The Impact of Animated Pedagogical Agents on Girls' and Boys' Emotions, Attitudes, Behaviors and Learning

Ivon Arroyo, Beverly P. Woolf, David Cooper
Department of Computer Science
University of Massachusetts Amherst
Amherst, MA, USA

<Ivon, Bev, DCooper>@cs.umass.edu,
Royer@psych.umass.edu, Minghui_Tai@yahoo.com

Winslow Burleson, Kasia Muldner
School of Computing, Informatics
and Decision Systems Engineering
Arizona State University
Tempe, Arizona, USA

Winslow.Burleson@asu.edu,
Katarzyna.Muldner@asu.edu

Abstract. We report on the reactions of males and female students to the presence of animated pedagogical agents that provided emotional and motivational support. One hundred high school students used agents embedded in an Intelligent Tutoring System for Mathematics and randomized controlled evaluations compared students with and without learning companions. The results indicate that affective pedagogical agents improve affective outcomes of students in general and particularly so for female students, who reported being more frustrated and less confident while solving math problems prior to using the tutoring system. We discuss issues of incorporating gender into user models and of generating responses tailored to gender.

Keywords: Intelligent Tutoring Systems; gender differences; pedagogical agents; motivation and affect; empirical evaluation; mathematics education.

I. INTRODUCTION

If tutoring systems are to interact naturally with students, they need to understand students' affect and to act upon students' affective states. Progress has been made regarding modeling students' emotions [1-4]. However, little research has explored how learning environments should respond to students' affect and how individual differences between students impact this process (for recent research see [4, 5]).

Animated pedagogical agents have the potential to help students not only learn but also to feel better about their learning experience, by engaging students through social interactions and tailoring of instructional content. However, pedagogical agents to date have mainly targeted cognitive rather than affective aspects. Some efforts have been made to create affective agents [6], but evaluation of their impact is still preliminary.

Here, we report on an evaluation of pedagogical agents (see Figure 1) in real school settings, with about 100 students from a public high school in Massachusetts. One of our main findings is that gender has a key impact within the context of tutoring systems for mathematics. The next section describes research on gender differences and our initial results, followed by a description of the intelligent mathematics tutor that was used as a testbed. Sections 4 and 5 describe the tutor's emotional support and the research study. Section 6 describes the benefits from this support and Section 7 provides a discussion of the value of including gender in a user model, especially when emotion of the user is involved.

II. WHO NEEDS AFFECTIVE SUPPORT WHILE LEARNING MATHEMATICS?

Past research suggests that female students have higher affective needs in certain disciplines; for example, in early adolescence gender and ethnic differences exist in mathematics self concept (a student's belief about their ability to learn mathematics) and mathematics utility (the student's belief that mathematics is valuable to learn [7,8]). Specifically, girls have less liking for math, more negative emotions and more self-derogating attributions about their math performance [7]. This poor affective relationship to the subject is likely one reason why females do not choose advanced math classes and later science careers in college [8], as compared to males who maintain a more positive relationship to the subject. Thus, helping girls in particular to foster a positive affective relationship to mathematics is highly relevant.

Our first goal was to test the hypothesis that girls feel worse about math for our student population, which included 108 students from High School 1 and High

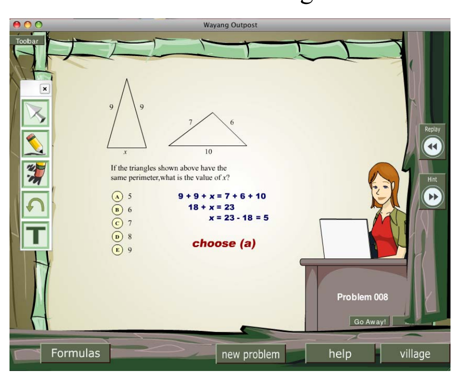

Figure 1. The Wayang Outpost Tutoring System

School 2, two public high schools in MA, with 50% females/males and 15% having some documented level of learning disability. We focused our analysis on the affective needs of female students by analyzing mean differences between them and male students over a battery of affective pretest questions given to students before tutoring. These questions covered general attitudes towards mathematics, such as likes/dislikes of the subject, the importance of math, and questions about how students feel (e.g., anxiety/ confidence, frustration, boredom, excitement) when they solve math problems.

