AIED 05 WORKSHOP 4

Amsterdam, 18-22 July, 2005

Workshop on Motivation and Affect in Educational Software

WORKSHOP

12th International Conference on Artificial Intelligence in Education, Amsterdam, the Netherlands

Table of contents

Motivation and Affect in Educational Software

Organizing Committee

Cristina Conati, University of British Columbia, Canada: conati@cs.ubc.ca
Benedict du Boulay, University of Sussex, UK: B.Du-Boulay@sussex.ac.uk
Claude Frasson, University of Montreal, Canada: frasson@iro.umontreal.ca
Lewis Johnson, USC, Information Sciences Institute, USA: johnson@isi.edu
Rosemary Luckin, University of Sussex, UK: R.H.Luckin@sussex.ac.uk
Erika A. Martinez-Miron, Univ. of Sussex, UK: E.A.Martinez-Miron@sussex.ac.uk
Helen Pain, University of Edinburgh, UK: helen@inf.ed.ac.uk
Kaska Porayska-Pomsta, University of Edinburgh, UK: kaska@inf.ed.ac.uk
Genaro Rebolledo-Mendez, Univ. of Sussex, UK: G.Rebolledo-Mendez@sussex.ac.uk

Aims

Motivation and affect (e.g., basic affective reactions such as like/dislike; specific emotions such as frustration, happiness, anger; moods; attitudes) often play an important role in learning situations. There have been various attempts to take them into account both at design time and at run time in AIED systems, though the evidence for the consequential impact on learning is not yet strong. Much research needs to be carried out in order to better understand this area. In particular, we need to deepen our knowledge of how affect and motivation relate to each other and to cognition, meta-cognition, learning context and teaching strategies/tactics. This workshop is intended bridge the gap existing between previous AIED research, particularly in motivation and meta-cognition, with the ever-increasing research in emotions and other affective components.

By bringing together researchers in the area, the workshop will be a forum to discuss different approaches with the aim of enriching our knowledge about how to create effective and affective learning environments. Also, it is expected to be a forum on which to address the appropriateness of defining bridges that could bring about new ways of relating cognitive and affective aspects of learning. At the end of the workshop we expect to reach agreements on which are the relevant emotions in learning contexts, as well as in the terminology been used so far (e.g. affect, emotion, motivation).

Call for Papers

We invited papers, which present either finished, or work in progress or theoretical positions in the following areas:

- Affective/motivational modelling.
- Affective/motivational diagnosis.
- Relevant aspects of motivation and affect in learning.
- Strategies for motivational and affective reaction,
- Integrative models of cognition, motivation, and affect.
- Personal traits, motivation, and affect.
- Learning styles, learning domains and learning contexts.
- Learning goals, motivation, and affect.
- Influences of dialogues in affective computing.
- Use of agents as affective companions.
- Interface design for affective interactions.

The workshop is focused on exploring the following questions:

- Which emotions might be useful to model (e.g. basic affective reactions such as like/dislike; specific emotions such as frustration, happiness, anger; moods; attitudes)?
- How do individual traits influence the learner's motivational and affective state?
- How are motivation and emotional intelligence related?
- How do cognitive aspects affect the learner's emotional intelligence and vice versa?
- Is it important and feasible to standardise cognitive and emotive terms (cognition/meta-cognition, emotion/meta-emotion)?
- How do different learning contexts –e.g. individual, collaborative- affect the student's emotions (and vice versa)?
- Does emotional intelligence mature with age?
- How are emotions affected according to the learning domain?
- What are the most appropriate ways to assess motivation and affect?

An affective agent-based virtual character for learning environments

Elisa BOFF
Departamento de Informática, Universidade de Caxias do Sul
Rua Francisco Getúlio Vargas, 1130 - 95070-560 Caxias do Sul, RS, Brazil
eboff@ucs.br

Eliseo Berni REATEGUI
Departamento de Informática, Universidade de Caxias do Sul
Rua Francisco Getúlio Vargas, 1130 - 95070-560 Caxias do Sul, RS, Brazil
ebreateg@ucs.br

Rosa Maria VICCARI
Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
rosa@inf.ufrgs.br

Abstract. This paper presents a virtual character placed in an educational environment with the main purpose of stimulating cooperative learning among students. The character can communicate with users in natural language; it can suggest them appropriate contents and activities, as well as recommend student tutors that may be helpful at given learning tasks. For such, a data mining mechanism is used to find patterns in students' behavior and interactions, taking into account different social-affective aspects. The paper discusses such social-affective issues, and proposes a multi-agent architecture. Preliminary results of initial experiments are discussed, conclusions and directions for future work are presented.

Introduction

Personifying a website means adding a human component to it, for example, by using a virtual character to communicate with and present recommendations to the user. The main advantage of this approach is that the use of the human figure can help us draw the user's attention, emphasizing recommendations or highlighting any other message we may want to deliver. Many projects have used virtual characters in presentation tasks in a wide range of applications, such as e-commerce, e-learning and helpdesks. Pandzic classifies them in six categories [1]: entertainment, communication, navigation, broadcasting, commerce and education. Rist et al [2] propose a different classification based on several levels of interactivity between characters and users. The characters may present information to the user as if they were television presenters [3], without promoting a lot of interactivity. Another way they may interact with users is through dialogues in natural language, answering questions, bringing relevant information and guiding the user in the system. The third type of interaction may happen in environments where the character "lives" with other characters in a virtual world. These characters bring information to the user, playing specific roles and interacting with one another. Finally, the users may enter such virtual worlds, interact and question the characters living in it.

The idea of developing a computational character that talks to the user is old. In 1967 Weizembaum had already developed ELIZA [4], a program that was capable to engage in conversation with the user, simulating the role of a psychiatrist. Although the use of the human figure in interfaces is still a somehow controversial topic [5], scientific results are coming to surface indicating that, at given circumstances, personification may augment user involvement and reduce workload [6]. Rickenberg and Reeves [7] found that the presence of a character in a website may increase the user's confidence, even though it may also augment the user's anxiety (which may have positive as well as negative aspects). De Angeli et al. [8] present arguments supporting the use of personified characters, explaining that by introducing social stimuli, virtual characters may improve the communication between user and computer. Virtual characters can also be found in educational environments. In [9] virtual teachers guide the students in online interactive environments. Scotty et al. [10] show how the interaction with animated and static virtual characters can affect students learning. Elliot et al. [11] also investigated the importance of the affective dimension in virtual characters, developing studies about pedagogical agents that are sensible to the states of emotion of the students and are capable of reasoning about affective aspects in problem resolution.

This paper presents an agent-based virtual character whose purpose is to stimulate cooperative learning among students by motivating their interaction and mutual assistance. The virtual character was integrated to a learning environment, communicating with the students in natural language, suggesting readings according to the activities being performed. When a student needs assistance to learn a given topic, the character is capable of finding other students that may play the role of a tutor. This is done by monitoring the students' actions and using a data mining tool to find patterns that can guide the search for the best instructor in a given situation. Such tutor recommendation service explores the social-affective dimension through the analysis of emotional states and social behavior of the users.

The next section presents social-affective issues that have to be considered in students' interaction and behavior to enable the identification of suitable tutors. Section two details the multi-agent architecture of which our virtual character is a component. Then preliminary results are discussed, conclusions and directions for future work are presented.

1. Social and Affective Issues

Current research has shown the potentiality of cooperative learning, demonstrating that group work is fundamental for the cognitive development of the student [12]. In this perspective, motivating the students to interact can lead to an effective learning practice. By identifying students with good teaching abilities and recommending them to other students with difficulties, we promote group formation and cooperative learning.

According to Andrade [13], a group can be formed due to similarity and empathy of its members or to the necessity of support for the accomplishment of some task. The latter can be motivated by prestige or status, economic benefits or the necessity and desire of contribution. Andrade also says that the affective states of the individuals have significant importance in the interaction process, and affirms that some dimensions of personality seem to have certain connections with the social performance in the interaction.

For Scherer [14], the affective states are divided in five categories: Emotion, Mood, Interpersonal Stances, Attitudes, and Personality Traits.

Emotion is the episode of evaluation of an external or internal event as being of major significance, relatively brief of synchronized responses for most organic systems. Emotion's examples are anger, sadness, joy, fear, shame, pride, elation and desperation.

Mood is a diffuse affective state that consists in a subjective feeling change, with low intensity, but long duration without apparent cause. Dipert [15] considers that moods differ from emotions most strongly in not having an intentional object. Their causes are typically

conceptual or evaluative (things are or are not going well). He mentions some examples of moods: cheerful, gloomy, irritable, listless, depressed, and buoyant.

Interpersonal Stance is an affective position in relation to the other person in a specific interaction. Distant, cold, warm, supportive and contemptuous are examples of interpersonal stances.

Attitudes are relatively tolerant, affectively colored beliefs, preferences and predisposition in relation to objects or people. Examples of attitudes are liking, loving, hating, desiring and valuing

Personality Traits are emotionally laden, stable personality dispositions and behavioral tendencies, typical of a person. For example: nervous, anxious, reckless, morose, hostile, envious and jealousy.

The cognitive approach to the modeling of emotions considers that different emotional states are attained according to evaluations based on world stimuli and the behavior of the individual. In this area, Ortony et al. [16] worked in the construction of a classification model for emotions and in the description of the reasoning process that involves them. At present, there are four basic ways to recognize emotions through a computer system: voice; facial expressions; physiological signs (blood pressure, skin conductivity, etc...); and, observable behavior. The latter corresponds to the observation of the user's interaction with the system, e.g. chosen options or typing speed.

Beyond the affective aspects, the service of recommendation of students tutors must also take into account social aspects of the students. In this field, Social Networks are a very popular approach for the analysis of human interactions. The most important concept of such approach is centrality: if an individual is central in a group, he/she is popular and gets a great amount of attention from the group members.

The work of Yang & Tang [18] investigates the relationship between the performance of the students and their position in three types of social networks: friendship, advice and adversity networks. An individual that is central in a friendship network has greater chances to have access to resources that can be important for his/her academic success. Therefore, there is a great possibility that this individual gives help and is helped. The advice networks consist of relationships through which the individuals can share information related to their work [19]. A student who is central in an advice network is capable of accumulating information, knowledge and experience on tasks related to the resolution of problems. Adversity network include those relationships that involve negative exchanges [18]. If a person is central in an adversity network, this person has little possibility to get information or knowledge from others.

2. The Multi-Agent Architecture

Our virtual character has been integrated to an educational environment to assist students in learning algorithms. The multi-agent architecture proposed here was based on an initial model presented by [20]. This model has many agents to provide services to an ITS (Intelligent Tutoring System) in different tasks. There are agents to select exercises, to support users' communication, to log users, to store users and profiles, and other. The architecture includes a *Social Agent* that interacts with a *Mining* and a *Recommender Agent*, whose goal is to identify suitable tutors and to offer their help. The interaction with the system is supported by a *Virtual Character*. All agents communicate through an *Agents Environment* using a specific interaction protocol. *Student Agents* are used to represent each student and to communicate. Our multi-agent society is presented in Figure 1.

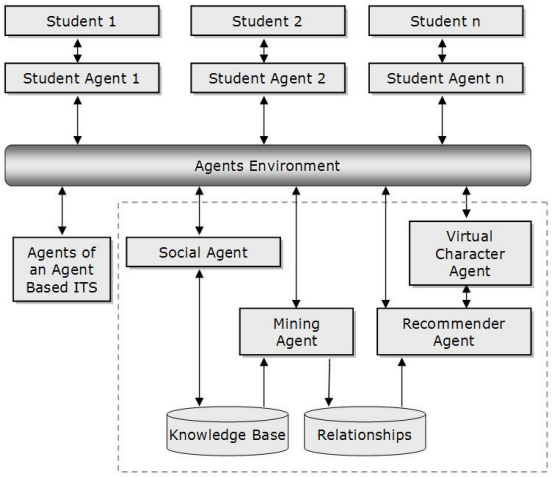


Figure 1. Multiagent Architecture

2.1. The Social Agent

The main goal of the social agent is to collect data about students' interactions and to provide such information for the mining agent. The information collected can be classified in six different categories: Social Profile; Acceptance Degree; Sociability Degree; Mood State; Tutorial Degree and Performance.

The Social Profile (SP) is built during the students' interaction through a synchronous mechanism (e.g. chat tool). During the students' interaction, some information are collected. For instance, number of times that a student had the initiative to talk with another; number of times that a student answered a communication request; individuals with whom the student interacts or has interacted, and number of interactions; individuals with whom the student interacts regularly, and number of interactions.

Based on Maturana [17] we defined the Acceptance Degree (AD), which measures the acceptance between students. Maturana affirms that the acceptance of other individuals is a necessary condition for social development. Such data is collected through a graphical interface that enables each student to indicate his/her acceptance degree for other students. This measurement may also be considered from a point of view of Social Networks. As the AD is indicated by the students themselves based on their affective structures, the measurement can indicate different emotions, such as love, envy, hatred, etc. The average of all ADs received by a student influences his/her Sociability Degree (SD).

The Mood State (MS) represents our belief in the capability of a student to play the role of a tutor if he/she is not in a positive mood (although the student may have all the technical and social requirements to be a tutor). We consider three values for the MS: "bad mood", "regular mood" and "good mood". These states are indicated by the students in a graphical interface through corresponding clip-arts.

After a helping session, a small questionnaire is submitted to the student who got assistance, with the purpose of collecting information about the performance of the tutor. The questions made are based on concepts from Social Networks and Sociometry, and may be answered by four qualitative values: "excellent", "good", "regular", and "bad". The questions are presented below:

- How do you classify the sociability of your class fellow?
- How do you classify the help given by your class fellow?

The answer to the first question together with the average of the ADs of a student, form his/her Sociability Degree (SD). This grade indicates how other individuals see the social capability of this student.

The Tutorial Degree (TD) measures a student's pedagogical capacity to help, to explain and teach. This value is obtained from the answers given for the second question of the questionnaire above and from the marks the tutor got when he/she studied the contents for which he/she was asked for help. These grades were called Performance (P) and were used in the computation of the TD because, when a tutor is not able to help another student, it does not necessarily mean that the student is a bad tutor. He/she may simply not know very well the content for which his/her help was requested.

The data collected is stored in a knowledge base, which is accessed by a mining agent in order to identify behavioral patterns. Next subsection explains this mining process.

2.2. The Mining Agent

The main task of the mining agent is to extract profiles from the data collected and to store them in a knowledge representation mechanism called item descriptors. These descriptors store information about how the social-affective characteristics described in item 2.1. may affect a student tutoring ability. Users' features may be classified in the descriptors as:

- demographic: data describing an individual, such as age, gender, occupation, address;
- behavioral: data describing tutoring and social capacity, navigation and study patterns.

It has been shown that both types of data are important when building a user profile [21] and inferring user's needs [34]. Demographic material is represented here in attribute-value pairs. Behavioral information is represented by actions carried out by the user, such as the selection of a topic for reading. Emotional states and social behavior can either be inferred or collected explicitly in the questionnaires.

While attributes used to define demographic features are typically single-valued, behavioral data is usually multi-valued. For instance, a person can only belong to one age group (demographic), but he/she may be friendly and patient at the same time (behavioral). Nevertheless, both types of information are represented in our model in a similar way.

The descriptor lists every term that is correlated to it. We use confidence as a correlation factor in order to determine how relevant a piece of information is. This is the same as computing the conditional probability $P(d_j|e)$, i.e. the probability that the item represented by descriptor d_j holds a possible answer given evidence e. Therefore, the descriptors can be learned through the analysis of actual records. For each item for which we want to define a recommendation strategy, a descriptor is created with the item defined as its target. Then, the confidence between the target and other existing demographic features and behavioural data is computed. This process continues until all descriptors have been created. For the recommendation of tutors, descriptors are built indicating the features of good and bad instructors.

2.3. The Recommender Agent

Given a list of possible user tutors $U=\{u_1, u_2,..., u_m\}$, the recommendation process starts with the gathering of demographic and behavioural information about each of them. Next, the data collected for each user is matched against a descriptor d_j which lists the most important features of good instructors, according to the terms $T=\{t_1,t_2,...,t_k\}$ stored in the descriptor. The system computes a score for each student that ranges from not similar (0) to very similar (1), according to the formula:

$$Score\ (d_j) = 1 - \prod_{kji} (Noise\ (t_p))$$

where $Score(d_j)$ is the final score of the descriptor d_j; $Noise(t_p)$ is the value of the noise parameter of term t_p, a concept used in noisy-OR probability models [23] and computed as $1 - P(d_j \mid t_p)$. The individual with the highest score is selected to assist the student needing

assistance. That expression contains an assumption of independence of the various t_p - which the designer of a practical system should be trying to achieve in the choice of terms.

This method is based on the assumption that any term matching the user's terms should increase the confidence that the descriptor holds the most appropriate recommendation. In a real-life example, let us suppose that we have a certain degree of confidence that a student who has shown a good ability in answering factorial exercises is our best bet to help another student who is having problem with the subject. Knowing that that same student is friendly and is in a good mood should increase the total confidence on his recommendation as a tutor.

2.4. The Virtual Character Agent

The virtual character agent has the goal of communicating with the user through a natural language mechanism, identifying when to recommend a tutor to a student needing assistance and triggering the recommendation process. The agent uses the Artificial Intelligence Markup Language (AIML) to represent its conversational knowledge [24], employing a mechanism of stimuli-response. The stimuli (sentences and fragments that may be used to question the agent) are stored and used to search for pre-defined replies.

The most important tags of AIML are:

<aiml>: indicates the beginning of a document;

<category>: marks the units in the knowledge base

<pattern>: keeps a pattern that is searched for in sentences that the user may enter to communicate with the virtual character;

<template>: contains the possible answers to the users' input.

In addition to the existing AIML tags, new ones were created to manage the agents' emotional appearance. For instance, we used the tag <humor> to control the image changes reflecting different moods of the virtual character (happy, receptive, annoyed, etc).

This agent also has the ability to detect that a student is having some difficulties in performing a certain task (e.g. the student is solving a given exercise more than once, the student is spending too much time at a given activity, etc.). At this moment, the agent triggers the search for a tutor by sending a message to the recommender agent. When the most suitable tutors logged in are identified, the virtual character is the one to suggest to the student needing assistance that other class fellows may help him with the given task. By taking into account different aspect of the affective and the social states of the students, we classify our virtual character as an affective subsystem.

3. Validation and Discussion

This paper presented a social-affective virtual character that interacts with the users in order to motivate group formation among students and promote collaborative learning. We used mining algorithms to identify suitable students that can play the role of a tutor, and to recommend them to other students needing assistance. Our tutor recommendation mechanism explores the social-affective dimension through the analysis of emotional states and social behaviour of the users.

An Environment for the Learning of Algorithms (3A) has been developed at the Department of Computer Science of the University of Caxias do Sul in conformity with our architecture. The main goal of the system has been to make the courses more dynamic, increase the interest and participation of the students and provide an environment where they may interact in order to improve their knowledge. The environment presents students with the regular contents of algorithms, it proposes exercises, provides a forum for discussion and a tool for the testing and running of algorithms. Having been developed as a dynamic website, the system enables teachers and administrators to modify contents easily. The system

promotes the communication among students through the use of a virtual character that has the ability to find suitable student tutors and suggest them to individuals showing difficulty in learning a given topic. Figure 2 shows the interface of the system.

Figure 2. System Interface

The menu on the left shows the main sections of the learning environment (Contents, Exercises, Algorithm tool, Forum and Links). In the central area, the section selected in the menu is displayed. In the example, the item selected is "*Comando de Leitura*", or "*The Command Read*". On the right, we may see the virtual character, which provides a textbox to let the user enter questions and remarks. Below, there are other textboxes used to let students chat with each other. On the upper part of the screen, there is a panel with different resources allowing the users to insert information about their social-affective states and to choose school mates for interaction.

An experiment was carried out in order to identify the impact of using a virtual character in the user interface. We simply built different versions of the system, one of them without any character and the other ones mediated by a virtual character developed according to a different set of features (gender, age, cartoon or photographic style, animated or non-animated style). We noticed in the experiment a clear preference of the students for the interfaces containing the virtual characters, specially one which was "inhabited" by a known professor. This evidence can reinforce the importance of embodying social-affective features in educational systems.

The 3A environment has started to be used in 2 courses at the Department of Computer Science. Descriptors were built manually in order to get the system to recommend contents and tutors. The data collected so far has not been sufficient for us to carry out conclusive experiments as to whether the system is making tutoring recommendations appropriately. However, the results obtained so far show that the use of Social Profile, Mood State, Performance Acceptance, Sociability and Tutorial Degree in tutor recommendation, is a promising alternative. Other experiments carried out and reported in [25] also show that the item descriptors approach has a good performance in terms of processing time and accuracy, when compared with collaborative filtering, one of the most popular approaches in recommender systems.

One of our biggest challenges now concerns the automatic inference of students' affective states. For the moment, we are using questionnaires and graphic interface controls to let the user indicate such states. Thus, little is done to automatically infer the social-affective information necessary for tutor recommendation. This will be one of our main research efforts in the near future.

References

[1] Pandzic, I.S. Life on the Web. Software Focus Journal. Wiley, NY, 2001, 52-58.

[2] Rist, T., André, E., Baldes, S. A flexible platform for building applications with life-like characters. 2003 International Conference on Intelligent User Interfaces. Janeiro 12-15, 2003, Miami, FL, USA.

[3] Rist, T., André, E., Müller. J. 1997. Adding Animated. Presentation Agents to the Interface. Proc. of IUI 97, 79-86, N.Y.: ACM Press.

[4] Weizenbaum, J. Contextual understanding by computers. Communications of the ACM, vol 10, num. 8, agosto 1967.

[5] Kaasinen, E. Usability Issues in Agent Applications:What Should the Designer be Aware of. USINACTS 01.06.98, VTT Information Technology, Finland, 1998.

[6] Blom, J. & Monk, A. One-to-one e-commerce. who's the one? Proceedings of CHI 2001: Conference on Human Factors in Computing Systems. Seattle, Washington, 31 March - 5 April 2001.

[7] Rickenberg, R. and Reeves, B. The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces. CHI 2000 The Hague, Amsterdam, The Netherlands. 1-6 APRIL 2000.

