Evidence for Gender Specific Approaches to the Development of Emotionally Intelligent Learning Companions

Winslow Burleson
Computer Science and Engineering/
Arts, Media and Engineering
Arizona State University
Winslow.Burleson@asu.edu

Rosalind Picard
MIT Media Lab
Picard@media.mit.edu

Abstract:

A 2 x 2 experiment investigated the effect of elements of an affective learning companion's emotional intelligence on seventy-six 11-13 year-old participants during a challenging problem solving activity. The experiment contrasted use of an agent showing sensor-driven non-verbal mirroring with one showing pre-recorded non-verbal interactions and, separately, affective support vs. task support interventions. The effect of emotional intelligence, in terms of the presence of active listening, delivery of appropriate interventions, and type of non-verbal interactions on participant's experience, including: frustration, perseverance, intrinsic motivation, and meta affective skill were examined. Hypothesized effects of interacting with a more vs. less emotionally intelligent agent did not hold true at the group level, however significant gender differences were found. Discussing these, this paper contributes new evidence on the importance of appropriately coordinating the relationships between affect and task based intervention and non-verbal mirroring with respect to the affective state of girls and boys.

Introduction:

Social bond between teachers and learners and affective support has been shown to have considerable impact on learners' performance and motivation. Wentzel has shown that caring bonds between middle school children and their teachers are predictive of learners' performance [1]. Lester has shown that intelligent tutoring systems that employ agents elicit a social presence or "persona effect" that increases learners' engagement [2]. Beyond the persona effect, Bailenson and Yee have shown that non-verbal mirroring in the form of behavioral mimicry can increase the likeability and persuasive effect of a virtual agent [3]. Bickmore and Picard have developed interaction and evaluation strategies to increase empathetic and caring relationships between agents and participants [4]. Providing participants a choice, in terms of the ethnicity and gender of an agent-tutor, has also been shown to have beneficial impacts on learners' impressions of the agent and on their own performance; similarly, matching learners' gender and ethnicity also leads to more positive impressions and performance [5].

One of the ways to develop a social bond with learners is to provide assistance. Systems that provide affective support at times of user frustration have been shown to reduce frustration [6]. Emotional support is an important factor in learning activities. In fact, in a study of expert human tutors' interactions with their students it was found that up to half of these interactions are focused on support of the learner's affective state [7], yet most intelligent tutoring systems provide predominantly task based support.

Dweck has shown that supporting learners by encouraging them to "think of the mind as being like a muscle and believe that they can increase their intelligence through effort, even when experiencing frustration" helps learners in their approach to and perseverance in challenging learning activities [8]. Her message supports learners' development of "meta-affective skill" –the ability to coordinate meta-affective knowledge (knowing a strategy based on affect, such as "when you feel frustrated it helps to think of the mind as a muscle"…) with meta-affective

experience (a conscious reflection on what an emotion, such as frustration, is doing to you, or may do to you).

Inspired by these findings, and many others that point to the importance of supporting the emotions of people, we undertook the design of an automated companion that could sense and respond to certain aspects of human emotion in a learning context. Because no automated system today can reliably sense all the emotions that occur during learning, and no system is smart enough to know how to respond appropriately all the time to the affective information that is sensed, this undertaking is an extremely ambitious one, and illuminates the challenges in creating successful versions of such future technologies. The research described in this paper implements just a few of the multitudinous possibilities for intelligently sensing and responding to learner affect, but provides the first experiment that we know of to implement real-time character responses to affective cues based on theory of how to support learners. The rest of this paper describes the experiment, main hypotheses and findings, together with discussion and recommendations about future experiments in this area.

Experiment:

A multi-modal real-time affective agent research platform [9] that incorporates a facial expression camera, pressure mouse, skin conductance sensor, and posture chair to engage in sensor driven non-verbal mirroring was built and used to begin to develop elements of an affective learning companion's emotional intelligence (Figure 1). This system collects data from the sensors that relates to the users' affective states. The data is both processed off-line with a classifier to determine affective state [9] and processed in real-time via a system server to influence the character's interactions with the user. The system server coordinates the user interface, activity, behavior engine and character interactions. The behavior engine processes the real-time data from the sensors to determine non-verbal interactions that are in turn displayed by the character engine. The character's behaviors include speaking, nodding, smiling or fidgeting the mouth, shifting its posture forward or backward, changing its color and fidgeting very slightly. These are the main behaviors controlled in this experiment, even though the character is capable of much more (e.g., turning its head to watch the actions of the learner, or walking around.)

An experiment was conducted with seventy six 11-13 year old girls and boys, who interacted with the agent and sensing system in the context of a challenging problem solving activity, the Towers of Hanoi activity with 7 disks [9]. The character followed one of two strategies for its non-verbal movements: (1) *Sensor driven non-verbal mirroring*, in which the four sensors were used to create a 4-second delayed behavioral mimicry of elements of facial expression, agitated swaying proportional to mouse pressure, reddening skin tone relative to skin conductance values, and leaning forward/back posture mirroring that of the learner; or (2) *Pre-recorded interactions* generated from the recorded files of the "most average" pilot participant interactions (determined using the standard deviation of each behavioral channel to categorize five naturally-occurring pilot files as being most average across the behavioral channels; for each new participant one of these five naturally-occurring files was randomly selected and used to present a pre-recorded control condition that exhibited a similar range of non-verbal behavioral expressions to those receiving sensor-driven non-verbal interactions.) Thus, both cases involved non-verbal movements by the character, but in only the first case were these synchronized to the learner's current sensor outputs. In a series of pilot studies participants were found to be unaware of the agent's 4-second delayed mirroring. In addition to non-verbal interactions, the character also practiced one of two interventions: (1) *Affective support intervention* included adaptive *"active listening"* strategies [6] and support of meta-affective skill based on Dweck's message ("the mind is like a muscle and you can increase your intelligence, through mental exercise") or (2) *Task*

