback. This means that every single child may need to deal with particular dilemmas in order to tackle his needs directly and discover his weak points in order to deal with them in the right way. Within the classroom environment this is very difficult to address. Teachers are not able to focus on every single child and discuss separately his/her entire thoughts - time and space do not allow them to do so. In addition, giving every child enough time to revise the whole experience and to self reflect is very difficult because of the mentioned constraints.

Computers offer much in addressing the problems of classroom teaching. They offer the required privacy and the safe environment which children can explore, and gain the appropriate experience especially when guidance is provided. They allow the inclusion of many different dilemmas that the child can interact with and learn from, and hopefully transfer what they learn to the real world. Most importantly they can offer adaptation that provides personalized teaching and feedback. On this basis, an educational narrative-based system is introduced that tackles the classroom teaching problems and follows the constructivist view in order to teach in the citizenship and ethics domain. The student will be able to learn by interacting with the surrounded environment and by being involved in moral activities, where he has to take actions and decisions that affect himself and others. Such interaction helps students to move from the *making moral judgments state* to the *taking moral actions state*, from the knowing state to the doing state, which we consider a very important step in moral education.

2. Educational Theories

Constructivist teaching is based on the belief that students learn best when they gain knowledge through exploration and active learning. Education is centered on themes and concepts and the connections between them, rather than isolated information [9]. Students are encouraged to think and explain their reasoning instead of memorizing and reciting facts. Piaget suggested that the educator must provide students with opportunities for personal discovery through problem solving, rather than indoctrinating students with norms [13]. Moral dilemmas are taught at schools following Piaget, using the Socratic Method of questioning. In this model, according to [8], the teacher asks a series of questions that lead the students to examine the validity of an opinion or belief. This is a powerful teaching method because it actively engages the learner and forces critical thinking, which is just what is needed in examining ethics, values, and other character issues. The method is also dramatic and entertaining, and it triggers lively classroom discussion. The Socratic Method displays its strengths when the students made a bad choice. Through discussion, students should then be forced to face the contradictions presented by unethical actions. This method requires a delicate balance between letting the students make decisions, and demonstrating the limits in their reasoning. Finally, "raising the ante", which is defined as raising the stakes and introducing consequences, is a tactic followed if the student sticks with an unethical choice.

Simpson in her article emphasizes the importance of stories in our lives and their role in tightening the human relationships [6]. It can be seen that stories and moral dilemmas help in emphasizing the moral behavior and the strength of the actor. This hits directly the human emotional level and has an impact on the reader or listener. By this way stories are acting as a source of inspiration and direction for moral conduct; help to incorporate ideals into people's lives. In addition, encouraging students to think critically is a very important issue [1] as it allows an appropriate amount of choices during ill-structured investigations that lead to the development of inquiry skills. In [4], it has been shown that even in domains where it is impossible to make sharp distinctions between "good" and "bad" solutions due to the lack of ideal solutions or a domain theory, the solution differences are meaningful. The Socratic Method can be an example where the student's answers to the questions reflect their own beliefs and thoughts. We aim through this work to allow the student to interact with moral problems and direct them to focus on the underlying justice or human welfare considerations of the proposed situation.

3. Related Work

Intelligent tutoring systems get a share in assisting learning in ill-defined domains, for example in the cultural domain to enhance good cultural discussions [2]. In the domain of ethics and citizenship, analyzing ethical dilemmas using diagrams has been discussed in [4]. And analyzing bioengineering ethics cases through framing has been developed in [12].

Other attempts appear in the utilization of intelligent tutoring and interactive narrative techniques in narrative_based learning environments in which narrative is a key instrument in facilitating teaching and learning, for example the IN-TALE system developed for military training skills [17]. Although IN-TALE provides an adaptive practice environment, it does not provide hints, give feedback, or attempt to re-mediate incorrect understanding of concepts. Another system is FearNot! [18], a character-driven learning environment developed for the anti-bullying domain. Although the environment emphasizes highly autonomous agents that promote empathic relationships with the student, it does not have any kind of student model. ISAT system developed by [3] attempted the missing student model issue in the IN-TALE and the FearNot! systems, however, it is not clear in the ISAT system if it is always the case that the choice of the next story plot depends on the skill model or sometimes the director may choose a plot for its dramatic relevance to the story instead. If this is the case, this means that the director agent is not continuously considering the skill model in its choices, which may allow the choice of a plot that can dramatically fit but does not the current student's skills and the educational targets.

4. Architecture

The main vision of our work is to integrate interactive narrative and intelligent tutoring that makes use of a student model, see $Fig.1$, where learning objectives are always un-

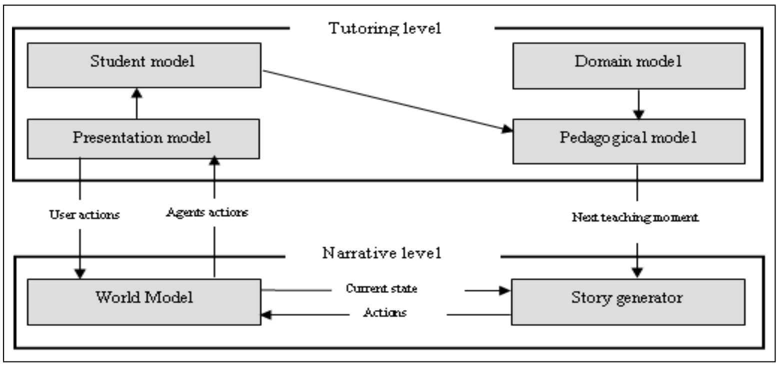


Figure 1. An architecture for a narrative-based system that includes intelligent tutoring

derpinned by effective story telling. Each module in the system is fully separated from the other parts; all the parts can communicate through set of production rules. Based on this architecture a narrative-based educational narrative system called AEINS has been developed. For more detailed information about the general approach AEINS takes and how it works can be found at [15].

The architecture reflects the importance of having intelligent tutoring facilities in order to keep the educational process as effective as possible. The intelligent tutor makes use of a domain model, a pedagogical model, a student model, and a presentation model to reason about the learning process and the student behavior. The narrative consists of the story world and the story generator, which interact with the intelligent tutor. The following subsections introduce these components and explain their role in the learning environment. Based on this architecture, a prototype called AEINS (Adaptive Educational Interactive Narrative System) has been developed, see $Fig.2$. AEINS is an inquiry-based learning environment, that helps 8-12 year old children to acquire certain skills in the citizenship and ethics domain. AEINS generates a continuous story, where moral dilemmas (so-called teaching moments) form the focal points of the story. The story generation is similar to the GADIN system [11], where it is based on STRIPS planning that is currently able to provide a plan from any current state to the next dilemma.

The teaching moments are a crucial part of the story generation process. They have certain prerequisites that must be fulfilled before the execution of the teaching moment takes place. Each teaching moment represents a part of the whole story and focuses on certain domain concepts. The interaction of the student with the teaching moments is monitored and evaluated. Feedback and explanation are part of the teaching moments, they are presented according to the students actions and choices. The understanding students gain through this process is situated in their experience and can best be evaluated in terms relevant to their experience. Some teaching moments use Kohlberg's moral dilemmas for their story lines, an example can be found in [15].

The teaching moments have two types of preconditions; tutoring and narrative. The teaching moment is selected according to the current student model (tutoring precondition). Then the teaching moment is presented after the system satisfies the story preconditions (narrative prerequisites) through the planner (story generator). Within the teaching moment the learning tasks are tightly coupled with the narrative where specific skills (goals) have to be acquired by the student. The interaction of the student with the teaching moments is monitored and evaluated.

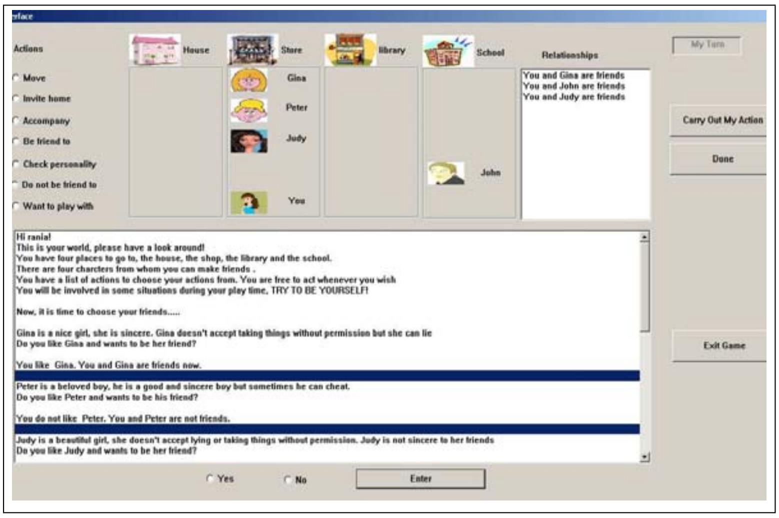


Figure 2. AEINS Interface

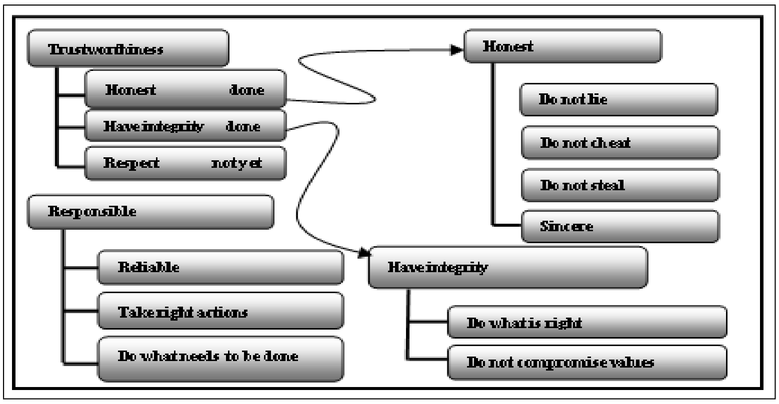


Figure 3. Domain model representation

4.1. Domain Model

The domain model describes the various concepts in the ethics domain and their relationships. The principles of character education [7] are used to structure the domain model. The domain is represented in the form of hierarchical frames that describe the progression of the learning concepts. $Fig.3$ shows that some values are independent such as trustworthiness and responsibility, but others like trustworthiness and honesty are interdependent. Put it differently, Mastering the trustworthiness value requires honesty, integrity, and responsibility values to be mastered first. The domain model also contains a repertoire of teaching moments.

4.2. Pedagogical Model

Wenger has defined a learning pedagogy in which learning can be viewed as successive transitions between knowledge states. Based on Wenger [**?**], we developed the pedagogical model in the form of production rules. These rules are used to give the system specific cognitive operations to reason about the student and the teaching process. The model specifies how a student ideally would use the system and how the system reacts to his actions. In addition, the model updates the student model and chooses the next state in the teaching process, considering the student's model history and the domain model. A representation of the model is given below.

```
Trigger: action ("student", "TM1", "agree_to_lie")
   and action ("student", "TM1", "insist_to_lie")
   and action ("student", "TM1", "lie_for_friend_sake")
   and action ("student", "TM1", "agree_lying_is_bad)
Action: skill_mastery ("student", "do_not_lie", "acquired",
0.6)

Trigger: teaching moment TM1 has not been presented
   and teaching moment TM2 has not been presented
   and the be_sincere value is not held by the user
   and the do_not_lie value is held by the user
Action: set priority to teaching moment TM2
```

The above representation denotes that if (a) a specific pattern of teaching moments (TM_1 and TM_2) has not been presented to the student yet and (b) learner holds certain values (sincere) and does not hold others (do not lie), the action part of the rule executes (teaching moment TM_2 has priority over teaching moment TM_1). If several rules have been satisfied, this results in having more than one teaching moment suitable to be presented next to the learner. In this case, one of these teaching moments is chosen randomly. The chosen teaching moment is presented after the system satisfies the story preconditions (narrative prerequisites) through the story generator (planner). Within the teaching moment the learning tasks are tightly coupled with the narrative where specific skills (goals) have to be acquired by the student.

The pedagogical model provides explicit and implicit feedback, for example if the student seems not to be aware of stealing as mis-value, one of the agents will verbally state "don't you know that stealing is wrong and you can go to jail" or in another situation feedback can be implicit, for example if the student was following a good model, the student's friend can say something as "your mum is proud of you!!", which implicitly denotes positive feedback.

4.3. Student Model

AEINS builds a model of the student's learning characteristics by observing, recording the student's actions and comparing them to the pre-specified model. The student model is represented by rules similar to those of the pedagogical model but associated with confidence value that quantify the reliability of (or degree of confidence in) the rule. Each concept requires a number of skills to be mastered in order to set the state of the concept to learned. Given the following representation, a model has been developed that infers the character stereotype.

```
IF skill ("student", "do_not_lie", "acquired", X1)
   and IF skill ("student", "sincere", "acquired", X2)
   and IF skill ("student", "do_not_steal", "acquired", X3)
   and IF skill ("student", "responsible", "acquired", X4)
THEN concept-learned ("student", "honest", "held", Z)
```

The above representation denotes that if the student acquires the above mentioned skills with confidence factors X_1, X_2, X_3, X_4 respectively (X_i values are obtained by the pedagogical model), then the rule confidence factor can be determined using the combination function: CF(rule)= min (X_1, X_2, X_3, X_4)*Z.

4.4. Presentation Model

To interact with the story, AEINS offers a GUI as shown in $Fig.2$ where the student is able to take actions, such as move, invite and so on. The student is then able to click on one of the characters' and places' pictures in the world. for example, the student can choose *"invite"* action and then clicks on *Ziad's* and *house's* pictures. The end result will be "invite Ziad to my house". Ziad has the freedom to accept or reject the student's invitation according to the rules that describe the non playing-characters actions.

The student gets engaged in a conversation that evolves depending on the student's actions. The aim is to enable students to test their own intuitions about certain moral value and to perform arbitrary experiments, in so doing it is believed that students will better understand the nuances of the domain. In addition to presenting the student with good models and examples, hopefully, after which they could model their own behavior. An example run is attached as an appendix to this paper.

An illustration for the story world and the story engine can be found in [16].

5. Conclusion

The paper presents a framework for using educational theories and moral dilemmas within an adaptive tutoring environment. An architecture for interactive narrative and intelligent tutoring that makes use of a student model has been introduced. Based on this architecture an adaptive educational interactive narrative system has been implemented in order to help teaching in the ethics domain. The teaching moments are selected and presented in a way that allows the student to draw analogies from one problem to another (if needed). In the current version of AEINS, the intelligent tutor uses an overlay student model; no representation of misconceptions of alternative views exists. This is actually one of the further steps we intend to consider in our future work. AEINS uses the Socratic Method as its teaching pedagogy because of its capability of forcing the learner to face the contradictions present in any course of action that is not based on principles of justice or fairness.

The early evaluation provides encouraging results, four students aged 8-12 years had tried the system. The students were of different origins and had different cultural backgrounds, a good primary sample. AEINS interacted with every single student on different basis according to his or her student model. The rule representation has proven its validity and usefulness in the fast interaction between the student model and the system (deleting and updating rules). The advantages of incorporating moral dilemmas within the system provides the appropriate privacy and allows students to investigate different experiences and was able to develop new thoughts. AEINS seems to be an appropriate environment that can provide transferability of skills to the students' real life. One of the students stated " I felt as if I am in real world and these characters are really talking to me, their reactions were very believable." Future work involves a larger evaluation in order to obtain statistics significant results.

References

[1] A. Avner and C. Moore and S. Smith. Active external control: A basis for superiority of CBI, Journal of Computer-Based Instruction, **6**4,115–118, 1980.

[2] A. Ogan, E. Walker, V. Aleven and C. Jones. Toward Supporting Collaborative Discussion in an Ill-Defined Domain., Intelligent Tutoring Systems (ITS2008 Conf.), B. P. Woolf and E. Aïmeur and R. Nkambou and S. P. Lajoie (Edts), 825-827, Springer (Pub.), Lecture Notes in Computer Science , **5091**, 2008.

[3] B. S. Magerko, player Modeling in the interactive Drama Architecture, Department of Computer Science and Engineering, University of Michigan, 2006.

[4] C. Lynch and N. Pinkwart and K. Ashley and V. Aleven, What Do Argument Diagrams Tell Us About Students' Aptitude Or Experience? A Statistical Analysis In An Ill- Defined Domain, 9th intenational Conference on Intelligent tutoring Systems, Proceedings of the workshop on Intelligent Tutoring Systems for lll-Defined Domains,Montreal, Canada,2008.

[5] D. A. Shapiro. Teaching ethics from the Inside-Out: Some Strategies for Developing Moral Reasoning Skills in Middle School Students, Seattle pacific University Conference on the social and Moral Fabric of School Life, 1999.

[6] D. E. Simpson, Dilemmas in Palliative Care Education, Palliative Medicine Journal,**1**2, 1998.

[7] D. H. Elkind and F. Sweet, Good Character, *http://www.goodcharacter.com/*.

[8] D. H. Elkind and F. Sweet, The Socratic Approach to Character Education, appeared in Educational Leadership, 1997.