[8] De Angeli, A.; Lynch, P. and Johnson, G. Personifying the e-market: A framework for social agents. To be presented at Interact 2001 - Eighth IFIP TC.13 Conference on Human-Computer Interaction, Tokyo, Japan, July 9-13, 2001.

[9] Shaw, E. and Johnson, W.L. Pedagogical agents on the Web. In Proceedings of the the Third International Conference on Autonomous Agents, ACM Press, NY, 1999, 283-290.

[10] Scotty D. Craig, Barry Gholson, and David M. Driscoll. Animated Pedagogical Agents in Multimedia Educational Environments. Journal of Educational Psychology, 2002, Vol. 94, No. 2, 428-434.

[11] C.D. Elliott, J. Rickel, and J. Lester. Lifelike Pedagogical Agents and Affective Computing: An Exploratory Synthesis. In M. Wooldridge and M. Veloso, editors, Artificial Intelligence Today, number 1600 in Lecture Notes in Computer Science, pages 195--212. Springer-Verlag, 1999.

[12] J. Greer, G. McCalla, J. Collins, V. Kumar, P. Meagher, J. Vassileva (1998) Supporting Peer Help and Collaboration in Distributed Workplace Environments, International Journal of AI and Education. 9, 159-177, available on the WWW at: http://cbl.leeds.ac.uk/ijaied/.

[13] Andrade, A. F. (2003) "Uma Aplicação da Abordagem Sociointeracionista de Vygotsky para a Construção de um Ambiente Computacional de Aprendizagem". Tese de Doutorado. PGIE-UFRGS. Brasil.

[14] Scherer, K. R. (2000) "Psychological models of emotion." In J. Borod (Ed.). The neuropsychology of emotion. Oxford/New York: Oxford University Press. pp. 137-162.

[15] Dipert, R. (1998) The Nature and Structure of Emotions. DRAFT. U. S. Military Academy, http://www.neologic.net/rd/

[16] Ortony, A.; Clore, G.L.; Collins, A. (1988) The cognitive structure of emotions. Cambridge University Press, Cambridge, UK.

[17] Maturana, H. (1998) Emoções e Linguagem na Educação e na Política. Belo Horizonte: Ed. UFMG.

[18] Yang, H.; Tang, J. (2003) "Effects of Social Network on Students' Performance: A Web-Based Forum Study in Taiwan". Journal of Asynchronous Learning Networks (JALN). Sloan-C, Volume 7, Issue 3, September.

[19] Sparrowe, R. T.; Liden, R. C.; Kraimer, M. L. (2001) "Social networks and the performance of individuals and groups". Academy of Management Journal, 44(2), p.316-325.

[20] Silveira, R. A.; Vicari, R. M. (2002) "Developing Distributed Intelligent Learning Environment with JADE - Java Agents for distance Education Framework.". ITS 2002. Lecture Notes in Computer Science, Biarritz, v. 2363, p. 105-118.

[21] Krulwich, B. 1997. LIFESTYLE FINDER: Intelligent User Profiling Using Large-Scale Demographic Data, Artificial Intelligence Magazine 18(2). 37-45.

[22] Claypool, M.; Brown, D.; Le, P.; and Waseda, M. 2001. Inferring User Interest. IEEE Internet Computing, 5(6):32-39.

[23] Pradhan, M., Provan, G. M., Middleton, B., and Henrion, M. 1994. Knowledge engineering for large belief networks. In Proceedings of Uncertainty in Artificial Intelligence, Seattle, Washington. Morgan Kaufmann.

[24] Wallace, R. (2003). "The Elements of AIML Style <http://www.alicebot.org/documentation>", ALICE A. I. Foundation.

[25] Reategui, E.; Campbell, J. A. and Torres, R. 2004. AAAI - The Nineteenth National Conference on Artificial Intelligence, Workshop on Semantic Web Personalization, San Jose, USA, July.

Empathetic Virtual Peers Enhanced Learner Interest and Self-Efficacy

Yanghee Kim
Department of Instructional Technology, Utah State University,
2830 Old Main Hill, Logan, UT 84322, USA
yanghee.kim@usu.edu

Abstract. This study investigated the effects of empathetic response (responsive vs. non-responsive) and gender (male vs. female) of pedagogical agents as learning companions (PALs) on learner interest, self-efficacy, social judgments of PAL persona, and learning. Fifty-six pre-service teachers were randomly assigned to one of four conditions. The results revealed that students who worked with a PAL who responded with empathy to learners' affect showed significantly higher interest (p < .05) and self-efficacy in the tasks (p < .05) than students with a non-responsive PAL. Also, students rated the persona of a male PAL more favorably than that of a female PAL (p < .05). The findings imply that PAL/learner relationships in computer-based learning might be consistent with human relationships in traditional classrooms.

Keywords. Pedagogical agents, Learning companions, Virtual peers, Motivation

Introduction

Social interaction among participants in learning contexts is seen as the primary source of cognitive and social development of learners [1]. This emphasis on social interaction for learning and cognition might suggest reframing the conventional design of educational software and adopting a new metaphor: computers as interacting partners. Along this line, pedagogical agents can facilitate social interaction in computer-based learning. Pedagogical agents are in general defined as animated life-like characters [2]. Social presence and social interaction may make pedagogical-agent-based environments unique from the conventional courseware. In particular, pedagogical agents as learning companions (PALs) are a specific use of pedagogical agents. PALs as virtual peers may be able to simulate peer interaction in traditional classroom and help learners attain the cognitive and affective gains in human peer-mediated learning.

To build social relations with learners, desirable pedagogical agents should be humanlike and have personas [3]. Similarly, to facilitate social interaction, PALs might need to be perceived as believable virtual peers. To make them believable, a crucial feature may be the PALs' capabilities to demonstrate affect [4].

Affect is an essential part of social cognition, allowing us to function successfully in daily social and intellectual life. It is well documented that affect influences our memory, decision-making, and information processing. In classrooms, the affective states of teachers and peers function as social contexts to influence learners' self-efficacy, motivation, cognitive gains, and behaviors [5, 6]. Also, gender difference manifests in affective expressions, empathetic accuracy, and affective behavior [7]. This affect/gender interaction becomes more salient in peer relationships than in learner/adult relationships. This impact of affective interaction among participants in traditional classroom may be applied to

PAL/learner relations [8]. Given that human/computer interaction is consistent with human-to-human interaction, it is plausible that learners may build empathetic relationship and emotionally interact with PALs in computer-based learning and that the empathetic interactions might be differentiated by PAL gender.

Very few studies have been done in this regard. Although some research on affective computing was conducted in gaming environments [9, 10], it is not clear that the findings from those studies might be generalized to learning environments consistently. Thus, this study was aimed at investigating whether gender and empathetic response of PALs would influence learners' affective and cognitive characteristics similarly to human relationships in conventional classrooms. A controlled experiment was designed to examine the impact of the empathetic response and gender of PALs on learner interest, self-efficacy, social judgments, and learning.

Method

1. Participants

Participants were 56 pre-service teachers enrolled in an introductory educational technology class in a large public university. The intervention was implemented as a mendatory course activity. About 80% of the participants were female; 20% were male. The average age of participants was 20.71 (SD = 2.92).

2. Materials

2.1. Instructional Intervention

The learning task was instructional planning, processed in five main stages--Introduction, Case Study, Blueprints, Plan, and Assessment--in each step of which PALs provided learners with context-specific information and suggestions. The stages were indicated by large buttons located at the top of the screens. Introduction briefly explained the learners' task. Case study described a scenario to teach Anna, a sixth grader, the economic concepts of supply and demand. In Blueprints, the participants wrote instructional goals or objectives in a text-box field. As additional information, two links were provided regarding Texas Standards and Benchmarks for supply and demand. In Plan, the participants wrote instructional strategies and activities. In Assessment, the participants described the assessment plans to test Anna's learning.

2.2. PAL Design

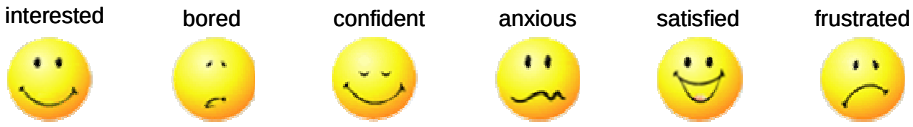
The PALs were developed using Poser 5, a 3D-image/animation-design tool, and Mimic Pro 2, a voice/affect-editing tool. The animation files created in Poser 5 were converted to Macromedia Flash movies for compression and were later integrated into the instructional intervention. The PALs, Chris, (one version male, one female) were designed to look about twenty years old, casually dressed and speaking informally, sometimes using slang. Male and female voices of college students were recorded. The participants estimated the PALs' age as an average of 20.39 (*SD* = 7.94).

2.3. Learners' Expressions of their Emotions

In between the stages, learners expressed their affective states at the moment by clicking an emoticon (i.e., icons expressing emotions). Six emotional states that typically occur in learning situations were drived from classroom emotion/motivation research and Affective Model suggested in Affective Computing Group [11]: interest, boredom, confidence, anxiety, satisfaction, and frustration. A panel of six emoticons appeared when the learners

attempted to move to the next stage. When the learners expressed their affect, the PAL responded to it or not, according to experimental conditions. Figure 1 shows the six emoticons.

Figure 1. Emoticons



3. Independent Variables

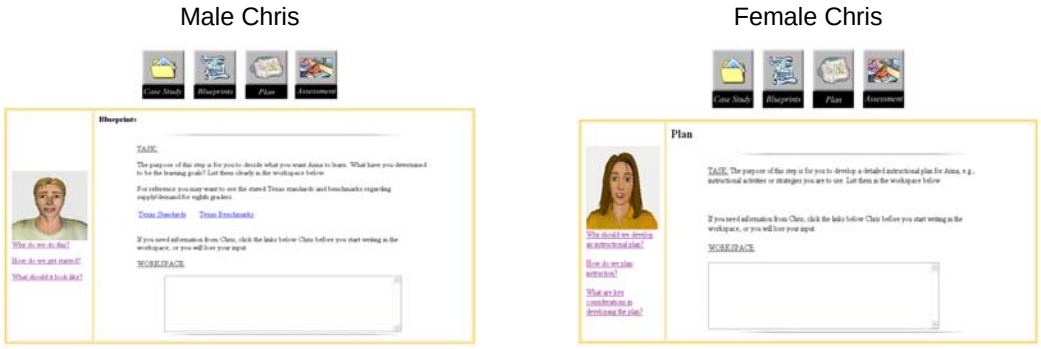
3.1. Empathetic Response

Empathetic response, i.e., whether the PAL responded with empathy to the learners' affect, was categorized as responsive or non-responsive. In the responsive condition, when a learner expressed his/her affect by clicking any of the emoticons, the PAL immediately responded to the learner's affect by verbal expressions. The affective responses were brief and did not affect the overall instruction time. In the Non-responsive condition, the PAL did not respond when the learner expressed affect. The module simply led the learner to the next phase. In both conditions, the amount of information provided by the PAL was identical.

3.2. Gender

Either a male or female PAL, each named Chris, was included depending on experimental conditions. Figure 2 presents the male and female PALs in the intervention.

Figure 2. The male and female PALs



4. Dependent Variables

4.1. Learning

The author was interested in examining the learners' engagement in interactions with the PAL, speculating that if learners were more engaged, they would recall more of the ideas presented by the PAL. Thus, learning was measured by an open-ended posttest question. Students were asked to write all the ideas and information that the PAL had conveyed about instructional planning. The number of legitimate ideas in the students' answers was counted and coded by two instructional designers. Inter-rater reliability was evaluated as Cohen's Kappa = .94.

4.2. Interest

Interest referred to learners' disposition toward working with the PALs and toward the task of instructional planning. Based on Anderson and Bourke's guidelines [12], a questionnaire with seven items was scaled from 1 (Strongly disagree) to 5 (Strongly agree): 1) How much are you interested in designing a lesson plan for E-learning? 2) How much are you interested in learning about designing a lesson plan for E-learning? 3) I was interested while doing the task, 4) I was attentive while doing the task, 5) I was absorbed while doing the task, 6) I was interested while working with Chris, and 7) I was attentive while working with Chris. Item reliability was assessed as coefficient α = .87. Learner interest was measured before and after the intervention to control for prior interest.

4.3. Self-Efficacy

Self-efficacy referred to learners' belief about their competency in the task of instructional design. Based on the guidelines from previous research [13], a questionnaire with five items was scaled from 1 (Strongly disagree) to 5 (Strongly agree): 1) How well can you write a lesson plan? 2) How sure are you that you can design a good lesson plan? 3) I can write a lesson planning on a topic of my subject, 4) I am confident in designing a lesson plan, and 5) I am competent in designing a lesson plan. Item reliability was evaluated as coefficient α = .95. Learners' self-efficacy beliefs were measured before and after the intervention to control for prior self-efficacy.

4.4. Social judgments

Social judgments refered to learners' judgments about the attributes of PALs as their learning partners [14]. Learners' social judgments were measured by a questionnaire consisting of three sub-measures: Facilitating learning, Human-like, and Engaging. Facilitating learning included 4 items: 1) Chris led me to think more deeply about the presentation. 2) Chris encouraged me to reflect what I was learning. 3) Chris kept my attention, and 4) Chris presented the material effectively. Item reliability was evaluated as coefficient α = .91. Human-like included 3 items: 1) Chris has a personality, 2) Chris' emotion was natural, and 3) Chris was human-like. Item reliability in each category was evaluated as coefficient α = .73. Engaging included 4 items: 1) Chris was expressive, 2) Chris was enthusiastic, 3) Chris was motivating, and 4) Chris was friendly. Item reliability was evaluated as coefficient α = .81. All items were scaled from 1 (*Strongly disagree*) to 5 (*Strongly agree*).

5. Procedure

The experiment was integrated as a class activity during a regular session. Participants were randomly assigned to one of four conditions by system programing. The participants logged on to the web-based instructional module. First, they entered demographic information. They then answered pretest questions, taking as much time as they needed. After that, they answered posttest questions. The session took approximately one and half hours, with individual variations.

6. Design and Analysis

The study employed a 2 × 2 factorial design. For learning, two-way ANOVA was conducted. For interest and self-efficacy, MANCOVA was conducted, with pretest interest and self-efficacy as covariates. For social judgments with three sub-measures, MANOVA was conducted. The significant level was set at $\alpha < .05$.

Results

1. Learning

There were no significant main and interaction effects for empathetic response and gender on learning.

2. Interest

The overall MANCOVA revealed a significant main effect for PALs' empathetic response: Wilks' Lambda = .53, F (5, 20) = 3.54, $p < .05$, partial $\eta^2 = .47$. Students who worked with the responsive PAL showed significantly higher interest toward the task and the PALs than students who worked with the non-responsive PAL.

3. Self-Efficacy

The overall MANCOVA revealed a significant main effect for PALs' empathetic response: Wilks' Lambda = .71, F (3, 31) = 4.29, $p < .01$, partial $\eta^2 = .29$. Students who worked with the responsive PAL showed significantly higher self-efficacy than students who worked with the non-responsive PAL.

4. Social Judgments

The overall MANOVA revealed a significant main effect for PALs' gender: Wilks' Lambda = .85, F (3, 46) = 3.08, $p < .05$, partial $\eta^2 = .15$. To identify the contribution of each sub-measure to the overall effects, univariate analyses (ANOVA) were conducted.

The results revealed a significant main effect for PAL gender on "facilitating learning": F (1, 48) = 3.8, $p < .05$. both male and female students who worked with the male PAL (M = 3.56, SD = .64) perceived the PAL as significantly more facilitating to their learning than students who worked with the female PAL (M = 3.14, SD = .82). The standardized effect size for this difference was Cohen's $d = 0.57$, which indicated a medium effect according to Cohen's guidelines.

The univariate results revealed a significant main effect for PAL gender on "human-like": F (1, 48) = 6.95, $p < .05$. Students who worked with the male PAL (M = 3.59, SD = .52) perceived the PAL as significantly more human-like than students who worked with the female PAL (M = 3.14, SD = .69). The standardized effect size for this difference was Cohen's $d = 0.74$, which indicated a medium-large effect.

The results revealed a significant main effect for PAL gender on "engaging", F (1, 48) = 4.11, $p < .05$. Students who worked with the male PAL (M = 3.79, SD = .52) perceived the PAL as significantly more engaging than students who worked with the female PAL (M = 3.51, SD = .43). The standardized effect size for this difference was Cohen's $d = 0.59$, which indicated a medium effect. There was no significant effect for affective response on agent persona.

Discussion

This study examined whether the empathetic response and the gender of a PAL influenced learners' affect and cognition. The results supported the positive impact of the PAL's

affective response on both interest and self-efficacy. Students were more interested in the task and the PALs and showed higher self-efficacy beliefs in the task when the PAL was empathetic to their affect than they did when the PAL was not empathetic. These results were parallel with the findings of human emotion research in classrooms. When students understood that their teachers cared about them, the students' motivation and self-concept were improved [15]. In the same fashion, when the PALs expressed care about the learners' affective states by verbally responding with empathy, the learners' interest and self-efficacy were enhanced. These affective gains from affective response were demonstrated in a gaming environment as well. Klein and colleagues [10] reported that users in a computer game stayed longer in the game when they received affective support by the computer. This positive impact of affective response may provide an implication for the design of PAL affect for learner motivation. To be effective, PAL affect should be tied to the learner's affect when possible. Rather than being simply a "happy" talking head, a PAL should respond to or deal with the learner's affect and flexibly adapt its affect to the learner's in order to motivate the learner.

However, the presence of PAL affective response did not influence learning. There was no statistically significant difference on learner recall between responsive and non-responsitve PALs. This result confirms the current knowledge of affective pedagogical agents research, which has supported the impact of agents on motivation but not on learning [16, 17]. It should be further investigated why enhanced motivation fail to bring enhanced learning. The one-time implementation might be a reason. Many affective computing studies are typically conducted on a short-term basis, unlike the classroom-based research that indicated a significant impact of affective interaction on both learners' affective and congitive outcomes. Social relations seems to be built over time; fostering virtual relations may require even more time. Long-term research may expand our understanding. Also, learning in the current study was measured simply by recall. However, the benefits of social empathetic interaction might be better exercised in learning transfer [18] or higher-order learning, such as reflective thinking and making a solid argument.

At any rates, the inconsistent impact of PAL affect on motivation and learning might suggest judicious implementation of agent emotions. Typically, in instructional settings, there are different goals and emphases, which might focus more or less on cognitive skill acquisitions or on affective gains. Given the findings of the current and previous studies, PAL affect can be effectively utilized in the context emphasizing motivational and attitudinal changes, but not necessarily in the context where knowledge and skill acquisition is the sole issue.

Regarding PALs' gender, the results support the positive impact of a male PAL on learners' social judgments of PAL persona. Both male and female students perceived the persona of the male PAL more favorably than that of the female PAL. This result indicated that PAL/learner relations might reflect the stereotypic expectations in the real world. Men are generally more influential than women [19], perceived to possess a higher level of expert and legitimate power [20]. In a similar fashion, the male PAL was perceived as more facilitating to learning, more human-like, and more engaging than the female PAL even though they provided identical informational and emotional support. Future research should examine how to overcome this infusion of real-world stereotypic expectations to computing environments.

Lastly, there were some limitations in the study. The measures of interest and self-efficacy relied solely on self-report. Brody [7] argued that affective experience is a feeling state known only to the individual; thus, the best way to really measure experience might

be to ask people to think about their feeling states. However, Brody also pointed out that people may not want to articulate socially unacceptable feelings to others. In order to gather more accurate information on learners' affective states, we may employ multiple methods to complement this study, including advanced technology to sense learners' emotional states [21]. Second, Given that voice is a significant indicator of social presence [22, 23], the PAL's interactions with learners were all scripted in order to include human voices, thus not allowing adaptive response. Further elaborated studies including PAL intelligence in emotional interactions are invited. Lastly, the study was conducted through one-time implementation. It is open to question whether the impact of PALs on interest and self-efficacy would sustain in the long term as well. Future research may be complement the limitations of the current study to confirm the findings.

References

[1] V. John-Steiner and H. Mahn, "Sociocultural contexts for teaching and learning," in *Handbook of psychology: Educational psychology*, vol. 7, A. Reynolds, M. William, and G. E. Miller, Eds. New York: John Wiley and Sons, 2003, pp. 125-151.

[2] W. L. Johnson, J. W. Rickel, and J. C. Lester, "Animated pedagogical agents: face-to-face interaction in interactive learning environments," *International Journal of Artificial Intelligence in Education*, vol. 11, pp. 47-78, 2000.

[3] A. L. Baylor and Y. Kim, "Simulating instructional roles through pedagogical agents," *International Journal of Artificial Intelligence in Education*, vol. 15, 2005.

[4] J. Bates, "The nature of characters in interactive worlds and the Oz project," School of Computer Science, Carnegie Mellon University, Pittsburgh, PA CMU-CS-92-200, 1992.

[5] R. E. Sutton and K. F. Wheatley, "Teachers' emotions and teaching: a review of the literature and directions for future research," *Educational Psychology Review*, vol. 15, pp. 327-358, 2003.

[6] B. Weiner, "Intrapersonal and interpersonal theories of motivation from an attributional perspective," *Educational Psychology Review*, vol. 12, pp. 1-14, 2000.

[7] L. Brody, *Gender, emotion, and the family*. Messachusetts: Harvard University Press, 1999.

[8] Y. Kim, "Learners' expectations on the desirable characteristics of learning companions," presented at the Annual Conference of American Educational Research Association, San Diego, CA, 2004.

[9] J. Scheirer, R. Fernandez, J. Klein, and R. W. Picard, "Frustrating the user on purpose: a step toward building an affective computer," *Interacting with Computers*, vol. 14, pp. 93-118, 2002.

[10] J. Klein, Y. Moon, and R. W. Picard, "This computer responds to user frustration; Theory, design, and results," *Interacting with Computers*, vol. 14, pp. 119-140, 2002.

[11] B. Kort, R. Reilly, and R. W. Picard, "An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion," presented at IEEE International Conference on Advanced Learning Technologies, 2001.

[12] L. W. Anderson and S. F. Bourke, "Assessing affective characteristics in the schools," 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

[13] A. Bandura and D. H. Schunk, "Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation," *Journal of Personality and Social Psychology*, vol. 41, pp. 586-598., 1981.