In general, our results showed that students do not find

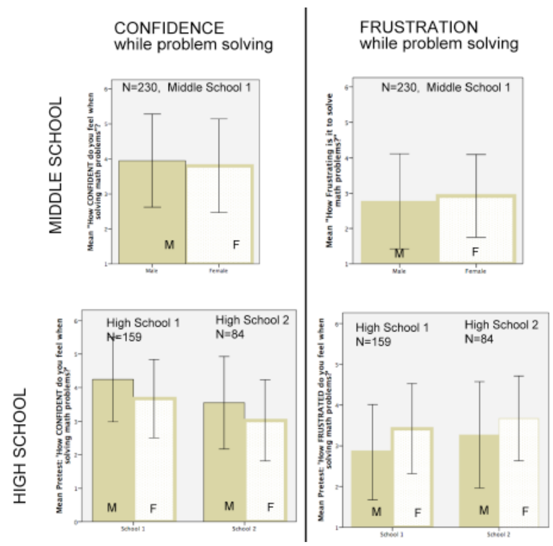

Figure 2. Results for a pre-tutor survey in two public schools: Girls develop negative feelings for mathematics, including *decreased confidence* (left) and *increased frustration* (right), in the transition from middle to high school.

math very interesting or exciting (they score lower than neutral in our 6-likert type scale assessments) [5]. Our results also suggested that both female students and low achieving students[1] have higher affective needs, as compared to other students [5]. Low achieving students disliked math more, valued it less, had worse perception of their math ability, and reported feeling worse when solving math problems than did high achieving students. The *math ability* of female students was similar to that of males according to our math pretest scores (Females M=.47 SD=.22; Males M=.50 SD=.19). However, female students reported feeling significantly *less confident* (*more anxious*) when solving math problems than male students (Females M=3.1 SD=1.4; Males M=3.8 SD=1.2; F(107,1)=7.8, p=.006) and more *frustrated* (Females: M=3.6 SD=1.31; Males: M=3.1 SD=1.06; F(108,1)=4.8, p=.03) when solving math problems. This was measured with surveys before using the tutoring system.

Low achieving females reported significantly worse feelings than low achieving males (e.g., for confidence, low achieving females M=2.7 SD=1.3; low achieving males M=3.4 SD=1.2, F(63,1) = 5.1, p=.03). In fact, high achieving females have very similar feelings to low achieving males (e.g., for confidence, high achieving females M=3.6, SD=1.1) while high achieving males feel better than high achieving females (for confidence, high achieving males M=4.4, SD=1.4). The range of values for these emotions was 1 (I feel anxious) to 6 (I feel very confident).

The above analysis confirms that our population for this study has the documented effect of females feeling worse about mathematics. In fact, Figure 2 further shows this same effect for a much larger population than the one covered for the data reported here, obtained for a second study across two different high schools (most subjects who were part of this study came from High School 2 and some from High School 1). It also shows that these negative feelings are not yet developed in 230 middle school students from School 1 (which contained both a middle and high school) just as the literature suggests. It seems that scaffolding female students' affect is an important target by itself: not only is female perception of mathematics worse than males but it is also unjustified, as it doesn't match their true mathematical abilities.

III. THE TESTBED TUTORING SYSTEM: WAYANG OUTPOST

The test bed for our research is Wayang Outpost, an intelligent tutor that helps students solve geometry problems, of the type that commonly appear on standardized tests [9]. To solve problems in the Wayang interface, students choose a solution from a list of multiple-choice options (typically four or five, see Figure 1). Wayang provides immediate feedback on students' entries by coloring them red or green in the interface. As students solve a problem, they can ask the tutor for "hints" that are displayed in a progression from a general suggestion to a bottom-out solution. In addition to this domain-based help, the tutor provides a wide range of affective support, see Table 1, delivered by its learning companions (LC): agents designed to act like peers who care about a student's progress and offer support and advice. Figure 3 shows the female and male version of the learning companion (both genders are available as we are interested in how LC gender influences outcomes).

The tutor has several user models, including one for the student's *cognitive* skills and a second for the student's *effort* and *affect*. A simple *effort* model is used to assess the degree of effort a student invests to develop a problem solution, mainly based on time per action. A linear regression *affect* model is used to assess a student's

[1] Low achieving students correspond to half of the students, determined by a median-split of math pre test scores

Figure 3. Jane, the female affective learning companion, and Jake, the male affective learning companion.

emotional state; this model is derived from data obtained from a series of studies [1, 4].