[9] D. Kelly. Framework for using Multiple Intelligences in an ITS, Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2003, D. Lassner and C. McNaught (Edts.), 2423–2430, 2003.

[10] G. Bolton, Acting in classroom drama: A critical analysis., Heinemann, London (Pub.), 1999.

[11] H. Barber and D. Kudenko, A user model for the generation of dilemma-based interactive narratives, AIIDEŠ07 Workshop on Optimising Player Satisfaction, Stanford, California, 2007.

[12] I. M. Goldin and K. D. Ashley and R. L. Pinkus, Teaching Case analysis through Framing: prospects for an ITS in an ill-defined domain, 8th international Conference on Intelligent tutoring Systems, Proceedings of the workshop on Intelligent Tutoring Systems for lll-Defined Domains, V. Aleven and K. Ashley and C. Lynch and N. Pinkwar (Eds.), t. Jhongli, Taiwan, 2006.

[13] J. Piaget, The moral judgment of the child ; The Free Press; New York, 1995.

[14] Moral Development and Moral Education: An Overview, http://tigger.uic.edu/ lnucci/MoralEd/overview.html

[15] R. Hodhod and D. Kudenko , Interactive Narrative and Intelligent Tutoring for Ill Defined Domains, In proceedings of a workshop on ITSs for Ill-Structured Domains Focusing on Assessment and Feedback. The 9th international Conference on Intelligent Tutoring Systems, Montreal, Canada, June 23-27, 2008.

[16] R. Hodhod, D. Kudenko, and P. Cairns, Serious Games to Teach Ethics. In proceedings of AI and Games Symposium held during AISB 2009 Convention: Artificial and Ambient Intelligence, Edinburgh, Scotland, UK, April 6-9, 2009.

[17] M. O. Riedl, A. Stern, D. Dini, and J. Alderman, Dynamic Experience Management in Virtual Worlds for Entertainment, Education, and Training. International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning, **4**2, 2008.

[18] R. Aylett, M. Vala, P. Sequeira, A. Paiva, FearNot! - An Emergent Narrative Approach to Virtual Dramas for Anti-bullying Education. International Conference on Virtual Storytelling 2007, 202–205, 2007.

Appendix
This appendix shows an example of a complete run. The system interaction is in normal font. The student's actions are in bold. Comments and illustrations are italicized.

At the very beginning the system allows the learner to enter his name and pick a character to represent himself in the game world. Then the system greets him/her and presents a brief introduction about the game world.
Hi rania! This is your world, please have a look around! You have four places to go to: the house, the shop, the library and the school. There are four characters from whom you can make friends. You have a list of actions to choose your actions from. You are free to act whenever you are ready Now, it is time to choose your friends.....
The system presents the characters by name and personalities (for example)
Gina is a nice girl, she is sincere. Gina does not accept taking things without permission but she can lie Do you like Gina and wants to be her friend?
The user has to choose and answer, as no free text is allowed in the current version of AEINS
yes
You like Gina. You and Gina are friends now.
AEINS continues with the rest of the characters, the student model is initialized according to the student's choices to his friends
AEINS asks the learner to either act or allow AEINS to act
Please choose an action or press done for the system's turn
The learner choose to act, he chooses to invite someone to his house (this is done by choosing one of the actions the student chooses "invite home"
The system now asks the user to choose whom he wants to invite and then press the "carry my action button"
Now CLICK on whom you want to invite to your home. Then press CARRY OUT MY ACTION button.
The learner chooses to invite Gina (this is done by clicking on Gina's picture)
You chose to invite Gina
Gina accepts your invitation
Gina is at your house now.
AEINS asks the learner to either act or allow AEINS to act Now choose another action or press done for the system's turn
The learner chooses to allow AEINS to act (this is done by pressing the done button)
At this point AEINS pedagogical model figures out what is the most suitable moral dilemma to present the learner with (at this stage this is done according to the learner's choice to his friends)
The chosen teaching moment is called TMa and has certain preconditions that need to be satisfied first.
Since it is AEINS turn to act, the planner executes the action with the highest preference (lowest cost) in the first level of the plan.
Gina decides to go to the store
Gina is at store now.
AEINS asks the learner to either act or allow AEINS to act
Now choose another action or press done for the system's turn
The learner chooses to allow AEINS to act: Judy asks the learner to be her friend

Judy likes you and wants to be your friend, do you agree?
You agreed to be friend to Judy
AEINS asks the learner to either act or allow AEINS to act
Now choose another action or press done for the system's turn
The learner chooses to allow AEINS to act: Judy is at the store, she calls the learner
Judy is calling you to come and meet them at the store. Do you like to go?
yes
You agreed to join Judy
AEINS asks the learner to either act or allow AEINS to act
Now choose another action or press done for the system's turn
The preconditions of the teaching moment have now been satisfied, the learner, two friends of his are all at the store.
Since the preconditions and actions are represented by variables, this allows different characters to join the learner next time.
The teaching moment starts as follows
Gina is looking around in the store, they saw the kind of chocolate they like.
Gina has no enough money to buy the chocolate
Gina has an idea!! they can take the chocolate if you helped them to distract the shop keeper attention
Do you think you should help your friend to take the chocolate bar?
yes
You agreed
AEINS reflects on the user's action by stressing on his chosen action
You helped Gina to take the chocolate
AEINS figures out that the learner's attitude is not a desired one.
Judy, the one who holds the not-steal principle, is the one who will lead the Socratic Dialog in this dilemma.
Judy does not like what you did Judy thinks this is not good. Do you agree?
no
You disagreed
The teaching moment unfolds according to the student's answers to Judy- Judy will follow the Socratic Dialogue and raise the ante if necessary
.
.
Lets say Judy reached the end of the conversation
Judy says that they will not be your friend any more if you insists on what you did do you want to ask Gina to return the chocolate back.
yes
You agreed
Judy admires what you did, the shopkeeper caught Gina and will punish him for what they did.
The teaching moment ends at this point and the student model is updated based on the student action in this dilemma
NOW, AEINS ASKS THE LEARNER TO EITHER ACT OR ALLOW AEINS TO ACT
Please choose an action to perform or press done for the system's turn
AEINS continues interacting with the learner based on the student model

Scenario Adaptation: An Approach to Customizing Computer-Based Training Games and Simulations

James NIEHAUS and Mark RIEDL
School of Interactive Computing, Georgia Institute of Technology
jniehaus@cc.gatech.edu, riedl@cc.gatech.edu

Abstract. Game-based learning environments often make use of pre-scripted scenarios to present educational and training content. However, a one-size-fits-all scenario may not address all of the abilities, needs, or goals of a particular learner. This paper presents a methodology for automatically adapting a game scenario to better suit such requirements. We describe initial steps toward an intelligent technology called a Scenario Adaptor that employs planning-like algorithms to add or remove events from a scenario that relate to learning objectives. The Scenario Adaptor attempts to "re-write" scenarios in order for the learner to achieve the desired set of learner-specific learning objectives. Scenario cohesion is maintained by integrating causal chains of events that lead to the scenario's outcome. We also briefly discuss scenario execution in two different game environments.

1. Introduction

Training scenarios - especially those utilized by military, intelligence analysts, and emergency responders - are often utilized to give learners hands-on experience with real-life problem solving tasks in game environments. Learners experience a simulated course of events, assess the situation, practice skills, and act to achieve the goals of the scenario. Because the scenario is simulated within the game world, errors can be addressed immediately by an instructor or automated system and there are few repercussions for failure. Thus training scenarios are ideal for task domains where actual failure can be dangerous or costly. Game-based learning environments often utilize scenarios to provide effective training experiences. In these systems, scenarios may be simply a setting for a simulation, or they may be a predefined set of events, a script, that is to occur during the simulation. This work focuses on script-based scenarios.

Training scenarios are often authored by domain and system experts. Not only must the scenario present an accurate picture of the task domain, it must also give the learner opportunities to practice the important learning objectives. Scenarios can be difficult and time consuming to author and implement, and the result of this process is often a small set of one-size-fits-all scenarios crafted for the typical learner. However, such scenarios may not adequately address a particular learner's abilities, needs, or goals. It may force the learner to practice concepts with which he is already quite familiar and ignore concepts which require more practice.

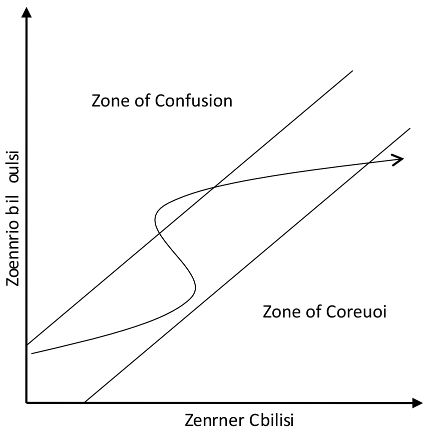


Figure 1. Graph demonstrating the Zone of Proximal Development.

The Zone of Proximal Development (ZPD) (see Figure 1) provides a framework for thinking about a learning experience in training scenarios. The ZPD is the trade-off between learner ability and scenario difficulty. The ideal situation is that, over time, the learner does not stray out of the ZPD into the zone of confusion or zone of boredom. While the ZPD is a developmental theory, it can be applied to training scenarios as well. Here, the ZPD is the difference between the ability the learner has *already demonstrated* and the most difficult challenge the learner is *able to overcome*. A training scenario (e.g., the curved line in Figure 1) is considered effective if falls within the ZPD.

How do we intelligently change scenario data to maximize time spent in the ZPD? We observe that script-based scenarios share many features with stories. Both “narrative” and “scenario” are descriptions of events or sequences of events. Whereas narratives often describe past events, scenarios describe - at some level of abstraction or specificity - events that are expected to unfold. Hence, intelligent narrative technologies such as narrative generation systems [1,2,3,4,5] and interactive narratives [6,7,8,9] may be an embarking point for scenario adaptation.

Interactive narrative systems have been explored as tools for education and training [7,9,4,8]. However, scenario adaptation is not the same as interactive narrative. Interactive narrative systems change the execution dynamics to react to the actions of a user/trainee within the scope of the original system parameters. Scenario adaptation is a process that occurs *before* execution that alters the parameters and the scope of the scenario. Interactive narrative can dynamically change the execution of a scenario, but only Scenario Adaptation can change what the scenario is *about*. The time just prior to scenario execution is the time in which the needs and abilities of the trainee can be most effectively taken into consideration. We view a Scenario Adaptor and interactive narrative systems as complimentary. The scenario adaptor provides the initial, individualized game scenario and the interactive narrative system manages execution in the game environment with the goal of achieving the adapted scenario. This is especially true of the interactive narrative systems based on planning (cf., [6,7]) because they employ similar narrative/scenario representations.

The work described here takes initial steps toward an intelligent system called a *Scenario Adaptor* that automatically customizes a scenario to suit a learner’s abilities, needs, or goals. Given a profile of a particular learner, the Scenario Adaptor customizes a scenario to target the ZPD. This allows the learner to practice underdeveloped skills and avoid the redundancy in areas in which she has shown proficiency. We define operations for deleting, adding, and replacing learning objectives in a scenario, we present an extension to partial order planning for scenario adaptation, and we discuss how this process may be used for game-based learning environments.

2. Related Work

Due to our perspective on scenario adaptation as resembling a process of "re-writing" existing scenario content, we note that there is a class of technologies – narrative generators – that may inform scenario adaptation. Though there are many systems that use many techniques for narrative generation [2,3,5,1,8], the systems that are most relevant to our problem of scenario adaptation are those based on planning formalisms. (e.g., [3,1]). Planning based narrative generators are given a goal state specification that must be achieved, and any narrative that does not achieve the goal state is rejected as incomplete. In scenario adaptation, it is necessary to drive a scenario to a goal state because the goal state specifies when the learning objectives are met. Failing to reach the goal state is a failure to meet the learning objectives.

Interactive narrative systems demonstrate how players or learners may interact with story and scenario content in complex simulation environments. While interactive narrative systems may adjust the simulation during execution to meet story or learning goals, scenario adaptation rewrites the goals of the scenario. Thus, we view these technologies as complimentary. The Crystal Island interactive narrative [8] attempts to guide the player through a story by modeling the player and automatically including hints and player aids. The ISAT system [9] is a training system for tactical field medical care that chooses interaction scenes to maximize players practice in related skills. The FearNot! system [4] uses virtual agents to play as characters in a simulation to teach children about school bullying.

The Mimesis system [6] employed a partial-order planner to generate plots for interactive narrative. The Automated Story Director [7] system represented stories as plans and uses a director and character agents to execute the story plans. The director agent to sends commands regarding the story to the character agents that executed these commands. The character agents also possess "local" freedom to act in a natural manner and engage the player. These systems demonstrate the use of plans as stories and the possibilities of mixing a planning approach with character agents.

There is currently much demand for computer-based training for non-kinetic skills such as social awareness, cultural awareness, and bilateral negotiation. These so-called "soft skills" are crucial, and also challenging to train because they occur in ill-defined contexts and situations. Many computer-based training systems targeting soft skills have been developed [10,11,12,13,9,8], and these systems employ scenarios to manage learning content. There is a need for a greater range of scenario contexts and thus non-kinetic skills scenarios are a natural target for scenario adaptation. Each system represents scenarios in a different manner, and the Scenario Adaptor must be sensitive to these representations in order to generate effective scenarios. Section 4 examines the means by which the Scenario Adaptor may be integrated with the game-based environments below.

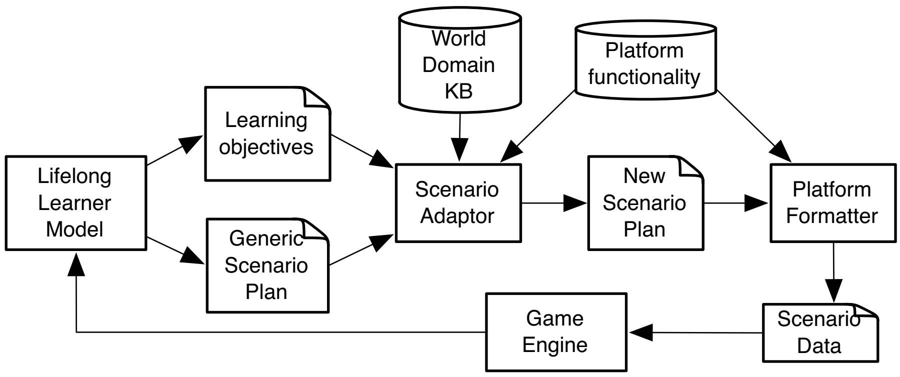


Figure 2. Process for adapting a scenario.

The Adaptive Thinking & Leadership (ATL) program [10] and DARWARS Ambush NK! [11] in particular are a virtual environments to train soft-skills. ATL and NK! feature multiplayer interaction with roles for an instructor and multiple student role players. Scenarios in ATL and NK! are script-based; they are carefully written ahead of time and then managed by the instructor and student role players. The manual process of customizing the scenario may be done for the course material, but the scenario is not reauthored to fit the needs of a particular group of students.

3. Scenario Adaptation

Due to our perspective on scenario adaptation as resembling a process of "re-writing" existing scenario content, our approach to the Scenario Adaptor extends on previous work on narrative generation and interactive narrative systems. Figure 2 depicts the process for adapting scenarios. Two external processes are required. First is a *Lifelong Learner Model* [14], an intelligent system that tracks a learner across many learning experiences (e.g., training scenarios), updating a learner model, and choosing the next training scenario that will help the learner advance. The second is a *Game Engine* that executes a training scenario with the learner as an interactive participant. A generic scenario plan and a set of learning objectives are combined to create a scenario adaptation problem using a world domain knowledge base. The generic scenario plan defines the scenario to be adapted, and the learning objectives define the goals for adaptation. The world domain knowledge base provides a library of possible world events, learning objectives, characters, objects, and other setting related information defining the dynamics of the simulation world and ensuring the adapted scenario is possible within the simulation. The scenario adaptation problem is solved by the Scenario Adaptor, described below, and the resulting scenario is formatted for execution within the game environment. Once the game is played, proficiency is assessed and reported back to the Lifelong Learner Model. See [15] for more about updating the lifelong learner model.

3.1. Scenario and Learning Objective Representation

Two basic requirements influence our representation of scenarios: coherence and pedagogical efficacy. In order for a scenario to be coherent, it must operate within the rules of the simulation. In order for a scenario to achieve pedagogical efficacy, it must provide the opportunity to learn the desired learning objectives. A simulation planning domain is defined to encompass the possibilities for events within the game world and a mapping between event sequences and learning goals is defined to reason about learning objectives.

We employ a specialized plan representation from decompositional partial order planning (DPOP) to represent the scenarios and associated learning objectives [16] (see Figure 4 for a schematic view). DPOP plans contain steps with preconditions and effects. The steps are related via causal links, denoting that one effect satisfies one precondition, ordering constraints, denoting that a step comes before another, and decomposition links, denoting that a step is part of another's decomposition. Abstract steps must be decomposed via decomposition recipes (or, simply, decompositions) to into other abstract steps, non-abstract steps, causal links, and orderings between steps.

