Evaluating Adaptive Feedback in an Educational Computer Game

Cristina Conati and Micheline Manske
Computer Science Department, University of British Columbia
2366 Main Mall, Vancouver, BC, V6T1Z4, Canada
conati@cs.ubc.ca

Abstract. In this paper, we present a study to evaluate the impact of adaptive feedback on the effectiveness of a pedagogical agent for an educational computer game. We compare a version of the game with no agent, and two versions with agents that differ only in the accuracy of the student model used to guide the agent's interventions. We found no difference in student learning across the three conditions, and we report an analysis to understand the reasons of these results.

Keywords. Educational games, student modeling, evaluation.

1 Introduction

Educational computer games (edu-games) are an increasingly popular paradigm embedding pedagogical activities in highly engaging, game-like interactions. While edu-games usually increase student engagement and motivation, there is still limited evidence on their pedagogical potential (see [1] for an overview). One possible reason for these results is that most edu-games are designed based on a one-size-fits-all approach, rather than being able to respond to the specific needs of individual students. We aim to overcome this limitation with intelligent pedagogical agents that can provide individualized support to student learning during game playing [2].

Providing this support is challenging because it requires a careful trade-off between fostering learning and maintaining engagement. Our long-term goal is to enable our agents to achieve this trade-off by relying on models of both student learning and affect [2]. In this paper, however, we analyse the performance of an agent that acts only on the basis of a model of student learning. In particular, we describe a study to evaluate the effect of improving model accuracy on the agent's pedagogical effectiveness.

Although there is widespread interest in educational computer-games, adaptive versions of these learning tools are still relatively new, and empirical evaluations of the learning benefits of having adaptive game components are rare (see next section). The evaluation we discuss in this paper focuses on Prime Climb, an adaptive edu-game for number factorization. Our evaluation is novel because it is the first in adaptive edu-games research to combine an analytical evaluation of the accuracy of the game's student model with an empirical evaluation of the effectiveness of adaptive interventions based on this

model. Although our study shows no advantage in having an accurate student model, our methodology allows us to provide insights into the reasons for this null-result, representing a step towards understanding how to devise effective adaptive edu-games. In the rest of this paper, we first discuss related work. Next, we describe Prime Climb and the versions of its agent and student model that we evaluated. We then present the study and its results, and discuss implications for future work.

2 Related work

Because of the highly motivating nature of electronic games, there has been growing interest in investigating whether they could be utilized to assist learning, especially for those children who lost interest in math or other science courses [15,16]. Results on the effectiveness of these educational tools, however, are mixed. There is evidence that these games can increase student engagement and motivation (e.g., 17, 18), but the results on their pedagogical potential are limited (e.g., [1],[15],[19]), unless the interaction is led by teachers and integrated with other instructional activities [e.g., 16]. There is also initial evidence that, for some students, educational games can be less engaging and motivating than more traditional e-learning tools [14].

One of the main reasons for these limitations of educational games is that learning how to play the game does not necessarily imply learning the target instructional domain. Learning happens when students actively build the connections between game moves and underlying knowledge. However, building these connections on one own is a form of exploratory or discovery learning, and there is extensive evidence that not all students are proficient in these activities , because they lack relevant meta-cognitive skills such as self-explanation and self-monitoring) [10, 20]. These students tend to perform better in more structured pedagogical activities [10], thus they may benefit from having some form of tutorial guidance when playing educational games.

In light of these findings, researchers have started investigating *adaptive* educational games, that is games that can autonomously tailor the interaction to the specific needs of each individual player. Although adaptive techniques have been successfully applied to other types of computer-based instructional environments [12], research on adaptive educational games is still in its infancy, and there are very few formal evaluations that explicitly target the pedagogical impact of adding adaptive functionalities to educational games. Both [9] and [13] showed that it is possible to devise user models that can capture student learning in educational games. The work in [9] relates to the educational game targeted by this paper and described in the next section. The work in [13] describes a model of student learning for Zombie Division, an educational game designed to help elementary school students learn about division. None of these works, however, show that having an adaptive component built on their student models supports learning. KMQuest [3], an adaptive edu-game for business decision-making, was shown to significantly improve student learning, but was not compared with a non-adaptive version. The Tactical Language and Culture Training System (TLCTS) supports language learning by combining an ITS component (the Skill Builder) and two games [4]. TLCTS is being actively used by the US military, and there is substantial evidence of its pedagogical effectiveness.