IV. PROVIDING AFFECTIVE SUPPORT THROUGH SPOKEN MESSAGES

Wayang's learning companions deliver approximately 50 different messages intended to motivate students, as well as provide meta-cognitive help for effective problem-solving strategies and effective use of Wayang's tools. Ultimately, the interventions will be tailored to Wayang's affective student model. Since we are currently validating the models and algorithms for deciding which intervention to provide when, for this study, the messages were based on the effort model only and emphasized the importance of effort and perseverance and the idea that intelligence is malleable instead of a fixed trait [10]. The characters provided this support by responding to the effort exerted by students rather than to the student's emotions. Characters were either unimpressed when effort was not exerted, or simply ignored the student's solution. They also offered praise to students who exerted effort while solving a problem, even if their answers were wrong, lessening the importance of performance in favor of learning.

The next three types of interventions described are verbal messages tailored according to Wayang's modeling of students' effort.

Attribution Interventions. Attribution theory proposes that students' motivation to learn is directly rooted in their beliefs of why they succeed or fail at tasks [16]. If students can be taught to alter certain beliefs, for instance to understand that failure is the result of a lack of effort instead of a lack of ability, then their motivation to learn and learning outcomes can be significantly improved [16]. We embedded several types of attribution training interventions into Wayang; these are generated when students face a new problem (see Table 1).

Effort-Affirmation Interventions. In contrast to the effort-attribution messages described above, which aim to change students' attitude towards effort during problem solving and are generated before the student actually starts problem solving, the *effort-affirmation* interventions acknowledge effort after students obtain a correct solution (see Table 1 for examples). The intention underlying the design of these messages is to both build a more realistic social bond between the companion and the student and to motivate the student. These interventions include:

Strategic Interventions. These messages target meta-cognitive strategies, with the goal of making students effective problem solvers and as a result more motivated for learning in general (see Table 1 for examples).

Type	Sample message
Attribution (General)	I found out that people have *myths* about math, like, they think that only some people are good in math. The truth is that we can all be good in math if we try.
Attribution (Effort)	Keep in mind that when we are struggling with a new skill we are learning and becoming smarter!
Attribution (No Effort)	We will learn new skills only if we are persistent. If we are very stuck, let's call the teacher, or ask for a hint from Wayang!
Attribution (Incorrect)	When we realize we don't know why the answer was wrong, it helps us understand better what we need to practice.
Effort Affirmation (Correct No-effort)	That was too easy for you. Let's hope the next one is more challenging so that we can learn something.
Effort Affirmation (Correct Effort)	Good job! See how taking your time to work through these questions can make you get the right answer?
Strategic (Incorrect)	Are we using a correct strategy to solve this? What are the different steps we have to carry out to solve this one?
Strategic (Correct)	We are making progress. Can you think of what we have learned in the last 5 problems?

Table 1. Sample behavior-based messages that characters speak to students

V. THE USER STUDY

The user study was designed to quantitatively analyze the benefit of learning companions on affective and cognitive outcomes. The subjects included 108 students from two high schools 1 (one low and the other high achieving) in the state of Massachusetts and involved 9th and 10th graders. Two-thirds of the students were assigned to a learning companion of a random gender, and one-third to the no learning companion condition. Students took a mathematics pretest before starting, and completed a survey that assessed their general attitude towards mathematics.[2] Four questions asked about student feelings towards problem solving before they began to

[2] The pre-test included 3 items for self-concept in math ability, e.g., students compared themselves to other students in their math ability and compared mathematics to other subjects; 3 items to address subjective mathematics liking/value). In addition, questions were asked about feelings about math problem solving in particular (4 distinct emotions).

work with the tutor, including interest/boredom, frustration, confidence/anxiety, excitement (e.g., how frustrated do you become when solving math problems).

For the next three days, students used Wayang instead of their regular mathematics activities. Approximately every five minutes, students were asked to provide information on one of the four target emotions (e.g., how frustrated do you feel?). The Wayang learning companions offered help and spoke the messages as described in the previous section, occasionally at the beginning of a new problem or after a correct or incorrect attempt to solve the problem. Several student behaviors were logged, e.g., success at problem solving, use of tools and help, students' self-report of their emotions as well as other behaviors, e.g., muting the characters (using a mute button), and help abuse and/or quick-guessing. After students used Wayang for three days, they took a mathematics post-test, and answered the same questionnaire they had received prior to using the tutor. In addition, the post-survey included five questions about the student's perceptions of the Wayang tutoring system (*Did you learn? Like it? Think it was helpful? friendly? Was it concerned about your learning?*).