[14] G. L. Clore, "Affective influences on social information processing," in *Handbook of Affect and Social Cognition*, J. P. Forgas, Ed. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2001.
[15] C. A. Wong and S. M. Dornbusch, "Adolescent engagement in school and problem behaviors: The role of perceived teacher caring," presented at Annual Meeting of the American Educational Research Association, New Orleans, LA, 2000.
[16] J. C. Lester, S. G. Towns, C. B. Callaway, J. L. Voerman, and P. J. FitzGerald, "Deictic amd Emotive Communication in Animated Pedagogical Agents," in *Embodied Conversational Agents*, J. Cassell, S. Prevost, and J. Sullivant, Eds. Boston: MIT Press, 2000.
[17] S. G. Towns, P. J. FitsGerald, and J. C. Lester, "Visual emotive communication in lifelike pedagogical agents," presented at The 4th International Conference, ITS98. Lecture Notes Computercience San Antonio, Texas, San Antonio, Texas, 1998.
[18] Y. Kim and A. L. Baylor, "Pedagogical agents as learning companions: The role of competency and type of interaction," *Educational Technology Research & Development*, in press pending revision.
[19] L. L. Carli, "Gender and social influence," *Journal of Social Issues*, vol. 57, pp. 725-741, 2001.
[20] L. L. Carli, "Gender, interpersonal power, and social influence," *Journal of Social Issues*, vol. 55, pp. 81-99, 1999.
[21] W. Burleson, R. Picard, K. Perlin, and J. Lippincott, "A platform for affective agent research," presented at Workshop on Empathetic Agents, International Conference on Autonomous Agents and Multiagent Systems, Columbia University, New York, NY, 2004.
[22] C. Nass and J. Steuer, "Computers, voices, and sources of messages: computers are social actors," *Human Communication Research*, vol. 19, pp. 504-527, 1993.
[23] R. E. Mayer, K. Sobko, S. Fennell, and T. Mautone, "Social Cues in Multimedia Learning: Role of Speaker's Voice and Conversational Style," presented at AERA, 2002.

Acknowledgements

This work was sponsored by National Science Foundation Grant # IIS-0218692 and the Pedagogical Agent Learning Systems lab at Florida State University. The author thanks Dr. Amy L. Baylor for the support.

Intrinsic Fantasy: Motivation and Affect in Educational Games Made by Children.

M. P. Jacob HABGOOD, Shaaron. E. AINSWORTH, Steve. BENFORD
The Learning Sciences Research Institute, The University of Nottingham
Jubilee Campus, Wollaton Road, Nottingham, United Kingdom

Abstract: The concept of intrinsic fantasy has been considered central to the aim of usefully applying the positive affect of computer games to learning. Games with intrinsic fantasy are defined as having "an integral and continuing relationship with the instructional content being presented", and are claimed as "more interesting and more educational" than extrinsic fantasy games [1]. Studies of children making educational games have shown they usually create extrinsic games for curriculum learning content. In this study, children were encouraged to create non-curriculum games, more easily distanced from the extrinsic preconceptions of formal schooling. Forty, 7-11 year olds took part in this study (17 boys and 23 girls), designing and making their own games at an after-school club. Despite non-curriculum learning content, no more intrinsic games were created than in previous studies. The children failed to create their own pedagogical models for non-curriculum content and did not see the educational value of intrinsic fantasy games. The implications for transfer and learning in intrinsic games are discussed whilst the definition of intrinsic fantasy itself is questioned. It is argued that the integral relationship of fantasy is unlikely to be the most critical means of improving the educational effectiveness of digital games.

Pioneering work by Malone and Lepper in the 80s used computer games as a platform for studying intrinsic motivations for learning. This work produced a taxonomy identifying four categories of individual motivations responsible for the positive affect created by computer games: challenge, fantasy, curiosity and control [1-7]. The theory behind the taxonomy suggests that the motivational effect of a challenge depends on engaging a player's self-esteem using personally meaningful goals with uncertain outcomes. Uncertainty can be achieved through variable difficulty levels, multiple level goals, hidden information and randomness. It also proposes that it is the emotional appeal of fantasy and the sensory and cognitive components of curiosity that provide their motivational effect in digital games. It further suggests that cognitive curiosity is aroused when learners discover that their knowledge is incomplete, inconsistent, or unparsimonious. The motivational effect of control is attributed to empowerment and self-determination, suggesting that it is affected by the range of choices offered by a game, the extent to which outcomes are dependant on the responses of the player, and the inherent power of these responses. In addition it proposes that the perceived level of control is more important to motivation than the actual level of control.

Whilst these theories did not focus on what makes games educational, they made a significant theoretical claim by attributing educational benefits produced by the positive affect of fantasy to the distinction between *intrinsic* and *extrinsic* fantasy. Based on a number of empirical studies it was proposed that, "in general, [intrinsic] fantasies are both more interesting and more educational than [extrinsic] fantasies" [7]. An educational game with intrinsic fantasy is defined as one in which "the skill being learned and the fantasy depend on each other" and "there is an integral and continuing relationship between the

fantasy context and the instructional content being presented" [7, p. 240]. Conversely an educational game with extrinsic fantasy is defined as "one in which the fantasy depends on the skill being learned but not vice versa". To illustrate this we could contrast the extrinsic fantasy of 'hanging a man' in a game of computer hangman, with the intrinsic fantasy of 'darts and dartboard' in a game of computer darts. The fantasy context of hangman is usually applied to a spelling exercise, guessing missing letters in a word, but could just as easily be applied to a mathematics exercise, guessing missing numbers in a sum. However in the fantasy context of darts, using mathematics is intrinsic to the strategy of reducing a score to zero without a remainder and cannot be removed without fundamentally changing the game.

Two previous studies of educational games made by children found that children rarely attempt to create educational computer games with intrinsic fantasies [8, 9]. This would seem to suggest that children do not naturally attempt to harness the motivational potential of computer games in the most effective way. The explanations offered for this included the extra difficulty in creating intrinsic games, a design tension between game and educational content, and the prevalence of extrinsic question and answer models in 'edutainment' and teaching practice in schools. However, these studies concentrated mainly on learning content from the traditional school curriculum such as fractions. In the current study children were encouraged to create computer games that taught non-curriculum learning content that might be distanced from the extrinsic teaching methods used in school. They were also provided with concrete examples of games containing both intrinsic and extrinsic fantasy as models for completing the exercise. In this way we hoped to gain a deeper understanding of children's attitudes and preferences for intrinsic and extrinsic fantasy and how the positive affect of computer games might be more effectively applied in creating educational digital games.

1. Method

The study was run within the context of a weekly after-school computer club at an English primary (ages 7 to 11) school. The club was advertised to the children as a chance to learn how to make computer games and spaces were allocated on a first come first served basis. Due to the overwhelming interest, the club was run twice a week with two separate groups of different age groups. The younger group was made up of 11 boys and 9 girls and the older had 6 boys and 14 girls. The children's ages ranged from 7 years 7 months to 11 years 4 months old. At the club each child had access to their own computer in the school I.T. suite and as far as possible both groups were exposed to the same instruction and direction over the course of the study.

The clubs began by getting the children to discuss their favorite computer games and what made them motivating to play. This gave an initial insight into the children's mindset and their answers mapped well on to the taxonomy of intrinsic motivations for learning. This was followed by eight weeks of tuition in Stagecast Creator, the software package used to create the games. Stagecast is a commercially available package that provides a visual programming environment specifically designed for children [10, 11]. Programming is achieved by creating 'before' and 'after' pictures that define how situations are changed by each programming rule. This visual method of programming is very easy to pick up and accessible to children as young as seven years old. A number of lessons were prepared in collaboration with a class teacher at the school, and these were delivered in whole class sessions by the principal researcher. These sessions covered the essential skills for making simple games as well as introducing them to different examples of games that can be created with the software.

Stagecast was also used to create the intrinsic and extrinsic example games for the study. In keeping with the context of a computer club, the learning content of the games was about the functionality of computer components. This choice also had the advantage of being obscure enough to prevent the children from directly plagiarizing the game ideas. Both examples covered exactly the same learning content, but contrasted intrinsic and extrinsic fantasy solutions.

In the intrinsic example players have to control a robot whose abilities are directly dependent on the functionality of the components he collects (see figure 1). In order to escape an alien base he must decode the codebooks for a number of security doors. Upon reaching the first codebook, the robot is told that he needs to do 'hard sums' in order to break the codes. Using the on screen information the player can work out that the robot needs to collect the missing processor first. The codebook for the second security door is some distance away from the door and the robot forgets the code as soon as he moves away from the book. The player can then work out that the robot needs to collect the memory first. In this way the learning content is firmly embedded within its application in a fantasy context, creating the 'integral and continuing relationship' of an intrinsic fantasy.

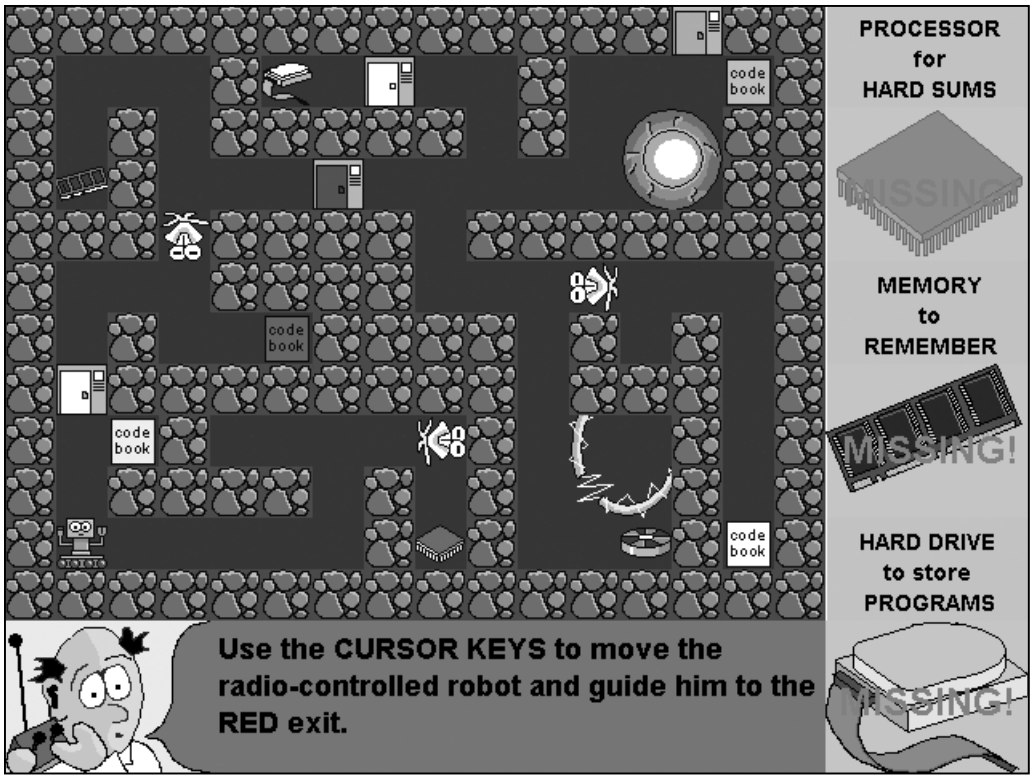


Figure 1: "Robot Rescue" the intrinsic example game.

In the extrinsic example the player's had to retrieve the same computer components by correctly answering multiple-choice questions about their functionality (see figure 2). The game is set within a skateboarding fantasy context, but this is entirely extrinsic to the learning material being presented. The fantasy context could be used for any kind of learning content as 'the fantasy depends on the skill being learned but not visa versa'.

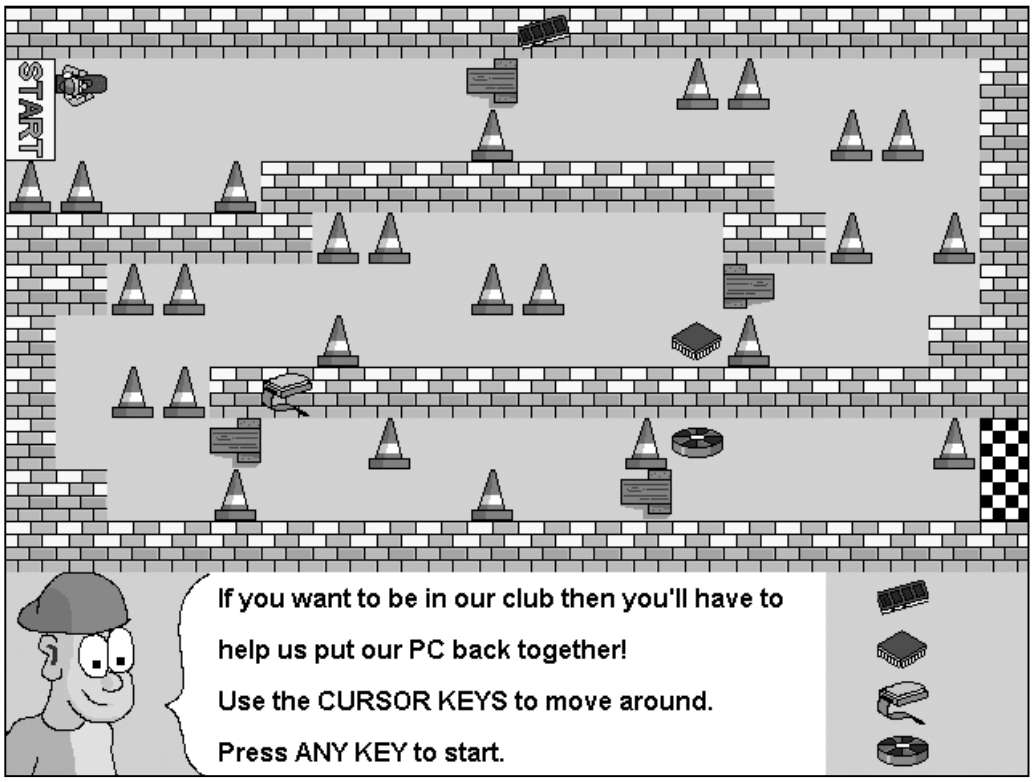


Figure 2: "Silly Cone Alley" the extrinsic example game.

At the end of the eight weeks of tutorials the children were presented with the two games as potential models for their own game designs, and given a chance to play and discuss them as a class. Afterwards, they were shown real computer components and asked to explain what they do in order to demonstrate that they had learnt something from the games. Next they were asked to design their own games to help players learn about a topic of their choice such as a hobby or interest they had outside of school. They were told that the learning in their games should be something that would be useful to the player in the real world after they had finished playing the game. They were provided with templates to structure the design process and taken through the procedure of completing them using the two games as examples. These included specific questions about the fantasy context of the game design (story and characters) as well as about what would make the game motivating (fun) to play. Two club sessions were then devoted to completing their designs, and stragglers were given time to finish off at lunchtimes. All of the children's designs were copied for later analysis.

For the remainder of the study the children worked independently on their games in the computer suite. The children were given periodic reminders that their games were supposed to help someone learn something, when it started to become apparent that many of them did not contain educational content. Towards the end of term they were given the opportunity to spend additional time working on their games in their lunch hour, giving them the potential of 10 hours total development time working on Stagecast by the end of the project. Copies of all the games were taken at the end of every session, providing a complete development history of the children's games. In the last week of term the children were interviewed in groups of 2 or 3 about their games and their learning content. These interviews were videoed and transcribed for analysis along with the class discussions about the example games.

2. Results

The children created a total of 34 game designs and 29 finished games. The mean number of hours spent developing the finished games was 7.0, with a minimum of 2.0 and a maximum of 10.0. Table 1 shows the number of games created in each year group along with the corresponding number of intrinsic, extrinsic and non-educational games. Games were considered to be 'educational' if they contained learning content that was identified by the game's author as relevant outside of a gaming context. This is because most games contained unintentional learning content (such as improving reaction times, or hand-eye co-ordination), which would be unrepresentative to count unless the child designer had included it deliberately.

Table 3: The number of games designed and created by children of each year group and the distribution of extrinsic, intrinsic and non-educational games.

Year Group	Games Designed/Created		Designs classified by researcher as:			
			Extrinsic	Intrinsic	Not Educational	Educational Misconceptions
7-8 Years	10	10	1	1	5	3
8-9 Years	10	8	0	0	5	5
9-10 Years	6	6	3	0	2	1
10-11 Years	8	5	5	0	3	0
Combined	34	29	9/34	1/34	15/34	9/34

26% of the games were extrinsic in design, nearly all following the same multiple-choice question and answer format used in the extrinsic model provided to them. Common themes were shopping adventures based around maths questions and horse riding games based on questions about riding tack. Only a single game had intrinsic learning content that was identified by its author as relevant outside of a gaming context. This game taught players not to touch dangerous electrical items, by making them switch off a number of electricity pylons in order to safely progress through a maze. 44% of the games and designs did not include instructional content at all, and were not considered to 'help player's learn anything' by their authors, despite most of them beginning with this intention. One example was a space adventure game that was supposed to teach players about the structure of space rockets. Unfortunately, despite producing a very polished game the author left out the learning content altogether. 26% of the games and designs were considered educational by their creators, but suffered from significant misconceptions in their choice and implementation of their learning goals. Examples include games that attempted to teach players how to drive cars, fly planes or swim by moving a character around the screen using the cursor keys.

3. Analysis: The Role of Intrinsic Fantasy in Educational Games

The children in this study put an enormous amount of energy and enthusiasm into the project and demonstrated that young children are capable of designing and making sophisticated computer games. Nonetheless, despite the use of an intrinsic model and a free choice of learning content, there was no increase in the number of intrinsic games over previous studies. However, a simple preference for familiar extrinsic learning models does not seem to fully explain their prevalence in the children's game designs. Many of the

children in this study found it extremely hard to create realistic learning models for non-curriculum content and the intrinsic model did not appear to help. Furthermore, the interview data shows that many children struggled to see the educational value of the intrinsic model or intrinsic reinterpretations of their own games. Our interpretation of these results is that using non-curriculum learning content did remove some of the extrinsic associations of schooling, but without an understanding or appreciation of the intrinsic model most children were left with no model at all. This then either resulted in children creating misconceived learning models of their own, or avoiding this difficulty by leaving out their intended learning content altogether.

This study has also raised questions about the definition of intrinsic fantasy, and the significance of fantasy as the key factor in creating a closer integration between a game and its learning content. The children's games have provided a diverse range of fantasy contexts with which to test the definition of intrinsic fantasy, and the results have often proved ambiguous. Three of the games in particular illustrate this point. In the first of these a quartz crystal, called Nile, must escape from a dungeon by collecting a number of rocks and minerals (see figure 3). The learning content is delivered by displaying the name of each mineral on the screen as Nile collects it. This elaborate fantasy is certainly 'depends on its subject matter' – even to the point of the main character being a mineral himself. However, the learning content is not needed to complete the game, and can be completely ignored, so it can hardly be intrinsic. The second game design takes place in a haunted school where players have to answer questions put to them by monstrous teachers. As the game is set in a school and the questions are all on school subjects so it could be argued that there is a 'continuing relationship between the fantasy context and the instructional content being presented'. Unlike the mineral game, the learning content is also integral to completing the game, as each question must be answered correctly to proceed. Yet this game uses the same multiple choice question format that was used in the extrinsic example game.

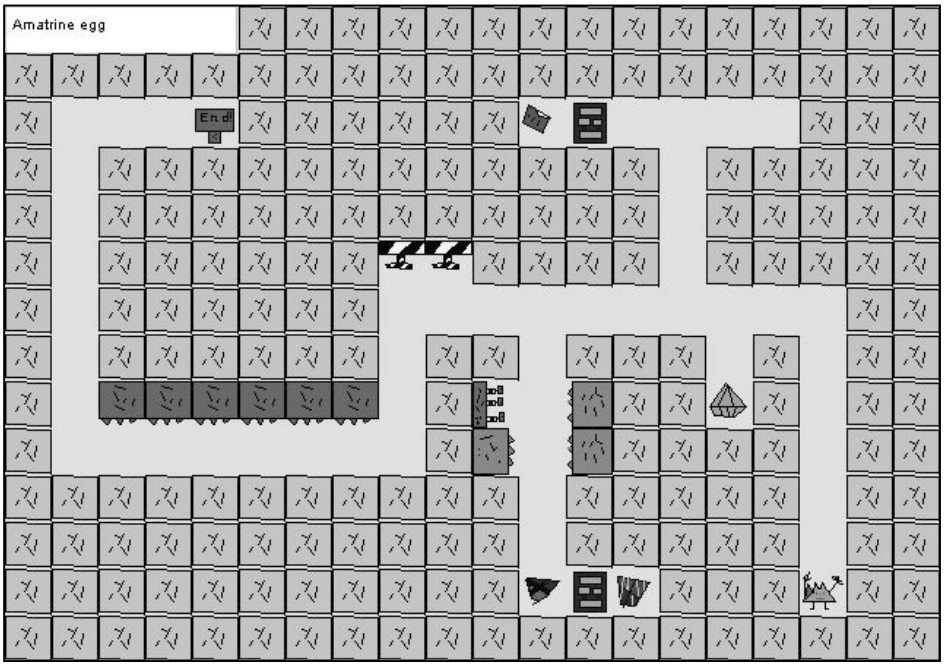


Figure 3: "Nile the Quartz Crystal" (boy aged 8 years 3 months).

The final game is about a collecting gold, whilst avoiding robots and live electricity pylons (see figure 4). Compared to the other two games the fantasy context is relatively thin, and its author freely admitted that there isn't really a story behind the game. In his interview the author was solely concerned with the rules and goals of his game rather than the fantasy context. However, he was the only child who was able to verbalise how he had embodied his learning material within the game and how it was relevant in the real world. He taught players not to touch dangerous electrical items by making them switch off electricity pylons before going near them in the game. As a result it was considered to be our only intrinsic fantasy game, but how much of a role can such a thin fantasy play in the integration of learning content? These examples, and others illustrate the ambiguity involved in classifying intrinsic fantasy games.