support intervention ("another way to think about this is to think about the small disks that are in the way. If you move these out of the way then you can move the disk that you want to move"). Both affective and task based interventions concluded with similar phrases, encouraging the learner to continue with the activity. The overall 2x2 design is summarized in Table 1.

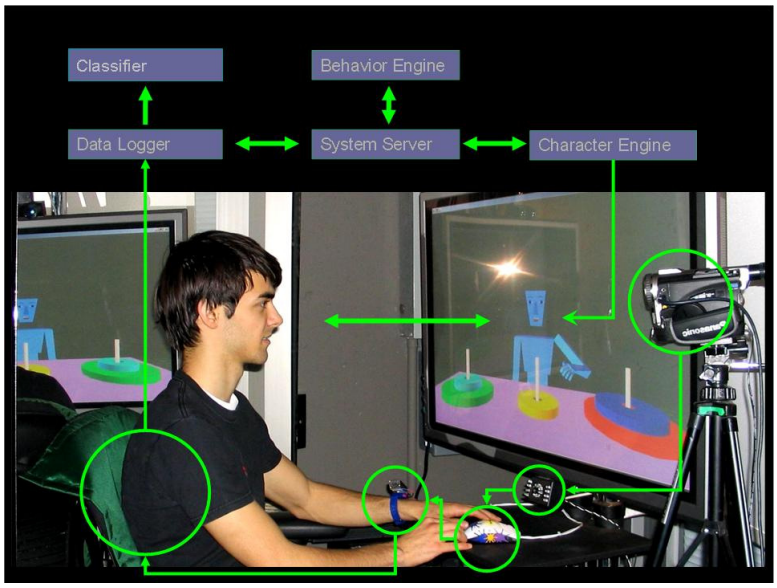

Figure 1. Affective Agent Research Platform with sensors listed from right to left: off-line video camera, facial expression camera, pressure mouse, skin conductance sensor, and posture chair.

	Sensor-driven non-verbal mirroring	*Prerecorded non-verbal interaction*
Affect support intervention	***Affect support, non-verbal mirroring*** 16 valid out of 20 assigned 8 girls valid out of 10 8 boys valid out of 10	***Affect support, prerecorded non-verbal interaction*** 14 valid out of 19 assigned 5 girls valid out of 8 9 boys valid out of 11
Task support intervention	***Task support, non-verbal mirroring*** 15 valid out of 18 assigned 7 girls valid out of 8 8 boys valid out of 10	***Task support, prerecorded non-verbal interaction*** 16 valid out of 19 assigned 9 valid out of 11 7 boys valid out of 8

Table 1. The 2 x 2 design contrasting *intervention* x *mirroring* conditions. Cells depict the number of valid participants included in the analysis of the hypotheses (See Participants section).

In this experiment, the affective learning companion was coded as being more emotionally intelligent when it engaged in *sensor driven non-verbal mirroring* and likewise when it provided *affective support interventions* than when it provided neither. These interventions and the mirroring condition were considered additive; an agent that provided both mirroring and affective support would be considered more emotionally intelligent than one that provided either separately, or neither.

Hypothesis

There were four specific areas to which this work had planned contributions:

First, it sought to extend Bailenson's use of Transformed Social Interactions [3], where he showed that when a participant wore an immersive head-mounted display that sensed head motions, and a virtual agent mimicked the head motions, it made the agent more likable and persuasive. Bailenson's findings were with college-aged participants and the persuasive message concerned security card usage. The extension of Bailenson's approach in this experiment includes four significant components: providing a new domain (a learning platform), including a different kind of persuasive message (meta affective skill based), addressing a new audience (11-13 year old learners), and using a new set of less invasive sensors to extend the mirroring beyond head movements.

- *H1: the affective learning companion is expected to be more persuasive (as measured by self report during introduction, perseverance measures, and intrinsic motivation measures), and users will form a stronger social bond (as measured by bye.button response and positive/negative impressions assessed with the Modified Working Alliance Inventory) with the affective learning companion, when sensor-driven non-verbal mirroring informs the affective learning companion's interactions than when pre-recorded non-verbal interactions are displayed.*
  * See the Methodology section and Table 3 for further explanation of the H1-H4 measures.

Second, this research sought to create new applications of Dweck's strategies of intervention [8] that facilitate learners' metacognitive strategy and meta affective skill. Additionally, the pedagogical benefits, of increased social bond [1] and persuasion [3], due to the approach taken in H1 was expected to leverage Dweck's message.

