

# Defining the Behavior of an Affective Learning Companion in the Affective Meta-tutor Project

Sylvie Girard, Maria Elena Chavez-Echeagaray, Javier Gonzalez-Sanchez, Yoalli Hidalgo-Pontet, Lishan Zhang, Winslow Burleson, and Kurt VanLehn

Arizona State University, Computing, Informatics, and Decision Systems Engineering,  
Tempe, AZ, 85281, U.S.A.

{sylvie.girard,helenchavez,javiergs,lzhang90,yhidalgo,  
winslow.burleson,kurt.vanlehn}@asu.edu

**Abstract.** Research in affective computing and educational technology has shown the potential of affective interventions to increase student's self-concept and motivation while learning. Our project aims to investigate whether the use of affective interventions in a meta-cognitive tutor can help students achieve deeper modeling of dynamic systems by being persistent in their use of meta-cognitive strategies during and after tutoring. This article is an experience report on how we designed and implemented the affective intervention. (The meta-tutor is described in a separate paper.) We briefly describe the theories of affect underlying the design and how the agent's affective behavior is defined and implemented. Finally, the evaluation of a detector-driven categorization of student behavior, that guides the agent's affective interventions, against a categorization performed by human coders, is presented.

**Keywords:** affective computing, affective learning companion, intelligent tutoring system, robust learning, meta-cognition.

## 1 Introduction

Research in AIED has taken interest in the potential of using interventions of affective nature in intelligent tutoring systems to improve learning [2, 19, 23] and motivation [8, 13, 20] and to reduce undesirable behaviors such as gaming [3-5] and undesirable affective states such as disengagement [17]. The interventions have been designed to either respond to student' specific behavior [14, 19], or to elicit a certain emotional state in the student [9], often by providing cognitive support and scaffolds within the learning environment.

The hypothesis of our project [24] is that affective interventions in a meta-cognitive tutor can help students achieve robust learning by being persistent in their use of meta-cognitive strategies during and after tutoring. In order to test this hypothesis, an affective intervention was designed, using an affective learning companion to convey the affective message. This article describes the design of the affective intervention. In the first section, a three-dimensional design space of affective interventions is outlined, along with our choice along each dimension. The second

section describes the implementation of the design using categorization of student behavior based on log data detectors. The last section describes an empirical evaluation of the classification accuracy.

## 2 Design of the Affective Intervention

### 2.1 Definition of the Affective Intervention

Over the past decade, numerous affective interventions have been designed and evaluated with respect to alternate techniques in the field of educational technology. In order to define a design space of the affective intervention for the AMT project, a review of current research was performed. The design space has three dimensions: mechanism for delivery of the affective intervention, timing of the intervention, and type of message delivered during the intervention. We briefly describe each dimension, then indicate where along it our design falls.

#### ***Mechanism: How Is the Intervention Message Conveyed?***

There are various ways to intervene affectively in tutoring systems, ranging from the presentation of an affective message via a user-interface component [2, 19], to the use of bio-feedback and affect-sensitive tutors that respond to the user's emotional state [9]. Some results [2,8,12,23] have shown the potential of using pedagogical agents, or Affective Learning Companion (ALC), to portray the affective message. These interventions involve design decisions concerning the different components of a pedagogical agent that can impact learning, such as the presence of facial expressions or deictic gestures [2,14], vocal intonation [6], gender [2,8,16], or ethnicity and student's cultural background [12,19].

In this phase of our project affective messages in the form of pop-up text messages are provided by a pedagogical agent, represented by an image with neutral facial expression. The agent is a humanoid comic-like gendered character, representing a student of a similar age to our target population (16-21 yrs olds). This decision took into account the results from [12] for the agent's image type, and [2, 23] where pairing students' gender to the agent's gender was found beneficial for user's self-concept and learning.