To represent both scenario events and learning objectives, we split the planning namespace into world state and learning state. Non-abstract world steps define the events in the scenario, and abstract learning steps define preconditions and effects on the learning state. The scenario initial and goal states contain both world and learner state. The decompositions of the abstract learning level steps are defined by the domain author such that experiencing the steps in the decomposition gives the learner the opportunity to practice the learning objective. Thus, the decompositions define the possible collections of events for achieving learning objectives.

Figure 4 is an example of a simple scenario with two learning objectives. In this example, the learning objective of 'Treat Victim for Arm Injury' is composed of the world steps 'Cleanse Wound' and 'Dress Wound'. The step 'Find Victim' is not part of a learning objective, but it serves the purpose of achieving coherence in the scenario. The victim must be found before he can be treated.

3.2. Adaptation Operations

Given our representation of scenarios and learning objectives, our definition of scenario adaptation is the task of adding, removing, or replacing abstract learning level steps to satisfy learner state preconditions and completing the plan by decomposing all abstract steps and satisfying all preconditions. The starting scenario is assumed to be complete in that all preconditions are satisfied with causal links and no links are threatened. The additional requirement of scenario cohesion indicates that most steps should contribute to a causal chain leading to the goal state (the main scenario outcome). Because learner time is valuable, the scenario should be no longer than required to achieve the learning objectives and promote causal cohesion. We identify three basic operations to adapt a scenario: deletion, addition, and replacement.

3.2.1. Deletion

Deletion is the process of removing a learning objective an its associated steps. Once a learning objective is no longer required by the learning level goal state or other learning level steps, it is deleted. This operation may break causal links, preventing the preconditions of steps from being satisfied and breaking causal chains leading to the goal state. New links, possibly from new steps, must be created to establish preconditions and form chains to the goal state. To delete a learning objective, the learning step and all steps which belong to its decomposition are first removed along with all associated causal links and orderings. This removal may leave an incomplete plan containing link threats, dead end steps, and open preconditions. Planning is performed to complete the plan.

3.2.2. Addition

Addition is the process of adding a learning step and associated world steps. New abstract learning steps may be required to satisfy new preconditions of the learning level goal state or of other learning steps. Addition requires that new world state preconditions are satisfied and new world steps are linked to causal chains to the goal state. To add a learning objective, the learning step and one of its decompositions are inserted into the plan. This insertion may create an incomplete plan containing link threats, dead end steps, and open preconditions. Planning is performed to complete the plan.

3.2.3. Replacement

Replacement is the process of replacing an existing learning level step and its steps with a new learning level step and associated world steps. Replacement is combines deletion and addition. First the existing learning level step and associated world steps are removed, along with all relevant causal links and ordering. Then, the new learning level step and its associated world steps are inserted. Planning is performed to complete the plan.

3.3. Scenario Adaptation Planning

Standard decompositional planning is not sufficient to support the operations in scenario adaptation as defined above. Decompositional planning provides two assurances: 1) that the events relating to the learning objectives are contained in the scenario and 2) that the preconditions of the events are satisfied. However, decompositional planning does not attempt to a) ensure a cohesive causal structure nor b) define methods for adapting existing complete plans. Events that are added as part of a decomposition may not contribute to the goal of the plan, leaving possible causal 'dead ends'. Such dead end steps break the cohesion of the scenario. Decompositional planning is defined in such a way as to only add steps, links, and decompositions to a plan until completeness is achieved. However in scenario adaptation as defined above, a new complete plan must be produced from an existing complete plan, and the removal of steps, links, or decompositions may be required to integrate new events into causal chains leading to the goal.

To satisfy preconditions, maintain plan consistency, and eliminate dead end steps we propose a Scenario adaptation planning algorithm. Figure 3 outlines the algorithm. Scenario adaptation planning performs the standard decompositional planning techniques of satisfying open preconditions, eliminating causal threats, and decomposing abstract steps. The algorithm also attempts to improve plan cohesion by eliminating dead end steps that do not contribute to a causal chain leading to the goal. This algorithm is utilized to complete scenario plans after performing the learning objective operations.

3.4. Example

Figure 4 is an example input scenario featuring a domain for medical field training serious game. The game world for this scenario contains a single victim, lying hurt in some

Scenario Adaptation Planning (problem=⟨plan, learning-objectives⟩, domain-kb)

1. **Termination** If the plan is inconsistent, fail. Otherwise, if the plan is complete, return.
2. **Plan Refinement** Choose a flaw from the plan. Switch on flaw type:
   - **Open Precondition:** Resolved in the standard manner via reusing a step or adding a new step to establish the precondition. (cf. [16])
   - **Causal Threat:** Resolved in the standard manner via promotion, demotion, or separation. (cf. [16])
   - **Abstract Step without Decomposition:** Resolved in the standard manner via inserting a decomposition from the library into the plan. (cf. [16])
   - **Dead End Step:** Choose to do one of the following
     * **Nothing:** Do nothing, leave the step as a dead end.
     * **Satisfy Precondition:** Link one of the step effects to a unifying open precondition.
     * **Replace Link:** Replace a causal link to a unifying precondition with a link from the dead end step.
     * **Remove Step:** Remove the step from the plan, if the step is part of a decomposition, remove the entire decomposition.
3. **Recurse**

Figure 3. Scenario Adaptation planning attempts to incorporate Dead End steps in the causal chains leading to the goal state.

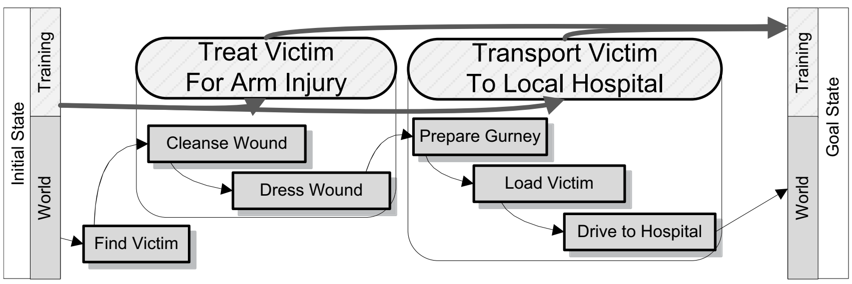


Figure 4. A representation of a simple scenario with learning objectives.

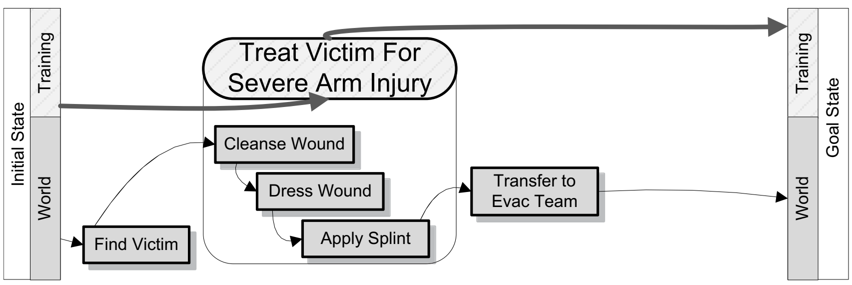


Figure 5. Scenario adapted by removing a learning objective and replacing a learning objective with a more difficult one.

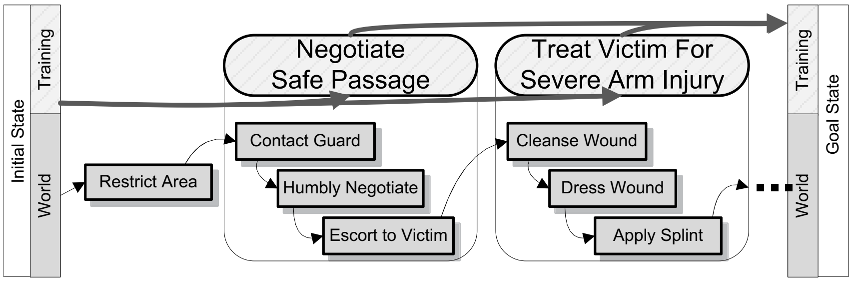


Figure 6. Scenario adapted by adding a learning objective.

wreckage. The learner must locate the victim, treat his injuries, and see him to safety. In this scenario, the learner is given the opportunity to practice two skills, treating a victim for an arm injury and transporting a victim to a hospital. The world steps such as ‘Cleanse Wound’ denote the actual events in which the learner acts or observes. The learning steps such as ‘Treat Victim for Arm Injury’ decompose into the world steps that provide the requisite practice. This scenario is an example of the input, a general scenario, given to the Scenario Adaptor.

Perhaps the learner is proficient in transporting victims but wishes to practice more difficult first aid procedures. The first adaptation requests a deletion of the ‘Transport Victim to Local Hospital’ learning objective and a replacement of ‘Treat Victim for Arm Injury’ with the more difficult ‘Treat Victim for Severe Arm Injury’. As per the deletion operation, the ‘Transport Victim to Local Hospital’ step and associated world steps and links are removed. As per the replacement operation, the ‘Treat Victim for Arm Injury’ step and associated world steps and links are removed, and a new ‘Treat Victim for Severe Arm Injury’ step is added. The planner is then invoked to repair the scenario. The results are depicted in Figure 5. Here, the learner locates the victim, treats a severe arm injury, and then a computer controlled evacuation team arrives to transport the victim to safety.

The learner may require practice in negotiation. The second adaptation requests an addition of the ‘Negotiate Safe Passage’ learning objective. The results of this adaptation are in Figure 6. Before the learner finds the victim, a computer controlled officer restricts the wreckage area. The learner must arrive on the scene and negotiate entrance to the wreckage before the victim can be treated. As per the addition operation, the new learning step is inserted into the scenario plan. The planner is invoked, and the learning step is decomposed into three world steps. In order to satisfy a precondition of ‘Contact Guard’, a new step, ‘Restrict Area’, is added from the operator library. Because, the world steps in ‘Negotiate Safe Passage’ are still dead end steps the link from the previous ‘Find Victim’ step to ‘Cleanse Wounds’ is replaced with a link from ‘Escort to Victim’. Because the ‘Find Victim’ step is a dead end step and no longer required, it is removed.

4. Adapted Scenarios for Training Games and Simulations

As depicted in Figure 2, the adapted scenario is executed in a game environment to expose the learner to the scenario content. Our technical approach enables us to adapt scenarios for different game engines by swapping planning domains. We are currently

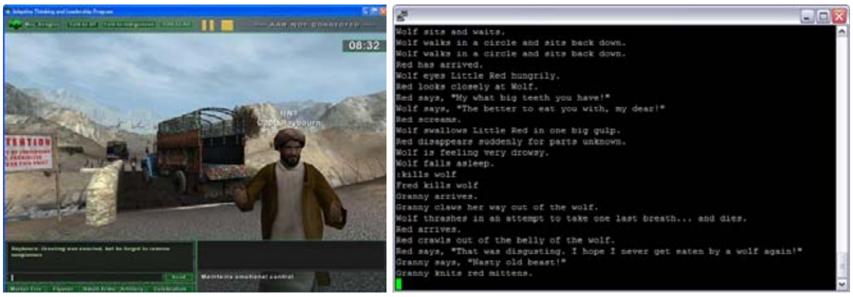


Figure 7. Screen captures of ATL [10] (left) and Automated Story Director in a MOO [7] (right), target platforms for the Scenario Adaptor.

targeting two game engines. The first is a multiplayer online game with human role-players. The second is a single-player game with NPCs.

In multiplayer online environments such as ATL [10] and DARWARS Ambush NK! [11], human readable scripts may be generated for the instructors and roleplayers. Narrative discourse planning [17,18] is being employed to express the scenario in human readable text. The scenario plan is input to a discourse planner customized for the purpose of generating scripts that can be read and enacted on by human role players. For multiplayer online games, enemy and neutral characters are played by trained role-players. The text is suggestive instead of providing exact dialogue to enable role-players to improvise and create a more seamless learning experience.

Because it is not always possible to have trained role-players to support the learner in his or her learning experiences, we also target single-player games in which enemy and neutral characters are controlled by intelligent agents. As proof-of-concept, our game engine is a MOO (a textual multiplayer environment) populated by semi-autonomous character agents. This is a technique used in the Automated Story Director drama management system [7], except we have replaced the "drama manager" with a script execution system that sends instructions to character agents to execute. In this approach, instructions are high-level descriptions of goals the characters should achieve. The agents decompose the instructions into primitive behaviors such as movement, speech (hard-coded lines), and gestures using a reactive planner. The semi-autonomous character agents are assumed to have a wide repertoire of behaviors that they can execute that correspond to the actions that can be included in a scenario from the Scenario Adaptor. Further description of the agents is beyond the scope of this but technical details of agent execution can be found in [7].

5. Conclusions

The adaptation of scenarios may make more more effective computer-based learning environments by creating more efficient training experiences. Individualization of scenarios has the potential to improve engagement. Learners may spend less time on tasks that are either too easy or too difficult and more time on tasks that are challenging but within reach. New algorithms are needed to adapt scenarios, and future work will finalize these algorithms and test learning outcomes.

References

[1] Mark O. Riedl and R. Michael Young. An intent-driven planner for multi-agent story generation. *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems, New York*, 2004.

[2] James R. Meehan. Tale-spin, an interactive program that writes stories. In *Fifth International Conference on Artificial Intelligence (IJCAI 77)*, 1977.

[3] Michael Lebowitz. Story-telling as planning and learning. *Poetics*, 14, 1985.

[4] Ruth Aylett Sandy Louchart Joao Dias and Anna Paiva. Fearnot! an experiment in emergent narrative. In *Proceedings of the 5th International Conference on Intelligent Virtual Agents*, 2005.

[5] Marc Cavazza, Fred Charles, and Steven J. Mead. Planning characters' behaviour in interactive storytelling. *Journal of Visualization and Computer Animation*, 13:121–131, 2002.

[6] R. Michael Young, Mark O. Riedl, Mark Branly, R.J. Martin, and C.J. Saretto. An architecture for integrating plan-based behavior generation with interactive game environments. *Journal of Game Development*, 1, 2004.

[7] Mark O. Riedl, Andrew Stern, Don Dini, and Jason Alderman. Dynamic experience management in virtual worlds for entertainment, education, and training. *International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning*, 4(2), 2008.

[8] Bradford Mott, Sunyoung Lee, and James Lester. Probablistic goal recognition in interactive narrative environments. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 187–192, Boston, Massachusetts, 2006.

[9] Brian Magerko, Bob Wray, Lisa Holt, and Brian Stensrud. Improving interactive training through individualized content and increased engagement. In *Proceedings of I/ITSEC 2005*, Orlando, FL, 2005.

[10] Elaine M. Raybourn, Ed Deagle, Kip Mendini, and Jerry Heneghan. Adaptive thinking & leadership simulation game training for special forces officers. In *I/ITSEC 2005 Proceedings, Interservice/ Industry Training, Simulation and Education Conference Proceedings*, Orlando, Florida, USA, November 2005.

[11] Elaine M. Raybourn, Bruce Roberts, David Diller, and Leslie Dubow. Honing intercultural engagement skills for stability operations with darwars ambush! nk game-based training. In *Proceedings of the 26th Army Science Conference*, 2008.

[12] John E. Deaton, Charles Barba, Tom Santarelli, Larry Rosenzweig, Vance Souders, Chris McCollum, Jason Seip, Bruce W. Knerr, and Michael J. Singer. Virtual environment cultural training for operational readiness (vector). *Virtual Real.*, 8(3):156–167, 2005.

[13] H. Chad Lane, Mark Core, David Gomboc, Arshish Karnavat, and Milton Rosenberg. Intelligent tutoring for interpersonal and intercultural skills. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2007)*, 2007.

[14] Judy Kay. Lifelong learner modeling for lifelong personalized pervasive learning. *IEEE Transactions on Learning Technologies*, 1(4), 2008.

[15] H. Chad Lane, Mark Core, Dave Gomboc, Mick Birch, John Hart, and Milton Rosenberg. Using written and behavioral data to detect evidence of continuous learning. In *Proceedings of the UMAP 2009 Workshop on Lifelong User Modeling*, 2009.

[16] R. Michael Young and Martha E. Pollack. Decomposition and causality in partial-order planning. In *Proceedings of the Second International Conference on Artificial Intelligence and Planning Systems, Chicago, IL*, 1994.

[17] R. Michael Young and Johanna D. Moore. Dpocl: A principled approach to discourse planning. In *Proceedings of the Seventh International Workshop on Text Generation*, Kennebunkport, ME, 1994.

[18] Mark O. Riedl. *Story Generation: Balancing Plot and Character*. PhD thesis, North Carolina State University, September 2004.

Assessment and Learning in Intelligent Educational Systems: A Peek into the Future

Valerie SHUTE[a], Roy LEVY[b], Ryan BAKER[c], Diego ZAPATA[d], and Joseph BECK[e]
[a] *Florida State University,* [b] *Arizona State University,* [c] *Carnegie Mellon University,*
[d] *Educational Testing Service,* [e] *Worcester Polytechnic Institute*

Abstract. New demands associated with living in a highly-technological and globally-competitive world require today's students to develop a very different set of competencies than previous generations of students needed. The general goal of education is to prepare young people to live independent and productive lives. Unfortunately, our current educational system is not keeping pace with these changes and new demands. The world is becoming increasingly complex and to make progress toward fixing educational woes, we need to have a good sense of bearings—where we are, and where we're heading. This paper is intended to provide such bearings, specifically in s of a fresh vision for education. We envision new modes of learning teaching using stimulating online environments such as games and simulations, coupled with an assessment infrastructure covering a broad set of competencies and other attributes to support learning. This represents a long view of the field to inform current R&D efforts.