- *H2.A: A learner's social bond (as measured by bye.button response and positive/negative impressions assessed with the Modified Working Alliance Inventory) with an affective learning companion will positively correlate with his or her perseverance (time from character's departure until participant clicks on a quit button or until time limit) and self-theories – adoption of internal beliefs that he or she can increase his or her own intelligence and the adoption of mastery orientation (as measured by Dweck's Self Theories of Intelligence and Goal Master Orientation survey instruments).*

- *H2.B: The level of persuasion (as measured by self report during introduction, perseverance measures, and intrinsic motivation measures) a learner experiences from the affective learning companion's metacognitive message (presented during the introduction) will positively correlate with the social bond (as measured by bye.button response and positive/negative impressions assessed with the Modified Working Alliance Inventory), with perseverance (time from character's departure until participant clicks on a quit button or until time limit), and will negatively correlate with frustration (self-reported at the time of intervention and in the post-activity survey).*

Third, this research sought to design interventions that would increase intrinsic motivation and reduce frustration by taking into account strategies for empathetic and caring relationship development [4] and "frustration handling" [6].

- *H3: An affective learning companion that exhibits emotional intelligence (active listening provided during the affect support intervention, appropriate interventions -- see Table 5 and related discussion of the congruence measure, and sensor-driven non-verbal mirroring rather than pre-recorded non-verbal interactions) will increase learner's intrinsic-motivation (as measured by voluntary re-engagement with the activity) and reduce frustration (self-reported at the time of intervention and in the post-activity survey).*

Fourth, the research intended to evaluate the impact of this system on learners' meta affective skill development. Meta affective skill addresses a learner's awareness of feelings during an activity. An affectively aware Learning Companion might facilitate a learner's awareness of their feelings.

- *H4: Metacognitive skill will be exhibited at higher levels when learners interact with emotionally intelligent agents (see H3) and will positively correlate with perseverance (time from character's departure until participant clicks on a quit button or until time limit), willingness to continue (as measured by self-report at the time of intervention), and intrinsic motivation (as measured by reengaging in the activity during the final 2 minutes of the protocol).*

Methodology

The methodology (Table 2) included a pre-test, administered to determine children's self theories of intelligence and their goal mastery orientation [8]. The learning companion presented itself saying "Hi there. My name is Casey. I'm a digital character..."; its introduction was the same (other than non-verbal interactions, determined by the *mirroring* vs. *pre-recorded* condition) across the *affective support* vs. *task support* conditions. The character engaged in either *non-verbal mirroring* or *prerecorded non-verbal interactions* throughout the time of its presence. The learning companion presented a slide show, during which it asked the learner several questions. The slide show was based on a script used by Dweck that has been shown to shift children's beliefs about their own intelligence toward incremental self theories [8]. The learning companion then presented the Towers of Hanoi activity and explained that it may have to leave before the learner completes the activity. The companion instructs the learner to, "Click on a disk to start whenever you want, I'll just watch and help if I can." The learner is given four minutes to engage with the activity before the character intervenes with either an *affect support* or *task support* based intervention. (see [9], for the exact intervention dialogues). During the intervention, self-report measures are obtained through the interaction with the companion when it asks face to face questions of the learner, e.g. "On a scale from 1 to 7, how frustrated are you feeling right now?" Then the character says that it will need to leave and tells them, "I have to go now. Thank you for letting me watch you do this activity. Watching you has helped me learn too. Sorry that I have to leave now." Then the companion encourages them to continue. Finally it says, "If you feel like you would like to stop there will be a few buttons in the upper right hand corner that you can press. Bye bye." Participants have the opportunity to respond by pressing one of three *bye.buttons*: "Ok, bye", "Ok, bye I was glad to have you here", or "Ok glad you are finally going"; presented in different random order for each participant, to control for presentation order effects. After they select one of the three bye-responses, or after 20 seconds elapses (when the bye-response choices disappear/time out so as not to distract the participant), the character

disappears and three quit buttons that the character previously discussed appear in the upper right corner of the screen offering the opportunity for the learner to end the activity. The three buttons appear with the labels: "I want to stop because I'm too frustrated to continue", "I've put in all the effort that I can and want to stop", and "I want to stop for some other reason".

Protocol Events for Subjects in all Four Conditions	Duration in minutes
Assent and consent forms	~3
Initial Survey Questions and Pre-Test (including Self Theories of Intelligence and Goal Mastery Orientation)	~10
Character arrives, introduces itself, the activity, and shows a Slide Show (based on Dweck's message [8])	~10
Participant engages in Towers of Hanoi activity	4
Character provides *affect support* or *task support intervention* Obtains self-report measures Introduces quit buttons Then says goodbye (offers *bye.button* response)	~2
Participants persist in Towers of Hanoi task with three "quit" buttons present	up to 15 minutes from the start of the activity
Post-activity survey of experience	~3
Neutral affect inducement video	1.5
Post-Test (including Self Theories of Intelligence and Goal Mastery Orientation)	~10
Modified Working Alliance Inventory	~2
Opportunity to reengage with Towers of Hanoi	2

Table 2. Experiment protocol with durations in minutes; the approximate values indicate that these events have participant interactions and therefore some variation in duration

At the time the learner clicks one of the three quit buttons, or 15 minutes after the start of the activity, whichever happens first, the learner is presented with post-activity questions about the experience e.g. "How many minutes would you say this activity took from the time you first moved a disk until now?", "Mark how much of the time you were frustrated", and others, (see [9] for complete list). After these questions the learner is presented with a 1.5 minute video clip of a seascape, as a neutral affect inducement [10] to help alleviate frustration that may bias answers to subsequent questions. The learner is then presented with post-test questions on self theories of intelligence and goal mastery orientation, followed by a modified Working Alliance Inventory [4] to gauge his or her impression of the character. Finally the learner is again presented with the Towers of Hanoi activity, still in its previous state, along with instructions indicating that, "It will be a couple of minutes before the next activity is ready. You can do whatever you want now, just stay seated here please." After two minutes another message appears and says: "Thanks for waiting." The experimenter informs them they are done, and conducts a debriefing. This final two minute period allows a learner to reengage in the activity, as an indication of intrinsic motivation.