#### ***Timing: When Is the Affective Intervention Happening in the Learning Process?***

The affective intervention can happen before any tutoring takes place, between learning tasks during the tutoring, and at different moments while a learner is performing a specific task or learning a specific set of skills. In order to describe when the affective intervention occurs, we first must describe the instruction.

The AMT software teaches students how to create and test a model of a dynamic system. The instruction is divided into three phases: (1) an *introduction* phase where students learn basic concepts of dynamic system model construction and how to use the interface; (2) a *training* phase where students are guided by a tutor and a meta-tutor to create several models; and (3) a *transfer* phase where all scaffolding is

removed from software and students are free to model as they wish. The tutor gives feedback and corrections on domain mistakes. The meta-tutor requires students to follow a goal-reduction problem solving strategy, using the Target Node Strategy [24], which decomposes the overall modeling problem into a series of “atomic” modeling problems whose small scope encourages students to engage in deep modeling rather than shallow guess-based modeling strategies. Using various measures of deep and shallow learning [5], an experiment demonstrated that requiring students to follow this strategy during training did indeed increase the frequency of deep modeling compared to students who were not required to follow the strategy. However, the effect was not strong, and the amount of deep modeling could certainly be improved. The goal of the ALC is to encourage students to do even more deep modeling.

The pedagogical agent conveying the affective message in AMT intervenes at three different moments of software interaction:

- *At the beginning and the end of the introduction:* These interventions aim to introduce the agent and its role in the instruction, as well as building rapport between the student and the ALC which has been shown in [7] to help keep students motivated and on task.
- *Between each modeling task in the training phase:* The main purpose of these interventions is to invite the student to reflect on his/her actions and decisions during the task, as well as maintain the interest of the student. As performing a given task can require from 3 to 15 minutes, the ALC intervenes after each task rather than intervening after a pre-defined number of tasks as in [1,2,23].
- *At the end of the training phase:* This intervention tries to convince the student to persevere in the use of the deep modeling strategy during the forthcoming transfer phase.

### **Type: What Type of Message Is Given/Transmitted During the Intervention?**

Finally, the third dimension of the intervention represents its affective or motivational content: what does the ALC say and what emotional tone does it use when saying it? Our design is based on the following policies:

- Baylor and Kim [6] showed that a combination of cognitive and affective interventions (the “Mentor”) led to better student self-regulation and self-efficacy than the presence of either type of intervention alone. Our meta-tutor and tutor already provide cognitive information without affect (like the “Expert” of [6]). To avoid boring redundancy, the ALC presents as little cognitive and meta-cognitive content as possible (just enough to maintain context) while presenting motivational messages (described below) in a friendly, encouraging manner.

The content of the intervention has been designed to help low-achievers and shallow learners get back on track and avoid gaming [3-5, 9, 19], while not interrupting high-achievers who might not benefit from an affective intervention [2, 19, 23]. It involves the following theories:

- *Dweck's "the mind is a muscle" theory* [10]: the more you exercise your mind, the more competent you become. Before the introduction phase, all students read a text introducing this theory. The between-task interventions reinforce the message by mentioning passages of the reading and referring to how different activities help to improve the brain's function.
- *Attribution theory* [21]: failures should be attributed to the difficulty of the task or lack of preparation, whereas success should be attributed to the student's effort.
- *Theory of reflection* [15]: Students have been found to be more receptive after completing a problem rather than during problem solving [15]. Every time a task is finished the ALC invites students to reflect on what they have experienced. It encourages them to replicate the action if it was positive or to change the action if it was negative.
- *Use of a meta-cognitive representation of student's modeling depth* [1, 22]: Alongside the ALC is a bar showing the depth of the student's modeling while working on the current task. That is, it shows the proportion of student actions that were classified as deep, based on the detectors described in [11]. ALC messages often refer to the modeling depth bar in combination with the other theories listed above.

The following section illustrates how we defined the ALC behavior by using learners' prior interactions with the system.