VI. RESULTS: WHO BENEFITS FROM AFFECTIVE FEEDBACK?

We carried out Analyses of Covariance (ANCOVA) for each of the affective and behavioral dependent variables (post-tutor and within tutor) shown in Table 2. Covariates consisted of the corresponding pretest baseline variable (e.g., when analyzing within or posttest report of confidence towards problem-solving, we accounted for the pretest baseline confidence). Independent variables corresponded to *condition* (either LC [Present/Absent], or Group [Jane/Jake/no-LC]). We analyzed both main effects and interactions for gender and condition over all student data (see last column of Table 2). In addition, because of the special affective needs of female students, we repeated the ANCOVAs for the female population only. Results in Table 2 with asterisks indicate a significant between-subjects effect, as revealed by the Fisher test for between-subjects comparisons.

Results suggest a general advantage of learning companions for affective outcomes, and no significant benefit of learning companions for cognitive outcomes (all students improved after using the software, regardless of condition). Table 2 shows that students reported significantly less frustration and more interest (less boredom) when learning companions were present compared to absent. At the same time, students receiving the female character reported significantly higher self-concept and liking of math. Given the short exposure of students to the software (4 sessions) we are impressed with the results.

Table 2. Results for Presence of Learning Companions

Variable	Between-Subjects Effects
Learning Improvement in Math Test (Posttest-Pretest)	Overall (Pre to Post)** LC Presence LC for Female Subjects Females vs. Males
Unproductive Behaviors (Help Abuse, Quick-guessing)	LC Presence LC for Female Subjects** Females vs. Males***
Productive Behavior (Spending time on Help)	LC Presence LC for Female Subjects Females vs. Males+
Post-Tutor Perceptions of the Software and Learning Experience	LC Presence LC for Female Subjects** Females vs. Males***
Enjoy/NotHaveFun When Solving Problems Emotion	LC Presence LC for Female Subjects+ Females vs. Males*
Frustrated/Not Frustrated When Solving Problems Emotion	LC Presence*** LC for Female Subjects*** Females vs. Males
Confident/Anxious When Solving Problems Emotion	LC Presence* LC for Female Subjects** Females vs. Males
Interest/Bored When Solving Problems Emotion	LC Presence+ LC for Female Subjects Females vs. Males
Mathematics Liking Attitude	LC Presence* LC for Female Subjects Females vs. Males
Mathematics Self-Efficacy Attitude	LC Presence* LC for Female Subjects Females vs. Males

*** Significant at p<.005 ** Significant at p<.01
* Significant at p<.05 + Near significant at p<.1

Key: *Overall* means pair-wise comparisons effects for Pre to Post improvement regardless of condition from pre to posttest; *LC Presence* means an effect for the presence of learning companion; *Females vs. Males* indicates an interaction effect that indicate a gender difference (females and males benefitted differently) from the presence/absence of learning companions; *LC for Female Subjects* means an effect for the presence of learning companion over the subset of female subjects only.

While learning companions afford affective advantages for all students (e.g. see *LCPresence* for "frustrated"), several significant effects in ANCOVAs indicated a higher benefit of learning companions for female students. In general, the effects are stronger for females than for males. Females' confidence is improved with learning companions but this is not the case for males. It is important to note that these gender effects on emotions (within or after the tutor) are not due to females starting out feeling worse, as our analyses account for that baseline pretest emotion as a covariate.

Girls perceived the learning experience with Wayang significantly better when learning companions were present, while the opposite was true for males, who reported better perceptions of learning with Wayang when learning companions were absent. Female students also had more productive behaviors in the tutor when the companions were present than when they were absent: they spent more time than males on problems where help was seen. They also had less unproductive behaviors: they "quick-guessed" and clicked fast through hints less when characters were present. At the same time, a significant interaction effect for LC Presence and Gender revealed that the opposite is true for males: they have less unproductive behaviors when LCs are absent.

VII. DISCUSSION

We have provided encouraging results for the general benefits of Affective Digital Pedagogical Characters within tutoring software, indicating a promising area of further research. However, while the benefits of these characters for female students are out of the question, our research has shown mixed results for the impact of affective characters on male students. The results presented here suggest that girls and boys might need to be considered separately, as what works for girls does not necessarily work for boys. The importance of including information on gender in a user model is not a mere hypothesis, but is based on extensive research on gender differences and learning [12, 13]. Some research suggests that girls and boys have different approaches to problem solving [14] and that they should be taught differently [13]. Moreover, boys and girls' learning styles are different, with girls tending to ask for help and boys using the teacher as a last resort. Some of that research has indicated the two genders respond to different motivational techniques: boys respond better to time-constrained tasks and pressure situations than girls. While this literature explores gender differences during standard classroom activities, we have found empirical evidence over the years that gender differences also exist when males and females use tutoring systems at the K-12 level. For instance, females are more "diligent" when using tutoring systems, showing behaviors that are more conducive to learning than those of male students (e.g., spending time on hints or accepting help when offered). Also, females report better general attitudes while learning with our tutoring systems [15], even without the characters.