However, in TLCTS the adaptive behaviors reside primarily in the ITS component, and the only results on how the games contribute to system effectiveness relate to increasing student motivation [5]. The Elektra project [6] is a large research initiative aiming at defining a general methodology and tools to devise effective educational games. The proposed methodology includes having cognitive and motivational student models to allow a game to react adequately to the individual learner's cognitive and motivational needs. One of the games built as part of the project for teaching the physics of optics was evaluated and the study results showed positive trends in students perceived effectiveness of the game's adaptive interventions. The results, however, failed to provide results on actual learning gains [6]. McQuiggan et al. [7] evaluate the impact of rich narrative in a narrative-based adventure game for teaching microbiology, but there is no adaptive component in this system.

3 The Prime Climb game, its agent and student model

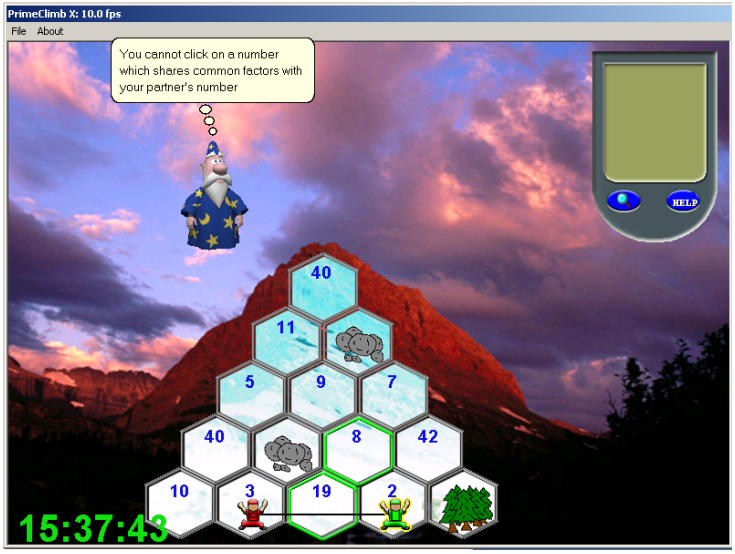

Figure1: The Prime Climb Interface

In Prime Climb, students in 6th and 7th grade practice number factorization by pairing up to climb a series of mountains. Each mountain is divided into numbered sectors (see Figure 1), and players must try to move to numbers that do not share common factors with their partner's number, otherwise they fall. To help students, Prime Climb includes the Magnifying Glass, a tool that allows players to view the factorization for any number on a mountain in the PDA device displayed at the top-right corner on the game interface (see Figure 1). Each student also has a pedagogical agent (Figure 1) that provides individualized support, both on demand and unsolicited, when the student does not seem

to be learning from the game. In the next subsections, we describe two versions of the agent, built through an iterative cycle of design and evaluation.

3.1 First Version of the Prime Climb Agent

To provide appropriate interventions, the agent must understand when incorrect moves are due to a lack of factorization knowledge vs. distraction errors, and when good moves reflect knowledge vs. lucky guesses or playing only based on game heuristics. Thus, Prime Climb includes a student model, based on Dynamic Bayesian networks, that assesses the student's factorization knowledge for each of the numbers involved in a Prime Climb session (*factorization skills* from now on) based on the student's game actions [8]. A first version of the agent gave hints at incremental levels of detail based on this model, as is commonly done in several ITS [21], with the goal of triggering student reasoning about number factorization as they play.

- The first (*focus*) level aims to channel the student's attention on the skill that requires help. For instance, the agent says *"Think about how to factorize the number you clicked on"* if the student model predicts that the student doesn't know how to factorize that number;
- the second (*tool*) level is a hint that encourages the student to use the magnifying glass to see relevant factorizations.
- The third (*bottom-out*) level gives either the factorization of a number or which factors are in common between two numbers [8].

Students can choose to progress through the various levels by asking for further help. Otherwise, the agent goes through the progression when it needs to intervene on the same skill more than once. Hints are provided regardless of the correctness of the student's move, if the student model assesses that the student needs help with the relevant number factorization skills.