Keywords. Assessment, data mining, diagnosis, games, learning, modeling.

Introduction

Educational systems (in the U.S. and around the world) face huge challenges that require bold and creative solutions to prepare students for success in the 21st century. How and what we learn is rapidly changing. No longer are students spending years in classrooms that have changed little in the past two hundred years, learning skills that have also changed little in the past two hundred years. Now, students are graduating with the understanding that their educational process has not yet been completed, rather, it is just beginning; lifelong learning is a reality for most citizens. Similarly, there is a shift in how instructional content is delivered. No longer is most content learned from an instructor at the front of a room; more and more content is being delivered electronically via software—in school and outside of school.

This paper peeks into the future of education, 20 years hence. We describe a future of assessment and learning that is intended to inform current and near-future R&D efforts. For example, we envision new modes of learning and teaching using rich online environments coupled with an assessment infrastructure covering a broad set of competencies and other attributes to support learning. Toward this end, we need to identify ways to fully engage all students through learning environments that meet their diverse needs and interests. Online games and simulations are potentially optimal venues. One big problem is that most online games currently lack an assessment infrastructure, especially in relation to educationally-valuable knowledge and skills. However, assessment results can and should have important implications for learning.

Assessing generally refers to the process of gathering important information about competencies and attributes, either in formal or informal learning contexts. This should lead to valid and reliable inferences, both diagnostic and predictive. Too often, classroom and other high-stakes assessments are used for purposes of grading, promotion, and placement, but not to engender learning (i.e., a typical educational cycle is: Teach. Stop. Administer test. Next loop, with new content). The stance we take on assessment is that it should: (a) encourage and support, not undermine, the learning process for learners and teachers (as well as online agents); (b) provide formative information whenever possible (i.e., give useful feedback during the learning process instead of a single judgment at the end); and (c) be responsive to what is known about how people learn, generally and developmentally. This vision of assessment has its primary goal to improve learning [4, 14, 18] which we find to be exciting, powerful, and absolutely critical to support the kinds of learning outcomes and processes necessary for students to succeed in the 21st century. This type of assessment is referred to as "formative assessment," or assessment *for* learning, in contrast to "summative assessment" (or assessment *of* learning).

Consider the following metaphor representing an important shift that occurred in the world of retail outlets (from small businesses to supermarkets to department stores, as suggested by Pellegrino, Chudhowsky, & Glaser [11]). No longer do these businesses have to close down once or twice a year to take inventory of their stock. Rather, with the advent of automated checkout and barcodes for all items, these businesses have access to a continuous stream of information that can be used to monitor inventory and the flow of items. Not only can business continue without interruption, but the information obtained is far richer, enabling stores to monitor trends and aggregate the data into various kinds of summaries, as well as to support real-time, just-in-time inventory management. Similarly, with new assessment technologies, schools should no longer have to interrupt the normal instructional process at various times during the year to administer external tests to students. Instead, assessment should be continual and invisible to students, supporting real-time, just-in-time instruction (cf. [15]). This comprises a main feature of our vision of assessment to support learning as embedded in rich online environments, 20 years hence.

There are two other key aspects of our vision that can inform a research agenda. First, assessment will cover not only the particular domain skills, but also key competencies and attributes that are important for success in the 21st century, but aren't being assessed today. Specifically, we propose assessing general cognitive competencies such as problem-solving ability and systems thinking, as well as non-cognitive competencies and attributes such as teamwork, motivation, frustration, and open-mindedness. Since these learner characteristics can affect learning, it is essential to start thinking about how we can assess them in valid and reliable ways. Second, rather than being used just for the current lesson, assessment data will be made widely available, for interpretation by a wide variety of researchers, and for use by a broad community of stakeholders.

In summary, our three themes for assessing learning, described below, include: (1) comprehensive profiles/models (*what* and *how* to assess), (2) seamless and ubiquitous assessment (*when* and *where* to assess), and (3) assessment information for decision making (*who* is using the assessment data). Each of these will be briefly explained.

Three Themes for Assessing Learning in Online Environments

Theme 1: Comprehensive Models

For the first theme, 20 years from now, we envision a well-mapped landscape of traditional and emerging competencies as well as other personal attributes. To achieve this goal, research needs to be conducted in the areas of identifying, modeling, and assessing these attributes.

Identifying attributes refers to determining which sets of learner characteristics help to direct or support a learner's education. In addition to domain-specific knowledge and skills (e.g., reading, math, and computer literacy skills), some examples of other relevant attributes include: creative problem solving, systems thinking, self-regulation, information-seeking skills, compassion, and ability to transfer skills to new contexts. Examples of other personal attributes include boredom, frustration, and excitement, which can have an impact on students' success. Research is needed to derive a taxonomy of relevant competencies and attributes that are optimally suited for our rapidly changing world as instantiated in a variety of contexts. Along the lines of research already underway in the "lifelong learning user modeling" community [6, 7], we can imagine combined and evolving profiles representing a comprehensive synopsis of what's known, what can be done, what is believed, what is preferred, etc.

Modeling refers to establishing conceptual and computational representations of each key competency and attribute, which will require considerable research. Existing modeling tools like Bayesian networks, artificial neural networks, genetic algorithms, and Markov decision processes are promising. However, we need an explicit research goal to develop powerful new tools and modeling techniques that are even more effective and efficient than those available today. Because models should (a) work across a range of students, (b) be validated by experts (when relevant), and (c) be able to be applied within new environments (e.g., games, simulations, computer tutors) additional research will be needed on the transfer, portability and integration of models across contexts. Toward that end, we see the need to explicitly include context information in the models (e.g., socio-cultural and instructional environment data).

Assessing students yields information that can be used for both formative and summative purposes through appropriate design and other analytical methods. Such a perspective aligns with viewing assessment as a dynamic agent in student learning over time. Research is needed on the possible automation and streamlining of these approaches. Additional research relating to longitudinal assessment is also necessary to support this dynamic perspective of the role of assessment in student learning and progress. Currently, recent innovations in measurement include the development of principled frameworks that explicitly integrate specific theories regarding the domain, cognition, and learning into task and assessment design [9] and analytic techniques that support inferences about students along multiple dimensions and at multiple grain sizes [8]. Such developments have focused on traditional, cross-sectional assessments, but have the potential for longitudinal measurement (e.g., learning trajectories over time) as occurs in student modeling.

In short, assessment should be driven by the definition of terms (identification) and the rules of interaction (model). In this way, assessments can be developed as the definitions and models are created.

Theme 2: Seamless and Ubiquitous Assessment

For the second theme, 20 years from now, we envision a continuous process that fuses assessment and learning, similar to the metaphor about the stores-inventory mentioned earlier. Seamless refers to the removal of the false boundaries between learning and assessment that characterize the current Teach/Stop/Test model. Ubiquitous refers to the constant nature and need to feed back the results and implications of assessment into learning, anywhere anytime. The current state of affairs is characterized by a few, illustrative examples of the rich potential for assessment to be fully integrated into the educational enterprise, and considerably more instances of a stark divide between assessment and other aspects of education. At present that divide is normally crossed by teachers. Our goal is to remove the load from the teacher, creating tools that are easy to incorporate in the daily lesson plan, and which include actionable information.

Research necessary to accomplish this integration includes development and evaluation of the design, implementation, and interpretation of the ensuing data from seamless and ubiquitous assessment and learning systems. Some existing examples of systems that implement embedded assessment are as accurate as some of the best instruments available. For example, the accuracy of the Reading Tutor [10], which listens to students read stories aloud, and uses automated speech recognition to assess their reading proficiency, is comparable to the one produced by the Woodcock Reading Mastery Test [3]. The Assistment system for mathematics education can predict future performance on standardized mathematics exams as well as the test itself can [16]. Another example of these systems is assessing creative problem solving while a player is immersed in a game environment, such as Oblivion [13]. Finally, ongoing research at the Naval Air Warfare Center Training Systems Division (NAWCTSD) includes embedded assessment within various simulations. Immediate next steps would include use of existing tools to collect and report on educationally-valuable competencies (e.g., using MS Word to track spelling errors to make inferences about mechanical or comprehension reading challenges). Seamless assessment will involve models and procedures for supporting inferences across contexts (e.g., time, domains, and developmental levels). Constructing principled methods of data management to enable the integration of these diverse sources of information and sharing them among stakeholders is an important new challenge, described next.

Theme 3: Assessment Information for Decision Making

Our third theme relates to the needs of various stakeholders (e.g., students, parents, teachers, administrators, policy makers, researchers) with regard to assessment information to enable informed evidence-based decisions. Important research relating to this theme concerns the determination of particular stakeholder requirements with respect to assessment information or score reports.

Recognizing that each stakeholder has different information needs, assessment models should provide each stakeholder with access to assessment information in meaningful forms. Current reports tend to convey assessment information in a one-size-fits-all manner rather than conveying the information stakeholders need to make a decision. In addition, research shows that assessment information can enhance decision-making processes at different levels (e.g., student, class, school, and district levels). Policy makers, for example, require aggregate information about the strengths and weaknesses of students to make informed evidence-based decisions [5]. As more complex environments evolve, research would be needed regarding the kind of information required by each stakeholder and the kinds of external representations that would communicate assessment information in effective ways. Transparency regarding the characteristics of the assessment models used in these new learning environments facilitates acceptance and wider adoption by the community. Table 1 shows information of interest to different stakeholders.

Table 1. Assessment information and usage across stakeholders

Stakeholder	Assessment Information	Assessment Usage
Policy makers, adminis-trators	• Validity and reliability of inferences based on assessments • Information collected on the students • Types of assessment claims supported	Policy makers need evidence to decide whether the current educational policies are effective and appropriate. Assessment data can provide this evidence, although information has to be summarized to be useful.
Teachers, mentors, tutors	• What is being learned (content, competencies, other attributes) o Relative to other students o Relative to student o Relative to standards	Teachers have a diverse set of needs (e.g., individual progress, sub-groups, the whole class). Progress can be measured in relation to the learner, a group, or criterion. Teachers can use assessment results to determine what works and what does not to inform future teaching.
Students	• Strengths and weaknesses of valued competencies. • Levels/types of other attributes.	Students can use assessment results to learn content, hone skills, and learn about learning.
Parents	• Same as teachers/mentors, but at simpler level of interpretation	Parents can use assessment results to answer questions such as: Does my child need help? Should I talk to the teacher? Should we switch schools?

Having briefly defined our three themes comprising our vision of assessment to support learning, we now focus on the benefits of and barriers to this vision.

Benefits and Barriers of Implementing This Vision

The envisioned shifts in content types and delivery mechanisms described in the Introduction of this paper present many challenges. For instance, how can we assess learners in these new skills? Are we able to perform better assessments using these new resources? Will assessment become more challenging with a variety of technological educational resources, or can we streamline the process? All of these challenges also represent exciting opportunities for improving the educational process. We begin with the benefits of this proposed approach.

Benefits of this Vision

Constructing seamless, ubiquitous assessments across multiple learner dimensions, with data accessible by diverse stakeholders, is expected to yield several direct educational benefits as well as other indirect ones. First, the time spent administering the test, handling make-up exams, and going over test responses is not particularly conducive to learning. Approximately 10% of class time is spent on assessment activities. Given the primacy of time on task as a predictor of learning, reallocating that 10% into activities that are more educationally productive is a potentially large benefit that would apply to almost all students in all classes, and would be equivalent to giving students graduating from high school an extra year of instruction.

Second, by having assessments that are continuous and ubiquitous, students are no longer able to "cram" for an exam. Although cramming provides excellent short-term recall, and is a viable strategy for passing an exam, it is a poor route to long-term retention and transfer. Thus, standard educational policy is to assess students in a manner that is in conflict with their long-term success. By providing a continuous assessment model, the best way for students to do well is to do well every day; although this statement sounds tautological, it is not how most classes are structured. By moving students toward a model where they will retain more of what they learn, we are enabling them to better succeed in cumulative domains, such as mathematics.

The third direct benefit is that this shift in assessment mirrors the shift from evaluating students based on the number of years they have sat at a desk to evaluating students on the basis of acquired competencies. A growing number of U.S. states are requiring students to pass a high-stakes final exam in order to graduate from high school. While we do not especially resonate with the model of a pencil and paper, high-stakes test for which students must prepare, this shift toward ensuring students have acquired "essential" skills fits with our proposal of continuous assessment. Many educators would argue that certain milestone assessments are needed to ensure quality across larger populations, and thus such assessments would have to be based on the same principles. In line with our proposed vision, the next steps would entail broadening the set of educationally valuable competencies and attributes to be more aligned with current (and near future) educational needs.

In addition to the direct benefits to education, there are substantial indirect benefits as well. First, our ability to instruct students effectively is fundamentally limited by our ability to assess them. If students have varying degrees of proficiency at cognitive and non-cognitive attributes, and this varying background predicts how well a student will respond to a given intervention, then in order to provide optimal instruction it is necessary to accurately assess students. Furthermore, for us to understand the efficacy of different educational objects (e.g., intelligent educational games, specific instructional modules) it is necessary to have precise understanding of the student's knowledge before and after being exposed to the intervention. Thus, an ability to assess students will also enable us to better evaluate the educational objects with which future students will spend much of their educational time. Second, our current capacity to assess students is often limited in that it is based on a relatively small number of test items. As we move to a seamless assessment model, we will be able to more accurately assess students since we will have access to a much broader collection of the student's learning data. More accurate assessments enable us to suggest suitable educational objects to students as well as accurately evaluate those objects' efficacy. In addition to the various benefits described above, there are other educational issues that our proposed vision can help to address or resolve.

21st Century Skills. Students will need to develop a different set of competencies than those in the current schools. The issue is not that the 21st century is that different from the 20th, it's that it is different from the 19th century upon which much of "modern" schooling is based. In particular, so-called "soft skills" (e.g., teamwork, computer literacy, and presentation skills) are expected to become more important in education than they are now. Again, research is needed to identify attributes as well as their relation to learning outcomes and processes. Given this perceived increase in importance, it is important that we research good methods to assess students in these capabilities and to figure out how to take into account various cognitive and non-cognitive abilities when designing instruction. Moreover, given the growing importance of lifelong learning, we must find methods of assessing those cognitive and non-cognitive factors that are likely to be predictive of learner success so as to best guide the learner. Finally, as we are envisioning seamless and ubiquitous assessment in the context of lifelong learning, this vision can readily lead us to seamless and ubiquitous learning integrated with job performance support systems. In all cases, we must have a means of knowing whether the student has improved. Therefore, our approach of building comprehensive models of learner competencies and attributes, and then developing assessment techniques to infer levels of those constructs is necessary in a shifting educational landscape.

Broader Emergence of Educational Technology. Currently, assessment within educational software is typically handled on a system-by-system basis. To measure a specific construct (e.g., persistence, help-seeking) requires a substantial amount of effort to construct a model that is particular to the system in question. The amount of investment required to develop such a model—for a single construct for a single system—could easily require hiring a full time graduate student for an entire year. Thus, the current approach does not scale to the increasing numbers of electronic learning environments. Our vision and associated research agenda of building comprehensive models of general learner characteristics, and constructing them in such a way as to transfer across systems, avoids this problem. Aside from reducing the costs of electronic educational objects that would have been created, our vision will also increase the number of such artifacts that are built since a broader set of content creators will be able to participate.

Structuring the Data Deluge. Given a world where learners are using a variety of electronic educational objects, and those objects are continuously assessing learner progress on a variety of measures, it is possible to become drowned in details. Therefore, we recommend that assessment designers think about who the potential consumers are of this knowledge, and determine how they can distill the assessment content down to be of use to each stakeholder. If this is the responsibility of individual designers, it would be helpful to provide them with a framework for orientation – a shared data dictionary that prevents duplication of efforts and streamlines nomenclature and categorization. Otherwise it will be extremely difficult to aggregate information across individual contributions. As we described earlier, our envisioned taxonomy would first have to be established by corresponding research and then disseminated (and perhaps governed) by a body similar to other shared standards as coordinated by IEEE, ISO, IMS, or SCORM.

By making assessment information available to a broader variety of members of the educational establishment, the likelihood that the learner will succeed is improved. For example, young learners could benefit from their parents being informed of learning deficiencies and providing additional help or motivation. Teachers could benefit from seeing a summary of areas of weakness in the class above and beyond a report for each student; such a report would enable an immediate alteration of teaching methods. This highlights the importance of mechanisms that facilitate the communication of data in a way that is desired by and meaningful to stakeholders. So, by considering the social processes of learning outside of software, the assessment technologies described herein are intended to enhance the learner's experience and support network, resulting in effective, efficient, and enjoyable instruction.