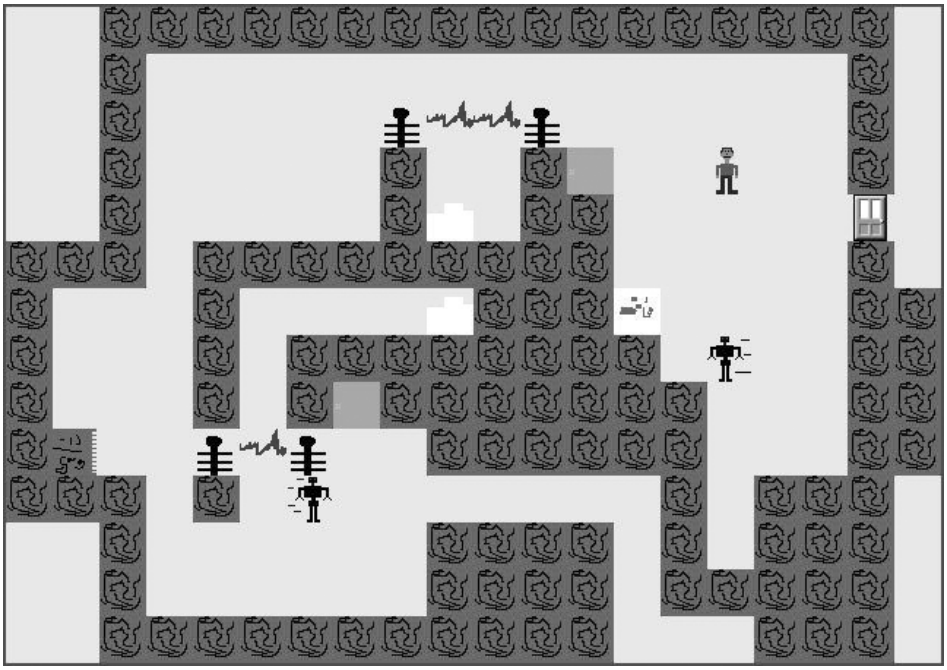


Figure 4: "Get Out" (boy aged 8 years 1 month).

Even our original darts example doesn't stand up to much scrutiny as an intrinsic fantasy game. If we were to swap the fantasy context of throwing darts for the fantasy context of firing elephants out of a cannon, how would this change the game? If there was an 'integral and continuing relationship' between the fantasy and the learning content then the game should no longer function. However, providing that the rules and the scoring mechanism remain the same then the game should remain just as integrated and just as educational. Consequently, we suggest that fantasy is not the main factor in creating effective integration and that integration of the learning content and the rule-system of a game is a more significant factor. Such rule systems are commonly referred to by game developers as the core mechanics of a game and are defined by Salen & Zimmerman as the "mechanism through which players make meaningful choices and arrive at a meaningful play experience" [12](p. 317). Core mechanics are the procedural mechanisms of a game that provide the essential interactions required to create a meaningful gaming activity. So the core mechanic of the classic game of Breakout is in controlling the horizontal position of one object in order to repeatedly intercept another moving object and keep it bouncing around a confined space. Whether the game uses the fantasy context of a bat and ball or (as in a later interpretation of the game) a space ship and energy bolt, it makes no difference to the fundamental gaming activity.

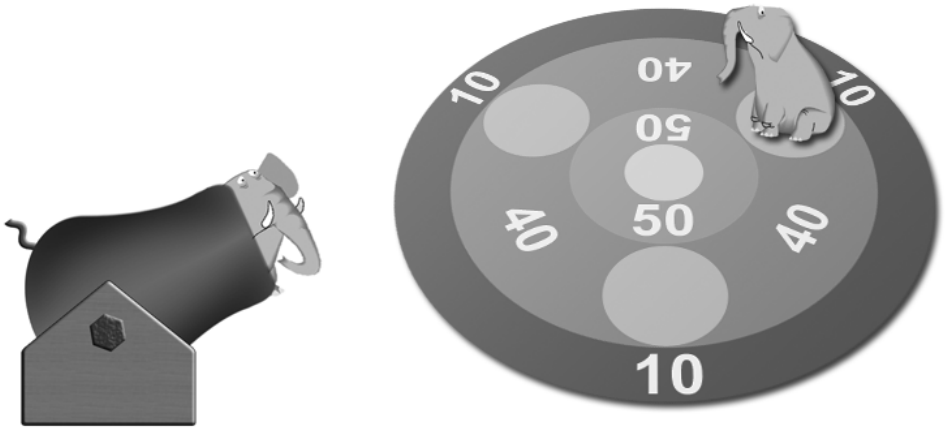


Figure 4: An alternative fantasy context for darts?

Our experiences in this study suggest that children have great difficulty in understanding the pedagogy behind intrinsic fantasy games and as a result often do not see them as educational at all. This raises important questions about how this viewpoint affects children's learning and particularly the transfer of learning content in intrinsic educational games. However, this study has also raised doubts about the "integral and continuing relationship" of fantasy as a critical means of improving the educational effectiveness of digital games. We suggest that intrinsic fantasy is a misnomer, which merely clouds our ability to distinguish the effect of fantasy elements within games from the more precise distinctions between them. Continuing to use the term intrinsic fantasy in this context would therefore only stand in the way of gaining a deeper understanding of how to successfully harness the positive affect of computer games for educational use.

4. References

[1] Malone, T.W., *Toward a Theory of Intrinsically Motivating Instruction.* Cognitive Science, 1981. **5**(4): p. 333-369.

[2] Malone, T.W. *What Makes Things Fun to Learn? Heuristics for Designing Instructional Computer Games*. in *Association for Computing Machinery Symposium on Small and Personal Computer Systems*. 1980. Pal Alto, California.

[3] Malone, T.W., *What Makes Computer Games Fun?* BYTE, 1981. **5**: p. 258-277.

[4] Malone, T.W., *Guidelines for Designing Educational Computer Programs.* Childhood Education, 1983. **59**(4): p. 241-247.

[5] Malone, T.W., *What Makes Computer Games Fun? Guidelines for Designing Educational Computer Programs*, in *Intelligent Schoolhouse*, D. Peterson, Editor. 1984, Reston Publishing Company: Resten, VA. p. 78-92.

[6] Lepper, M.R. and T.W. Malone, *Intrinsic Motivation and Instructional Effectiveness in Computer-Based Education*, in *Aptitude, Learning and Instruction: III. Conative and affective process analyses*, R.E. Snow and M.J. Farr, Editors. 1987, Erlbaum: Hilsdale, NJ. p. 255-286.

[7] Malone, T.W. and M.R. Lepper, *Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning*, in *Aptitude, Learning and Instruction: III. Conative and affective process analyses*, R.E. Snow and M.J. Farr, Editors. 1987, Erlbaum: Hilsdale, NJ. p. 223-253.

[8] Kafai, Y.B., *Minds in Play*. 1995, Hillsdale, NJ: Lawrence Erlbaum Associates.

[9] Reiber, L.P., N. Luke, and J. Smith, *Project KID DESIGNER: Constructivism at work through play.* Meridian: Middle School Computer Technology Journal, 1998. **1**(1).

[10] Gilmore, D.J., et al., *Learning graphical programming: An evaluation of KidSim.* 1996: Nottingham, UK.

[11] Smith, D.C., A. Cypher, and L. Tesler, *Novice Programming Comes of Age.* Communications of the ACM, 2000. **43**(3): p. 75-81.

[12] Salen, K. and E. Zimmerman, *Rules of Play*. 2004, Cambridge, MA: The MIT Press.

Does Aesthetics Affect the Users' Perceptions of VLEs?

Rosamelia PARIZOTTO-RIBEIRO & Nick HAMMOND
Department of Psychology – University of York,
YO10 5DD York, England

Abstract. This article presents the results of an empirical study about the role of aesthetics on screen layout design of Virtual Learning Environments (VLEs). There are many theories about what aesthetics is and how it can be achieved. However, the application of aesthetic aspects into interactive computer systems is just showing its first results (e.g. Automatic Teller Machines, web pages and VLEs). This study involved 98 participants who navigate on the environment to carry out two tasks using a simulation of a VLE interface. They evaluated the environment aesthetics, usability and hedonic characteristics. The results showed significant relationship between aesthetics and usability. The comparison between different screens layout suggests that perception of aesthetics are not directly related with the participants' background, schooling, computer literacy, gender or age.

1. Introduction

The use of Virtual Learning Environments to deliver online learning is becoming a well established practice used by Educational Institutions and private companies. Nowadays, computer programs such as VLEs have much higher usability than before. However, the concerns with the aesthetic aspects of its interface are at an early stage and there are many important issues to be addressed in order to motivate the learning process as well as users' perception of pleasure and affect.

Results from various studies suggest that a new approach for interface design, including VLEs, is no longer a matter of achieving beauty or fashion but giving motivation and satisfaction, affecting the user perception of the environment. A recent book 'Emotional Design – Why we love (or hate) everyday things' [1] gives a new perspective to products based on affective design, emphasizing their aesthetics and pleasure of use. The theory proposed presents three levels of emotional design: i) visceral design is concerned with appearance and it doesn't depend on cultural aspects, ii) behavioural design deals with pleasure and effectiveness of use and iii) reflective design considers the message, meaning of the product and it is culturally dependent.

This theory can also influence computer interface design, especially to VLEs, where the users may need additional motivation to be fully engaged on a distance learning course.

2. Computers and Aesthetics

The literature reports several studies related to learnability and aesthetics. It was found that aesthetically pleasing layouts have a definite effect on the student's motivation to learn [2] and that good graphic design and attractive displays contribute to the transfer of information [3]. In

other words, good design helps the user to comprehend the information in a better and easier way.

In a study conducted by [4], subjects who used the lesson with good design principles completed the lesson in less time and had a higher completion rate than those who used the lesson with poor design principles and [5] indicated that organization and visual interest are important criteria in judging the readability and studyability of the real screens. Screens that are plain, simple, unbalanced, and bare are perceived as undesirable.

Recent research on the visual aesthetics of computer interfaces suggests that aesthetics is a strong determinant of users' satisfaction and pleasure [6] and that visual attractiveness of the site affects users' enjoyment as well as perceptions of ease of use [7]. This is supported by [8] who found that careful application of aesthetic concepts can aid acceptability, learnability, comprehensibility and productivity. Acceptability was investigated by [9] and [10] where the studies showed very high correlations between users' perceptions of interface aesthetics and usability.

3. Previous Studies

The empirical study conducted by [11] using a VLE found a positive relationship between interface aesthetics and perceived usability. Another of our studies tested the importance of five design principles (unity, proportion, homogeneity, balance and rhythm) as a simple way to achieve aesthetics of computer interface [12].

The results showed to be statistical significant for all five design principles. This gives evidence that the users prefer the screens that apply the design principles. It also tried to find if aesthetics (achieved through applying the design principles) were related to the users' perceived usability of the environment. The stimuli material used comprised three different sets of screens (home, content and e-mail) each one with four different layouts. Therefore, a total of 12 screens had been evaluated on their aesthetic aspects and after, on their perceived usability.

The results were statistical significant and showed a positive correlation between aesthetics and perceived usability in all three sets and also in all four different screen layouts.

4. Empirical Study

The present study was conducted to: (i) confirm previous study results and (ii) verify if the results would have major differences after the user's interaction with the proposed VLE. Participants interacted with the interface, performing two tasks, before evaluating the aesthetics and usability of the interface used. It was made necessary to develop a prototype of a VLE environment to be the stimuli material so that the participants could interact with the VLE in order to perform the tasks.

The aesthetic aspects had two different sets of screen layouts, one that had all the design principles applied (high aesthetics) and other that violated all of them (low aesthetics). The usability aspects had on set with delays and error messages (low usability) on another without delays (high usability). The four conditions (Table 1) generated by aesthetics and usability aspects were the following:

- condition 1 (1A): high aesthetics and low usability;
- condition 2 (1B): high aesthetics and high usability;
- condition 3 (2A): low aesthetics and low usability;
- condition 4 (2B): low aesthetics and high usability.

4.1 Method

4.1.1 Design

The experiment had a 2 x 2 between-subjects design. The aesthetics factor had two levels (high and low aesthetics) and the usability factor with two levels (high and low usability). It gave four conditions as shown in Table 1. Each participant took part in just one condition and the tasks were performed using the Latin square design to counterbalance the order effect.

4.1.2 Participants

The experiment was completed by 98 participants enrolled or working at a Higher Education Institution in Brazil. The participants had different backgrounds, schooling, computer literacy and age. The study was controlled for gender (49 male and 49 female) and the mean age of the participants was 27.9 years.

4.1.3 Material and Apparatus

The study was carried out in a controlled environment. . It was used three PCs, allowing three participants to take the experiment at the same time. Each PC had 17'' monitors calibrated to display the same colours values, contrast and brightness. They were positioned in such a way that did not allow sight of other participant computer screen.

The stimulus materials intended to simulate the most common parts of an interactive VLE environment in four different conditions for aesthetics and usability (Table 1). There were three main screens: the homepage which the participant would see first, the content material which they would navigate and the email where they would read and write the messages. Also, they were shown a slide with two thumbnails, one with high aesthetics and the other with low aesthetics to rate the overall impression of the layout.

Table 1: Study conditions.

		Aesthetics	
		High	**Low**
Usability	**Low**	Condition 1 (1A)	Condition 3 (2A)
	High	Condition 2 (1B)	Condition 4 (2B)

The last part of the experiment was done on paper questionnaire and intended to collect information regarding the hedonic characteristics of the inteface design. The questionnaire used the bipolar verbal anchors. It was used 12 out of 21 hedonic characteristics proposed by [13] to better suit the needs of the present study.

4.1.4 Procedure

As part of the ethics procedure, each participant read a letter of information about what kind of study they were about to participate in and that they had the right to withdraw from the study at any moment. They also received verbal information about the aim of the study before signing

the consent form allowing the data collected to be used for scientific and academic purposes. The data collected was anonymous and the identity of the participant was removed. Once the participants started the experiment, there were intermediary screens with instructions guiding them from the beginning to the end. The tasks were designed with the objective of getting the participant to navigate through the environment and, at the same time, getting them involved with it. The VLE prototype was developed to assure that they would follow the same path and just perform the task asked by this study, by enabling just one link at a time.

On the first task the participant played the role of a student taking an online course who had received email from a friend asking his or her opinion about a particular topic of the content material. She or he was asked to go to the content material to find the answer and then reply to the email based on their aesthetic opinion.

For the second task, the participants were told that they were tutors of online courses who had received an email from a group of students asking about the content of weeks 6 and 8. They were asked to go to the calendar and reply to all the students with the correct answer.

The interface used 'Previous' and 'Next' buttons to advance and return pages. They could navigate through the material and spend as much time as they needed to complete the task. When answering the emails, it was possible to return to the content material or calendar if they forgot the answer or wanted to check it again.

After they finished the two tasks they were asked to evaluate the environments' aesthetics and usability using a five-point Likert scale. First, they evaluated the aesthetic aspects, varying from (1) Unattractive to (5) Attractive, and then the usability aspect, varying from (1) Difficult to use to (5) Easy to Use.

To verify if the participants would have different opinions when both screens layouts were presented together they were shown two layouts (Figure 1) with two thumbnail each of the screens used in the study, screen A (applying all the principles) and screen B (violating all of them). They simply had to choose between screen A or B in terms of their preference for the aesthetic aspects and on the second slide do the same in terms of usability. It gave the participant a chance of comparing them before rating the one they perceived to be more attractive and then, the one they found easier.

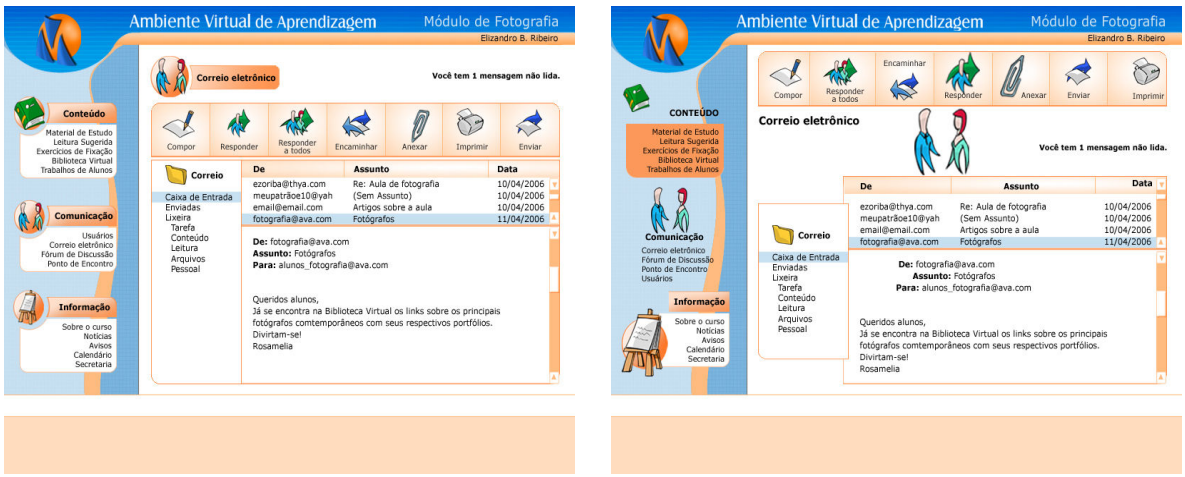


Figure 1 – Thumbnails of screen A (high aesthetics) and screen B (low aesthetics).

Finally, they would evaluate 12 the hedonic characteristics of a given screen on a 5 point scale paper questionnaire. The closer to 5 was the participants' ratings the stronger was the affect caused by the screen layout.

4.1.5 Results

There was a positive correlation (r = .446, N = 98, p < .001) between ratings of aesthetics and of usability, after using the environment. The correlation were even higher between the ratings of aesthetics and usability (r = .679, N = 98, p < .001) when the participants had the opportunity to compare between good and bad screen layouts. Figure 2 presents a graph with the mean ratings using a five-point Likert scale.

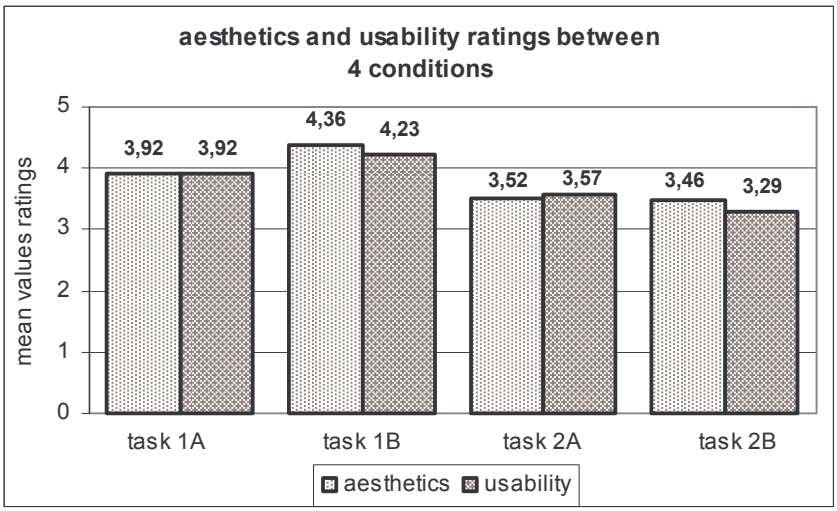


Figure 2: Mean ratings for aesthetics and usability on 4 conditions.

Figure 3 shows a graph of the users' aesthetics and usability preferences when comparing the two screen layouts, one that had applied all the principles (good layout) and the other that had violated the design principles (bad layout). These final evaluations were done independently of the conditions that the participant had been submitted.

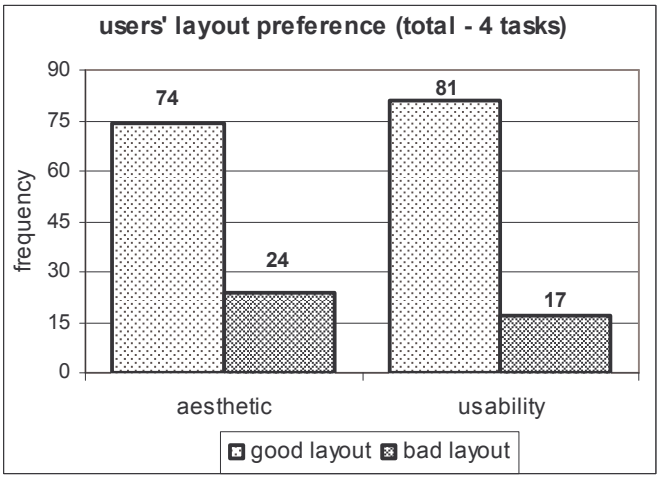


Figure 3: Users' perception of aesthetics and usability based on good or bad layout.

Figure 4 shows the mean value of participants' perception towards the hedonic characteristics taking into account the screens layouts. It had 3 main factors: identification, stimulation and pragmatic.

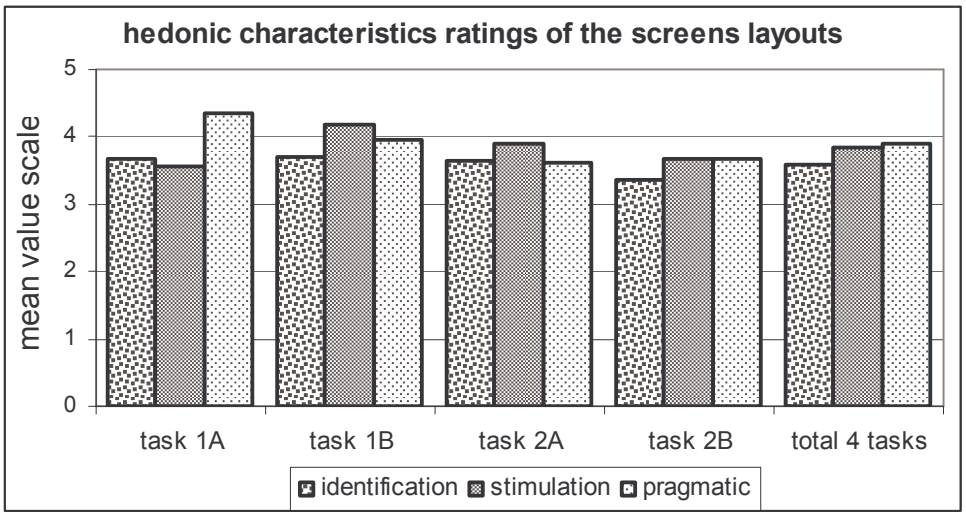


Figure 3: Users' perception of hedonic characteristics of the screens layouts.