Measures

The measures used in this study use self-report surveys, learner's responses during dialogue with the learning companion, and learner's behavioral activities e.g. duration of engagement and re-engagement in the activity. An abbreviated set of the measures is presented in Table 3 (see Burleson 2006 for a complete list and discussion of their precise implementations). The learners' *positive/negative impressions* of the character were obtained using a *Modified Working Alliance Inventory* self-report survey that was based on Bickmore's research on users' social bond with agents [4]. The *self theories of intelligences* and *goal mastery orientation* instruments were

developed by Dweck [8] and have been used extensively in her research on learners' approach to and perseverance in challenging learning activities. The *Flow/Stuck* measure was a composite developed from a self-report survey based on the theory of Stuck, a state of non-optimal experience and Flow, a state of optimal experience [9]. Frustration was measured through dialogue based self-report at the time of intervention and in a post-activity survey. We did not employ specific learning measures in this study; instead we focus on affect and affective learning (use of affective strategies and related behaviors). While the Tower of Hanoi is so well studied that at times it is considered a "toy problem" with respect to traditional learning measures, the focus of this research on affective learning and strategies during frustration deals with the Towers of Hanoi activity as a very real (and frustrating) experience for learners who have not encountered it before. Based on our findings in this study we now have plans to conduct studies of affective learning companions in conjunction with intelligent tutoring systems that have explicit learning measures in Science, Technology, Engineering, and Math (STEM) topics.

Measures	**Method**
Self report measures during introduction: *Persuasion* Self report measures at time of intervention: *frustration* *effort* *willing to stick with it* *able to use the strategies (presented in the intervention, affective or task strategies)*	Self reported as part of dialogue with the character (Character asks a question, student selects a dialogue response or Likert response that indicates their answer)
Perseverance	Measure of time from character's departure until participant clicks on a quit button or until time limit
Social bond measures: *bye.button response* *positive/negative impressions*	Measured by the pressing of one of three bye buttons Modified Working Alliance Inventory (bond dimension)
Post-activity frustration	Composite scale from post-activity self report survey
Meta-affective skill	Composite scale from post-activity self report survey
More Flow/less Stuck	Composite scale from post-activity self report survey Flow is a state of optimal experience Stuck is a state of non-optimal experience
Dweck pre/post test measures [8] *goal mastery orientation* *self theories of intelligence*	Composite scale from pre/post-activity self report survey
Intrinsic-motivation	Measure of whether or not learners reengaged in the activity during the final 2 minutes of the protocol

Table 3. An abbreviated list of the measures used in this experiment (see [9] for comprehensive list)

Participants

The participants were seventy-six 11-13 year old children from three semi-rural schools in western Massachusetts, randomly assigned to the conditions as shown in Table 1. Attrition eliminated 10 students due to a variety of factors: some participants needed to leave prior to completing the activity, due to unexpected all-school meetings or changes in transportation schedules; power failures due to storms and some equipment failures occurred; a few participants were unresponsive to the character interactions (e.g. did not answer several questions even after the experimenter instructed them to do so, so the timing of the introduction and the intervention were inconsistent with respect to other participants); and one participant was identified by her teacher as a student with special needs (her rapid response times to the self report questions also indicated that she did not take time to read the questions.) Five additional students were excluded

from the analysis due to their prior knowledge of the Towers of Hanoi learning activity and/or because they completed the activity. The analysis below was then conducted on the remaining sixty-one participants.

Results of the Investigation of the Hypotheses across Gender:

On the whole the results from the investigation of H1-H4 across gender did not support the hypotheses (see chapters 6.1-6.4 in [9] for a detailed explanation and discussion of these findings). However, there is some indication that the lack of significant findings may, in some cases, be because of the boys and girls behaving in opposite ways with respect to the conditions. Initial investigations with respect to gender, described below, indicated that there were several differences that showed significant interactions. These interactions may be due to developmental differences with respect to boys' and girls' emotional intelligence, which are particular to the 11-13 year old age group, which were not adequately understood and incorporated into the initial design of the experiment. These differences may have contributed to unanticipated variance that could interfere with support for the primary hypotheses. Since these differences are likely to generalize to a broader population, they are important to consider in future evaluations of affective technologies with boys and girls aged 11-13. Thus, while the comparisons below were unplanned, and do not carry the same weight as the planned comparisons, we think researchers will find these data of interest for current and future efforts to build emotionally intelligent learning companions for this age group.

Results of the Investigation of the Hypothesis with regard to Gender:

With an interest in explaining the general lack of support for the primary hypotheses H1-H4 and to further explore the initial gender findings, this section presents exploratory analysis with respect to gender differences and gender effects. Here, too, H1– H4 were not generally supported for either gender separately. However, there were significant ($p<0.05$) exploratory findings that have gender specific implications for the development of emotionally intelligent learning companions (Table 4).