Challenges and Barriers to this Vision

Though the vision and agenda of research discussed herein have a number of direct and indirect benefits for the science and practices of learning, there are several challenges that will need to be addressed for the agenda to be successfully completed, and for it to achieve its full potential influence on the scientific community and on educational practitioners. These are now described below.

Generalizable Educational Models. The vision we have outlined depends to a significant degree on the success of models in generalizing among educational objects, competencies and other learner attributes. However, the study of generalizability of these types of models is still in its infancy. There exist examples of the study of the generalization of models between learning objects, involving stratified cross-validation ("leave-out-one-learning-object-cross-validation") [cf. 1], but the methodology used is generally overly simplified, and does not explain why models can generalize in some cases but not others. The scientific literature on transfer learning, from the machine learning community is a valuable resource for understanding the transfer of models (e.g., to new environments), but has not generally been applied to the types of models developed by the educational research community. Selecting from and applying this literature to the type of educational models proposed here is likely to increase the success of the proposed research agenda.

In addition, there is significant variation in the design of educational objects and how the competencies and constructs advocated here manifest themselves within such objects. This may lead to the need for meta-models, drawing from the cognitive modeling literature, that express the competencies and constructs at higher levels that can be automatically translated to the low-level features of the environment often found in machine learned models of educational constructs.

Considering the Stakeholder. We have presented a need for more focused research on the dissemination of information to a wide variety of stakeholder communities. We cannot expect that school administrators, teachers, or parents will become experts in complex data analysis. Hence, we will have to develop tools for communicating assessment information to these stakeholders in their own language, and tools that these stakeholders can use to explore the deluge of data available. This effort will require deploying methods from the interaction design and human-computer interaction communities, in order to develop reporting and communication tools that are useful, usable, and desirable to members of these communities. In general, the educational practice community (including teachers and administrators) must be included in the development of this vision and the research agenda proposed here. Traditional assessment methodologies such as tests have a long and rich history of usage by these communities, and have been successful at addressing specific summative assessment needs. It is our view that the methods proposed in this paper have the potential to be more efficient, less disruptive, and better able to assess more complex competencies and educational constructs than existing educational measurement methods that are widely used today. However, they must be designed in collaboration with these partner communities in order to achieve wide usage and high effectiveness. In particular, these communities must be involved in the choice of competencies and constructs to assess, or the models may make little impact on educational practice.

Walled Gardens. At the moment, developers of educational software have little incentive to cooperate in making the educational content interoperable. For instance, consider a student using learning objects A and B developed by two different software companies. The content in learning object A is prerequisite to the content in learning object B, hence students predominantly experience learning object A first. Learning object A distills information about the student that can make learning object B more effective, resulting in learning object B being highly effective. What is the incentive for the developer of learning object A to share that information to learning object B? In a competitive commercial environment, sharing information has the potential for asymmetrical impact, where the information-receiving learning object appears significantly more effective than the information-donating learning object. Additional issues to be resolved relate to intellectual property; e.g., the question of revenue generation, compensation, and possibly royalties.

This problem can be addressed in part through using the educational data mining method of learning decomposition [2] to infer when learning object B's effectiveness was likely enhanced by receiving information from learning object A (i.e., by analyzing object B's effectiveness both when object A's information is present and missing). This may increase the incentive for sharing information, as the developers of learning object A can point to their software's benefits on students' future learning.

Privacy. As with any large-scale data management project, incorporating data that can potentially identify individuals and which gives a broad range of information about individuals, privacy concerns must be accounted for. The positive intentions of educational practitioners and technology experts notwithstanding, any large quantity of data provides risks that inadvertent errors or intentional abuses can lead to privacy violations. Hence, all efforts must be taken to ensure that data is as anonymous as possible, including removal of obvious personal information such as names and birthdays (and its replacement with unique personal identifiers which cannot be reverse-engineered to link to a person), and scrubbing of potential identifying information. Such practices are already standard in large public educational data repositories such as the Pittsburgh Science of Learning Center DataShop, TalkBank, and the Kingsbury Center. All research supported by this initiative should study these existing examples and attempt to match or improve on the privacy practices used by these repositories.

Summary

What are the critical research and development questions that can begin to move us toward the vision we have described? The principal goals will be to figure out (a) which attributes to value, assess, and support for 21st century success, and (b) how to accomplish the design and development of robust assessments which would ultimately be embedded within online systems (e.g., educational games). Modeling, assessing, and supporting students in relation to an expanded set of competencies and attributes is intended to allow students to grow in important new areas, function productively within multidisciplinary teams, identify and solve problems (with innovative solutions), and communicate effectively. Critical research and development issues include those related to assessment and modeling—particularly in support of student learning and also that can be delivered in a cost-effective way. Such research is needed given changes in (a) the types of learning we are valuing today (and in the near future), as well as (b) the new, broader set of contexts in which learning is taking place.

Additional research and development will need to be done in terms of effective and efficient adaptive technologies that are closely coupled with valid diagnostics, and research is needed that facilitates the linking of results from various forms of assessment, e.g., to support the creation of developmental or vertical scales. Tools for critically evaluating theories and models will also be required. Finally, controlled evaluations need to be conducted on advanced, online educational systems to determine what works, for whom, and under what conditions.

In conclusion, we foresee three major funding targets to move these ideas (and the field) forward. This includes research on: (1) Understanding the full complement of characteristics that are brought to bear in learning - what are they, how do they relate, how do we get evidence about them, and how do we take that evidence to inform learning? (2) Fusing assessment and learning - what are the new sources of assessment, how do they flow to, from, and with learning, and how can we tear down conceptual and practical barriers between assessment and learning? and (3) Rendering assessments useful to all parties - who makes what decisions, what information do they need, how does assessment provide evidence for those decisions, and how to best communicate the complicated results of assessment to each party?

References

[1] Baker, R. S. J. d., Corbett, A. T., Roll, I., Koedinger, K. R. (2008) Developing a generalizable detector of when students game the system. User Modeling and User-Adapted Interaction, 18, 3, 287-314.

[2] Beck, J.E. (2006). Using learning decomposition to analyze student fluency development. Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems (pp. 21-28). Jhongli, Taiwan.

[3] Beck, J. E., & Sison, J. (2006). Using knowledge tracing in a noisy environment to measure student reading proficiencies. International Journal of Artificial Intelligence in Education, 16, 129-143.

[4] Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 80(2), 139-148.

[5] Honig, M. I. & Coburn, C. E. (2008). Evidence-based decision making in school district central offices: Toward a research agenda. Educational Policy, 22(4), 578-608.

[6] Kapoor, A., & Horvitz, E. (2007). Principles of lifelong learning for predictive user modeling. In Proceedings of the Eleventh Conference on User Modeling (UM 2007), 37-46.

[7] Kay, J. (2008). Lifelong learner modeling for lifelong personalized pervasive learning. IEEE Trans on Learning Technologies, 1(4), 215-228.

[8] Leighton, J. P., & Gierl, M. J. (Eds.) (2007). Cognitive diagnostic assessment for education: Theory and practices. Cambridge University Press.

[9] Mislevy, R. J., Steinberg, L. S., & Almond, R. A. (2003). On the structure of educational assessments. Measurement: Interdisciplinary Research and Perspectives, 1, 3-67.

[10] Mostow, J., & Aist, G. (1997). The sounds of silence: Towards automated evaluation of student learning in a reading tutor that listens. Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-97) (pp. 355-361). Providence, RI.

[11] Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). Knowing what students know: The science and design of educational assessment. Washington, DC: National Academy Press.

[12] Shute, V. J. (2007). Tensions, trends, tools, and technologies: Time for an educational sea change. In C. A. Dwyer (Ed.), The future of assessment: Shaping teaching and learning (pp. 139-187). New York, NY: Lawrence Erlbaum Associates, Taylor & Francis Group.

[13] Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), The Social Science of Serious Games: Theories and Applications. Philadelphia, PA: Routledge/LEA.

[14] Stiggins, R. (2006). Assessment for learning: A key to motivation and achievement. Edge: The Latest Information for the Education Practitioner, 2(2), 1-19

[15] Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID* model, Educational Technology, Research and Development, 50(2), pp. 39-64.

[16] Feng, M., Beck, J., Heffernan, N. & Koedinger, K. (2008) Can an Intelligent Tutoring System Predict Math Proficiency as Well as a Standardized Test? In Baker & Beck (Eds.). Proceedings of the 1st International Conference on Education Data Mining. pp.107-116. Montréal 2008.

Incorporating game mechanics into a network of online study groups

Ari BADER-NATAL [1]
Grockit, Inc.
ari@grockit.com

Abstract. The recent movement towards publishing open educational resources has increased the variety and quantity of learning materials available to students outside of the traditional classroom environment. Several core characteristics of the classroom environment, however, are difficult to offer through a web-based interface, including: (1) interaction and camaraderie among a cohort of peers, (2) the ability to get "real-time" answers to pressing questions, and (3) a motivating force to keep the student engaged over time. An online learning environment can approximate the value of peer cohorts and live question-answering by supporting (and encouraging) synchronous interactions among individuals studying a common topic. A learning system can motivate participation and collaboration by incorporating elements of game mechanics in the activity. We discuss Grockit, a recently-launched website that combines a virtual study group format with multi-player game dynamics to provide an engaging live collaborative learning environment for geographically-dispersed learners.

Keywords. learning games, study groups, motivation, collaboration

Not all learning happens in classrooms. Self-motivated learners have a wealth of options, vastly enriched by widespread access to the internet. Now, in addition to local libraries, students have access to open educational resources such as MIT's OpenCourseWare initiative [5] and educational podcasts such as those found on iTunes U [1]. But several key aspects of classroom learning are still not readily available, including (1) a peer group to learn with and from, (2) the opportunity to ask someone a question and get an immediate response, and (3) the structure and motivation that helps students remain committed over time. Synchronous communication among learners on an educational site could theoretically enable a virtual community of peer learners. The approach that we have been pursuing at Grockit (grockit.com) has been to leverage various game mechanics to affect how students choose to participate and interact with each other. With respect to AIED systems, the goal of the current approach is to reduce the traditional challenge of identifying and remediating learner misunderstandings to the much simpler task of enabling and motivating peers to do this for one another.

As of June 2009, Grockit has built support for two groups of learners. The first is for students studying for the GMAT, a common business school entrance exam. The second is for students studying for the SAT, a national college entrance exam. These types of groups offer a good testbed for live online learning communities, as there are a large number of students who share the common learning goal of mastering the skills that

[1] Corresponding Author. Grockit, Inc. 777 Florida St. #300. San Francisco, CA 94110. http://grockit.com

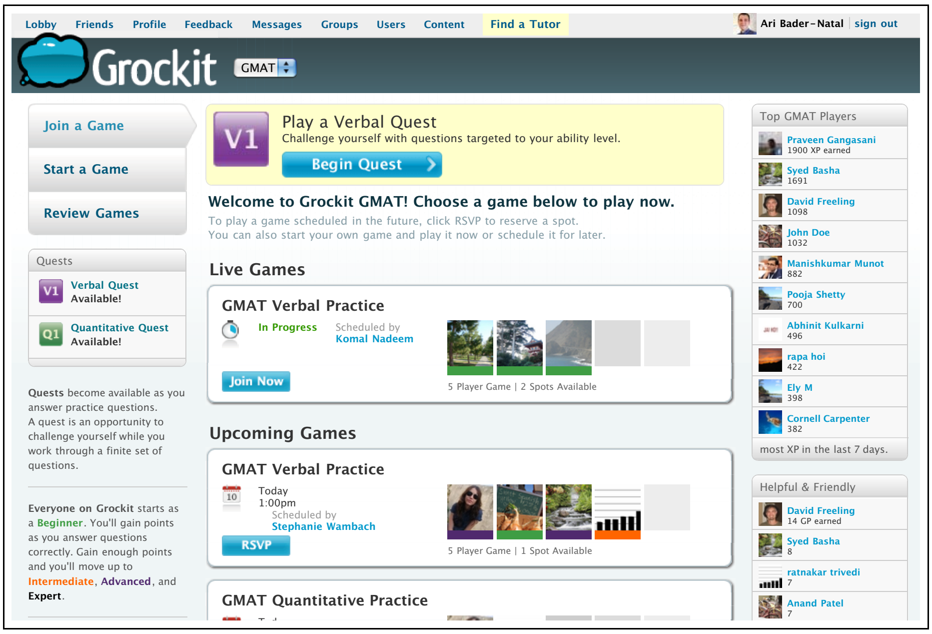


Figure 1. Each learning community has a "game lobby" which serves as an entry point, displaying in-progress and upcoming study sessions. (Note that all screenshots were captured in June, 2009). Leader-boards are displayed on the left, and available quests are displayed on the right.

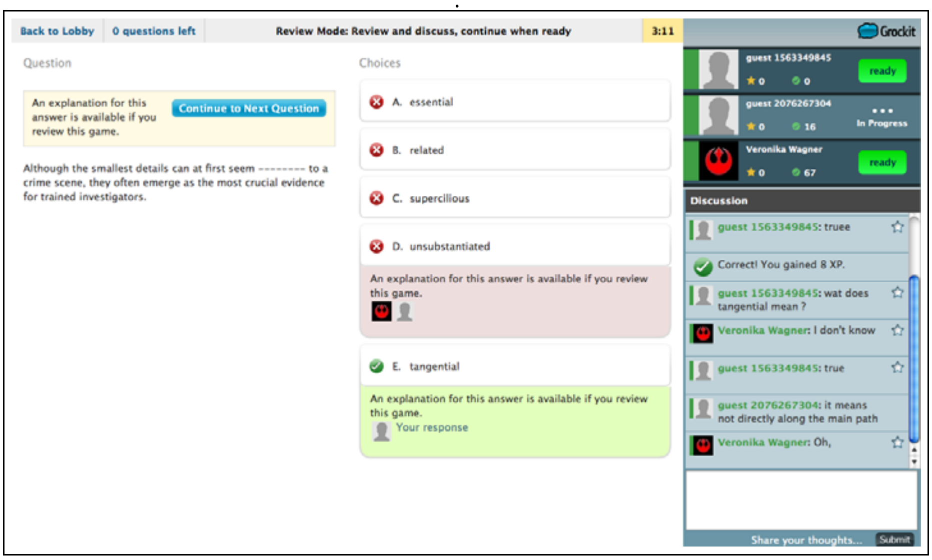


Figure 2. In each game, a series of questions are presented, such as the one above. After all players answer the question, they are provided with an opportunity to see, and discuss, the correct answer before moving on to a new question. The discussion pane includes both player discussion and system feedback (e.g. point awards). Within each study group session (referred to as a "game"), discussion among students ("players") takes place in a chat pane.

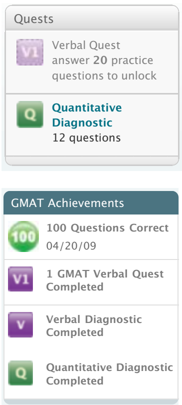


Figure 3. Diagnostics and Quests (top), and Badges (bottom).

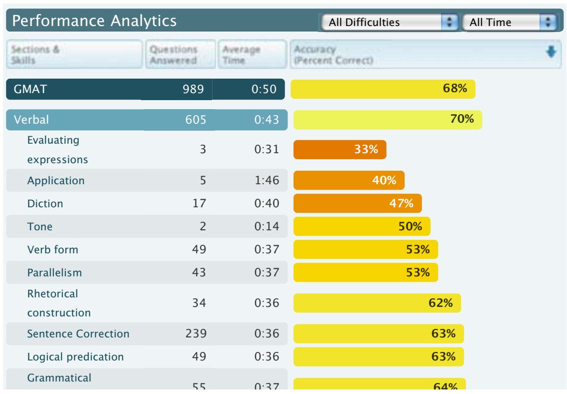


Figure 4. Player performance statistics are reported at the test, section, and skill levels. Color-coded bars help students focus their attention (for future games and reviews) on the skills for which the student had the lowest percentage accuracy.

are tested with the exam. At any time of the day, there are generally several groups of students collaborating (and site activity is quickly increasing).

For the GMAT site, students connect from around the world, generally from a home or shared computer. As aspiring business school students are often several years out of school, they are less likely to have friends or co-workers currently studying for the same exam. The characteristics of the standardized test-based groups (i.e. large numbers of students independently studying the same material) appear to be a good match for the live online study group, though future Grockit groups will not be restricted to test-oriented domains.

Over the past several months, we have incorporated into Grockit a variety of game-like mechanisms to motivate students to productively engage in the learning community:

Points and Leaderboards Grockit currently incorporates two point systems, which serve as reputation-like indicators: Experience Points (XP) and Grockit Points (GP). A player earns XP by answering questions. The point value for an answer is a function of the difficulty of the question (as estimated by a one-parameter Item Response Theory model) and the accuracy of the response. While this XP metric reflects the player's participation and performance as a *learner*, the GP scale reflects the player's activity as a *teacher*. When a peer finds a player's comment or explanation particularly helpful, they have the opportunity to award GP to that player. In the right-hand column of the Game Lobby (illustrated in Figure 1), "recent top player" leader-boards are displayed for XP and GP earned, and players frequently compete to appear on these lists.