5. Discussion

Our previous study [11] showed that the five design principles (unity, proportion, homogeneity, balance and rhythm) are perceived by the users as important aspects of the screen layout design. These gave indication that they could be used as a measure for achieving aesthetics.

The static screens layouts study also found a positive correlation between aesthetics and perceived usability in all 3 different sets and in all 4 different layouts. However the difference between groups were not statistical significant corroborating the findings from [10] and giving indication that at the visceral and behavioural levels of emotional design theory [1], the perception of aesthetics and its relation to usability do not depend on culture. The visceral level is first level of emotional design, what nature does. It is concerned with appearance and does not depend on cultural aspects. The behavioural level relates to the brain process that processes and control everyday behaviour. It deals with pleasure and effectiveness of use.

The present study used an interactive environment and the participants evaluated the screens layout aesthetics (developed using the design principles) and the usability of an interactive environment after they completed the tasks.

The positive correlation found between aesthetics and usability shows the participants' preference for screens layouts that apply the design principles giving evidence that those screens were perceived as aesthetically pleasing and as having better usability. It also implies that the design principles are a good measure for aesthetics.

We also found a positive correlation between aesthetics and usability in the four different conditions. The variance between different groups was not statistically significant. This suggests that the perception of aesthetics and usability may not be dependent upon culture.

The correlation between aesthetics and usability were higher when the participants could compare between different screen layouts. The big difference on the frequency of participants' preference gives clear evidence that they were able to distinguish good from bad interface screen layout. It was not related with computer literacy, age or gender. Although the

results were similar for schooling and cultural background, further investigation needs to be done. Finally, the hedonic characteristics suggest three factors that may have influence on users' perception the screen layout towards the environment.

6. Acknowledgements

I, R. Parizotto-Ribeiro, express my thanks to CNPq (Brazil) for sponsoring this PhD research and those who helped to have this study done. I much appreciated my supervisor's valuable corrections and comments on this article.

7. References

[1] Norman, D. Emotional Design: Why we love or have everyday things. Basic Books, New York, USA, 2004.

[2] Toh, S.C. (1998). Cognitive and motivational effects of two multimedia simulation presentation modes on science learning. PhD dissertation, University of Science Malaysia.

[3] Aspillagae M. (1991). Screen design: a location of information and its effects on learning, Journal of Computer-Based Instruction, vol.18 (3), pp.89-92.

[4] Szabo, M. and Kanuka, H. (1998). Effects of Violation Screen Design Principles of Balance, Unity and Focus on Recall Learning, Study Time, and Completion Rates. ED-Media/ED-Telecom 98 Conference Proceedings, Association for the Advancement of Computing in Education.

[5] Grabinger, R.S. (1991). Computer screen designs: viewer judgements. Educational Technology Research and Development, vol.41 (2), pp.35-73.

[6] Lavie, T. and Tractinsky, N. (undated). Assessing Dimensions of Perceived Visual Aesthetics of Web Sites [online]. Retrieved 02 Janeiro, from http://burdacenter.bgu.ac.il/publications/LavieTractinsky.pdf

[7] Heijden, van der H. (2003). Factors Influencing the Usage of Websites: The case of a Generic Portal in the Netherlands. Information and Management, vol.40, pp.541-549.

[8] Ngo, D., Teo, L. and Byrne, J. Modelling Interface Aesthetics. Information Sciences (2003), vol.152, pp.25-46.

[9] Kurosu, M. and Kashimura, K. (1995). Apparent Usability versus Inherent Usability Experimental analysis of the determinants of the apparent usability, CHI'95 Mosaic of Creativity Proceedings, pp.292-293.

[10] Tractinsky, N. (1997). Aesthetic and Apparent Usability: Empirically Assessing Cultural and Methodological Issues, CHI'97 Proceedings pp.22-27.

[11] Parizotto-Ribeiro, R., Hammond, N., Mansano, J,. and Cziulik, C. (2004). Aesthetics and perceived usability of VLEs: preliminary results, HCI 2004 Proceedings, pp.217-221.

[12] Parizotto-Ribeiro, R. and Hammond, N. (2004). What is Aesthetics anyway? Investigating the use of design principles, Aesthetic Approach to Human-Computer Interaction, Proceedings of the NordCHI 2004 Workshop, pp37-40.

[13] Hassenzahl, M. The interplay of beauty, goodness and usability in interactive products. HCI Journal: In press.

Affect Assessment in Educational System Using OutSite Factors

Sylvia TZVETANOVA[1]
Ming-Xi TANG[1]
[1]*Design Technology Research Centre, School of Design, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*

Abstract. This research presents an approach to the improvement to the hedonic qualities of user performance in web interfaces based on the principles of emotional design. It addresses the issue of how emotional factors can be integrated into the communication between users and the web systems. The paper describes factors that can be used in emotion recognition for adaptation upon the user's emotions in online Intelligent Tutoring Systems. The integration of cognitive methods for web system design and computational techniques for the representation of emotional elements of users forms the core of the system architecture in which aesthetics and emotions are combined in computational terms for supporting users in a more emotional way.

Introduction

To date researchers have tried to create intelligent computers considering human cognition, reasoning, learning, perception, and language. Cognitive Science considers that emotion plays an essential role in human intelligence and behavior, and this indicates the need to rethink the role of emotion in building intelligence in computers [13]. Gratch and Marsella [7] believe that emotions are an essential part of human live and impact on their beliefs, inform decision-making and in large measure provide guidance on how to adapt their behaviour to the world around. Managing the effective communication between human and computer is a major task for interface designers as it is considered that memory is related to emotions [2]. As emotion is essential for human cognition and memory [13], many of the studies of emotions in computers are addressing the problem of how to improve the student's performance in Intelligent Tutoring Systems (ITS) using affective computing techniques. The concepts of Emotionally Intelligent Tutoring System (EITS) which uses emotions to improve learning efficiency were described by Ochs and Frasson [11]. Burleson and Picard [3] have purposed an affective companion that is keeping the user to have optimal experiences and sustain their motivation during studding. Benchetrit and Frasson [1]: purposed an idea for an agent that is based on the teaching method of the Bulgarian doctor and psychotherapist Lozanov [10], however the method of the doctor is not confirmed itself and has not been practiced widely.

In this work, we present a model of emotional factors for adaptation of web tutoring interface upon the user's affect. The model has the goal of establishing the standards for hedonic web interfaces that can tackle the affective state of the users. The purpose of the implementation is to provide emotional and usability support of the user with hedonic qualities, as we expect that the user performance will increase due to the emotional features of the system.

It is a complex task to "introduce" emotions to computers. In order to achieve the tasks of emotional tutoring interface several major issues are involved:

- Obtaining affective information about the user: make the system appropriately assume the affective state of the learner;
- Make reaction to the assumed affective state.

Many kinds of knowledge are involved in order to achieve the accurate results such as Pedagogy, Human-Computer Interaction, User Modeling, Psychology, and so on. Our approach is multidisciplinary based on the design research, practice and principles.

1. Affect Assessment Approach

1.1 Basic Emotions

Emotion recognition is a complex task and requires deep understanding of different knowledge fields. In an attempt to reengineer the state of online tutoring systems it is essential to look first at existing pedagogy approaches. Gathering knowledge from observation of experienced teachers can help implement emotions in the computer teaching systems. Computer system lack the ability to make individual approach to gather emotional knowledge about students while supporting teaching expertise and subject domain needed. There are a few major questions in making a belief about emotions in web based tutoring system: Which are the basic emotions involved in positive studying as the number of emotions can be large? What are the factors of these emotions in order to detect them? How can we make a computer detect those factors?

In the learning process memorization is done by associating emotions to the material. It is well known that the positive emotions are involved in the constructive learning process [9]. The positive emotions also increase the motivation and the learner is more focused on the material [8]. The goals of the EITS are to soften negative emotions to positive ones, and preserve the learner positive experiences in order to facilitate the constructive learning process. The motivation have been important for ITS as well, Vicente and Pain [16] are assigning to the motivation in ITS the variables: control, challenge, independence, fantasy, confidence, sensory interest, cognitive interest, effort, and satisfaction.

The number of possible expressed emotional elements involved in a web interface can be huge, with many variations. We have focused on the basic emotions. Many psychologists have tried to define the basic list of emotions. One of the well known definitions is the one of Plutchik [14], i.e., acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise and all the other emotions are mixtures, or compounds of the primary emotions. The cognitive theory of emotion developed by Ortony Clore and Collins [12] defines emotions as valenced (positive or negative) reactions to situations consisting of events, actors and objects. Further more, the valance of the emotion occurred depends on the desirability of the situation and one's goals and preferences. Rosalind Picard, Barry Kort and Rob Reilly have purposed emotion sets involved in the un-learning and constructive learning processes: anxiety-confidence, boredom-fascination, frustration-euphoria, dispirited-encouraged, terror-enchantment [9]. Based on the theories above we are making assumption about the possible emotions involved in the interacting process with web based tutoring system, which shall be evaluated further, see Table 1.

Table 1. Basic emotions that we assume occur during interaction with educational system

Negative	**Positive**
Distress	Joy
Fear	Confidence
Boredom	Fascination
Unsatisfied	Satisfied

1.2 Affect Assessment Approach by Reasoning the OutSite Factors

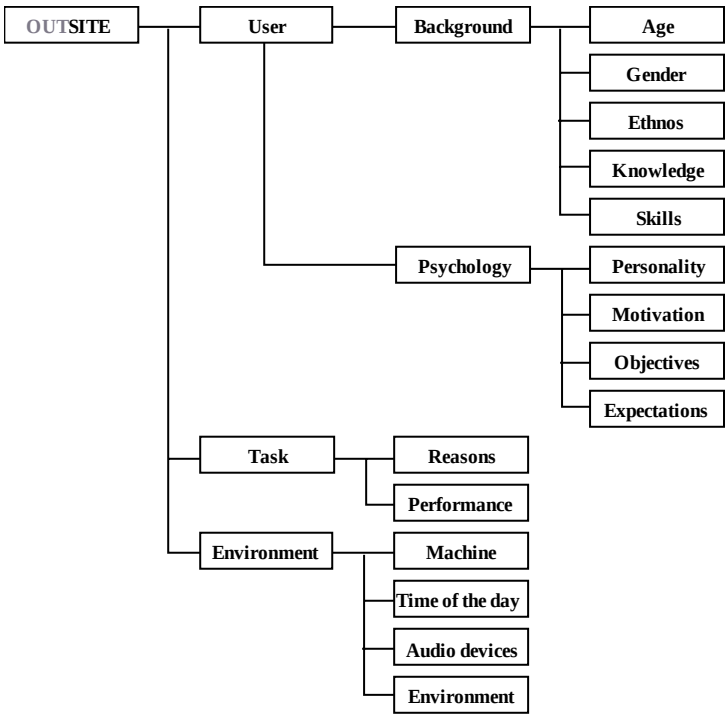
To make the computer recognize emotions, many researchers are focusing on affective wearables: physical devices that users wear with, and the computer is used to read their emotional parameters. The computer might recognize emotions as well by interacting with different agents. The human model is based on the assumption that people do not necessarily recognize emotions just by signals seen or heard, they also use a high level of knowledge and reason about their goals, situations, and preferences. It is predictable the person's emotions if their goals and perception of relevant events are known [12]. Implemented in a computational model this can be achieved by using agents, artificial intelligence techniques reasoning on goals, situations, and preferences [5], [4]. For example, if the system can reason about the reactions of a user from the input that the system receives, (assumption made based on the time of the day, speed of reading, browsing path and so on) appropriate content will be provided and displayed in a way adapted for the emotion or the mood of the user. Negative or disturbing emotions should be softened to stimulate the positive ones, the users are less effective at learning when distracted, bored, and stressed or feeling extremely excited. For the system adaptation we have chosen to use basic emotional features, which we consider reasonably sufficient and accurate for an experimental study.

To represent to the computer the emotional signs - the emotion factors that are involved in the interaction with educational tutoring system need to be collected. We have tried to collect Human–Computer Interaction (HCI) factors, outside the system (OutSite factors – which we use short from 'outside the web site'), that are related to the user, task and environment, which also have impact on the user's emotions. The recognition of emotions will be done using a reasoning system which takes the OutSite factors as the input, i.e., the input and assumptions regarding the goals, situations, and preferences of the user.

In the diagram of the OutSite factors (Fig. 1) are presented the factors that we collected and we believe that might have relations with the user's emotions when interacting with a web based system. They are classified in a tree structure with tree major branches: factors related to *the user*, *the task* that has to be accomplished by visiting the web site, and *the environment*. In the diagram the user related factors are mainly about their backgrounds and psychological profiles. *The background* involves the age, gender, ethnos, knowledge and skills of the user. *The gender* and *the ethnos* can have impact on a person's emotions according the Hofstede theory of masculinity and femininity [6]; *the age* has relation to the subject preferences of the students. *The personality type* (introvert or extrovert) has relation to their aesthetic preferences. For example extroverts prefer more attractive and dynamic interfaces and the opposite, i.e., the introverts, like more useful and less distracting interfaces [5]. The perception of colors in web sites is linked to one's emotions and personality [15]. The motivation is a subcategory of psychology, as in this section we combine the cognition elements involved in the process. The objectives are the personal reasons to attend the course, they have relation to emotion because the reasons can vary in valance: the student have been forced by the parents, the student like the course and want to enjoy it, the student has motivation to learn something new and so on.

The expectations can be used as a measurement tool of the happiness at the end of the course: a high expectation may cost disappointment.
The emotion recognition will be made by evaluation of the probability. The collection of information about the OutSite factors will be made by their input into the system only. The basic scenario is that the user in order to access the tutoring system will have to fill few questions to get a password. The questions will be related to age, and personal preferences. The other factors can be obtained by web programming such as location, task performance, and environment. Self report method will be more limited as we assume that it may take time for the student to report himself and also this may cost distraction. The system has own observations on how fast the pages are browsed respectively the student has get the materials and test results, and from this information can reason about the task performance. In the case when the user is browsing more slowly in time T+1, than they were clicking in time T, then a belief can be assigned in the system that they might be distracted, annoyed or less focused. As different emotions have priorities in humans [13], the same way the computer database also needs to have priorities that organize the importance and intense of the assumed situations. Such reasoning and judgment of the user's behavior is an intelligent way to tackle online adaptation upon visitor's emotions. Furthermore this simple model can mimic the human reasoning, by providing the system with knowledge about real satiations.

Figure 1. OutSite Factors that are related to the user's emotions in a branching hierarchy configuration



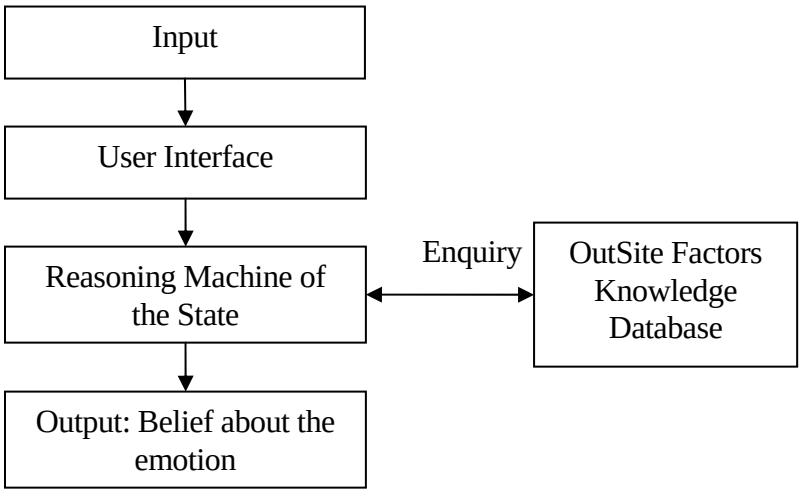
2. Related Work and Application Scenario

2.1 Application

In the previous section we have describe the factors that can be used to detect the affect. Here we are purposing a simple system configuration of how the OutSite factors shall take place into EITS (Fig 2.). A multiple input will be taken to get various responses from users. The occurred situation is to be analyzed by the reasoning system. When OutSite factors are determined, an

InSite element or elements have to be composed for the system to respond to that emotional belief of the user in an appropriate way. For example the user is self reporting that they has high motivation to accomplish the course. This input is passed to the reasoning machine, that reference the OutSite knowledge database about the priority of this factor and the description of it and the emotions attached to this factor. Based on that information the reasoning machine can make a belief about the emotional state of the user. Some previous state might have been reordered in the system, for example earlier at time T.

Figure 2. Adaptation Process

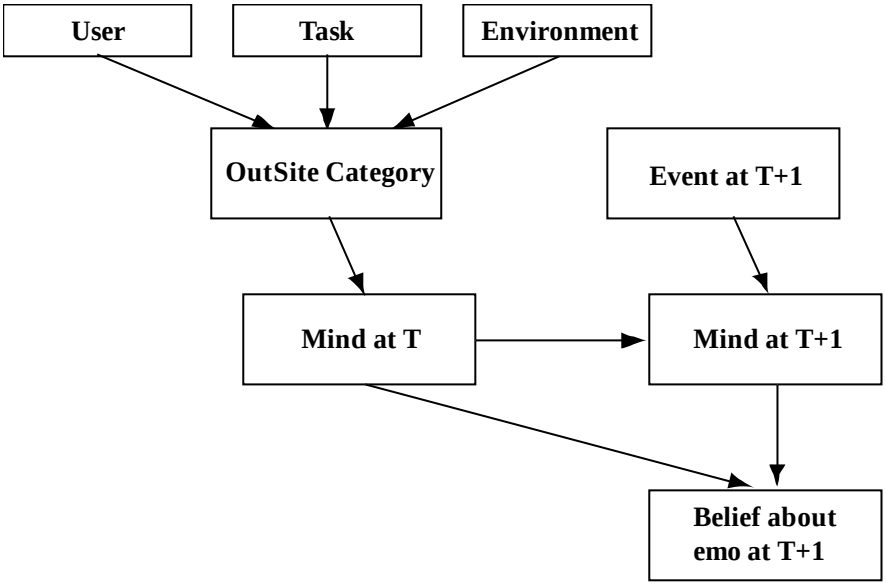


The problems of implementation of those factors are that 1) It is difficult to predict the initial emotion of the student, that they might have already set to a certain mood due to other outside factors related to their private life; 2) The emotion factors are having different weights: the possibility that some factor may influence the emotions. 3) The purity of emotions: the case when the emotions overlap [13].

2.2 Related Work

Carofiglio, de Rosis and Grassano observed the problem of overlaped emotions [4]. A computational model for making an artificial triggering of emotions over time and represent to an educational agent was made by using Dynamic Belief Networks (DBSs). Events occurred during a time interval (T,T+1) are observed to make a belief of the new state and reason about emotions that might be triggered by these events. The calculation of the intensity of emotions made by the uncertainty in the agent's beliefs about the emotional state and the weight assigned to this emotion. The variation of intensity in the emotion is assigned by the probability that certain factor will take place, times the weight of this factor (Fig. 3).

Figure. 3. Monitoring of Emotional Belief based on Carofiglio, de Rosis and Grassano



3. Future Work

We have taken a symbolic approach to improve the usability of web interface by adding an emotional response system to a web interface design based on a simple list of emotional elements. The OutSite factors have to be implemented in tutoring system using decision making approach. The future work is the InSite factors to be determined and adapted upon the belief of the student's emotions. While the system can not change the OutSite factors, it has a full control on the InSite factors. This approach is to be evaluated with interface and graphical designers when the system is fully implemented with an intelligent web based tutoring system as an example. We will find out whether the model is accurate for measuring the improvement on the study abilities of the students. The evaluation will be made by usability and design methods such as semantic differentiation. This approach has been used by designers to measure the experiences of new products.

Acknowledgments

This research aims at adding a new dimension to the existing research projects in the Design Technology Research Centre in the School of Design of the Hong Kong Polytechnic University, which are still largely based on digital and computational models, whose inference control mechanisms are not based on human cognition, rather, they are based on symbolic models of mathematic and mechanic thinking. This research is supported by a PhD studentship from the Hong Kong Polytechnic University.