Exploratory Test	**Significance**	**Girls Mean and Range**	**Boys Mean and Range**
ANOVA found that girls were more likely than boys to "think it will help to know that your mind is like a muscle and that you can increase your learning through effort"	p=0.03 F=4.9	3.7 range = 1-4	3.3 range = 1-4
ANOVA found that girls felt they would be better able to use strategies presented in the intervention than boys	p=0.003 F= 9.4	6.1 range = 1-7	4.8 range = 1-7
ANOVA found that girls persevere more than boys	p=0.016 F = 6	14.3 min range = 0-15min	12.2 min range = 0-15min
ANOVA found that boys that received task support responded more positively (bye.button) and had more positive impressions of the character (Modified Working Alliance) than boys that received affect support	p=0.02 F=6.1 p=0.036 F=4.8		1.1 affect resp. 1.7 task resp. range = 0-2 28.0 affect imp. 32.6 task imp. range = 6-42
ANOVA found that girls that received affect support responded more positively and had a trend (p = 0.09) toward more positive impressions of the character (Modified Working Alliance) than girls that received task support	p=0.04 F=4.8 p=0.09 (* trend) F=3	2.0 affect resp. 1.4 task resp. range = 0-2 33.2 affect imp. 29.2 task imp. range = 6-42	
ANOVA found that post-activity frustration had a congruence x gender interaction	p=0.043, F=4.3	6.9 high cong. 9.4 low cong. range = 2-28	11.5 high cong. 9.4 low cong. range = 2-28

Table 4. Summary of selected exploratory tests contrasting measures between girls and boys.

H1: As mentioned above, conducting an exploratory analysis of H1 for the separate gender groups did not show general support for H1 for either group. Here is a summary of the statistically significant gender differences found between *girls* and *boys*. *Boys* were less likely to "think it will help to know that your mind is like a muscle and that you can increase your learning through effort" ($p = 0.03$, $F = 4.9$). *Girls* indicated they would be able to use the strategies presented in the intervention to a greater extent than boys ($p=0.003$, $F =9.4$). As others have also found, we found that *girls persevere* longer than *boys* ($p=0.016$, $F = 6$).

The *intervention* had opposing effects for *boys* and *girls* with respect to the *bye.button* response, with boys responding more positively in the *task support* condition than *boys* in the *affect support* condition and *girls* having the opposite relationship with respect to these two conditions. *Boys* also had more positive impressions of the character that provided *task support* than the character that provided *affective support*, while *girls* showed a trend toward the opposite response. These are the types of differential findings that may explain the lack of significant results in the analysis across genders.

H2: Conducting an exploratory analysis of H2 for the separate *gender* groups did not show support for H2 for either group, although *boys* self-report using more *effort* than *girls.*

H3: Analysis did not find H3 to be generally supported for either *girls* or *boys;* however several interesting findings were made. The investigation of H3 focuses on three components of emotional intelligence (active listening, appropriate interventions, and sensor-driven non-verbal mirroring). Active listening is present in the *affective support intervention* and not present in the *task support intervention*. The appropriateness of the *affect-support* or *task-support* intervention provided with respect to a participant's level of *frustration* was encoded in the *congruence* measure presented in Table 5.

Two caveats should be noted with respect to the *congruence* measure. First, when considering those individuals experiencing Low levels of frustration, Low frustration could have meant a Flow state, in which case an intervention would probably be unwelcome. However, it could have alternatively meant boredom, in which case any intervention may have been welcome. Thus, referring to this condition as low or high *congruence* (see * in Table 5) is a rough approximation based on limited information. More complex affective state recognition is an open challenge. Second, because the *affective support intervention* is adaptive, it may be more appropriate for individuals experiencing low levels of *frustration* than the *task support intervention* for these same individuals (see ** in Table 5). While there are a few *frustration x intervention* conditions, acknowledged in the table, that might be more appropriate or less appropriate than the *congruence* measure indicates this measure was still found to be productive in assessing the impact of these elements of an agent's emotional intelligence.

Level of frustration	**Type of intervention**	**Congruence measure**
High	Affective intervention	High congruence
High	Task based intervention	Low congruence
Low	Affective intervention **	Low congruence *
Low	Task based intervention	High congruence *

Table 5. Congruence is a function of *frustration* and *intervention* that encodes the appropriateness of the *intervention* provided with respect to a participant's level of *frustration* (*, ** see explanation in the preceding paragraph).

In the assessment of the relationship between the character's *emotional intelligence* (*intervention, congruence,* and *mirroring*) and *girl* participants' *intrinsic motivation* the only significant finding was an interaction between *intervention x congruence* (p =0.02, F = 6.288). *Girls* who received *affect support* and had lower levels of *congruence* (i.e. *girls* that received *affect support* and were less frustrated) did not have as much *intrinsic-motivation* as those who had higher levels of *congruence* (i.e. *girls* who received *affect support* and were more frustrated). On the other hand, *girls* who received *task support* and had lower levels of *congruence* (i.e. *girls* that received *task support* and were more *frustrated*) had more *intrinsic-motivation* than those that had higher levels of congruence (i.e. *girls* who received *task support* and were less frustrated). For girls that were frustrated either intervention increased their *intrinsic motivation* over those that were less frustrated (at the time of intervention) (this will be specifically discussed in the next section). For boys the only measure of *motivation* that was effected by the character's *emotional intelligence* (*intervention, congruence,* and *mirroring*) was their willingness to *stick with it* which showed a trend toward significance (p=.065, F=3.8) suggesting that boys may be more willing to *stick with it* when they received *sensor driven non-verbal interactions* than when they received *pre-recorded non-verbal interactions*.