Performance Statistics In addition to displaying XP and GP points achieved, student's can view (and opt to share) their performance statistics (see Figure 4). This data provides students with test-, section-, and skill-grained feedback on their performance (and can be restricted to include data from a specified time and/or difficulty range). Students can select a set of skills from the list, then review detailed explanations of the questions tagged with these skills.

Subsequent activity (Person count)	*Always available* (86)	*Every 20 questions* (100)	*Every 5 questions* (73)	*Never available* (101)	p-value
subsequently participated in a game	54.7% (47)	69% (69)	61.6% (45)	60.4% (61)	0.25
subsequently participated in a game with others	31.4% (27)	**55%** (55)	42.5% (31)	42.6% (43)	0.01 *
subsequently logged in again	43% (37)	**59%** (59)	47.9% (35)	38.6% (39)	0.03 *
subsequently participated in game discussion	24.4% (21)	**41%** (41)	27.4% (20)	24.8% (25)	0.03 *
subsequently reviewed questions	66.3% (57)	72% (72)	68.5% (50)	71.3% (72)	0.83

Table 1. Testing the effect of Quest availability on subsequent participation. The null hypothesis is that the four populations of students have the same true proportions, and the alternative is that the proportion is different in at least one of the populations. Starred p-values indicate significance at the $\alpha = 0.05$ level, two-tailed.

Diagnostics, Quests, and Badges By answering questions in collaborative games, players work towards "unlocking" certain fixed-duration solo activities (seen in Figure 3a). A *Diagnostic* includes a fixed set of questions designed to generate an initial assessment of the student's current ability level. A Rasch Model [3] is used to construct and evaluate this assessment. After completing a Diagnostic, *Quests* can be unlocked through subsequent collaborative game-play. Quests are designed to provide students with targeted practice, including questions for which the Rasch Model predicts a probability of response accuracy near $p = 0.5$. "Medium" difficulty challenges such as these have been found to increase intrinsic motivation [4] and to facilitate a "flow" experience [2]. Students are awarded *badges* (Figure 3b) for various accomplishments, such as the completion of Diagnostics and Quests.

One interesting recent finding came from a small randomized experiment run over a two week period in June, 2009 (shown in Table 1). Students who completed a Diagnostic were randomly assigned to one of four groups, determining if and when a Quest was available to them ("unlocked"). For three of the five outcome measures of interest, there was a statistically significant difference among the groups, with the group for which Quests were unlocked every 20 questions outperforming the others (visible in Figure 4a). Perhaps the added question spacing introduces an attainable goal that motivates increased participation. Game mechanics, such as this one and those outlined above, offer a rich toolset for motivating students in learning and collaboration, and help increase engagement in the virtual study group environment.

Acknowledgements

Writing about the game-like aspects of Grockit has been complicated by the rapid pace of system development. Many thanks to Farbood Nivi and the entire Grockit team for this added challenge.

References

[1] Apple, Inc. iTunes U. http://www.apple.com/education/mobile-learning/.

[2] Mihaly Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. Harper&Row, New York, 1990.

[3] Gerhard H. Fischer and Ivo W. Molenaar, editors. *Rasch Models: Foundations, Recent Developments, and Applications*. Springer-Verlag, New York, 1995.

[4] Mark R. Lepper. Motivational considerations in the study of instruction. *Cognition and Instruction*, 5(4):289–309, 1988.

[5] Massachusetts Institute of Technology. MIT OpenCourseWare. http://www.mit.edu/ocw.

MiBoard: Multiplayer Interactive Board Game

Kyle B. DEMPSEY[a1], Justin F. BRUNELLE[b], G. Tanner JACKSON[a],
Chutima BOONTHUM[c], Irwin B. LEVINSTEIN[b], and Danielle S. McNAMARA[a]
[a]*University of Memphis, Psychology/Institute for Intelligent Systems*
[b]*Old Dominion University, Computer Science*
[c]*Hampton University, Computer Science*

Abstract. Serious games have recently emerged as an avenue for curriculum delivery. Serious games incorporate motivation and entertainment while providing pointed curriculum for the user. This paper presents a serious game, called MiBoard, currently being developed from the iSTART Intelligent Tutoring System. MiBoard incorporates a multiplayer interaction that iSTART was previously unable to provide. This multiplayer interaction produces a wide variation across game trials, while also increasing the repeat playability for users. This paper presents a demonstration of the MiBoard system and the expectations for its application.

Keywords. Intelligent Tutoring Systems, Games, Serious Games

Introduction

Serious games have developed into a serious force in the educational realm. Delivering content to students via entertaining and challenging games has become a legitimate avenue for curriculum developers [1]. Students stand to benefit when developers use games to deliver curriculum because students likely become more engaged, are likely to spend more time on task, and are likely to return for subsequent learning sessions [2]. Serious games must balance entertainment, education, motivation, deliberation, adaptability, and affordability, by identifying which features of the game effectively promote learning while providing interactive entertainment for the student.

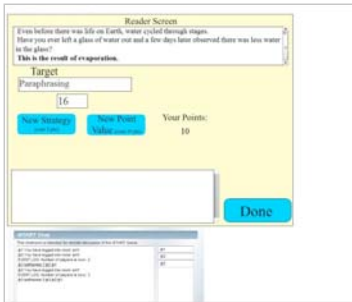
Here, we discuss a development project for a serious game called MiBoard. MiBoard is being developed as a serious games extension of the Intelligent Tutoring System, iSTART [3,4]. iSTART (Interactive Strategy Trainer for Active Reading and Thinking) is an automated tutor that teaches users to effectively self-explain texts using reading strategies. iSTART provides curriculum delivery at a one-on-one level without prohibitive cost. However, the current version of iSTART does not adequately address motivational factors for students who are required to use the system over long periods of time (e.g. months). By developing MiBoard, we believe that we are both providing variety in the classroom and improving motivation for long-term use that will result in the users increasing time-on-task (increasing the overall effectiveness of the system), as well as experiencing improved affect towards the system as a whole [5].

[1] Corresponding Author

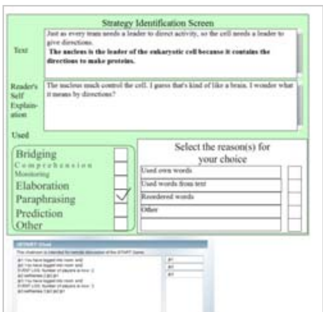
1. The Game

MiBoard is an online multiplayer board game that requires players to successfully produce self-explanations as well as identify the strategies used in other players' self-explanations (Comprehension Monitoring, Paraphrasing, Prediction, Elaboration, and Bridging). Within MiBoard, players earn points when a majority of players identify that the same strategy is used within another player's (the *reader's*) explanation. Players can spend these points during the game to change task parameters or activate special "in game" features (e.g., take an extra turn, freeze another player, draw an extra card, etc.). MiBoard does not provide feedback for the players' self-explanations. Instead, players receive feedback from the other players in the game through both modeling of self-explanations as well as through a chat room discussion.

a. Game Screen



b. Reader Screen



c. Strategy Identification Screen

d. Summary Screen

Figure 1. MiBoard: screen shots

2. Gameplay

MiBoard consists of a computerized representation of a game board that allows players to track their game progress in comparison to the other players. Players are represented on the board as tokens and take rotational turns as either the reader (responsible for producing the self-explanation) or the guesser (responsible for identifying which strategy is used in the self-explanation). Throughout the game, players encounter four different screens: the game screen, the reader screen, the strategy identification screen, and the summary screen (see Figure 1).

2.1. The Game Screen

The game screen encompasses the main game board and the choices that can be made during the game portion of the program, incorporating both token movement and game strategy. After completing their turn as the reader, the player first has an opportunity to use a Power Card (if applicable). Power Cards are one of the major motivational tools of MiBoard that allow the player to take a modicum of control over the movement of player tokens around the board. Power Card actions (e.g., freezing an opponent for their next turn) are activated by spending points. Next, the player rolls a die, moves, and draws an Event Card, cueing them to move forward, backward, or draw a Power Card. The *reader* tag passes to the next player at the completion of a turn. This sequence continues until one player reaches the finish, a congratulatory screen is displayed, and players are offered the opportunity to start a new game.

2.2. The Reader Screen

Each player takes turn as a *reader* by using an assigned strategy to self-explain a target sentence. The target text appears in context (with the previously presented text) where the reader may review the previous sentences at any time while self-explaining the target sentence. Players may at this point use their accumulated points to alter the potential point value for their self-explanation or alter the given reading strategy. Once a player has read the target text and submitted their self-explanation, the players all enter the strategy identification screen.

2.3. The Strategy Identification Screen

After the reader has submitted a self-explanation, the players are shown the strategy identification screen. The players are shown the target text (with context) and the self-explanation and must decide which dominant strategy the reader used in the self-explanation. After indicating which reading strategy they thought was used in the reader's self-explanation, the guessers and reader are moved to the summary screen.

2.4. The Summary Screen

At the summary screen, the system displays how everyone voted and awards points if at least half of the players agree on the reading strategy that is used. If the reader is part of the majority, those in the majority are awarded points based upon the points associated with the strategy for that turn, with the guessers receiving half of the points assigned to the category and the reader getting all of the points assigned. If there is a majority that does not include the reader, the guessers are awarded a smaller point total. By reducing the overall points available for an incorrect response by the reader, the players are encouraged to help the reader understand the strategies over the course of the game. If all of the players agree on the reading strategy used, an agreement bonus is awarded, and players are given the option of going directly back to the game screen.
If there is disagreement, the players enter the discussion stage where they are required to resolve their disagreements.

After the discussion phase, players are sent back to the strategy identification screen and are again given the chance to vote on which strategies were included in the self-explanation. After re-voting, players are awarded points for convincing other players to side with their original vote, and the self-explanation is complete. If a player is able to convince another player to choose their strategy, the player is awarded points. This feature allows for players to recoup some points.

3. Discussion

A major development of MiBoard is the use of the players as the comprehension check. Currently, there is no automated check against the target reading strategy. Requiring the players to police themselves introduces an interesting aspect that not many games have explored. By having the players discuss their strategy use, but not giving them as many points as they would get for all initially agreeing, the players should be more motivated to understand and apply the reading strategies because doing so will enhance game performance.

When completed, MiBoard will be the end result of the educational and motivational developments in iSTART. MiBoard is a dynamic, competitive environment available to a wide market that requires readers to understand and apply knowledge of reading strategies in order to succeed against other players. By increasing motivation, we expect users to display higher levels of engagement in the system as well as display a stronger desire to initiate a session with MiBoard. Therefore, as players compete in the game, they are likely to engage in the same amount and level of practice as in iSTART, but at the same time will be required to apply their knowledge of strategies by judging others' self-explanations. We expect these aspects of the games to have substantial and meaningful benefits for students' ability to understand challenging texts.

Acknowledgements

This research was supported in part by the Institute of Educational Sciences (IES R305G020018-02; R305G040046, R305A080589) and National Science Foundation (NSF REC0241144; IIS-0735682). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES or NSF.

References

[1] D.S. McNamara, G.T. Jackson, & A.C. Graesser. Intelligent tutoring and games (ITaG). In Y. Baek, Ed. *Gaming for Classroom-Based Learning: Digital Role Playing as a Motivator of Study*. (submitted). IGE Global Press.

[2] Garris, R., Ahlers, R. & Driskell, J.E. Games, motivation, and learning: A research and practice model, *Simulation and Gaming* **33** (2002), 441-467.

[3] D.S. McNamara, I.B. Levinstein, & C. Boonthum, iSTART: Interactive strategy trainer for active reading and thinking, *Behavior Research Methods, Instruments, and Computers* **36** (2004), 222-233.

[4] I.B. Levinstein, C. Boonthum, S.P. Pillarisetti, C. Bell, C., & D.S. McNamara, iSTART 2: Improvements for Efficiency and Effectiveness, *Behavior Research Methods Instruments and Computers* **39** (2007), 224-232.

[5] P.A. Alexander, P.K. Murphy, B.S. Woods, K.E. Duhon, & D. Parker, College instruction and concomitant changes in students' knowledge, interest, and strategy use: A study of domain learning, *Contemporary Educational Psychology* **22** (1997), 125–146.

I Was Playing When I Learned: A Narrative Game for French Aspectual Distinctions

Nicole HALLINEN[a1], Erin WALKER, Ruth WYLIE,
Amy OGAN, & Christopher JONES
Carnegie Mellon University

Abstract. While intelligent tutoring systems have been successful at promoting learning, students may fail to benefit if they are not motivated or engaged. We added game elements to an intelligent tutoring system for teaching the distinctions between French past tense forms, creating a narrative game in which players edit articles in a virtual journalism office. In a study with 38 students, we compared the game version to a non-game version of the system and assessed motivational responses. Students using the narrative game version found the activity significantly more engaging.

Keywords. Educational games, Computer-aided language learning, motivation

Introduction

Intelligent tutoring systems are extremely effective instructional interventions [1]. However, students may exhibit a lack of motivation while working with these systems, displaying negative behaviors like gaming and hint abuse [2]. Educational games may have the potential to motivate students to better attend to the material and spend more time on-task [3]. However, there has so far been little empirical evidence that this is the case [4]. Additionally, games have been theorized to distract from the learning content. Therefore, we propose incorporating game elements into an intelligent tutoring environment to test whether they enhance students' engagement with the learning content. Habgood and colleagues theorize that creating endogenous fantasies in which the narrative context and instructional material are intrinsically linked should promote learning and motivation [5]. Thus, our design decisions were guided by an effort to link intelligent tutoring properties (e.g. feedback) with relevant narrative and game elements, rather than introducing the elements as add-ons to the system. In a first study we evaluated the effects of introducing game elements into an intelligent tutoring environment on learning and student engagement. We hypothesized that game elements will increase student motivation with respect to the learning environment. Accordingly, students will attend more to the domain content and show increased learning in the educational game version of a tutor. In this paper, we describe the tutor and its transformation into an educational game, and evaluate these hypotheses.

[1] Corresponding Author: Nicole R. Hallinen; Carnegie Mellon University, SCS, HCII; 5000 Forbes Avenue; Pittsburgh, PA 15213

1. Learning Environments

Two computer-based language-learning environments were developed to explore the effectiveness of game elements in promoting student engagement and learning from educational tasks. In these learning environments, students received a short set of instructional material about the usage of *passé composé* and *imparfait*, read French paragraphs, and determined which of these two forms should be used for several highlighted verbs in these paragraphs. A pilot think-aloud study suggested that a binary selection of verb form made it too easy for students to guess the correct answer. Consequently, two distractor answer choices with incorrect subject-verb agreement conjugations were added to each selection.

After completing all of the selections for a given paragraph, students submitted their answers and received feedback in the following multi-level structure: (i) first, students were told how many verbs they correctly conjugated and were given a list of grammar rules related to their errors, (ii) upon the first incorrect re-submission, participants then saw a version of their paragraph with errors circled in red and feature-focusing questions (e.g. "Does this verb describe a completed action or one that is still in progress?") that appeared when a student moused over the circle, and (iii) a bottom-level hint in which the errors were circled in red and the feature-focusing question was followed by a verb rule (e.g. "The *passé composé* is typically used to describe completed actions"). Students continued submitting their answers and receiving feedback until all verb selections were correct. In total, students completed 5 paragraphs with a total of 35 verb selections. While engaging in verb selection, students had access to a help button that returned them to the instructional materials.

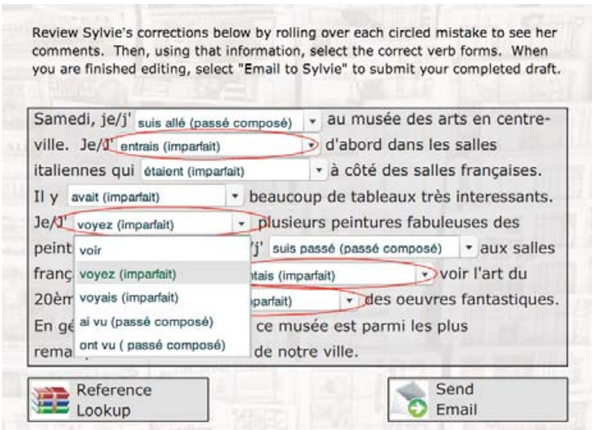


Figure 1. Game and Tutor Environment: Paragraph featuring drop-down verb selection and feedback circles.

Using this intelligent tutoring system as a base, we created a narrative game environment in which participants assume the role of a journalist working to edit articles at a newspaper. Because properly choosing verb forms is necessary to advance within the game framework, the content and game context are intrinsically related. Additional game features include time pressure (the game occurs between a time-scaled 9 am – 5 pm office day), levels (expressed as job titles and promotions in the journalism office), and points (conveyed as earnings in the narrative). The feedback structure was tied to the narrative metaphor as corrections were suggested through an imaginary email interface in which the player's boss edits his or her article submissions. The help button was labeled "Reference Lookup" as a journalist may consult references while writing. Both environments were created in Flash with Cognitive Tutor Authoring Tools (CTAT) components to facilitate data collection [6].

Figure 2. Game environment: Office layout including selection of articles using legends and user task bar featuring personalization, job title, salary, and time information.