References

[1] Benchetrit, O., Frasson, C.: Controlling Emotional Conditions for Learning. Workshop on Social and Emotional Intelligence in Learning Environments, In conjunction with the 7th International Conference on Intelligent Tutoring Systems, Maceio - Alagoas, Brazil, August 31st (2004)

[2] Bower G.: How Might Emotions Affect Learning?. In Handbook of Emotion and Memory: Research and Theory, edited by Svenake Christianson, Lawrence Erlbaum, Hilsdale, NJ (1992)
[3] Burleson, W., Picard, R. W.: Affective Agents: Sustaining Motivation to Learn Through Failure and a State of Stuck. Workshop on Social and Emotional Intelligence in Learning Environments, In conjunction with the 7th International Conference on Intelligent Tutoring Systems, Maceio - Alagoas, Brazil, August 31st (2004)
[4] Carofiglio, V, de Rosis, F., Grassano, R.: Dynamic Models of Mixed Emotion Activation. In Animating expressive characters for social interactions edited by D. Canamero and R Aylett, John Benjamins, in press
[5] Conati, C.: Probabilistic Assessment of User's Emotions in Educational Games. Journal of Applied Artificial Intelligence, Volume 16, Numbers 7-8 / August-September, Taylor & Francis (2002) 555 - 575
[6] Dormann, C., Chisalita, C.: Cultural Values in Web Site Design. Eleventh European conference on cognitive ergonomics, ECCE11 Catania, Italia (2002)
[7] Gratch, J., Marsella, S.: A Domain-Independent Framework for Modeling Emotion. Journal of Cognitive Systems Research, Volume 5, Issue 4 (2004) 269-306
[8] Isen, A. M.: Positive Affect and Decision Making. Handbook of Emotions, second edition, Guilford Press (2000)
[9] Kort, B., Reilly, R., Picard, R. D.: External Representation of Learning Process and Domain Knowledge: Affective State as a Determinate of its Structure and Function. MIT Press (2000)
[10] Lozanov, G.: Suggestology and Outlines of Suggestopedy. Gordon and Breach (1978)
[11] Ochs, M., Frasson, C.: Optimal Emotional Conditions for Learning with an Intelligent Tutoring System. Workshop on Social and Emotional Intelligence in Learning Environments In conjunction with the 7th International Conference on Intelligent Tutoring Systems, Maceio - Alagoas, Brazil, August 31st (2004)
[12] Ortony, A., Clore, G. L., Collins, A: The Cognitive Structure of Emotions. Cambridge University Press (1988)
[13] Picard, R. D.: Affective Computing. Cambridge, Mass.: MIT Press (1997)
[14] Plutchik, R.: A General Psychoevolutionary Theory of Emotion. Emotion: Theory, research, and experience: Vol. 1. Theories of Emotion edited by R. Plutchik & H. Kellerman, New York: Academic (1980)
[15] Santini, Jean-Andre: Personnalité et emotion dans les sites web: evaluation textuelle et colorimétrique. DEA D'Iinformatique anée 2001-2002 Université de la Mediterranée (2002) 211-225
[16] Vicente, A., Pain, H.: Informing the Detection of the Students' Motivational State: An Empirical Study. Intelligent Tutoring Systems, 6th International Conference, Biarritz, France and San Sebastian, Spain, June 2-7 (2002) 933-943

Creating a pedagogical model that uses student self reports of motivation and mood to adapt ITS instruction

Carole R. BEAL & Hyokyeong LEE
Information Sciences Institute, University of Southern California
4676 Admiralty Way, Marina del Rey CA 90292 USA
cbeal@isi.edu

Abstract. Our project focuses on the design, implementation and evaluation of a ITS pedagogical model that considers student motivation, mood and cognitive processes in making instructional decisions in the domain of secondary school mathematics. Students complete integrated self report assessments of motivation and mood. Cognitive skillls such as math fact knowledge, spatial cognition, and prior math achievement are also assessed. The pedagogical model adapts instruction (problem selection, problem difficulty, topic area, choice of activity, choice of help type, and availability of help) following a model of human tutoring expertise that balances motivational and cognitive goals.

Introduction

Much research in educational psychology points to the important role of motivation in student learning, in addition to cognitive processes [1]. Students who do well in school combine effective cognitive strategies for mastering the material with effective self regulation of motivation. Such self regulation strategies include goal setting, accurate appraisals of the reasons for their performance, and effective management of the emotions associated with learning activities, such as boredom and test anxiety. Other research shows that effective human instructors balance attention to the goal of promoting students' cognitive understanding of the domain content, and supporting the students' motivation.

In response to the recognition that learning involves both motivation and cognition, intelligent tutoring systems (ITS) researchers are now beginning to consider the role of student motivation. However, much of the research on user motivation while working with ITS systems has focused on the assessment of the student's motivational traits and emotional states, that is, how to tell if the learner is engaged, bored, frustrated, etc. Although assessment is critical, we must also consider the question of how an ITS should respond to different student motivational and affective states.

1. Research on student motivation and mood

The challenge for ITS designers is that the optimal pedagogical responses to various student motivational states -- in terms of student learning outcomes -- are not always obvious or intuitive. For example, instructional decisions that are meant to reduce learner frustration, although well-intentioned, may not really benefit the student. Most current theories of

learning argue that some element of frustration or cognitive conflict is an essential part of knowledge change; thus, the challenge is not to provide instruction that attempts to avoid student frustration but rather to help the student manage frustration and discouragement appropriately. In the domain of mathematics learning, the most effective teachers do not directly attempt to ameliorate student frustration; rather, they convey that some degree of uncertainty and failure is to be expected when learning demanding material, and emphasize these feelings will fade as the student becomes more proficient [2].

Another important consideration is that students with diverse personal characteristics (gender, ethnicity, prior achievement) often react differently to the same instruction and feedback. For example, common-sense might suggest that a discouraged student should be provided with supportive and encouraging feedback. However, female students often interpret such feedback as an implicit indication that the teacher feels that they lack ability in the subject [3]. Similar findings have been reported for students with learning disabilities. Instructor praise for doing well on difficult material actually led to declines in motivation for high-achieving African-American students, but not their low-achieving classmates [4]. The high-achieving students felt that they were already working very hard, and the instructor's praise increased their concern that they would not be able to sustain their successes.

Students' moods also influence their learning. Students experience both negative and positive emotions at school [5]. Anxiety was one of the most frequently experienced emotions, not only in relation to taking exams, but also being in class or studying at home. Students also reported experiencing anger, boredom and shame about their performance. High-ability students experienced academic boredom when demands were too low. A number of students reported that they experienced boredom when they were not able to keep up with demands. Students also experienced positive emotions at school, for example, enjoyment of learning, hope, pride, and relief. Although mood and academic motivation interact, the constructs are distinct: Mood shows more fluctuation, and may reflect aspects of students' out-of-school life (e.g., social relationships with peers and parents) more than is the case for academic motivation.

2. The Wayang-West ITS

The goal of our project is to implement and evaluate a pedagogical model that is based on the strategies used by expert human tutors and classroom instructors to help students learn while also sustaining student motivation: strategies that have been empirically validated in terms of student learning outcomes [6, 7]. The context for the research is the Wayang-West ITS for secondary school mathematics, specifically SAT-Math problem solving. (The SAT exam is a high-stakes math achievement test taken by most high school students in the United States as part of the college admission process.) Mathematics is the academic subject that is most often associated with both variations in student achievement and motivation.

The Wayang-West is a web-based ITS that includes multiple components to evaluate students' prior knowledge, and cognitive abilities that are predictors of mathematics learning and achievement:

> Pre and post tests. Students take a mock SAT-M exam before and after the tutoring activity. Multiple versions of the mock tests of equal

difficulty are available. Each test includes items that require tutored and non-tutored skills, to provide an internal control.

Cognitive assessments. At the start of the activity, students complete on-line versions of a standard test of math fact retrieval proficiency, and spatial cognition. Performance on these cognitive tasks is correlated with math achievement.

Tutoring module. Students work on a series of SAT-M problems provided by the College Board (the test author). The format of the SAT-M problems is multiple-choice. Students can answer at any point, they can skip a problem, or they can request multimedia help. Two forms of help are available for most problems: one form emphasizes traditional algorithms, whereas the other emphasizes visualization strategies.

Transfer tests. Wayang-West includes three virtual adventures: simple narratives animated in Flash about environmental science themes, that include complex, multi-step math problems, e.g., calculating the ratio of legal to illegal harvesting of teak forests in Indonesia. These items require mathematics skills tutored in the ITS but the skills must be applied in the context of real-world content about endangered species topics and thus serve as an indication of transfer of learning.

We have now extended the Wayang-West ITS to include integrated motivation and mood self report instruments. Self reports are an efficient source of information about student states and do not require expensive or intrusive instrumentation or equipment that cannot be used in public school classrooms. (Also, self reports constitute something close to ground truth: if a student says that she is in a bad mood, or that she does not feel that she is any good in math, we are inclined to take her at her word.) The motivation instruments are based on prior work with the Online Motivation Questionnaire (OMQ) [8]. The OMQ is a well validated instrument that has been used to assess motivation in several academic domains, including mathematics. Its structure reflects current work indicating that motivation is context dependent and multidimensional, rather than being a static trait-like characterstic. In addition, the instrument includes an attribution assessment that is completed after the activity.

We modified the instrument to reflect the content of our specific task (preparing for math achievement tests, versus math classroom learning) and performed a week-long evaluation test with two geometry classes at an urban high school. (The detailed empirical results will be reported in [9].) The results made clear that the original instrument was too lengthy to complete in its entirety within the constraints of the class schedule. We therefore adapted a number of items and eliminated redundant items. The instrument is now intended to be completed as part of a User Profile, and includes questions to assess:

Learning goal orientation. Prior work suggests that students typically adopt either a performance goal orientation (e.g., wanting to earn a high grade), or a learning goal orientation (e.g., wanting to master the material). Students with a learning goal orientation use deep learning

strategies, seek challenge, and devote effort to demanding tasks. In contrast, students with a performance goal orientation are more likely to use shallow learning strategies, regard mistakes as evaluative threats, and withdraw when the task becomes difficult [10, 11]. In our experience, the learning-performance goal orientation must be supplemented with a third orientation that we call, "dis-engaged". These students indicate that they do not really care about doing well in school or learning the material; their goal is simply to get through the activity.

Incremental-entity beliefs. Students tend to believe either that intelligence is a fluid quality that can be enhanced through practice and effort, or that it is a fixed quantity [10]. These beliefs are especially strong in the domain of math learning. Students indicate their agreement with statements such as "I think that how smart people are stays pretty much the same throughout their life". In our prior work, students with incremental intelligence beliefs were more likely to use the multimedia help resources available in the ITS than students with entity beliefs.

Mathematics motivation. Math motivation includes beliefs about self efficacy, value of math, and enjoyment of math that have been found to mediate students' academic achievement [1]. Students with low self-efficacy do not expect to do well in math, even if they have done well in the past. Female students tend to have lower self efficacy in math than their male peers, even though females receive high grades on average in math classes. Students provide ratings in response to items such as, "This year I expect to < fail, barely pass, pass, do pretty well in, be one of the top students in > math class.".

Daily mood reports. Wayang-West includes a brief mood report instrument that is completed by students each day before they begin the tutoring activity. Items include questions such as, "I'm having a < great day, OK day, bad day >".

3. The Wayang-West pedagogical model

The goal of the Wayang-Weest pedagogical model is to promote student learning outcomes by balancing the goal of providing instruction with the goal of maintaining student engagement and motivation. The student who has lost interest will not attend to the instruction or use the resources that are available. Conversely, even the student who is interested will not learn if the material and help features selected for presentation are not well chosen. A major goal is to encourage students to engage in effort-based learning [12]. Human capability is open-ended and ability is created through sustained and targeted effort. To achieve this goal, we need to consider a teaching strategy which takes both cognition and motivation. Providing students with opportunities of experiencing successes through effort

enables them to have high self-efficacy in math and builds confidence that they can succeed in learning the material.

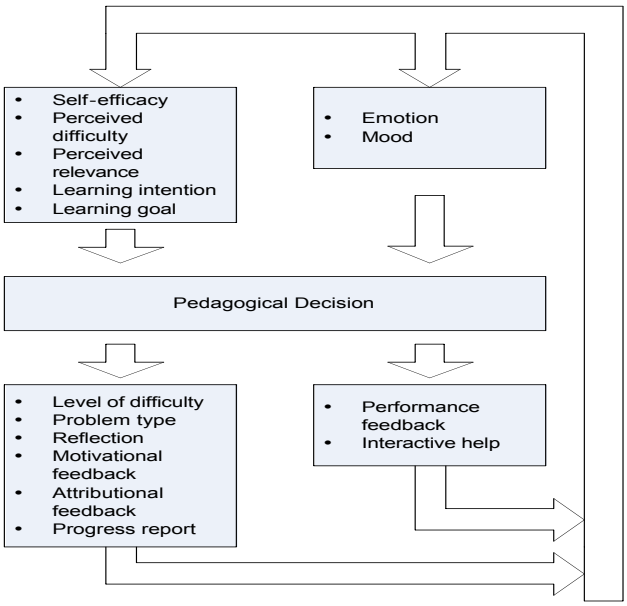


Figure 1: Pedagogical decisions in Wayang-West

The skilled human teacher has a wide range of tasks, strategies, and approaches that can be used to engage and instruct the student. As indicated in Figure 1, the ITS is more limited in the options that can be varied in the model. These include: the selection of problems to be presented to the student by difficulty (easy, moderate, difficult, in relation to student prior achievement); the selection of problems by domain topic (require the student to continue working on math problems about right triangles, shift to another topic such as corresponding angles); tutorial activity (continue with same activity, e.g., tutoring, or offer choice of another activity, e.g., adventure problems); the type of multimedia help provided to the student (visually-oriented help; algorithmic help); the availability of multimeida help (on demand, or after encouraging the student to attempt the problem independently first); provision of progress feedback, and other features to support self explanation and collaboration.

Some of the instructional decisions can be made before students enter the tutoring activity. For example, the problem selection mechanism can be adjusted on the basis of student prior knowledge, as indicated by the pre test results, and the increase in difficulty over problems can be adjusted in relation to student self efficacy. The purpose of adjusting the level of problem difficulty according to students' current performance is to enhance students' self-efficacy. Based on pre-test result, self-efficacy, perceived task difficulty, tutor may select appropriate level of problem for the student. Adjusted level of problem may increase the probability of success in solving the problems. For high achieving students, random problem selection mechanism is appropriate. For low achieving students, the ITS might show easy problems at first and progressively increase difficulty. Accumulated experiences of successes may improve students' self-efficacy in math.

The model can also offer alternate activities in response to student mood. For example, a tutor who sees that a student is having a bad day might suggest starting the session with a review of some material that the student has already mastered (i.e., providing easy material as a gradual transition into the learning activity). Human tutors offer choice to

engage reluctant students: thus, the student might be offered the opportunity to work on the virtual adventure problems; the fantasy element of the adventures is appealing to students (e.g., saving an orangutan in a rainforest fire, reconstructing a wildlife rehabilitation facility in a village destroyed by mudslides, investigate an illegal logging operation in a teak forest). Fantasy elements in software have positive effects on learning and motivation [13]. The student can also be encouraged to use the self-explanation feature in Wayang-West: after completing a problem, the student can make notes about the problem, why it was difficult, and what insights were needed to solve it. Of course, at some point the student must be directed to the tutoring activity in the ITS; it would not be pedagogically appropriate to respond to student low motivation or negative affect with avoidance of the learning activity. The point is that providing students with transitional activities and choices, even if limited, can help to engage the disengaged or discouraged learner.

Other aspects of the pedagogical model can be dynamically modified in response to student behaviors while working with the ITS. In particular, the student's help seeking can provide an indirect measure of engagement and motivation, in that requests for help are automatically recorded and available for analysis. Students vary significantly in their use of help resources: Some researchers have found that good students use and benefit from ITS help more than weaker students [14]. However, this may have been because the quality of the help was better matched to good students [15]. Others have noticed that some students have high help seeking rates because they are just searching for the right answer, clicking rapidly through the help features [16].

Our model makes the decision about whether to provide help in response to a request on the basis of a) the student's prior help seeking, e.g., if the student has requested all the hints on previous problems, the model may recommend help fading with the message to try the problem independently first; b) whether the student has see a problem requiring the same skills before, e.g., the model might suggest the student review an earlier worked example; c) the student's prior knowledge: students with low skills should not be prevented from viewing help; d) the degree of interactivity in the help, e.g., rather than showing all hints in sequence, the student might be required to complete part of a problem before receiving more help. Interactive instruction makes students feel supported cognitively and affectively [17]. The pedagogical model can also adjust the probability of providing one type of help or another (visually-oriented, algorithmic). If students do not understand or benefit from one type of explanation, the model can increase the probability of providing the alternative type of explanation. Changing the type of explanation can also be used to re-engage the student with low motivation.

Student progress reports can also be used by the pedagogical model to address variations in learning, motivation and mood. Students want to know how they are doing, but they need more than objective information about the accuracy of their answers. Human instructors sandwhich information about objective performance (such as test scores, number of problems correctly solved, etc.) with motivational feedback: "I know you can do this," and "Keep up the good work". Positive feedback can enhance students' self-efficacy but it works only when students feel that they can meet the task demands. Effort feedback is critical. This is in line with the principles of effort based learning and may change students' attributional styles from a belief in native ability to the value of effort [10].

Student progress reports should also should have only individual progress information. Social comparison (information about how the student is doing compared to others in the class) can motivate some students, but is not generally very effective,

especially for low-achieving students. Their relatively poor progress may easily make them feel more helplessness and students are more likely withdraw effort to protect themselves from the implication that they lack ability [18].

4. Conclusions and future work

We are currently implementing the pedagogical model that considers student cognition, motivation and mood into the Wayang-West ITS for high school mathematics. The pedagogical model uses student self report data and observed student behaviors while interacting with the system to select appropriate teaching strategies for individual students. The strategies reflect research on effective human teachers' teaching methods in classrooms and in one-on-one tutoring, along with factors that influence on students' motivation and cognition. We will be collecting additional data with high school students and will present empirical data about the effectiveness of the model at the workshop.

Acknowledgments

The original Wayang Outpost ITS was developed with support from National Science Foundation HRD 0120809 (Beal, Woolf & Royer). The Wayang West project is currently supported by National Science Foundation grants HRD 0411532 and REC 0411886 (Beal). The views represented in this paper do not necessarily reflect the position and policies of the funding agency. We would like to thank project members Erin Shaw, Yuan-Chun Chui, Hannes Vilhjamsson, and Lei Qu for their help, and Lewis Johnson and Mark Lepper for their assistance with the design of the pedagogical model.

References

[1] Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. Contemporary Educational Psychology, 25(1), 82-91.
[2] Turner, J. C., Midgley, C., Meyer, D. K., Gheen, M., Anderman, E. M., Kang, Y., & Patrick, H. (2002). The classroom environment and students' reports of avoidance strategies in mathematics: A multimethod study. Journal of Educational Psychology, 94(1), 88-106.
[3] Beal, C. R. (1994). Boys and girls: The development of gender roles. New York: McGraw Hill.
[4] Martin, D. (2000). Mathematics success and failure in African-American youth. Mahwah NJ: Erlbaum.
[5] Pekrun, R., Goetz, T., & Titz, W. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. Educational Psychologist, 37(2), 91-105.
[6] Heffernan, N. T. (2001). Intelligent tutoring systems have forgotten the tutor. Computer Science Dept., Carnegie Mellon University.
[7] Lepper, M. R. et al. (1993). Motivational techniques of expert human tutors. In S. P. Lajoie & S. J. Derry (Eds.), Computers as cognitive tools (pp. 75-105). Hillsdale NJ: Erlbaum.
[8] Boekaerts, M. (2002). The On-Line Motivation Questionnaire: A self-report instrument to assess students' context sensitity. In P. R. Pintrich & M. L. Maehr (Eds),

New Directions in Measures and Methods (Advances in Motivation and Achievement, vol. 12). New York: JAI.
[9] Beal, C. R., Shaw, E., Chui, Y., Lee, H., Vilhjalmsson, H., & Qu, L. (2005). Enhancing ITS instruction with integrated assessments of learner mood, motivation and gender. 12th International Conference on Artificial Intelligence in Education, To appear.
[10] Dweck, C. S., & Legget, E. L. (1988). A social-cognitive approach to motivation and personality. Psychological Review, 95(2), 256-273.
[11] Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects of ability, conscientiousness, and goal orientation. Journal of Applied Psychology, 89(2), 231-247.
[12] Resnick, L. B., & Hall, M. W. (200). Principles of learning for effort-based education. http://curriculum.dpsk12.org/pol_prin_learn.htm
[13] 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(4), 625-633.
[14] Wood, H., Wood, D. (1999). Help seeking, learning and contingent tutoring. Computers and Education, 33, 153-169.
[15] Aleven, V., McLaren, B., Roll, I., Koedinger, K. (2004). Toward tutoring help seeking: Applying cognitive modeling to meta-cognitive skills. In the Proceedings of the 7th International Conference on Intelligent Tutoring Systems, ITS 2004.
[16]. Aleven, V., & Koedinger, K. R. (2001). Investigations into help seeking and learning with a cognitive tutor. In R. Luckin (Eds.), Papers of the AIED-2001 Workshop on help provision and help seeking in interactive learning environments, 47-58.
[17] Stevens, L., Van W. W., Stokking, K., Castelijns, J., Jager, A. (2000). Interactive instruction to prevent attention problems in class. Learning Environments Research, 3(3), 265-286.
[18] Burnett, P. C. (2003). The impact of teacher feedback on student self-talk and self-concept in reading and mathematics, 38(1), 11-16.

When do Students Interrupt Help? Effects of Individual Differences

Cecily HEINER, Joseph BECK, and Jack MOSTOW
Carnegie Mellon University
Project LISTEN, Newell Simon Hall, 5000 Forbes Avenue Pittsburgh, PA 15213

Abstract. When do students interrupt help to request different help? To study this question, we analyze a within-subject experiment in the 2003-2004 version of Project LISTEN's Reading Tutor. From 168,983 trials of this experiment, we report patterns in when students choose to interrupt help. To improve model fit for individual data, we adjust our model to account for individual differences. We report small but significant correlations between a student parameter in our model and gender as well as external measures of motivation and academic performance.

1. Introduction

In a companion paper [1], we study when students interrupt help and we report a model to predict the probability that students will interrupt help as shown in Equation 1. This model shows that students interrupt help increasingly often over the first thirty exposures, and then their interruption rate reaches an asymptote. We defer all discussion of the framework of our experiments and participants to that paper. In this paper, we expand our model by adding a student parameter. We show that this student parameter improves model fit. Then we correlate this variable with test scores that measure student attitudes and other affective variables. This paper deals primarily with variables which model personality and character traits; in our conclusion, we suggest directions for future work which might model affective states.

$$P(i) = a * (e^{b * prior\_exposures}) + c$$

Equation 1: Base Model

2. Adding a Student Parameter

To account for individual differences, we added a student parameter, *s*, to our model as shown in Equation 2. Conceptually, a student with a high *s* parameter interrupts more often than a student with a low *s* parameter. This student parameter, *s*, alters the asymptote of the graph and is related to a student's interruption rate, a value that should be between zero and one. To insure that values for *s* would be consistent with this idea, we set the initial value for the student parameter *s* at 1, and imposed the limits that *s* must be less than or equal to 1 and greater than -.5. Within this range [-.5, 1], SPSS fit a single student parameter for each student.

$$P(i) = a * (e^{b * prior\_exposures}) + c + (s * (1 - c))$$

Equation 2: Student Parameter Model

3. Comparing Model Fits

Table 1 uses mean squared errors and r^2 to compare the two models with a simple baseline called the overall interruption rate.. The overall interruption rate model simply predicts that 43% of all help will be interrupted, since this is the average interruption rate when all of the data is aggregated together. Table 1 shows that the biggest reductions in mean square error and improvements in r^2 come from applying a base model that takes time into consideration by accounting for the amount of previous help. Adding a student parameter improved the model moderately.