The relationship between the character's *emotional intelligence* (*intervention, congruence,* and *mirroring*) and participants' *frustration* was also assessed separately for *girls* and *boys,* using covariates of *age, school, pre-test self theories*, and *frustration* response at the time of intervention. *Girls* were reported to be less *frustrated* than *boys* at the end of the activity. There was also an interaction of *congruence* x *gender* (p= 0.043, F=4.327.) *Boys* and *girls* that received interventions with lower *congruence* had similar levels of *post-activity frustration* mean = 9.4; *boys* that received *interventions* that had higher *congruence* had a mean = 11.5, indicating more *frustration*, while *girls* that received *interventions* that had higher *congruence* had a substantially lower mean = 6.9, indicating less frustration.

For *girls* an interaction between *intervention x mirroring* (p = 0.001, F=16.3) was highly significant indicating that *girls* who received *non-verbal mirroring* and *affect support* had lower *post-activity frustration* than *girls* who received *non-verbal mirroring* with *task support*. *Girls* without *non-verbal mirroring* had the opposite relationship with the interventions – *girls* who received *affect support* had higher levels of *post-activity frustration* than *girls* with out *non-verbal mirroring* who received *task support. Girls* showed no main effect differences in *post-activity frustration* with respect to the type of *intervention* (*affect support* vs. *task support*) that they received.

For *boys,* there were significant (p= .009, F= 8.4) differences when grouped by *intervention.* Boys showed twice as much post *post-activity frustration* if they received the *affective support* than if they received *task support.* The higher levels of *congruence* were also detrimental for *boys;* they showed almost twice as much frustration for higher levels of *congruence* when compared to *boys* with lower levels of *congruence*. There was also a trend toward significance (p= .061, F= 4.0) for *mirroring:* boys that received *non-verbal mirroring* reported a third less frustration than *boys* that did not receive mirroring.

There was a significant interaction between *congruence x non-verbal mirroring* (p= .047, F= 4.6). *Boys* without *non-verbal mirroring* that had more *congruent* interventions reported levels of frustration approximately twice the level of frustration experienced by *boys* in other *congruence x mirroring* conditions (i.e. *boys* that had low levels of *congruence*, with and without *mirroring*, and *boys* that received high levels of congruence with mirroring).

H4: Here, too, further analysis did not find H4 to be generally supported for either *girls* or *boys*; however several additional interesting findings were made. For *girls* the *affective support* intervention was positively correlated to the *meta-affective skill* (p=.040, r =.37) and (more *Flow*/less *Stuck*) (p=.006, r=.52). Neither of these correlations was significant for *boys*.

In contrast to the result that no significant correlation between *meta-affective skill* and *Flow/Stuck* was present when assessed across both genders, the assessment with only *girls* shows significant correlation between *meta-affective skill* and *Flow/Stuck* (p = .010, r=.49). The assessment of only *boys,* for these same measures also shows a significant correlation (p=.021 r=-.40), but for boys this is a negative correlation. This is another clear instance where the grouping of the genders mixes different gender effects, yielding no significance when assessed together.

For *boys* controlling for *age, school, self theories, mirroring* and *intervention,* through partial correlation shows that there is significance (p = .048, r=.34) for *meta-affective skill* correlating with *perseverance*, while there is no significance for *girls.* With these covariates, neither *gender* shows significance for *Flow/Stuck* correlating with *perseverance*.

Meta-affective skill and *Flow/Stuck* were investigated with measures of *motivation* (*stick with it, strategies*, *post test goal mastery orientation*, "I would like to try this activity again", and *intrinsic-motivation*) using the same covariates. For *girls* there was no significance for *stick with it*; significance was found for both measures with respect to *strategies*, (p= .027, r=.4766) for *meta-affective skill*, and (p= .009 r=.5691) for *Flow/Stuck.* For *girls* there was also significant correlation to changes in *goal mastery orientation* for the *meta-affective skill* measure (p=.008, r=.57) indicating that *girls* that report higher levels of *meta-affective skill* also report higher levels of *mastery orientation*.

Controlling for the same variables (*age, school, self theories, mirroring and intervention*) there was no significant difference in *meta-affective skill* when the *girls* that showed *intrinsic-motivation* (measured by their reengagement in the task after the post-test surveys) were compared with those that did not reengage. There was a trend toward significant differences between these two groups of *girls* in terms of their *Flow/Stuck* (p = .067, F=3.8). Those that reengaged also had slightly higher levels of *Flow*/slightly lower levels of *Stuck*; both groups were fairly high on this measure, so there may also have been a ceiling effect – i.e. the differences may have been greater. In similar tests *boys* showed no significant differences across these groups and measures.