2. Method

Thirty-eight undergraduate students enrolled in second, third, or fourth semester French participated in this study. All were proficient English speakers, although 8 reported that English was not their native language. All had seen past tense in class, but instructors identified it as a difficult topic. Students earned bonus course credit for participating.

The study had a between-subjects design with two levels of learning environment (game and non-game). Dependent measures for learning were comprised of verb form selection questions isomorphic to the intervention as well as open-ended grammar questions. Motivational dependent measures included items on ease of use, enjoyment, engagement, difficulty of content, and perceived learning and utility in preparing for grammar tasks, assessed on 5-point Likert scales. Participants took a pretest and were randomly assigned to either a game or non-game learning environment for 30 minutes or until they had finished all exercises. Finally, participants completed a posttest and demographic and motivational measures. Each session lasted about 50 minutes.

3. Results

As analysis of the grammar cloze questions indicated no significant difference in learning between conditions, this section describes results from the motivational measures. We collapsed the motivation measures into 3 constructs: engagement in the system, perceived difficulty of the content material, and perception of learning. Table 1 contains the means and standard deviations for these measures. Students did indeed show more engagement in the system in the game condition than in the non-game condition (t=-2.09, p=0.044). Reported easiness of content material trended toward more challenging in the game condition (t=1.72, p=0.095). Perception of learning (collapsed from perceived amounts of learning and preparation for the posttest) was not significantly different (t=-1.00 p=0.324).

Table 1. Means and standard deviations of 3 motivation measures scored on a 1-5 Likert scale. (* $p < 0.05$)

Motivational Measure	Game Condition	Nongame Condition
Engagement in the system* (5 = very engaging, enjoyable, easy to use)	3.26 (SD = 0.76)	2.70 (SD = 0.90)
Perceived easiness of content material (5 = very easy content material)	2.37 (SD = 0.83)	2.89 (SD = 1.05)
Perception of learning (5 = a lot of preparation and learning)	3.65 (SD = 0.67)	3.42 (SD = 0.79)

4. Discussion

We hypothesized that using a narrative game environment to learn French verb forms would promote learning and increase motivation. While we did not obtain the desired learning results, a series of transfer questions requiring participants to translate and generate past tense sentences will be analyzed in an effort to understand possible learning effects. The significantly higher level of engagement with the software exhibited by the game condition is encouraging and confirms common intuitions about the motivational benefits of educational games [3]. This benefit did not appear to come at the expense of efficiency or quality of learning. We suggest that this motivation to interact with game environments as characterized by high levels of engagement, enjoyment, and perceived challenge may encourage students to continue gameplay and ultimately experience greater learning gains. Qualitative measures of motivation and data log files of each participant's interactions with the learning environment were recorded. Analyzing the results of these data will likely clarify the implications of the quantitative results described above. This investigation provides a good first step toward evaluating the effects of combining intelligent tutoring and educational games.

Acknowledgements: This work was supported in part by the Pittsburgh Science of Learning Center which is funded by the National Science Foundation award number SBE-0354420. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

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

[2] Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z. (2004). Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System". *Proceedings of ACM CHI 2004* , 383-390.

[3] Lepper, M.R. and Malone, T. (1987). Intrinsic Motivation and Instructional Effectiveness in Computer-Based Education. In *Aptitude, Learning and Instruction: Cognitive and Affective Process Analyses*, 255-286.

[4] Hays, R. (2005). The Effectiveness Of Instructional Games: A Literature Review and Discussion. Technical report.

[5] Habgood, M. P. J., Ainsworth, S. E., Benford, S. (2005). Endogenous Fantasy and Learning in Digital Games. *Simulation and Gaming*, **36(4)**, 483-498.

[6] Aleven, V., Sewall, J., McLaren, B. M., & Koedinger, K. R. (2006). Rapid authoring of intelligent tutors for real-world and experimental use. In *Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies* (ICALT 2006), 847-851.

Embedded Assessments of Science Learning in Immersive Educational Games: The SAVE Science Project

Brian C. NELSON[a,1], Diane Jass KETELHUT, and Catherine SCHIFTER[b]
[a] *Arizona State University*
[b] *Temple University*

Abstract. Keeping students engaged in science after fourth grade appears to be a major challenge. While some of this challenge is related to how the concepts to be learned become more complex and abstract, the expanding use of standards-based high stakes testing complicates the issue. The typical model for assessing whether students understand scientific concepts and inquiry is through standardized multiple-choice tests. An alternative model, centered on performance-based assessments, is often criticized as being too time-consuming and lacking standardization. Situated Assessment in Virtual Environments for Science (SAVE Science) is exploring a third model for science assessment, utilizing a standardized set of activities embedded in a game-like virtual environment to assess student knowledge and understanding of scientific inquiry. Through assessment quests situated in a virtual environment, we propose that we can capture individual student behaviors that give insights into evolving student understanding that cannot easily be captured through multiple-choice tests. In this demonstration, we provide a brief introduction to the SAVE Science project, and demonstrate one of the assessment quests created for the project.

Keywords. Game-based assessments, science education

Introduction

Students lose efficacy in science as they advance through school. An item on the 2000 National Assessment of Educational Progress assessment measuring this asked U.S. students whether “all can do well in science if they try.” 82% of fourth-graders agreed with this statement, while only 64% of eighth-graders and a mere 44% of twelfth-graders agreed. One source of data for students on whether they can succeed in science comes from traditional tests (Brookhart & DeVoge, 1999), but do these tests accurately measure students’ knowledge of both science content and processes? According to the National Research Council (2005), science processes are most often assessed by asking students to define words such as “hypothesis” and “scientific method.” However, knowing the definitions of these words is not synonymous with understanding how to do the processes.

Typically, student assessment has taken a two-pronged approach using either traditional methods, such as standardized multiple-choice tests, or performance assessments, such as portfolios. Published reviews are available that detail the debate

[1] Corresponding Author: Brian Nelson, Mary Lou Fulton College of Education, Arizona State University, PO Box 87011, Tempe, AZ 85287-0611 USA. E-mail: Brian.Nelson@asu.edu

between the proponents of each of these styles of assessments (i.e. Mehrens, 1998; Moore, 2003). To summarize: Proponents of performance assessments view them as capturing student understanding better than standardized tests, which they feel measure only decontextualized knowledge. Opponents argue that performance-based assessments are not cost effective, cannot be compared from teacher to teacher due to individual grading differences, and are inconclusive about what the tasks are actually measuring (Stecher & Klein, 1997).

There is a third avenue of assessment that advancing technology has opened up (Ketelhut, 2006). Many forms of interactive technology incorporate a database that records all student interactions that take place within the environment. For example, hypermedia consists of a series of information spaces that the user negotiates by activating links that take him or her from one space to another. The user can choose a relatively linear pathway through the spaces; or, if he or she so chooses, the pathway can be much more complex. Analyses of databases underlying hypermedia that record the features of these learner paths have illuminated student purposes and strategies while navigating the learning process (Barab, Bowdish, & Lawless, 1997; Marchionini, 1989). This behavioral data can enrich our perceptions of student understanding.

1. Prior Research

In recent studies, we have investigated the use of educational Multi-User Virtual Environments (MUVEs) as platforms for situated science inquiry, exploring design, functionality, and learning impact of such environments (e.g. Ketelhut, 2007; Nelson & Ketelhut, 2007; Nelson, 2007). Our conclusion is that student choices in virtual environments can offer new details about their understanding, giving information not just about their solutions but also about their evolving problem-solving strategies.

For example, we conducted a series of studies based in the River City virtual environment, designed to engage teams of three middle-school students in a collaborative scientific inquiry-based learning experience. In this MUVE, students conduct their scientific investigations in a virtual historical town—populated by themselves, digitized historical artifacts, and computer agents—in order to detect and decipher a pattern of illness that is sweeping through the virtual community. Students manipulate a digital character, called an avatar, in order to explore the town; they conduct virtual experiments to test their scientific hypotheses about the causes of the River City epidemic. River City incorporates a server-side database that supports a wide variety of coding and analysis techniques.

In analyzing behavioral data in the database, we have uncovered some interesting aspects of student behavior and understanding. For example, we analyzed student inquiry-based behaviors (as defined by the National Science Education Standards (NRC, 1996) to encompass data-gathering and research) to see if the curriculum was impacting their understanding of how to diagnose and solve the problems. We discovered that these inquiry behaviors increase from visit to visit (Ketelhut, 2007), and that these data-gathering behaviors in-world predict for gain scores on pre-post content and process assessments which offers evidence for the validity of using in-world behaviors as indicators of understanding. These results give insights into student understanding that are not shown on more typical pre-post assessments or without the use of other more researcher-intensive data collection methods, such as participant observations. They indicate that behavioral data can lend insights into the evolution of student understanding that are not easily uncovered with other means. Thus, we posit that developing complex thinking requires not only complex media, but also complex assessment. Further, behavioral assessments, such as these facilitated by virtual environments, offer a potential compromise between performance assessments and objective tests. Unlike performance assessments, virtual environment-based assessments have the potential for standardization in terms of reliability and validity; unlike typical high stakes tests, virtual environment-based assessments are not decontextualized (Baker, Niemi & Chung, 2007).

2. Our Current Study

To explore whether virtual environments can augment current, traditional assessment of student understanding in science, we have recently begun the SAVE Science project. SAVE Science centers on use an innovative system for evaluation of learning in STEM disciplines. SAVE Science is designing and implementing a series of virtual environment-based assessment adventures (or quests) used for assessing both science content and inquiry in grades 7 & 8 in the School District of Philadelphia (SDP), Pennsylvania. These will generate knowledge to inform teachers and students about student understanding on science content studied in classroom based curricular units. This form of assessment has the advantage of placing a problem in an authentic context for students to solve. Steele (2005) reports that if students can learn to connect any concept they are learning to real-world situations, not only will it make the concept more meaningful but also help make it easier to understand and remember. Emerging technologies allow for virtual recreation of scientists' labs and other "real" contexts.

We are currently designing our first set of assessments for the 7th grade curriculum in Philadelphia. The modules will make use of a novel assessment rubric based on student interactions within an authentic context-based science curriculum, embedded in a virtual environment. We hypothesize that doing so will enable us to capture and analyze different patterns of scientific understanding amongst the students in a classroom. Further, we are investigating two "conditions for success" for using game-based learning assessments: how best to handle the effects of cognitive load students experience while conducting assessment activities in complex virtual environments, and how to engage teachers with virtual environment-based assessment data. We hope that these modules will help students better understand their own knowledge and learning processes, thus promoting increased efficacy and interest in science; direct teacher practices by providing them with more detailed data to inform meaningful differentiated instruction; and provide recommendations for use of virtual environments to standardize performance assessments, leading to the development of innovative assessments of student understanding of science content and inquiry.

3. Example Assessment Quest

In our session, we will demonstrate our first assessment quest: Sheep Trouble (Figure 1). This quest centers on an embedded assessment of student understanding of adaptation of an organism to a given environment over time. In the quest, students enter a virtual farmstead set in a mythical medieval setting. They are asked to gather evidence for why a recently imported herd of sheep are becoming sickly and weak, while 'native' sheep are healthy. In gathering evidence, students are given the opportunity to demonstrate both competency at conducting scientific inquiry and their understanding of the relationship between specific physical adaptations and a given environment. As students conduct their assessment quest, both 'stealth' and direct data are collected about their level of understanding and application of the science concepts the quest is designed to assess.

Figure 1: *Scientopolis* Virtual Environment

References

[1] E. Baker, D. Niemi, & G.K. Chung, Simulations and the transfer of problem-solving. In E. Baker, J. Dickieson, W. Wulfeck & H. O'Neil (Eds.), *Assessment of Problem Solving Using Simulations* (pp. 1-18). New York: Lawrence Erlbaum Associates. 2007.

[2] S. Barab, B.E. Bowdish, & K.A. Lawless, Hypermedia navigation: Profiles of hypermedia users, *Educational Technology Research & Development*, **45(3)** (1997), 23-41.

[3] S.M. Brookhart & J.G. DeVoge, Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education*, **12(4)** (1999), 409-425.

[4] D.J. Ketelhut, The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in River City, a multi-user virtual environment. *Journal of Science Education and Technology*, **16(1)** (2007), 99-111.

[5] D.J. Ketelhut, Do students with higher self-efficacy exhibit greater and more diverse scientific inquiry skills: An exploratory investigation in "River City", a multi-user virtual environment. Unpublished dissertation, Harvard University, 2006.

[6] G. Marchionini, Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science*, **40(1)** (1989), 54-66.

[7] W. Mehrens, Consequences of assessment: What is the evidence? *Education Policy Analysis Archives,* **6(13)** (1998). Online: *http://epaa.asu.edu/epaa/v6n13.html*.

[8] W. Moore, Facts and assumptions of assessment: Technology, the missing link, *THE Journal* **30(6)** (2003), 20-26.

[9] National Research Council, *America's Lab Report: Investigations in High School Science.* National Academies Press, Washington, D.C., 2005.

[10] National Research Council, *National Science Education Standards: observe, interact, change, learn.* National Academies Press, Washington, D.C., 1996.

[11] B. Nelson & D.J. Ketelhut, Scientific inquiry in educational multi-user virtual environments, *Educational Psychology Review* **19(3)** (2007), 265-283.

[12] B. Nelson, Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment, *The Journal of Science Education and Technology* **16(1)** (2007), 83-97.

[13] B. M. Stecher & S. P. Klein. The cost of science performance assessments in large-scale testing programs. *Educational Evaluation and Policy Analysis*, **19(1)** (1997), 1-14.

[14] M. Steele, Teaching science to middle school students with learning problems. *Science Scope* (2005), 50-51.

Mily's World: Math game involving authentic activities in visual cover story

Dovan RAI [a], Neil T. HEFFERNAN [a], Janice D. GOBERT [b] and Joseph E. BECK [a]
[a] *Computer Science, Worcester Polytechnic Institute (WPI)*
[b] *Social Sciences & Policy Studies and Computer Science, WPI*

Abstract. Motivational effect of computer games is appealing to education designers who want to retain learning gain of intelligent tutoring system. Integrating these two, we have created *Mily's World* -- a math-tutor game which involves authentic activities wrapped in a visual cover story. This is a research and development effort. We have made our design guidelines based on the analysis of cognitive and affective aspects of educational games as well as their challenges and limitations. Unlike conventional games focusing on speed and drill activities, we aim to present math as an engaging, intuitive, and creative activity.

Keywords. Intelligent educational games, visual learning, authentic learning

Introduction

Computer games, due to their popularity among the youth, are viewed as potential tools to motivate and engage students in learning. However, empirical studies have shown them to be inferior to Intelligent Tutoring Systems (ITS) when it comes to learning gain. Thus, combining features of ITS (e.g.: student modeling and scaffolding) and motivational attributes of games has developed into a new area of research.

Mily's world is a math-tutor game with a sequence of 8th grade co-ordinate geometry problems wrapped in a visual cover story. Students have to help story characters solve problems and move the story forward. The complexity of problems and the richness of story grow as the students proceed. Students will receive tutor help and scaffolding as they stumble on problems and misconceptions.

Mily, a 13-year old animal-loving girl, is the protagonist in this game. She has a puppy. Students are engaged in many different math-related tasks. For example, they calculate Mily's height and the distance between her and her puppy based on the co-ordinates of their heads. This is a very simple warm-up to make students familiar with the environment. As they proceed, students have to help Mily decide the name of the puppy and then help create a kennel. When students give the correct answer of slopes and equations, the kennel's wall and roofs are built gradually and then a new kennel pops up. The puppy develops a bad habit of chewing socks; so, Mily ties him to a rope. Students have to help her find the co-ordinates of safe places to place socks where the puppy cannot reach them. Other tasks include making a slide, visiting pond, etc.

This project is based on visual learning paradigm [1] with two main goals: to make intuitive visual presentation of mathematical concepts (e.g., co-ordinates, length, angle, and slopes) in real world contexts; and to motivate students with game-like elements such as an imaginative cover story, stimulating problems, interactive animated feedback, etc.

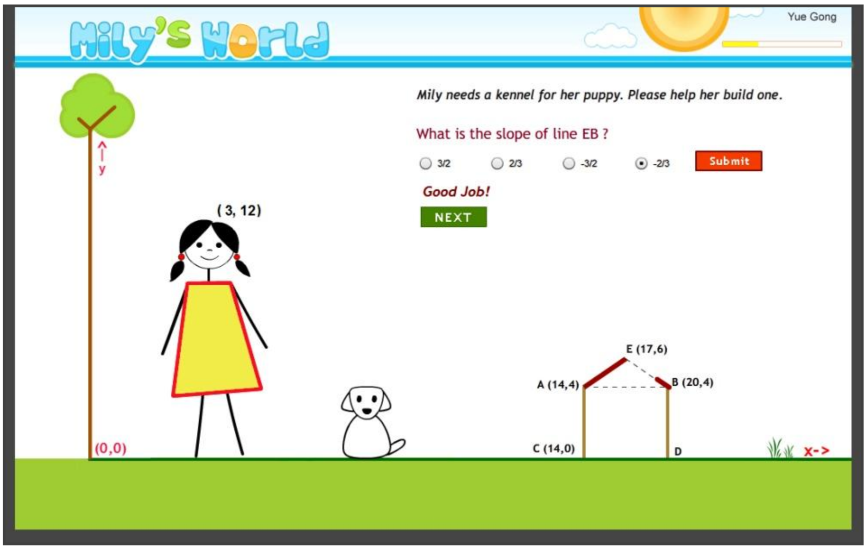


Figure 1: Screenshot of Mily's World

Mily's world can be viewed as a math-tutor with game-like properties. We consider tutors as effective learning tools that lack motivational features. Games, on the other extreme, have high motivational features but poor learning. We are curious how learning and motivation change if we incrementally add game-like properties on tutors.