Table 1: Sample revised hinting sequence triggered by a student not knowing the factorization of a number

Focus	Think carefully how to factorize the number you clicked on.
Definition 1	Factors are numbers that divide evenly into the number. Here's an example.
Definition 2	Factors are numbers that multiply to give the number. Look at this example.
Tool	You can use the magnifying glass to see the factors of the number you clicked on.
Bottom-out	You fell because x and y share z as a common factor. x can be factorized as $x_1*x_2*...*x_n$. y can be factorized as $y_1*y_n*...*y_m$.

An empirical study showed that this first version of the Prime Climb agent generated better student learning than the game with no agent [8]. A follow-up analysis of the student model used in this study showed limited accuracy (50.8%), due to various

limitations of the model, discussed in [9]. The fact that an agent based on this model could still trigger learning indicates that even hints based on an almost random model are better than no hints at all. However, there was still room for improvement in the post-tests of the agent-condition (the post-test average was 77%), suggesting that a more accurate student model may yield even more substantial learning gains.

3.2 Second version of the Prime Climb Agent

Following the results of the evaluation of the first version of the Prime Climb agent, we devised a new version of its student model that addressed the limitations uncovered by the study and that achieved an accuracy of 78% in assessing student factorization knowledge [9]. We also changed the agent's hinting strategy. We added a fourth hinting level (*definition*), to provide reteaching of the *factorization* and *common factor* concepts via definitions and examples. The original set of hints did not include an explanation of these concepts, thus students who still needed to understand them could only do so via some form of discovery learning during game playing. There is ample evidence, however, that for many students discovery or inquiry based learning is less effective than more structured instruction in the early stages of learning [10]. This effect may be more prominent with edu-games, when students are too busy playing to engage in non-spontaneous learning processes. Table 1 shows a sample revised hinting sequence.

Figure 2: Sample example that the agent presents to accompany Definition 1 in table 1

As the table shows, we provide two different factorization definitions, because there is no common unifying definition for this concept. The agent alternates which definition to give first, and gives the second the next time it needs to provide an unsolicited hint on the same skill. Figure 2 shows a screenshot of an example that accompanies Definition 1 in Table 1. The examples at this level are general (i.e., do not relate to the number targeted by the current hint) and serve both to solidify the

student's understanding of the definition and as a template for finding the factors of other numbers that the student sees on the mountain. *Definition* hints are given before the *tool* hint the first time the student goes through the hinting sequence, as shown in Table 1. Subsequently, they are given after the *tool* hint, because at this stage the student may just need a trigger to put together the definitions and examples seen earlier in order to find the answer by herself. All hints and examples were designed based on the experience of the second author, a former elementary school teacher (and award-winning university teaching assistant), and then extensively pilot-tested.

In the rest of the paper, we describe a study that we ran to test if and how the more accurate model we developed for Prime Climb impacts the effectiveness of the Prime Climb agent with this new hinting strategy.

4 Study design

The study was run in two local elementary schools with sixth grade students, with the constraint that each study session had to be held during a class period (40 minutes) to avoid disrupting regular class schedule. The students were randomly assigned to one of three conditions: *No Agent*: game with no agent nor any other form of adaptive support (13 students); *Old-model*: game with the pedagogical agent and the original version of the student model (14 students). *New-model*: game with the pedagogical agent and the new, more accurate, version of the student model (17 students).

The morning of the study, all students wrote a pre-test in class, designed to assess the students' factorization knowledge of various numbers involved in the Prime Climb game. The rest of the study was conducted with pairs of students in a separate room, due to constraints on computer availability. The two students were excused from the class for that period and joined the experimenters in a room provided by the school for the experiment. Following the set-up that had been successfully adopted in [8], each session was designed to last at most 30 minutes so that there would be sufficient time for students to get to the study room and return to their class for the next period. Students were told that they would be playing a computer game, and received a demo of Prime Climb. They were told that the game contained a computer-based agent that was trying to understand their needs and help them play the game better. Next, students played with one of the three versions of Prime Climb for approximately 10 minutes. We had to limit playing time to 10' to allow for sufficient time for post-tests and post-questionnaires, because they could not be taken during regular class hours. It should be noted that, although these were relatively short sessions, sessions of the same length in the study on the older version of the Prime Climb agent [8] were sufficient to show learning effects. Each student played with an experimenter as her partner, to avoid confounding factors due to playing with partners with different knowledge and playing behaviors. Experimenters made sure that students obtained help only from the pedagogical agent. After game play, all students wrote a post-test equivalent to the pre-test, and students in the *old-model* and *new-model* conditions filled out a questionnaire on their impressions of the agent.