Discussion of Results with regard to Gender:

The exploratory investigation yielded several interesting results that support strong and potentially important recommendations for further study. This section will summarize the results of the gender specific analysis presented in the previous section and argue for the importance of a deeper understanding of the impact of *mirroring* and of *affect* and *task support,* as these relate to the *frustration, meta-affective skill* and *Flow/Stuck* of the 11-13 year-old sampled population. In particular this section argues for the need for better understanding of the gender differences in the impact of the elements of a learning companion's emotional intelligence and for the importance of the appropriate "coordination" of these elements with each other, for both girls and boys.

As presented in the previous section there were a few differences in the pattern of the social bond that girls and boys develop with the character, with respect to the type of *intervention* the participants received. Boys responded more positively to the character and had more positive impressions of the character that provided *task support* than the character that provided affect

support; girls had the opposite pattern. Differences in the social and emotional skill developments of girls and boys at these ages (11-13 year olds), with girls typically maturing earlier than boys, may have contributed to these differences.[1] Boys also self-report using more *effort* than girls. This finding and the *frustration* finding for girls discussed later in this paragraph, may have influenced different levels of interest in this activity, for girls and boys. There were very few differences found in the *motivation* measures with respect to the different elements of the character's emotional intelligence for either girls or boys. It was found that the girls that were more frustrated at the time of intervention also showed higher levels of *intrinsic motivation*, regardless of *intervention.* A possible explanation for this may be related to how much a participant cares about the activity. Girls that care more about doing this activity may also find it more frustrating. Independent of the frustration and independent of the type of *intervention* they receive, the caring may also lead to their increased *intrinsic-motivation*. In contrast to the girls, boys showed a strong difference in their levels of frustration due to the type of *intervention*, with much lower levels of frustration occurring in the *task support* conditions. This is probably related to the *social bond* differences discussed above, in which boys responded better to the character in the *task support* intervention. Likewise it is likely related to the finding that boys and girls that received interventions that were had lower *congruence* had similar levels of post activity *frustration*; while boys that received *interventions* that had higher *congruence* had higher levels of *post activity frustration* and girls that received *interventions* with higher *congruence* had substantially less frustration.

One of the biggest gender differences was found in the relationship between *meta-affective skill* and *Flow/Stuck.* In contrast to the result that no significant correlation between *meta-affective skill* and *Flow/Stuck* was present when assessed across both genders, the assessment with only girls shows a strong correlation between *meta-affective skill* and more *Flow/*less *Stuck*, while for boys*,* these measures show a strong correlation in the opposite direction. This is a clear instance where the grouping of the genders clearly mixes different gender effects, yielding no significance when assessed together. One possible hypothesis for the discrepancy in gender at this age is that girls aged 11-13 may be better able to assess their own emotions than boys. If girls are better at assessing their emotions then they may be better able to use their meta affective skill to lead themselves to more *Flow/* less *Stuck*. Boys on the other hand may report that they have meta affective skill but may actually be less able to recognize their own emotions; thus, even though they have some meta affective skill, they may not be as capable at applying it to their own experiential benefit.

While girls showed no main effect difference in the level of frustration based on the type of *intervention*, a further analysis indicated that this masked a more complex relationship that showed highly significant differences due to the interaction of the type of *intervention* and the presence of *mirroring*. These differences can be explained in terms of the "coordination" of the different elements of the character's emotional intelligence. Girls that experienced an *affective support* intervention in conjunction with *non-verbal mirroring* (condition 1) had lower levels of frustration than girls who received either *affective support* without *non-verbal mirroring* or girls who received *task support* with *non-verbal mirroring*. Condition 1 is a condition in which the *mirroring* and *intervention* are "coordinated" so that the character displays higher levels of emotional intelligence (as defined in this experiment as the presence of *intervention, congruence,* and *mirroring*) than in the other two conditions. One might argue that girls that received *task support* without *mirroring* were also in a "coordinated" condition that presents a character with

[1] The effects for both males and females who received affect support from machines in non-learning environments have previously been positive, but all of those results were for participants over age 18, and were delivered in different contexts.

higher levels emotional intelligence*;* they could also argue that in this condition girls experienced similar low levels of frustration when compared to the girls in condition 1. Extending this argument one might then argue that the existing capabilities of Intelligent Tutoring Systems, to provide *task support* without *mirroring* have similar benefits to girls, and the effort to develop *affect support* and *mirroring* are unwarranted. However the importance of *affect support* for girls is bolstered by the exploratory analysis of H4 showing that girls that receive *affective support* have higher levels of *meta-affective skill* and more *Flow*/less *Stuck* (these relationships were not found for boys). *Meta-affective skill* correlated significantly with beneficial changes in *goal-master orientation* and there was a trend toward significance in the positive relationship between *Flow/Stuck* and *intrinsic-motivation*. The findings from H3 and H4, taken together, support an argument not only for the further development of *affective support* and its benefits for girls*,* but also for the appropriate "coordination" of the elements of the character's emotional intelligence. These findings indicate that there are important opportunities to increase girls' *meta-affective skills*, increase their experience of *Flow* and decrease their experience of *Stuck,* increase their *mastery orientation,* and increase their *intrinsic-motivation.*

Data from the boys also support the argument for coordinating the elements of the character's emotional intelligence. The significant interaction between *congruence x non-verbal mirroring* indicated that the boys that experience more *congruent* intervention without *mirroring* also experienced twice as much post activity *frustration* as boys in the other three *mirroring x congruence* conditions. This particular form of discordant emotional intelligence displayed by the character (i.e. more *congruent* intervention without *mirroring*) seems to have had a negative impact on these boys.