This paper provides further evidence that the impact of affective support is gender dependent. Males in general did better without the affective and motivational feedback from learning companions. However, it is also true that other analyses beyond the scope of this paper [5] indicate that low-achieving students (half of which were males) did benefit from learning companions to a large extent, with mixed results for high achieving students. This makes us believe that it is mostly high achieving males that did not benefit from them. We expect to conduct further studies with larger number of students in the future, in order to provide nuanced recommendations about affective feedback in tutoring software, for a combination of gender and ability student profiles.

ACKNOWLEDGMENTS

This research was funded by two NSF grants: (1) #0734060, HRD GSE/RES, *What kind of Math Software works for Girls? The effectiveness of motivational and cognitive interventions,* Arroyo, I. (PI) with Royer, J.M. and Woolf, B.P. and (2) #0705554, IIS/HCC *Affective Learning Companions: Modeling and supporting emotion during teaching*, Woolf, Burleson (PIs) with Arroyo, Barto, and Fisher. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] Arroyo, I., D. Cooper, W. Burleson, B. Woolf, K. Muldner, and R. Christopherson. *Emotion Sensors go to School*. Proc. of *Conference on Artificial Intelligence in Education (ITS'09)*. 2009. 17-24.

[2] Conati, C. and H. Maclaren. *Evaluating A Probabilistic Model of Student Affect*. In Proc. of *Intelligent Tutoring Systems*. 2004. 55-66.

[3] D'Mello, S., R. Picard, and A. Graesser, *Towards an Affect Sensitive Auto Tutor*. IEEE Intelligent Systems, 2007. 22(4): p. 53-61.

[4] Cooper, D., I. Arroyo, B. Woolf, K. Muldner, W. Burleson, and R. Christopherson. *Sensors Model Student Self Concept in the Classroom*. In Proc. of *International Conference on User Modeling, Adaptation and Personalization (UMAP'09)*. 2009. 30-41

[5] Woolf, B.P., I. Arroyo, K. Muldner, W. Burleson, D. Cooper, R. Dolan, and R. Christopherson. *The Effect of Motivational Learning Companions on Low Achieving Students and Students with Disabilities*. In Proc. of *Intelligent Tutoring Systems*. 2010. 327-337.

[6] Graesser, A., P. Chipman, B. King, B. McDaniel, and S. D'Mello. *Emotions and Learning with AutoTutor*. In Proc. of *Artificial Intelligence in Education*. 2007. 569-571.

[7] Eccles, J.S., A. Wigfield, R.D. Harold, and P. Blumenfeld, *Age and gender differences in children's self and task perceptions during elementary school.* Child development, 1993. 64: p. 830-847.

[8] Catsambis, S., *The gender gap in mathematics: Merely a step function?*, in *Gender differences in mathematics* 2005. p. 220-245). .

[9] Arroyo, I., C.R. Beal, T. Murray, R. Walles, and B.P. Woolf. *Web-Based Intelligent Multimedia Tutoring for High Stakes Achievement Tests*. In Proc. of *International Conference on Intelligent Tutoring Systems*. 2004. 468-477.

[10] Dweck, C.S., *Motivational processes affecting learning.* American Psychologist, 1986. 41: p. 1040-1048.

[11] Shute, V.J., *SMART: Student Modeling Approach for Responsive Tutoring.* User modeling -user-adapted instruction, 1995. 5: p. 1-44.

[12] Beal, C.R., *Boys and girls: The development of gender roles.* 1994, New York: McGraw-Hill.

[13] Sax, L., *Why Gender Matters: What Parents and Teachers Need to Know about the Emerging Science of Sex Differences*. 2005.

[14] Fennema, E., T. Carpenter, V. Jacobs, M. Franke, and L. Levi, *A Longitudinal Study of Gender Differences in Young Children's Mathematical Thinking.* Educational Researcher, 1998. 27(5): p. 6-11.

[15] Arroyo, I. and B. Woolf. *Inferring learning and attitudes from a Bayesian network of log file data*. In Proc. of *Conference on Artificial Intelligence in Education*. 2005. 33-40.

[16] Weiner, B. (1972) *Attribution Theory, Achievement Motivation, and the Educational Process.* Review of Educational Research, 42, 2, 203-215