5 Results

5.1 Impact on learning

We measure learning gains as the difference between post-test score and pre-test score. The study hypotheses are the following:

H1: *Students in the new-model condition will learn significantly more than students in the old-model condition.*

H2: *Students in conditions with the agent will learn more than students in the no-agent condition.*

Table 2 shows the results by condition. An ANOVA using learning as the dependent variable, condition as main factor, and pre-test scores as covariate (to control for student incoming knowledge) shows no significant differences between the three conditions.

Table 2: Pre-test, post-test and learning gain results by condition (maximum test score is 30)

	Average score (st. dev)		
	No-Agent	**Old-Model**	**New-model**
Pre-test	20.62 (2.83)	25.53 (1.81)	25.77 (1.72)
Post-test	19.39 (3.41)	25.40 (1.88)	25.35 (1.84)
Learning	-1.23 (1.33)	-0.13 (0.42)	-0.41 (0.64)

Thus, we have not been able to prove either of our two hypotheses. The fact that we did not manage to reproduce the results in [8], i.e., to show that having a pedagogical agent is better than not having one (H2 above), is especially surprising, given that, compared to the agent in [8], the new agent used in the study had a more accurate model and an improved set of hints, carefully designed by an experienced teacher. Students in the current study did have a higher level of knowledge than students in [8], scoring an average of 83% on the pre-test compared to 60% in [8], so it was indeed harder to see an effect of pedagogical interventions with this student population. But there were still several occasions in which agent interventions could have triggered learning (as we will discuss in the next sub-section). We investigate two possible reasons for the null effect of the improvements we made to both the agent and its model: (1) in this study, the new model was not more accurate than the old model; (2) elements of the new hinting strategy obstructed learning.

5.1 Comparison of Models' Accuracy

The accuracy of the old and new model reported in previous sections referred to model assessment of student factorization skills at the end of the interaction,

compared with student post-test performance [9]. A measure that is more informative for understanding model impact on learning (or lack thereof) is accuracy during game playing, influenced by how quickly the model stabilizes its assessment of student knowledge. We can't determine this accuracy on all the target factorization skills, because we do not have a ground-truth assessment of how the related knowledge evolves during game playing. We can, however, restrict the analysis to skills for which the student's answer did not change from pre-test to post-test, i.e., the related knowledge was constant throughout the interaction. Since there was little learning in the study (see Table 2), this selection covers a substantial fraction of our data points.

Table 3: Confusion matrices (# of raw data points) for the accuracy of the old model (left) and new model (right)

	Old Model			**New Model**		
	Test assessment			**Test assessment**		
Model Assessment	***Known***	***Unknown***	***Total***	***Known***	***Unknown***	***Total***
Known	369	84	453	354	27	381
Unknown	19	4	23	54	76	130
Total	388	88	476	408	103	511

Table 4: Confusion matrices (percentages) for the accuracy of the old model (left) and new model (right)

	Old Model			**New Model**		
	Test assessment			**Test assessment**		
Model Assessment	***Known***	***Unknown***	***Total***	***Known***	***Unknown***	***Total***
Known	77.5%	17.6%	95.1%	69.3%	5.3%	74.6%
Unknown	4%	0.9%	4.9%	10.5%	14.9%	25.4%
Total	81.5%	18.5%	100%	79.8%	20.2%	100%

The logs files from the old-model and new-model conditions included, for each student action, the model's assessment of the student factorization knowledge after that action. We searched these log files for all episodes in which a student encountered a number with the same pre-test and post-test results (*known* vs. *unknown*), and compared these results with the model's assessment for that number at

that point (also expressed in terms of *known* vs, *unknown*). Table 3 and Table 4 show the confusion matrices (with raw data and percentages, respectively) for the two models across students and all relevant episodes. We calculate from these matrices two standard measures of accuracy: recall (fraction of all *unknown* data points that the model classifies as such) and precision (fraction of all data points that the model classifies as *unknown* and that are actually unknown). Recall and precision are important from the pedagogical point of view, because they define, respectively, how good the model is at detecting situations in which the student's knowledge is low, and how good the model is at generating interventions that are justified.