Table 1: Models and Mean Square Errors

Model Name	**Mean Square Error**	**r^2**
Overall Interruption Rate	.24	-
Base Model (Equation 1)	.19	0.24
Student Parameter Model (Equation 2)	.17	0.30

4. Correlating the Student Parameter against External Measures

The student parameter, *s*, in the model is a variable that may relate to other measures of a student, including process variables and test scores. We considered the following process variables: help request rate, help interruption rate, disengagement (measured as the percentage of questions that students answer hastily [2]), and percentage of time picking stories. We were surprised that we did not find correlations with those affective variables, especially disengagement or help request rate.

For test scores, we considered pre- and post-test scores and gains for the Elementary Reading Attitude Survey (ERAS) [3] and a fluency test. ERAS is a twenty item instrument with ten items each for recreational and academic reading attitudes. The fluency test measures how many words per minute a student reads according to a trained tester; more fluent readers score higher. Small, significant negative correlations exist between the *s* parameter and the ERAS academic and motivational test scores. So, *s* relates to attitudes towards academic and recreational reading. Additionally, a small but marginally significant correlation exists between fluency pre-test and the student parameter *s*, so, *s* may also be related to proficiency. Thus, a more fluent reader is less likely to interrupt. Table 2 displays the correlations.

Table 2: Student Parameter Correlations

Test Name	Pearson Correlation	Signficance
Fluency Pre-Test	-.155	.072
ERAS Recreational Pre-Test	-.267	.002
ERAS Academic Pre-Test	-.283	.001

In order to determine the relationship between *s* and gender, we ran an independent samples T-test and found the mean *s* value for girls (-0.057) differs from the mean *s* value for boys (0.037) at $p<0.001$. Therefore, girls are less likely to interrupt than boys, the difference is significant, and the *s* parameter is related to gender.

4. Future Work and Conclusion

In this paper, we have introduced spoken help interruption as an observable, analyzable outcome variable that is an affective indicator. We refined a mathematical model that predicts help interruption to include a student parameter *s*, and we have shown that *s* is correlated with attitudes towards reading and gender. Attitude and gender are fixed traits, and they influence the asymptote in our model, the portion of the curve that represents more stable behavior. Interesting future work may involve considering those variables which model affective state and determining how they can further improve model fit. Such future work would need to take into account temporal regions of data and would ideally show correlations with other observable behaviors indicative of affective state.

Acknowledgements

This work was supported by the National Science Foundation, under ITR/IERI Grant No. REC-0326153. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation or the official policies, either expressed or implied, of the sponsors or of the United States Government.

Thanks to Ryan Baker, Hao Cen, and Jonathan Brown for comments on drafts of this paper. Additional thanks to students, teachers, and administrators in schools where the Reading Tutor was deployed when this data was collected.

References (website: www.cs.cmu.edu/~listen/)

1. Heiner, C., J. Beck, and J. Mostow. *When do Students Interrupt Help? Effects of Prior Exposures and Individual Differences*. in *Artificial Intelligence in Education*. 2005. p. To Appear Amsterdam, The Netherlands.
2. Beck, J.E. *Using response times to model student disengagement*. in *Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments*. 2004. p. 13-20 Maceio, Brazil.
3. Kush, J.C. and M.W. Watkins, *Long-term stability of children's attitudes toward reading*. The Journal of Educational Research, 1996. **89**: p. 315-319.

Knowledge Weaving: Supporting Effective and Affective Interactions in the Call Centre.

Nicola J. Millard,
British Telecommunications PLC, Adastral Park, Ipswich, U.K.

Abstract. To be able to talk to customers, solve problems and develop relationships, call centre advisors need to be equipped with technologies that will help them understand who the customer is, allow them to speak knowledgeably about a complex and often ever expanding set of products and services and negotiate a minefield of legalities, processes and procedures. Systems to support this knowledge are often separate and distinct from the user's task. However, in a call centre, both advisors and customers require fast accuracy. There is always too much for advisors to remember, yet simply providing knowledge via systems is no guarantee that these systems will be used. This paper considers how knowledge, motivation, affect and problem solving styles need to be weaved into the advisors' daily tasks. It looks at how systems designs need to support efficient and effective interactions as well as encouraging curiosity, active learning and discovery.

1. Call Centres and Knowledge.

One of the main roles of a call centre advisor is to serve as an interpreter between customers and the data held within the company systems. However, customers don't just call for mediated access to databases but for expert and empathic assistance. Advisors respond to these needs through a mixture of their knowledge of the customer, the customer's context, the company's language, process and services and the structure and content of the support systems. As products and services become more complex, call centre advisors cannot know everything and increasingly rely on knowledge systems to support them in their task.

2. The Challenges of Creating Knowledge User Interfaces for Call Centres.

User interfaces for call centre knowledge systems need to present a large amount of complex information whilst also supporting the emerging dialogue with the customer [1]. The process of giving knowledge to the customer is not emotion neutral [2,3]. Call centre problem solving can involve confusion or frustration, as well as joy and satisfaction if the advisor can help the customer. This also involves a number of "memories" [4] including the advisor's head, from databases, intranets or from other sources, e.g. paper notes.

Creating effective systems to support the advisor's task is critical. However, it is often assumed that knowledge, once acquired, will be used willingly, appropriately and efficiently [6]. Advisors, under pressure of call handling times, often use their own (often tacit) knowledge in preference to any of the systems supporting them. In a study by the author, advisors used the tacit knowledge in their heads for 71% of the time and the knowledge management system only 7% of the time. The problem is more about trust, motivation and affect than about simply providing knowledge at the advisors' fingertips.

3. Increasing Technology Acceptance.

The reasons for negligible user uptake of knowledge management systems are often because they are not embedded in the user's task, capture common knowledge rather than tacit knowledge and offer little personal benefit to the user [5]. Making a product more usable does not necessarily ensure that it is used [6]. For the knowledge system to be accepted, advisors needed to be convinced that it was useful and relevant to their task. However, perceptions are also influenced by more emotional factors such as enjoyment, satisfaction and fun [7, 8]. Traditional ways of motivation, such as mandating or incentivising the usage of the system can result in advisors looking at the incentive rather than the reason for it. Motivational psychology notes that recognition by peers (through 'social influence'), rather than financial rewards, can be a primary motivator for use [9].

From the advisors' perspective, knowledge management may be more about *finding* the knowledge rather than about actually committing it to memory [10]. Behavioural decision theory would suggest that the cost/benefit equation for encoding the location of the knowledge is different to that of the knowledge itself. If the perceived cost of retrieval is too high (i.e. it takes too long or is too difficult to find) then advisors are more likely to store a local copy, in the form of paper notes, or rely on sources other than the knowledge system.

4. Implications for User Interface Design.

These factors of effect, affect and motivation were all considered when the author was asked to redesign the UI of an existing call centre knowledge management system. Observation of the advisors found that they have two explicitly different task drivers depending on whether they are *online* with a customer (where they are maintaining a dialogue, solving problems and controlling the call) or conducting *offline* tasks (where they are often browsing through briefings and news or communicating with their team or manager).

In '*Online' or 'Goal' mode*, the emphasis is on effectiveness and efficiency. Low emotional arousal is preferred since any increase is likely to prove detrimental to the advisor's capacity to problem solve. Anything that prevents achievement of these goals will result in frustration and negative affect. The implication for UI design in this mode is simplicity with optimal support for the achievement of tasks. If the system also matches the advisor's problem solving style, then they will be more likely to maintain a positive attitude towards it and usage is likely to increase, with a resulting increase in familiarity and information retention [11]. Trust is often founded on consistency in navigation and clarity in information presentation plus the ability for the user to form a mental map of the presentation space.

Whilst online, the knowledge portal had to address the efficient access of knowledge, reduction of advisor cognitive load and increasing the usability and structure of the knowledge that advisors had to read off the screen. Individual search strategies were based upon training and personal preferences and, since time was a factor when they were on the phone, they were unlikely to explore unfamiliar search mechanisms. The design involved usage of space- semantic relationships to logically group related links together.

In *'Offline' / 'Action' mode* is where effectiveness and efficiency are less important, volatility is a key factor and playfulness and spontaneous action is frequently experienced along with high arousal. If arousal decreases, people are liable to get bored. This implies a more emotional approach to UI design, emphasising playfulness, curiosity, active learning and discovery [12, 13]. Knowledge management is not simply about passively providing information to the user, it is about engaging curiosity and allowing users to learn whilst

responding to their emotional state. Increased ease of use of a system motivates the user to explore the system functionality, which may in turn increase intrinsic motivation, and result in greater enjoyment of the activity. However, during this period of time there is no motivated impetus for learning, i.e. there is no specific problem to be solved.

The 'offline' interface was designed along the lines of the computer game, 'The Sims'. This fosters an atmosphere of discovery and implies that an element of play or exploration is allowed without the user being nervous that they will "break things".

5. Results and Discussion.

The redesigned knowledge system was taken back to the call centre advisors. They were asked to evaluate it both on a general usability level but were also asked to describe their feelings towards the system. Words such as "easy", "friendly", "encouragement", "clarity", "easier", "relaxed" and "fun" appeared frequently in the evaluation. Users asserted that they would want to use the new knowledge interface more than the old one which the advisors described as "boring", "confusing", "intimidating" and "complex".

However, this was under experimental conditions, so the research team returned to complete an ethnographic evaluation of the knowledge system use three months after changes had been implemented to assess whether the system was *actually* being used during calls. This evaluation showed that usage had increased from 7% to 49%. It is difficult to conclude that the design alone was responsible for the changes in usage pattern as there is likely to be influence simply from increased familiarity with the tool. Further study is required to look at the factors contributing to increased technology acceptance in this case.

References:

[1] Millard, N., Hole, L. and Crowle, S. (1997), From Command to Control: Interface Design for Customer Handling Systems, in Howard, S. et al (eds), *Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction*, Chapman & Hall/ IFIP, 294 - 300.

[2] Kalbach, J. (2004), Feeling Lucky? Emotions and Information Seeking, *Interactions*, 11:5, September/October, New York: ACM Press.

[3] Kuhlthau, C.C. (1993), *Seeking Meaning: A Process Approach to Library and Information Services*, Norwood, NJ: Ablex.

[4] Ackerman, M.S. and Halverson, C. (2000), Re-Examining Organisational Memory, *Communications of the ACM*, 43:1, 58-64.

[5] Alavi, M. and Leidner, D. (2001), Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues, *MIS Quarterly*, 25:1, March, 107-136.

[6] Nielsen, J. and Levy, J. (1994), Measuring Usability: Preference vs. Performance, *Communications of the ACM*, 37:4, 66-75.

[7] Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. (1989), User Acceptance of Computer Technology: A Comparison of Two Theoretical Models, *Management Science*, 35, 982-1003.

[8] Venkatesh, V. (2000), Determinants Of Perceived Ease Of Use: Integrating Control, Intrinsic Motivation, And Emotion Into The Technology Acceptance Model, *Information Systems Research*, 11:4, 342-365.

[9] Malholtra, Y. and Galletta, D.F. (2003), Role of Commitment and Motivation in Knowledge Management Systems Implementation: Theory, Conceptualisation and Measurement of Antecedents of Success, *Proceedings of Thirty Sixth Hawaii International Conference on System Sciences*.

[10] Dix, A., Howes, A. and Payne, S. (2003), Post-Web Cognition: Evolving Knowledge Strategies for Global Information Environments, *International Journal of Web Engineering Technology*, 1:1, 112-126.

[11] Jordan, P.W. (2000), *Designing Pleasurable Products: An Introduction to the New Human Factors*, Taylor & Francis, September.

[12] Logan, R.J. (1994), Behavioural and Emotional Usability: Thomson Consumer Electronics, in Wiklind, M. (ed), *Usability in Practice*, Cambridge, M.A: Academic Press.

[13] Hassenzahl, M., Beu, A. and Burmester, M. (2001), Engineering Joy, *IEEE Software*, January/February).

VALERIE: A Virtual Agent for Environmental Learning, Reacting and Emotional Interaction

Marco PALEARI, Christine LISETTI & Mikko LETHONEN
Multimedia Department, Eurecom Institute,
2229 Route des Crêtes, BP 193 06904, Sophia Antipolis, France

Abstract. In this paper, we describe some work in progress related to the development of a cognitive-affective architecture for an Intelligent Tutoring System, in our case a French language tutor avatar. A Belief-Desire-Intention strategy is implemented in conjunction with a simple model for the affective state of the avatar.

1. The VALERIE System

Intelligent Tutoring Systems (ITS) have shown positive pedagogic results with studies at Carnegie Mellon University showing students performing on average 43% better than those who didn't use such computer aided systems. It also offers particular advantages in resource constrained environments. The VALERIE framework draws on this experience and seeks to augment ITS strategies with affective Belief-Desire-Intention [1] reasoning mechanisms. Previous work supports the effectiveness of affective feedback in such human machine interaction [2][3][4][5].

VALERIE is based on two important pedagogic objectives: to provide feedback and give support (in conjunction with the implicit objective of having the student perform well). These goals are mapped in term of three possible coping strategies: (a) give positive feedback, while the user is performing well, (b) give positive support, while the user is performing poorly with the agent assuming it is trying to do well; (c) give an annoyed or strict comment when poor user performance and the agent assumes the user is not trying to improve.

The VALERIE system is based on the MAUI framework [6] and is implemented using a custom avatar interface developed using the Haptek People-Putty software and Microsoft Text-to-Speech. The system employs a *question rounds* strategy where the user is asked to answer 10 simple questions on French grammar taken from a predefined question database. The questions are characterized according to their level of difficulty with four possible answers, one of which is the correct answer. At the end of a question round, the user sends the answers to the agent. The agent subsequently stores information about the student's answering time, provides the correct answers to the user, and takes the time the student used to correct wrong answers. The agent can then compute a `reaction_frame` which is essentially characterized by two values, one for each goal. From this `reaction_frame` the agent computes its new affective state as well as the appropriate feedback and emotion features to be expressed (via an avatar). The reaction frame is computed using three steps: (a) compute new parameters based on the current internal state and user model; (b) compute feedback and support values based on a linear function of the user score and its parameters; (c) if the user performed badly, estimate if the user's effort in answering the questions is sufficient. The values used are `low_value_bad`, `low_value_good`, `bad` and `good`. A parameter called `strict_threshold` is also used which references an accepted minimum user's effort in answering the questions. If a user's effort is assumed to be below this threshold, the agent is strict and the support value is negative. VALERIE adapts the suitable parameters in order to

maintain this state, similarly when in a happy mode. This approach is described in several studies, for example [4].

In VALERIE, the emotional response of the agent is directly linked to the goal with the largest intensity, with the verbal action of the agent being chosen by searching for the dialog text that best represents the reaction frame. This facilitates adaptive utterance responses to different situations. Parameters, when computed, are adapted to the user through the beliefs the agent has in its user profile, with the most important factor of adaptation being the `user_average`. For example the system would be much less demanding when interacting with one user that scores four correct answers on average than with one that usually scores nine out of ten. The relatively simple Belief-Desire-Intention architecture implemented generates Beliefs based on the `user_average` and `user_effort`, and employs three Desires and three Intentions. A form of multilayer emotional or affective state with personality, mood and emotions is subsequently superimposed on this BDI framework. This architecture is subsequently implemented in a custom interface developed through Haptek People-Putty with a corresponding set of basic expressive emotional behaviours.

2. Experimentation & Results

The VALERIE architecture was tested to evaluate agent responses and the resulting emotive state under specific user interaction scenarios. The first of three tests was designed to show the adaptability of the system to the user, the second to see how the internal state, or mood, influences agent decision, and the third demonstrates three different designed "personalities". In all tests, the parameter values were in the range of [0, +100] for positive only values, and [-100, +100] for either positive or negative values.

User Average: The first test was designed to show the influence of `user_average` on agent `reaction_frames`. Tests were undertaken with three times a score of 6, twice with a score of 8, twice with a score of 6, and three times a score of 4. Depending on the `user_average`, the responses of the avatar were very different. With a starting average score of 4, the reactions were always positive or neutral and the agent tended to provide feedback to the user, while users with an average of 8 gave responses which were always supportive. Results show the adaptability of VALERIE to the `user_average`. In particular the internal state of the agent and the `reaction_frame` ([`feedback`][`support`]) changes depending on the user's initial average. In a particular case when the agent is getting angry due to a user being awarded bad scores, the emotion expressed can still be positive as the agent assessed `user_effort` as being sufficient. This is important in that it shows how VALERIE can hide his internal state with a positive emotion when it is estimated that constructive feedback may lead to long-term user improvement.

Agent Internal State: The second test was designed to show how the system's internal state can influence its responses based on the recognised behaviour of someone in a bad mood usually sees things worse than they are. Two runs with the same scores as previously were undertaken but with the initial internal state of the agent changed. The first run had the `internal_state` set at - 40, placing VALERIE in a "bad mood", and the second run at a value of + 20, i.e. relatively content. It can be seen in results shown in table 1 how the internal state significantly changes the interaction between VALERIE and the user. For the same inputs, when VALERIE is in a bad mood the agent gets angry while, when in a good mood, it becomes empathic, indicating two opposing emotional states. It is interesting to also look at the two selected dialog texts for this interaction which indicate a constructive pedagogical approach:

- *You disappointed me; I know you can do much better.*
- *Oh, don't worry; I 'm sure you can do much better. Let's try again, ok?*

Psychological studies support behavioural believability from adopting this strategy. An agent that does not demonstrate some degree of emotional extremes, i.e. angry or happy, would appear cold and unbelievable. It is argued that the fundamental tenet of agent believability rests on its ability to express a range of emotional responses.

Table 1: Results for internal state (mood) influence

Initial internal state = - 40

score	feedback	support	internal state	emotion shown
2	3	**-71**	-42	anger
3	2	**-56**	-52	annoyed
4	1	**42**	-59	happy
5	0	**36**	-65	happy
6	0	**30**	-70	happy

Initial internal state = + 20

score	feedback	support	internal state	emotion shown
2	9	**62**	21	empathy
3	13	**41**	15	happy
4	16	**20**	13	content
5	**20**	17	12	neutral
6	**24**	13	12	neutral

Agent Personality: The last test was designed to show the influence of personality on the avatar's deliberative reasoning process. VALERIE's personality depends on the initial values of the parameters used to compute the reaction frame and not on a complete personality regarding, for example, the Five Factor Model of personality [7]. To run this test, three different personalities were designed: Peter, who is very demanding, Alan, who is predominantly positive, and Mary, who is particularly emotional. Results have shown that the three avatar responses are significantly different with all agents initially getting angry, but in different ways. For example Alan, being the positive avatar, tries to be as supportive as possible, and gets angry only after the user scores particularly bad. Peter's behaviour is different where he becomes easily annoyed and subsequently maintains a more continuous emotional state. Mary was never neutral, with her emotions alternating from happy to angry and back to happy, following the user's results.

These experiments demonstrate how simple deliberative affective agent interfaces can simulate simple personalities. This presents the foundations for future work on a more complex model for social affective artificial personalities where models of the user's affective state would be employed.

References

[1] **Rao, M. Georgeff (1991).** *Modeling Rational Agent within a BDI-Architecture.* 2nd Int. Conf. on principle of knowledge representation and reasoning, KR91, San Mateo, CA, September 1991.

[2] **Elliott, C., Rickel, J. & Lester, J. (1999)** *Lifelike Pedagogical Agents and Affective Computing: An Exploratory Synthesis.* Artificial Intelligence Today, LNCS 1600, Springer Verlag.

[3] **Frijda N. H. (1986).** *The Emotions*. Cambridge University Press, 1986.

[4] **Picard R. (1997).** *Affective computing,* MIT Press, Cambridge, MA, 1997

[5] **Lisetti C. L. and Gmytrasiewicz P. (2002).** *Can a Rational Agent Afford to be Affectless: a Formal Approach*. In proceeding of Applied AI in 2002.

[6] **Lisetti C. L. and Nasoz F.(2002).** *MAUI: A Multimodal Affective User Interface.* In Proceedings of the ACM Multimedia International Conference 2002, (Juan les Pins, France, December 2002).

[7] **Costa, P., & McCrae, R. (1995).** *Primary traits of Eysenck's P-E-N system: Three and five factor solutions*. Journal of Personality and Social Psychology, 69, 308-317

Affect Detection and Metaphor in E-Drama[1]

Li Zhang, John A. Barnden and Robert J. Hendley
School of Computer Science, University of Birmingham, Birmingham, B15 2TT

Abstract. We report work in progress on adding affect-detection to an existing e-drama program, a system for dramatic improvisation in simple virtual scenarios. The system allows a human director to monitor improvisations and make interventions. To partially automate directors' functions, and to allow for automated affective bit-part characters, we have implemented a prototype affect-detection module. It is aimed at detecting affective aspects (emotions, moods, rudeness, value judgments, etc.) within users' speech. The detection is necessarily shallow, but the work accompanies basic research into how affect is conveyed linguistically, with a special focus on metaphor.

Introduction and Relationship to Other Work

Improvised drama and role-play are widely used in education, counselling and conflict resolution. Researchers have explored frameworks leading to e-drama systems in which virtual characters (avatars) interact under the control of human actors. The springboard for our research is an existing system (*edrama*) created by Hi8us Midlands Ltd. This has been used in schools for creative writing, careers advice and teaching in a range of subject areas. Hi8us' experience suggests that e-drama helps school children lose their usual inhibitions, because they are not physically present on a stage and are anonymous. In the *edrama* system, the characters are completely human-controlled, their speeches are textual, and their visual forms are static cartoon figures. There is generally also a human director, who can intervene by sending messages to actors and introducing bit-part characters. However, this places a heavy burden on human directors. One of our main research aims is to partially automate the directorial functions. Affect detection is an important element of directorial monitoring. Accordingly, we have developed a prototype affect-detection module. It has not yet been used explicitly for directorial monitoring, but is instead being used to control a simple automated bit-part character, EmEliza. This automatically identifies affective aspects of the other characters' speeches, makes certain types of inference, and makes appropriate responses to help stimulate the improvisation. Within affect we include: basic emotions such as anger; more complex emotions such as embarrassment; meta-emotions such as desiring to overcome anxiety; states such as mood and hostility; and value judgments (of goodness, etc.).