Conclusion:

There are four primary experimental contributions of this research. First, the experiment demonstrated that the primary hypotheses were not supported for this age group when genders are combined. Second, further analysis illuminated opposing reactions, based on gender, to help explain the outcomes. Third, affective interventions were positively associated with girl's meta-affective abilities, higher levels of Flow, and lower levels of Stuck. Fourth, it was demonstrated that the various elements of a character's emotional intelligence should be presented in a "coordinated" manner. Inconsistencies between the presence or absence of non-verbal social mirroring and the presence or absence of other elements of emotional intelligence (congruence or affective support intervention) were associated with both girl's and boy's frustration.

In the experiment conducted, the type of intervention (affect support or task support), the level of congruence of the intervention with respect to a learner's frustration, and the presence or absence of social non-verbal mirroring played several important and different roles with respect to girl's and boy's frustration, meta-affective abilities, increased Flow/reduced Stuck, and intrinsic motivation. If these findings are confirmed by further studies and if they generalize to broader populations than the participants used in this study, then as Intelligent Tutoring Systems, and other systems that use relational agent strategies, advance to incorporate greater levels of emotional intelligence, developers and researchers should be able to make considerable advances to their systems and to learners' experiences by incorporating these elements of emotional intelligence. At the same time developers and researchers must be careful to appropriately coordinate the diverse elements of emotional intelligence and be well aware of the differences in the impact of these elements on boys and girls aged 11-13.

Acknowledgements:

We acknowledge contributors to the system development: Ken Perlin and Jon Lippincott, Ashish Kapoor, John Rebula, David Merrill, Rob Reilly, Barry Kort, Danielle Martin, Devin Neal, Louis-Philippe Morency, Carson Reynolds, Phil Davis, Karen Liu, Marc Strauss and Selene Mota. This material is based upon work supported in part by the MIT Media Lab's Things That Think consortium, in part by the National Science Foundation under Grant No. ITR 0325428, and in part by Deutsche Telekom. Any opinions, findings, 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.

References:

1. Wentzel, K. (1997). "Student Motivation in Middle School: The Role of Perceived Pedagogical Caring, Journal of Educational Psychology." Volume 89, Issue 3, p. 411.

2. Lester, J. C., S. A. Converse, S. Kahler, S. T. Barlow, B. A. Stone, R. Bhogal, (1997). "The persona effect: Affective impact of animated pedagogical agents." CHI'97, New York, NY, ACM Press.

3. Bailenson, J. N. and N. Yee (2005). "Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments." Psychological Science 16: 814-819.

4. Bickmore, T. and R. W. Picard (2004). "Establishing and Maintaining Long-Term Human-Computer Relationships." Transactions on Computer-Human Interaction 12(2): 293 - 327.

5. Baylor, A. L., E. Shen, and Huang, X.(2003). "Which pedagogical agent do learners choose? The effects of gender and ethnicity." E-Learn World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education. Phoenix, Arizona.

6. Klein, J., Y. Moon, R. Picard. (2002). "This Computer Responds to User Frustration: Theory, Design, Results, and Implications." Interacting with Computers 14: 119-140.

7. Lepper, M. R., M. Woolverton, M., Mumme, D. L., and Gurtner, J. L. (1993). "Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors." Computers as cognitive tools, S. P. Lajoie and S. J. Derry. Hillsdale, NJ, Erlbaum: 75-105.

8. Dweck, C. S. (1999). Self-Theories: Their role in motivation, personality and development. Philadelphia, PA, Psychology Press.

9. Burleson, W. (2006), "Affective Learning Companions: Strategies for Empathetic Agents with Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive and Meta-Affective Approaches to Learning, Motivation, and Perseverance," MIT PhD Thesis, September 2006. (http://affect.media.mit.edu/publications.php)

10. Rottenberg, J., Ray, R. D., and Gross, J. J. (in press). Emotion elicitation using films. To appear in J. A. Coan & J. J. B. Allen (Eds.), The handbook of emotion elicitation and assessment. London: Oxford University Press.

Author's Biographical Sketch:

QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.

Winslow Burleson is an Assistant Professor at Arizona State University in the Department of Computer Science and Engineering and the Arts, Media, and Engineering graduate program. His technical interests focus on human-computer interaction applied to creativity, innovation, well-being, design engineering, and educational technology. He received his PhD in Media Arts and Sciences from the MIT Media Lab, Affective Computing Group. He is a member of the ACM.

Arizona State University, 699 South Mill Avenue Room 395, Tempe, AZ 85281, Winslow.Burleson@asu.edu

(pls crop image as needed)

Rosalind W. Picard is Professor of Media Arts and Science, Director of Affective Computing Research, and Co-Director of the Things That Think Consortium at the MIT Media Laboratory. Her interests include the development of new technologies and theories that advance basic understanding of affect and its role in human experience. She holds a Doctorate of Science from MIT in Electrical Engineering and Computer Science, and is a Fellow of the IEEE and a member of the ACM and AAAI.

MIT Media Laboratory, E15-448; Cambridge, MA 02142-1308, Picard@media.mit.edu