The old model has very poor performance in both recall (4.5%), and precision (17.4%). With 73.7% recall and 58.5% precision, the new model clearly outperforms the old model. We conclude that we can reject lack of difference in model accuracy as a reason for the null result with respect to H1 (more learning in the new-model condition than in the old-model condition). We now explore the second reason, i.e., that elements of the agent's hinting behavior obstructed learning.

5.2 Effects of the agent's hinting behavior

One factor that may disrupt learning is how often the agent intervenes, influenced by the student model. The last row of each confusion matrix in Table 4 shows that the breakdown of *known* and *unknown* data points is approximately 80%:20% for both conditions, indicating that the underlying student knowledge is the same in both groups (confirmed by a lack of significant differences in their pre-test scores). However, the last column in Table 4 shows that the old model judges factorization skills to be unknown 4.9% of the time, compared to 25.4% for the new model. Thus, the new model causes the agent to intervene much more often. In fact, there is a significant difference ($p < 0.001$, as per a two-tail t-test) between the average number of hints each student received in the old-model condition (mean 7.6, st. dev. 3.6) and in the new-model condition (mean 16.3, st. dev. 5.5). This difference is mostly due to the model's assessment, given that students in both agent conditions rarely asked for hints (The requested hints were only 3.4% of all given hints. [8] reports similar results with respect to student hints requests).

The fact that students in the old-model condition received very few justified hints explains why they did not learn from the interaction with Prime Climb. It should be noted that while the study in [8] used the same model as the old-model condition, in that study students likely learned because they had less factorization knowledge to start with, thus there were more occasions to generate learning, even for a model with limited recall/precision. As for the more frequent hints generated by the new model, although more of these are justified (58.4%) than the old model's hints (14.4%), students may not like to have their game playing interrupted by didactic interventions, especially when about 40% of these interruptions are not justified. This may have caused students to stop paying attention to the hints. To verify this conjecture, we looked at whether students in the new-model condition are taking the time to read the hints and accompanying examples.

Our log files do not contain the action of closing a hint, so we can't use the time between the appearance of a hint and its closure as an estimate for reading time. We

use instead the difference between the average time taken to make a move after getting a hint (12.82 sec., st. dev. 4.22), and the average time taken to make a move when there is no hint (9.14 sec., st. dev. 3.02). We obtain 3.42 seconds (st. dev. 2.62) as an estimate of the average amount of time each student spent reading a hint. The average adult reader can read 3.4 words per second [11]. With hints that were 22.5 words on average, an adult would take 6.62 seconds on average to read the hints. Thus, it is conceivable that students were not taking time to read the hints thoroughly and think about their meaning. This conclusion is supported by the fact that there are no significant correlations between the estimated time spent reading hints, or the number of hints received, and learning. As further evidence of lack of attention to hints, we compare the times between receiving a hint and performing an action for the *Focus* and *Definition* hints, the first time they are presented (see Table 5). The second row reports the number of hint words, not including the words in the accompanying examples.

Table 5: Average (and st.dev.) time (in seconds) between receiving a hint for the first time and acting

Hint Type	*Focus*	*Definition 1*	*Definition 2*
Words	19	26	27
Avg. time (st.dev.)	12.03 (4.53)	13.57 (4.41)	13.00 (3.75)

As expected, students spend more time between receiving a hint and performing an action with hints that involve examples (*Definition 1* and *Definition 2*) than with the *focus* hint. However, the additional time spent does not account for their higher number of words in *Definition* hints. For instance, *Definition 1* hint is 7 words longer than the focus hints, thus we would expect an average (adult) reader to spend approximately 2 seconds longer to read it, plus time to examine the example. Table 5 shows that students are not taking the time, and thus are probably not reading the hints thoroughly. If students are not finding the hints generated by the agent in the new-model condition useful, this should affect their perception of the agent. To see if this is the case, we look at the students' post-questionnaire answers.

5.3 Student's perception of the Prime Climb agent

The post-questionnaires on agent perception included six questions rated on a Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The average score (and standard deviation) for each question in the two agent conditions are shown in Table 6. We see that all the questions are in favor of the old-model condition, although the only difference that is statistically significantly is Q1: the agent in the old-model condition is rated as more helpful than the other agent ($p = 0.017$). This result is consistent with the picture that emerged from the previous sections: more students in the new-model condition received a hint, but they tended not read it, so the hint was not helpful to them. It is not surprising that more of these students rated the agent as "unhelpful", and that it received a quite high score for "intervening too often".