Much research has been done on creating affective virtual characters in interactive systems. Emotion theories, particularly that of Ortony et al. [5] (OCC), have been used widely. Prendinger and Ishizuka [6] used the OCC model in part to reason about emotions. Mehdi et al. [9] used OCC in their approach to the generation of emotional behaviour. Gratch and Marsella [10] presented an integrated model of appraisal and coping, partly to reason about emotions. However, few e-drama (or related) systems can detect affect comprehensively

[1] This work is supported by grant RES-328-25-0009 from the ESRC. We are grateful to our partners – Hi8us Midlands Ltd, Maverick Television Ltd and British Telecommunications plc. We have benefited from discussion with our colleagues W. Edmondson, S. Glasbey, M. Lee, A. Wallington and Z. Wen.

in open-ended utterances. Although Façade [4] included shallow processing for open-ended utterances, the detection of major emotions, rudeness and value judgements is not mentioned. Zhe and Boucouvalas [8] demonstrated emotion extraction using a tagger and a chunker to detect emotional words and to analyse emotions of the speaker. Unfortunately the detection focuses only on emotional adjectives, and does not address deep issues such as figurative expression of emotion. Also, the focus purely on first-person emotions is narrow. Our work is distinctive in several aspects. Our interest is not just in (a) the first-person case: the affective states that a virtual character X implies that it *has* (or *had* or *will have*, etc.), but also in (b) affect that X implies it *lacks*, (c) affect that X implies that *other* characters have or lack, and (d) *questions, commands, injunctions,* etc. concerning affect. We aim to make any relatively shallow detection that we manage to achieve in practical software responsive to general theories and empirical observations of the variety of ways in which affect can be conveyed in textual language [2], and in particular to the important case of *metaphorical* conveyance of affect [2,3]. Our system is in part aimed at providing guidance and inspiration to the theoretical study of affective language, as well as being an end in itself.

1. A Preliminary Approach to Affect Detection and Responding

Different dimensions of emotion are widely used in different emotion theories. Currently, we use an *evaluation* dimension (positive and negative), *affect labels* and *intensity*. Affect labels with intensity are used when strong text clues are detected, while the evaluation dimension with intensity is used when only fuzzy text clues implying affect are detected. At present, our implemented affect detection is based on textual pattern-matching rules that look for simple grammatical patterns or templates. This approach possesses the robustness and flexibility to accept ungrammatical fragmented sentences and to deal with varied positioning within speeches, but lacks other types of generality and can be fooled when the phrases are embedded as subcomponents in grammatical structures.

The language in the speeches created in e-drama sessions, especially by excited children, has many aspects that, when combined, challenge existing language-analysis tools. These include: misspellings, ungrammaticality, abbreviations, slang, use of upper case and special punctuation for affective emphasis, repetition for emphasis, open-ended onomatopoeic elements and occasional intrusion of wording from other languages. These characteristics of the language make the genre similar to that of Internet chat.

The transcripts analysed to inspire our initial knowledge base and pattern-matching rules for our implemented affect detection were produced from earlier Hi8us *edrama* improvisations based on a school bullying scenario. The actors were school children aged from 8 to 12. We are generalizing our current methods beyond the school-bullying scenario, for instance by drawing from the embarrassing-illnesses TV documentaries produced by Maverick TV Ltd (one of our partners).

To go beyond the limitations of the currently implemented text matching, we are exploring the use of the Plink parser in the GATE framework. In addition we will include an electronic thesaurus (e.g.WordNet) and existing dictionaries of affective items. The current dialogue management strategies are also quite simple and need to take more account of the intensity and frequency of each character's speeches. Ultimately we will work towards automated characters that are given goals for how they are meant to provoke other characters, and reason about how to achieve this.

2. Metaphorical Expression of Affect

The *explicit* metaphorical description of emotional states is common in ordinary discourse and has been extensively studied [2,3]. Examples of such description are "He nearly exploded" to indicate anger, and "Joy ran through me." Also, emotion, value judgments and other forms of affect are often conveyed *implicitly* via metaphor, as in "His room is a cess-pit", where affect associated with a source item (cess-pit) gets carried over to the corresponding target item (the room). In this work we are studying such language both theoretically and practically through the e-drama system itself and by further development of an independent metaphor processing system called ATT-Meta [1].

Physical size is metaphorically used in descriptions of negatively-valued types of people, as in "you are a big bully" (or similarly "you're a big idiot") and "you're just a little bully." The bigness can be literal but typically indicates the extent or intensity. Size adjectives may also be used to convey the speaker's attitude towards the object. "The big bully" expresses the speaker's strong disapproval [7] and "little bully" can express contempt, although "little" can also convey sympathy. These and other examples are not only practically important but also theoretically and implementationally challenging.

In the affective-computing domain there has been a considerable amount of attention to developing scientific models of human affective states. Such modelling is often important, but in studying how people understand mundane affective language and how they pitch their own affective utterances, what may be much more important is people's own *common-sense* views of how affect works in people, irrespective of scientific reality. This is one reason metaphor comes in, given the strong involvement of metaphor in many common-sense views of the mind.

3. User Testing

User testing with groups of secondary school students at several Birmingham schools will take place in autumn 2005, with a view to determining the effect of including automated affect detection. At the time of writing we are embarking on a pilot version of the user testing at Swanshurst School in Birmingham.

References

[1] Barnden, J., Glasbey, S., Lee, M. & Wallington, A. 2004. Varieties and Directions of Inter-Domain Influence in Metaphor. *Metaphor and Symbol*, 19(1), 1-30.

[2] Fussell, S. & Moss, M. 1998. Figurative Language in Descriptions of Emotional States. In S. R. Fussell and R. J. Kreuz (Eds.), *Social and cognitive approaches to interpersonal communication*. Lawrence Erlbaum.

[3] Kövecses, Z. 1998. Are There Any Emotion-Specific Metaphors? In *Speaking of Emotions: Conceptualization and Expression.* Athanasiadou, A. and Tabakowska, E. (eds.), Berlin and New York: Mouton de Gruyter, 127-151.

[4] Mateas, M. 2002. Ph.D. Thesis. Interactive Drama, Art and Artificial Intelligence. School of Computer Science, Carnegie Mellon University.

[5] Ortony, A., Clore, G.L. & Collins, A. 1988. The cognitive structure of emotions. Cambridge U. Press.

[6] Prendinger, H. & Ishizuka, M. 2001. Simulating Affective Communication with Animated Agents. In *Proceedings of Eighth IFIP TC.13 Conference on Human-Computer Interaction*, Tokyo, Japan, 182-189.

[7] Sharoff, S. (Forthcoming) How to Handle Lexical Semantics in SFL: a Corpus Study of Purposes for Using Size Adjectives. In *Systemic Linguistics and Corpus*. London: Continuum.

[8] Zhe, X. & Boucouvalas, A. C. 2002. Text-to-Emotion Engine for Real Time Internet Communication. In *Proceedings of International Symposium on Communication Systems, Networks and DSPs*, Staffordshire University, UK, pp 164-168.

[9] Mehdi, E. J., Nico P., Julie D. & Bernard P. 2004. Modeling Character Emotion in an Interactive Virtual Environment. In *Proceedings of AISB 2004 Symposium: Motion, Emotion and Cognition*. Leeds, UK.

[10] Gratch, J. & Marsella, S. 2004. A Domain-Independent Framework for Modeling Emotion. In *Journal of Cognitive Systems Research*. Vol 5, Issue 4, 269-306.

Assessing Dynamic Aspects of Learner Motivation in Simulation/Gaming Based Foreign Language Learning Environment

Sunhee CHOI[1] and W. Lewis JOHNSON[2]
[1]*Rossier School of Education, University of Southern California*
Los Angeles, CA 90089, U.S.A

[2]*Center for Advanced Research in Technology for Education (CARTE)*
Information Sciences Institute, University of Southern California
4647 Admiralty Way, Marina del Rey, CA 90292, U.S.A

Abstract: The purpose of this paper is to report results of a pilot study which assessed dynamic aspects of learner motivation using self-report instruments in a simulation/gaming-based Arabic learning environment – Tactical Language Training System (TLTS). The results indicate that learners with positive attitude and interest in learning a second language performed well. The results also imply that the interactivity and entertaining elements of simulation/gaming could help learners develop personal interest in learning even the most difficult language in the world.

Introduction

This paper reports on a pilot assessment of dynamic aspects of learner motivation using self-report instruments in a simulation/gaming-based Arabic learning environment – the Tactical Language Training System (TLTS). The paper focuses on the interactions among various motivational variables and its implication for learner performance.

The Tactical Language Training System (TLTS) is a simulation/gaming-based foreign language learning environment which aims to help learners rapidly acquire mission-oriented communicative competence in spoken Arabic language [11]. In the TLTS, an intelligent agent tutors learners through lessons, assesses learner progress and provides tailored feedback and pedagogical assistance using learner modeling and speech recognition technology (Skill Builder) [12]. Learners then practice their target language skills in an interactive, task-oriented social simulation in which they speak and use culturally appropriate gestures with autonomous, animated characters (Mission Practice Environment).

1. Overview of the Study

Six participants were recruited from the 4th Psychological Operations Group at Fort Bragg. Only one participant (Subject 002) had prior training in Arabic. All participants interacted with the system for four days including testing, and they had complete control over what to study and how much time to spend on different lessons and different components of the system. At the end of each day learners were asked to report their interest levels and the amount of mental effort they spent to process the learning material. Field observation notes ware also taken to record learners' behaviors and comments. After the training, participants

were also required to take the post test which consisted of a vocabulary test, an Arabic understanding test, and an Arabic speaking test, graded on a four-point scale.

The study also used questionnaires to access learners' personal interests and traits related to L2 learning motivation. Using a survey questionnaire, learner perception data were collected to gain the information about the usefulness, helpfulness, and interestingness of the system. Along with the learner perception survey questionnaire, Gardner's Attitude/Motivation Test Battery [7] was adopted before and after learners' interaction with the system to measure learners' initial L2 motivation and attitudes toward Arabic language, and possible changes after using the system. Combining the results from these multiple measures, it was expected to be possible to answer the question "whether the psychological state of interest experienced as the result of situational factors is the same or even comparable to the psychological state that is the outcome of individual factors" [2] (p. 48).

2. Results and Discussion

There were a total of ten lessons in the Skill Builder, each of which deals with different discourse functions and cultural topics. Learners' achievements are higher with the first four lessons (M = 2.47) than with all lessons put together (M = 2.05). The low performance could be attributed to the fact that learners experienced technical problems on Day 2 when most learners were focusing on the advanced lessons. Notable is the Subject 003's performance. He outperformed all other learners except Subject 002 who learned Arabic language before. There should be a lot of factors, both cognitive and affective ones, involved in the Subject 003's performance. The system's instructional benefits were also evidenced by the Subject 001's remarks. He said that Arabic is a category 4 language (the most difficult one) and it takes 18 months to achieve intermediate proficiency. He initially thought that expecting soldiers to learn some Arabic in such a short time (4 days) was not realistic, but he was surprised and excited that he was able to retain what he has learned with the TLTS (although he retained less than the other subjects).

To measure more stable characteristics of learner interest and motivation, five questionnaires were used to identify 'Interest in L2', 'Attitude toward Learning L2', 'Integrative Orientation', 'Instrumental Orientation', 'L2 Anxiety' before the training, and 'Attitude toward Arabic Language', 'Motivational Intensity' after the training. These measures have been used by Second Language Acquisition (SLA) researchers for more than 20 years, and their validity and reliability have been empirically attested [7]. Spearman's Rank-Order Correlation Coefficient, a nonparametric data analysis, was conducted at a 0.05 significance level in order to examine relationships among trait variables and L2 performance variables.

For pre-treatment motivational variables, statistically significant correlations were found between interest in L2 and integrative orientation (r = 0.941, p = 0.005), interest in L2 and instrumental orientation (r = 0.824, p = 0.044), integrative orientation and instrumental orientation (r = 0.941, p = 0.005), attitude toward L2 and basic level performance (r = 0.928, p = 0.008), and attitude toward L2 and overall performance (r = 0.812, p = 0.050). There is a negative relationship between L2 anxiety and performance although the correlation coefficient between was not statistically significant (r = -0.174, p = 0.742). It can be inferred from these results that when learners have negative attitudes or low interest in learning L2, they do not perform well. Therefore, learners with low interest and motivation might need more emotional assistance such as encouragement and positive in-time feedback to compensate their lack of their interest in the task [3].

For post-treatment motivational variables, 'Motivational intensity', learners' future plans or intention to learn or use Arabic was highly correlated with both basic level (r =

0.986, p = 0.000) and overall performance (r = 0.899, p = 0.015). However, no significant correlation was found between attitude toward Arabic Language and performance, which is interesting because the opposite was observed between attitude toward L2 (not a specific language) and performance. It might be the case that learners' perception of Arabic Language as a most difficult language is deeply rooted in their minds, but the TLTS somewhat convinced learners to want to learn the language in the future even though they still believe learning Arabic is not too much fun or easy. The results also imply that the interactivity and entertaining elements of simulation/gaming could help learners develop personal interest in learning even the most difficult language in the world.

Learners' situational interest and state motivation were measured at the end of every day by asking participants to self-report in a questionnaire (6-point Likert scale). The situational interest measure included questions asking the interestingness, usefulness, and helpfulness of the system. 4 questions related to mental effort were also included. Mental effort is one of the behavioral indexes of motivation [13], and it is assumed that the higher the interest, the more mental effort is exerted by the learner.

Overall, learners perceived the system very positively. What is also interesting about these data is that Subject 006 consistently displayed lower levels of situational motivation/interest compared to other learners. The subject scored highest in the L2 anxiety measure, and his high anxiety might have affected his perceptions of learning Arabic negatively. He also performed low in the posttest. Diagnosing these types of learner characteristics ahead could help implement pedagogic strategies that accommodate individual differences affecting learner performance.

References

[1] Adelman, H. & Taylor, L. (1999). *New Directions in Enhancing Educational Results: Policymaker's Guide to Restructuring Student Support Resources to Address Barriers to Learning,* The Center for Mental Health in Schools, UCLA, Los Angeles.

[2] Ainley, M., & Hidi, S. (2002). Dynamic measures for studying interest and learning. In P. R. Pintrich & M. L. Maehr (Eds.), New Directions in Measures and Methods (Advances in Motivation and Achievement, vol. 12). New York: JAI.

[3] Burleson, W. and R. W. Picard (2004), "Affective Agents: Sustaining Motivation to Learn Through Failure and a State of Stuck" Social and Emotional Intelligence in Learning Environments Workshop In conjunction with the 7th Intl. Conf. on Intelligent Tutoring Systems, August 31st, 2004.

[4] Conati, C., & MacLare, H. (2004). Evaluating a Probabilistic Model of Student Affect. *Proceedings of the Intelligent Tutoring Systems Conference*

[5] de Vicente, A., & Pain, H. (2002). Information the detection of the students' motivation state: An empirical study. In t*he Proc. of the Sixth International Conference on Intelligent Tutoring Systems*, Berlin.

[6] del Soldato, T., & du Boulay, B. (1995). Implementation of motivational tactics in tutoring systems. *Journal of AI in Education*, 6(4), 337-378.

[7] Gardner, R.C**.** (1985). *Social psychology and second language learning: The role of attitudes and motivation.* London: Edward Arnold Publishers.

[8] Halleck, G. B, Moder, C. L., & Damron, R. (2002). Integrating a Conference Simulation into an ESL class. *Simulation & Gaming,* 33 (3), 330-344.

[9] Hidi, S. (2000). An interest researcher's perspective on the effects of extrinsic and intrinsic factors on development. In C. Sansone, & J. M. Harackiewicz, Intrinsic and extrinsic motivation. The secret for optimal motivation and performance (pp. 309-339). New York: Academic Press.

[10] Hidi, S., & Anderson, V. (1992). Situational interest and its impact on an reading and expository writing. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.*), The role of interest in learning and development* (pp. 215-238). Hillsdale, NJ: Erlbaum.

[11] Johnson, W. L., Marsella, S., & Vilhjalmsson, H. (2004). The DARWARS Tactical Language Training System. Proceedings of I/ITSEC 2004.

[12] Johnson, W.L., Wu, S., & Nouhi, Y. (2004). Socially intelligent pronunciation feedback for second language learning. ITS '04 Workshop on Social and Emotional Intelligence in Learning Environments.

[13] Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in Education: Theory, Research, and Applications (2nd Ed.)*. Columbus, OH: Merrill Prentice Hall.

Evaluating Motivation and Affective Aspects of Educational Software

Julie MARSHALL and Ray ADAMS

Middlesex University, School of Computing Science,
Trent Park, Bramley Road, London, N14 9YX

Abstract. A sample of seven users aged 8 and 11 in schools in north London, UK were asked to complete a set task called "snap" based upon the topic of sound. Descriptions of sounds (pitch/volume) were to be matched with examples of sounds made. The users completed this task on Espresso Educational Software which teaches and tests all children aged 11 or below the entire national curriculum. Individually, whilst being video recorded, the users have evaluated their own performance and their opinions of this software. The qualitative results show that the children who were excited had more fun playing the game. There was no correlation between users' system-evaluation and users' self-evaluation as six out of seven children performed well even when some thought the game was difficult. All users were motivated to use snap on Espresso because of its perception of being a fun game which was sometimes causing high affect in anticipation. All the children were new to this task yet enjoyed playing the game, were comfortable using a computer and said they enjoyed a challenge.

Introduction and Aim

The discovery of motivational and affective factors of educational software usage is important to this study because these factors effect how widely recommended or frequently software is used. As the perception of using a computer changes from being an exciting treat to a commodity with exciting software and computers are becoming a part of everyday life, we run the risk of becoming reliant upon machines in order to achieve our work. Lee, Luchini et al (2004) prove that education is more effective when learning is enjoyable, through traditional or technological methods. They state that they are "overcoming the notion that a "toy" does not belong in the school and developing software that has real educational value while stimulating the learner." [1]

Children are highly emotive but not always motivated with activities so the use of a computer will be enough of an inspiration to achieve the task set.

1 The Educational Software – Espresso

Espresso is an online educational digital library, is widely used across the London boroughs and is a popular choice with teachers. The children enjoy the fun interactive multimedia as a method of learning and the teachers are able to test the child's understanding of both their work and online content using Espresso's activity and feedback modules. These modules supply the teacher with statistics on which to gauge the child's progression.

2 Methods

Observations were completed using a video recorder and mini disk recorder. This method was chosen to capture the emotional effects of the educational software usage. A classroom assistant and videographer were also present. Expressions and reactions were recorded to determine when a child was motivated. Interviews were taken whereby users answered a series of questions, one interviewer to one child. This method was chosen to capture their rating of their performance and the software. This rating was compared with the observed behaviour to match results in each performance.

3 Results

Child numbers 10 and 11 were female twins aged 8 who were rivals as they were competing for a better score than the other. They both found the task tricky but child 10 said she did quite well whereas child 11 said she performed badly but scored well receiving 7 and 5 out of 8 respectively.

Child 5: 7/8, 1 smile
Child 6: 7/8, 5 smiles
Child 7: 6/8, 0 smiles
Child 8: 3/8, 5 smiles
Child 9: 8/8 0 smiles
Child 10: 7/8, 3 smiles
Child 11: 5/8, 6 smiles

The child with a happier disposition smiled more regardless of whether they performed well but they all enjoyed the game. The children who smiled least during the game were concentrating the most as they thought scoring was the most important factor. They were competing on points rather than enjoying the game.

Table 1 Qualitative Results

Child no#	Emotion/ personality	Motivation	Performance
5	Uncomfortable, not smiling, blank expression	Wanted to complete the task and leave quickly. Game was not fun, nor related to school work.	7/8 confident on PC, hard on herself as it was difficult
6	Confident, happy, excited, smiles with every question, fidgets	Excitable, unfocussed in attention, wanted to complete the task but took 3rd attempt to get it	7/8 confident and quick on PC, enjoyed the task, did alright as 3rd go
7	Fidgets, didn't smile, complacent	Interested in playing game, likes to do things that are fun, but not fun	6/8 found game easy and he did ok
8	Embarrassed, smiling, shy	Wanted to play the game and do well	3/8 thought it was easy but did badly
9	Comfortable, talkative, concentrated	Wide eyed during the game, exciting, easy, fun, useful, smiled when correct answer given	8/8 couldn't believe it, was scared at first, but surprised
10	Happy, smiling, shy, fidgeting, concentrated	Excited about the game, happy, big smiles, had fun, smiled when correct	7/8 on 2nd go, it was tricky but did quite good
11	Smiling except when needing instructions, shy	Excited about the game, happy to do it, smiling when right, disappointed when wrong, likes ICT	5/8 needed help, unsure, thought she did badly

3.1 Correlations in Results

All of the children detailed in table 1 were capable of using the PC efficiently; they were all motivated to play the game and all performed well except one. No children from the sample had played the game prior to the testing. The less comfortable children were unemotive, whereas the happier children were facially expressive and excited. There was no correlation between how easy/difficult the game was perceived with their score.

The excited children needed more than one chance to view the game before they understood and could play. The happier children were realistic with their performance whereas the less comfortable children were less happy with their performance.

4 Conclusion

Discoveries made were that children are not all as excitable as expected and that this frame of mind hindered their concentration on understanding the task as quickly as those less emotive but that their performance was generally the same regardless. Emotive, highly motivated children with no inhibition had more fun during this test and were not emotionally restricted by the pressure of performance in front of a video camera. Once relaxed and confident later in the interviewing stage, laughter was also enjoyed by the persons interviewing and working the video camera while the children were testing the software. Those children with expectation in their performance felt that they performed badly even when they performed just as well as all the other children so nervous anticipation does not help positive self-perception.

Questions that should be asked are for students' interests and motivations; body language should be read; ability of students should be gauged; the topic should be related to their interests, motivations and abilities. The experience of learning should be made to be fun, exciting and memorable. Pictures, sounds, video clips and stories should be used to inspire creative thoughts and capture those thoughts regardless of form.

References

[1] Jeremy Lee, Kathleen Luchini, Benjamin Michael, Cathie Norris, Elliot Soloway, April 2004, Conference on Human Factors in Computing Systems, CHI 2004, Vienna Austria, p1375-1378, 1-58113-703-6

[2] Anderson, J. R., Bothell, D., Lebiere, C. & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, 38, 341-380