1. Mily's World as a tutor

Mily's world, as a math-tutor, is composed of different layers that lead to deeper learning of math concepts when compared to conventional materials. Our approach extends some of the existing research literature in this area, as follows:

• *Authentic learning:* Learning activities must be authentic activities, which must be embedded in realistic and relevant contexts [2]. In Mily's world, students are solving everyday problems like building a kennel and slide, etc.

• *Graphical representation:* Graphics are aesthetically appealing or humorous, attracting attention and maintaining motivation [4]. Successful problem-solvers build mental models from word problems [3]. Graphics help learners to develop mental models thus reducing the burden on memory and processing [4].

• *Storytelling:* This is not only more engaging but also helps to have better retention exploiting episodic memory [10].

• *Interactive immediate visual feedback and reward:* Students can tell what the error was and how it relates to the correct solution [7]. For instance, the puppy can chew socks if they are placed at wrong co-ordinates. Similarly, reward is given with positive visual feedbacks like a new kennel popping up.

Mily's world will incorporate tutor elements like scaffolding and mastery learning of ASSISTment, an existing math tutoring system. The story line will be made flexible to adjust to the mastery learning e.g. if they have difficulty in problem on skills required for kennel-building, they are required to make a new kennel or similar structure involving the skills.

2. Mily's World as a game

Conventional computer games are criticized for catering to a disproportionately high number of male users utilizing stereotyped masculine interests like violence. Although there are now an increasing number of computer games targeted towards female users, they tend to exploit stereotypical feminine interests like fashion. This game instead is designed to attract users by giving them an opportunity to use their mathematical knowledge and skills to create stuffs, help the characters, and move the story forward.

• *Competition and speed:* Most computer games take competition and speed as motivating factor. But in educational games, it can have an adverse effect. First, it can be too overwhelming. Next, emphasizing speed and instant recall of facts reinforce the idea that computation is all there is to mathematics. Speed is not an important factor. The children's emphasis on speed leaves little time or inclination to talk or reason about the mathematical problems involved [9]. Mily's world has a simple, relaxed, and friendly atmosphere.

• *Game and educational content:* According to the congruence principle, content and format of the story and graphics should correspond to the content and format of the concepts to be conveyed [4]. Story should mesh well with the mathematics in which players are engaged. Engagement with the story should naturally lead to engagement with the mathematics. If math is used just as a means to perform other non-mathematical, supposedly fun stuff, students will get the sense that math is not fun. They can feel that they are being fooled into having fun doing what they would prefer not to do if it were only math [9].

• Use elements of fantasy, control, choice, contextualization, and personalization [6, 8]. Students have emotional connection with the characters. They can name pets and get involved in stimulating activities.

• *Gaming the game:* Since this is game-like and enjoyable, there is a possibility that students might misuse the scaffolding to proceed to more interesting sequences. Hence some mechanism to keep track of scaffold usage would be desirable.

• *Integration of concepts:* In Mily's World, students have to answer various questions on co-ordinate, slope and equations to create a kennel. It can represent how different mathematical concepts are incorporated to create an integrated real-life structure.

3. Challenge: Cognitive Overload

Despite the intuitive appeal of constructivist approach and educational games, these ideas face challenge from cognitive overload that sets limitations esp. for novices.

Educational games should be cautious to strike a balance between stimulation and overload and should not get swayed towards adding fun stuffs. After all, educational games are not competing against regular games but instead with other educational materials to get students' attention. We have used the following two strategies but we are yet to experiment their effect.

3.1 Simple game environment: Regular games can afford to have new and complex environment with complicated rules. But for educational games, students can be overwhelmed by too much stuff going on. So we chose a simpler, more familiar environment and tasks.

3.2 Minimalist visual presentation: The presentation should be intuitive so that it helps to off-load rather than overload. While concrete scenario is desirable, making too realistic visual presentation can be counter-intuitive [5]. In Mily's world, a concrete scenario is presented with real objects like a dog and a kennel whereas graphics are minimalistic and it uses stick-like figures that are closer to symbolic representation. If the pictures are too realistic, e. g: if the house is shown in 3-D isomorphic form with patterned roofs, we assume that students will have difficulty to figure out slope. Kennel in wireframe structure is more intuitive as students can clearly see the connection between slope of the line and slope of the roof. It also gives better intuition of magnitude and direction of slope showing that two slopes of roof have same magnitude but opposite directions.

4. Hypotheses and conclusion

This is an ongoing work and we aim to have gain in three areas: motivation, learning and retention. We hypothesize that this tutor will improve students' attitude towards math. They will feel more comfortable with abstract mathematical concepts and will be more motivated. Students will see that math can be a tangible, creative activity rather than a dry, abstract exercise as they can see the characters and events in story affected by their mathematical performance.

We are also interested to see the effect in different student subgroups like visual-learners, girls and students with math anxiety.

References

[1] A. Arcavi, *The role of visual representations in the learning of mathematics,* Educational Studies in Mathematics, Volume 52, 2003, 215-241

[2] Shaffer, David Williamson, Resnick, Mitchel, *"Thick" Authenticity: New Media and Authentic Learning*, Journal of Interactive Learning Research, v10 n2 p195-215 1999

[3] Mary Hegarty, Richard E. Mayer, and Christopher A. Monk, *Comprehension of Arithmetic Word Problems: A Comparison of Successful and Unsuccessful Problem Solvers,* Journal of Educational Psychology (1995),Vol. 87,18-32

[4] Barbara Tversky and Julie Bauer Morrisony, *Animation: can it facilitate?,* Int. J. Human Computer Studies(2002), 57,247-262

[5] Robert L. Goldstone and Ji Y. Son, *The Transfer of Scientific Principles Using Concrete and Idealized Simulations*, the journal of the learning sciences (2005), 14(1), 69–110 2005

[6] Malone, T. W. (1980), What make things fun to learn? A study of intrinsically motivating computer games, Cognitive and Instructional Science Series, CIS-7, Xerox Palo Alto Research Center, Palo Alto

[7] Dugdale, Sharon, *Green Globs: A Microcomputer Application for Graphing of Equations.* (1981-08-00)

[8] Cordova, Diana I., Lepper, Mark R., *Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice* (1996)

[9] Megan Murray, Jan Mokros, and Andee Rubin, Where's the Math in Computer Games?, Through the Glass Wall project at TERC

[10] Daniel T. Willingham, *Why people love and remember stories*, http://www.aft.org/pubs-reports/american_educator/issues/summer04/cogsci.htm

[11] D Agosto, *Girls and gaming: A summary of the research with implications for practice* (2004), Teacher Librarian 31, (3), 8-14

[12] Ryan S. J. D. Baker, M. P. Jacob Habgood, Shaaron E. Ainsworth and Albert T. Corbett, *Modeling the Acquisition of Fluent Skill in Educational Action Games*, User Modeling, (2007)

Dynamic Guidance in Digital Games

Using an Extensible Plan-Based Representation of Exploratory Games to Model Student Knowledge and Guide Discovery Learning

James M. THOMAS and R. Michael YOUNG
Digital Games Research Center
Department of Computer Science
North Carolina State University, Raleigh, NC USA

Abstract. This paper describes a novel application of a task-based knowledge representation that seeks to facilitate automated guidance of task-based learning through the core mechanics of digital games.

Keywords. Digital games, exploratory ITS, narrative, planning, knowledge representation.

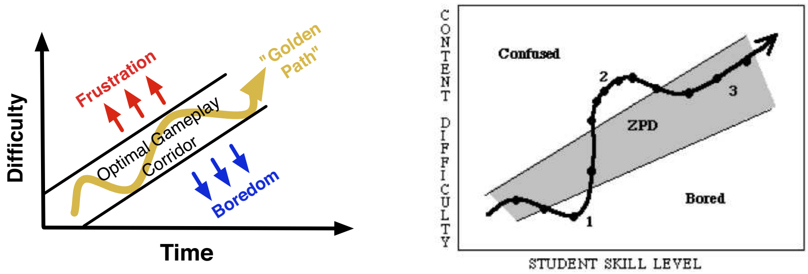
1. Introduction

Over the past few decades, Intelligent Tutoring Systems (ITS) have advanced to the point that many research-based systems exceed the teaching effectiveness of traditional classroom-based instruction [1]. Contemporaneous with this success, commercial digital game technologies have shown remarkable advancement in graphics, human-computer interaction, and financial significance. Although both fields have evolved to include a broad diversity of applications and genres, intriguing confluences of goals and challenges are apparent. Yet in the collaborative efforts to date to exploit this confluence, cynics say that the educational elements "suck the fun out" of games [2]. This paper proposes a knowledge representation structure and a general computational framework that leverage automated planning to integrate arbitrary sets of learning goals within the core mechanics of games. Our intention is to provide a general system that builds games where the learning and the fun are deeply intertwined.

2. Background

The core conceptual models posited by games and ITS to guide exploration are remarkably similar. Game researchers describe the goal of guidance is to lead the player down the "optimal game play corridor" [3], reproduced in Figure 1(a). Even a sporadic observer of the game industry will encounter multiple replicas of the chart shown in Figure 1(a) when the subject of learning or guidance is introduced [3,4].

Readers familiar with learning theory will recognize the similarity between the "optimal game play corridor" and Vygotsky's "Zone of Proximal Development" or ZPD



(a) Optimal Game Play Corridor (b) Zone of Proximal Development

Figure 1. Isomorphic Concepts of Guidance

[5,6]. This encouraging confluence has inspired an extensive research record to leverage the game paradigm within intelligent tutoring, from Smithtown [7] through many subsequent systems [8,9,10] with varied successes. However, de Jong recently noted that for exploratory or scientific discovery ITS, there is still no general approach to balance guidance and student initiative "in such a way that learning is supported effectively, but the inquiry process is not reduced to following cookbook instructions." [11]

Independent of the evolution in ITS, commercial games now implement a wide array of learning principles. Because games that are more difficult to learn make less money, game designers have become quite adept at embedding all necessary instruction into game play. "In essence a game manual has been spread throughout the early episodes of the game, giving information when it can be best understood and practices through situated experience." [12] What games lack, however, are deep models of an individual user's knowledge of the domain that can be applied to dynamically adjust to player behavior. Instead, producers of games rely on increasingly extensive and expensive play testing to statically calibrate the potential challenges in the game so as to reach the widest possible audience of players. Extensive testing is required at design time because unlike an ITS, a commercial game lacks the intelligence to guide a player past an arbitrary obstacle at run-time.

3. A Plan-Based Representation of the Core Mechanics of Digital Game Play

To capture the deep structure of core game mechanics, our representation is grounded in STRIPS-style [13] descriptions of the indivisible actions which affect the domain. Our system can begin by automatically deriving a set of meta-conditions from the known features of the preconditions, effects, constraints and parameters of all the tasks available in the domain. We then model the student's knowledge of the domain with a rough-grained five-valued scale to represent the system's estimate of the likelihood that the student knows about a particular facet of each domain task. For example, in a game that teaches the processes involved in aerobic cellular respiration, the student may know that one effect of the Krebs cycle is the production of CO_2 waste but not yet know another effect of the process is the production of H_2O. This could be represented in the student model by noting that the CO_2 production of a particular action in the Krebs cycle is known, while the effect that produces H_2O is still not known.

4. Integrating Game Mechanics With Learning

This representation is being implemented within a new system called Annie. Annie constructs an initial tutorial plan consisting of a plausible partially-ordered sequence of student and system-initiated actions that is designed to bring about a specific goal state for the world, including a particular state of task knowledge acquisition in the student model. The plan marks out the optimal game play path for the user prior to the start of the session, but it is continually revised based on student actions.

Annie's execution loop iterates each time an action is taken in the world, either by the student or the system. Following the action Annie consults an extensive library of general **diagnostic** templates to update its student model. These templates encode domain-independent plan reasoning diagnostics such as cases where a student seems to be ignorant of a precondition of a particular operator. Annie uses the updated student model in consulting a second extensive and domain-independent library containing **remediation** templates that can be used to generate narrative scaffolding. These templates are described more fully in [14].

Although these templates target a fairly primitive level of specific elements describing tasks, their hierarchical compositions mirror methods in which learning principles are currently embedded in games. An auxiliary goal for Annie is to take the learning principles currently embedded in games, like those featured in Gee's thoughtful and wide-ranging survey [12], and make them available at run-time as tools for dynamic adaptation to learners' needs. Gee describes 36 different learning principles, for which nearly a dozen map well to plan-based implementation structures.

An example of just one of these is the "Explicit Information On-Demand and Just-in-Time Principle" Gee describes this as the system giving explicit information "both on-demand and just-in-time, when the learner needs it or just at the point where the information can best be understood and used in practice". Commercial games achieve this by carefully scripting the progress of each segment of the game and extensively testing players to ensure that the right amount and type of help is given at the right times.

In games that are less tightly scripted and more exploratory, it is difficult for the system to deduce what the player is trying to do now or plans to do next. Without that knowledge it is difficult to choose the right help at the right time. Annie leverages its plan-based student model to solve this problem. As the student demonstrates correct or incomplete knowledge of the tasks in the domain, Annie applies its diagnostic templates to update its model of what the student knows. In addition, Annie can leverage the rich space it computes of all possible successful plans to rank the student's "knowledge gaps" by criticality and urgency. It does this by grouping the set of actions that could execute next, and the set that could be executed after one intermediate action, the set that could execute after two intermediate actions, and so on. This provides an urgency metric. Across these different urgency levels, the system can prioritize the most likely gaps in the student model based on its past observations of student behavior.

When Annie detects a knowledge gap diagnosed through its template library whose urgency is sufficiently high, a library of remediation templates is consulted to select the most useful plan revision template. Plan revisions may include things like changes to parts of the world that have not yet been explored, introduction of new characters in the game, or direct instructive interactions with the player through non-player controlled characters.

Space limitations prevent a broader exploration of the relationship between game play mechanics for learning and our knowledge representation. For a more complete discussion, see reference [15].

5. Conclusion

This paper has identified some of the difficulties inherent in building intelligent educational games, specifically the challenge of integrating pedagogy with core game play. We have briefly described a plan-based knowledge representation we believe can address this challenge for a subset of task-rich exploratory environments. We are building a system that leverages this representation to provide intelligent tutoring and plan to evaluate its teaching effectiveness in a game built using a commercial 3D game engine. An issue that may be explored in future work is whether the knowledge representation we have developed may be applied to manage student motivation and provide new mechanisms for evaluating student knowledge and performance.

References

[1] K. VanLehn. The Behavior of Tutoring Systems. *International Journal of Artificial Intelligence in Education*, 16(3):227–265, 2006.

[2] Marc Prensky and Jan-Cannon Bowers. Serious games debate. Serious Games Summit, October 2005.

[3] Maic Masuch. Is this the real thing? realism and fun in computer games. NSCU 'Future of Games' Lecture Series, August 2007.

[4] Hal Barwood. The language of games. Austin Game Writers Conference, October 2005.

[5] L.S. Vygotsky. *Thought and Language*. MIT Press Cambridge, Mass, 1986.

[6] T. Murray and I. Arroyo. Towards Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems. *ITS*, 2002.

[7] V.J. Shute and R. Glaser. A Large-Scale Evaluation of an Intelligent Discovery World: Smithtown. *Interactive Learning Environments*, 1(1):51–77, 1990.

[8] A. Bunt and C. Conati. Probabilistic Student Modelling to Improve Exploratory Behaviour. *User Modeling and User-Adapted Interaction*, 13(3):269–309, 2003.

[9] B.W. Mott and J.C. Lester. U-director: a decision-theoretic narrative planning architecture for storytelling environments. *AAMAS*, pages 977–984, 2006.

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

[11] T. de Jong. Technological advances in inquiry learning. *Science(Washington, D. C.)*, 312(5773):532–533, 2006.

[12] J.P. Gee. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1(1):20–20, 2003.

[13] R. Fikes and N. Nilsson. Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intellligence*, 2(3/4), 1971.

[14] J.M. Thomas and R.M. Young. Using Task-Based Modeling to Generate Scaffolding in Narrative-Guided Exploratory Learning Environments. *To appear in the Proceedings of the 14th International Conference on Artificial Intelligence in Education*, 2009.

[15] James M. Thomas and R. Michael Young. Guiding discovery learning with an extensible representation of actions in digital games. Technical Report DGRC-2009-01, Digital Games Research Center, North Carolina State University, Raleigh, North Carolina, 2009. Available as `http://dgrc.ncsu.edu/pubs/DGRC-2009-01.pdf`.