Interestingly, the agent in the old-model condition also scored quite poorly on this item, despite the fact that it intervenes much less than the other agent. This may be due to a general student dislike of any interruption of their game playing.

Table 6: Average responses (and st. dev.) in the post-questionnaire

Question	Old-model	New-model
Q1: the agent is helpful	3.60 (0.22)	2.56 (0.34)
Q2: the agent understands my needs	3.00 (1.05)	2.67 (1.41)
Q3: the agent helps me play better	2.80 (0.92)	2.44 (1.13)
Q4: the agent helped me learn factorization	3.20 (0.92)	2.56 (1.13)
Q5: the agent intervenes too often	3.20 (1.48)	3.89 (1.05)
Q6: I liked the agent	3.60 (1.07)	3.11 (1.36)

6 Discussion, Conclusions and Future Work

We have presented a study to evaluate the impact of adaptive feedback on the effectiveness of a pedagogical agent for an educational computer game. We compared a version of the game with no agent, and two versions with agents that differ only in the accuracy of the student model used to guide their interventions. We found no difference on student learning across the three conditions, so we combined an analysis of model accuracy during game playing with an analysis of log data on student relevant behaviors to understand the reasons for these results. We eliminated lack of difference in model accuracy as a possible cause for the null results, because the student model that was known to be more accurate in assessing student knowledge at the end of the interaction (new model) was also more accurate in assessing student knowledge during game playing. This model generated significantly more justified hints than the other model (old model). However, over 40% of the hints it generated addressed skills that students already had. This is likely one of the reasons why students seem to not pay attention to the hints, and thus failed to learn from the game.

Ironically, the old, less accurate model with simpler hints used by the first version of the Prime Climb agent (described in section 3), did generate more learning than the game with no agent [8]. This result is likely due to the combination of two factors. The study participants had low factorization knowledge, and thus there were more occasions for the few justified system interventions to have an effect than in the study presented here, where students scored 83% in the pre-test, on average. Because the system did not interrupt game playing often and because the hinting sequence was

shorter and simpler, students did not perceive it as intrusive, paid more attention to the hints and sometime they learned.

An obvious direction to improve the effectiveness of the adaptive hints' is to improve model precision, so that more of the agent's interventions are justified. However, students may resent being interrupted often during game play even when most interruptions are justified. Our results suggest a simple solution: some learning can be achieved with an inaccurate model, by favoring unobtrusiveness over intervening when it seems necessary. In Prime Climb, we could achieve this by lowering the probability threshold that dictates when a skill is considered known in the student model. A more interesting, although more challenging solution is to endow the model with the ability to reason about the expected effects of its interventions on both student learning and affect, to achieve a trade-off between maintaining engagement and promoting maximum learning. A decision-theoretic approach that combines a model of student learning with a model of student affect is one way around this issue [2]. We plan to explore both solutions, to determine their relative impact on game effectiveness. For the latter, we plan to combine the model of student learning described here with the model of affect we have been developing in parallel [21,22]. Another direction of investigation relates to the *form* of the agent's hints, i.e. how to devise pedagogical hints that can be perceived as less didactic and intrusive [e.g., 23, 24] and can thus be more acceptable for students during game playing.

References

[1] Van Eck, R., (2007). Building Artificially Intelligent Learning Games. *Games and Simulations in Online Learning: Research and Development Frameworks*. D. Gibson, C. Aldrich, and M. Prensky, Editors, Information Science Pub. 271-307.

[2] Conati, C. and M. Klawe (2002), Socially Intelligent Agents in Educational Games. *In Socially Intelligent Agents - Creating Relationships with Computers and Robots*. K. Dautenhahn, et al., Editors, Kluwer Academic Publishers.

[3] Christoph, N., J. Sandberg and B. Wielinga (2005). Added value of task models and metacognitive skills on learning. In *AIED '05 Workshop on Educational Games as Intelligent Learning Environments*.

[4] Johnson, W. L. (2007). Serious use for a serious game on language learning. In *Proc. of the 13th Int. Conf. on Artificial Intelligence in Education*, Los Angeles, USA.

[5] Johnson, W.L. and C. Beal (2005). Iterative Evaluation of a Large-scale, Intelligent Game for language Learning. In *AIED '05: Proceedings of the 12th International conference on Artificial Intelligence in Education*, Amsterdam, The Nederlands.

[6] Peirce, N., O. Conlan and V. Wade (2008). Adaptive Educational Games: Providing Non-invasive Personalised Learning Experiences. In *Second IEEE International Conference on Digital Games and Intelligent Toys Based Education (DIGITEL 2008)*, Banff, Canada.

[7] McQuiggan, S.W., Rowe, J., Lee, S. and J. Lester (2008). Story-Based Learning: The Impact of Narrative on Learning Experiences and Outcomes. In Proc. of *ITS 2008*, Montreal, Canada.

[8] Conati, C. and X. Xhao (2004). Building and Evaluating an Intelligent Pedagogical Agent to Improve the Effectiveness of an Educational Game. In *Proceedings of IUI '04, International Conference on Intelligent User Interfaces*, Island of Madeira, Portugal.

[9] Manske, M. and C. Conati (2005). Modelling Learning in Educational Games in AIED 05, *Proceedings of the 12th International Conference on AI in Education*. 2005. Amsterdam, The Nederlands.

[10] Kirschner, P., J. Sweller and R. Clark (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experimental and inquiry-based teaching. *Educational Pshychologist*, 2006. 41 (2): p. 75-86.

[11] Just, M. and P. Carpenter (1986). *The Psychology of Reading and Language Comprehension*, A. Bacon. 1986, Boston.

[12] Wool, B. (2008) Building intelligent interactive tutors, Morgan Kauffman.

[13] Baker, R.S.J.d., Habgood, M.P.J., Ainsworth, S.E., Corbett, A.T. (2007) Modeling the Acquisition of Fluent Skill in Educational Action Games. *Proceedings of User Modeling 2007*, 17-26.

[14] Rodrigo, M.M.T., Baker, R.S.J.d., d'Mello, S., Gonzalez, M.C.T., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sugay, J.O., Tep, S., Viehland, N.J.B. (2008) Comparing Learners' Affect While Using an Intelligent Tutoring Systems and a Simulation Problem Solving Game. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 40-49

[15] Randel, J.M., B.A. Morris, C.D. Wetzel, and B.V. Whitehill, *The effectiveness of games for educational purposes: A review of recent research.* Simulation & Gaming, 1992. 23(3): p. 261-276.

[16] Klawe, M. *When Does The Use Of Computer Games And Other Interactive Multimedia Software Help Students Learn Mathematics?* NCTM Standards 2000 Technology Conference. 1998. Arlington, VA.

[17] Alessi, S.M., Trollip, S.R. (2001). Multimedia for Learning: Methods and Development, 3rd ed. Allyn & Bacon, Needham Heights.

[18] Lee, J., Luchini, K., Michael, B., Norris, C., Solloway, E.(2004). More than just fun and games: Assessing the value of educational video games in the classroom. *Proceedings of ACM SIGCHI 2004*, Vienna, Austria, pp. 1375–1378.

[19] Vogel, J.J., Greenwood-Ericksen, A., Cannon-Bowers, J., Bowers, C.A.(2006).Using virtual reality with and without gaming attributes for academic achievement. *Journal of Research on Technology in Education* 39(1), 105–118.

[20] Conati, C. and J. Fain Lehman. *Toward a Model of Student Education in Microworlds* . 15th Annual Conference of the Cognitive Science Society. 1993: Hillsdale, NJ, Erlbaum.

[21] Conati, C. and H. Maclaren (2009). Empirically Building and Evaluating a Probabilistic Model of User Affect, *User-Modeling and User-Adapted Interaction*, (in press).

[22] Conati C. and H. Maclaren (2009). Modeling User Affect from Causes and Effects. To appear in *Proceedings of UMAP 2009, First and Seventeenth International Conference on User Modeling*, Adaptation and Personalization, Springer

[23] Arroyo, I., Ferguson, K. Johns, J., Dragon, T., et. al. (2007) Repairing Disengagement With Non-Invasive Interventions. AIED 2007: 195-202

[24] Ryan S. J. d. Baker, Albert T. Corbett, et al. (2006): Adapting to When Students Game an Intelligent Tutoring System. Intelligent Tutoring Systems 2006: 392-401.