### 3 Implementing the ALC's Behavior

While students learn, their motivation and attention to detail can fluctuate. In the context of a problem solving activity requiring modeling skills, the depth of the modeling techniques used by students can also vary. The ALC should adapt to these fluctuations, presenting different affective messages depending on the student's recent behavior. Simply mapping the student's behavior onto competence would not suffice, so we defined several behavioral classifications such as "engaged," "gaming" and "lack of planning." We then defined log data detectors relevant to each behavioral classification. We also paired affective messages with each behavioral classification. In the first subsection, the detectors that measure the user's behavior are described. The second sub-section then describes the behavioral classification, how they were created and how they are mapped to the detectors' output.

#### 3.1 How to Detect Shallow Modeling Practices?

The detectors process a stream of user interface activity (log data) and output behavioral measures. The detectors require no human intervention and run in real time, because they will eventually be used to regulate the system's responses to the student. Our detectors extend the gaming detectors of [4] by including measures relevant to depth of modeling and other constructs.

Nine detectors were defined. The first six detectors were based on classifying and counting segments in the log, where a segment corresponds roughly to a correct step in the construction or debugging of a model. Each segment holds the value of the detector that best represents the situation, for example a student showing both a `single_answer` and `good_method` behavior would be defined as following a `good_method` behavior for this segment. The output per task for each detector is a proportion: the number of segments meeting its criteria divided by the total number of segments in the log for the task. Based on an extensive video analysis of student's past actions and HCI task modeling techniques [11], six segmental detectors were defined:

- `GOOD_METHOD`: The students followed a deep method in their modeling. They used the help tools<sup>1</sup> provided appropriately including the one for planning each part of the model.
- `VERIFY_INFO`: Before checking their step for correctness, students looked back at the problem description, the information provided by the instruction slides, or the meta-tutor agent.
- `SINGLE_ANSWER`: The student's initial response for this step was correct, and the student did not change it.
- `SEVERAL_ANSWERS`: The student made more than one attempt at completing the step. This includes guessing and gaming the system.
- `UNDO_GOOD_WORK`: This action suggests a modeling misconception on the students' part. One example is when students try to run the model when not all of the nodes are fully defined.
- `GIVEUP`: The student gave up on finding the answer and clicked on the "give up" button.

A limitation of the above detectors is the inability to distinguish between a student trying hard to complete a step but making a lot of errors versus a student gaming or guessing a lot. This led to the development of two additional detectors based on earlier work in detecting robust learning and gaming [5, 9, 18, 23]: (1) the time spent on task and (2) the number of times the learner misused the "run model" button. While the former is self-explanatory and commonly used in ITSs, the latter is specific to the AMT software. As students construct a system dynamics model, they can reach a point where all elements are sufficiently defined to "run the model" (the model is correct in terms of syntax) and therefore test whether its semantics corresponds to the system they were asked to model. Students clicking on this button before the model's syntax is correct, or clicking repetitively on the model without making changes once it is correct in syntax but not in semantics, is considered shallow behavior that shows a lack of planning, a lack of understanding of the task to perform, or a tendency to guess/game the answer rather than think it through.

---

<sup>1</sup> Two help systems are available to users: (1) referring back to the instructions always available for viewing, and (2) looking at the problem situation where all details of the dynamic system to model are described.

The ninth and last detector is a function of the six segmental detectors. It is intended to measure the overall depth of the students' modeling. Although it is used as an outcome measure in the transfer phase, it helps drive the ALC during the training phase. It is based on considering two measures (GOOD\_ANSWER, VERIFY\_INFO) to indicate deep modeling, one measure (SINGLE\_ANSWER) to be neutral, and three measures (SEVERAL\_ANSWERS, UNDO\_GOOD\_WORK, and GIVE\_UP) to indicate shallow modeling.

In order to facilitate writing rules that defined the students' behavioral category (e.g., engaged, gaming, etc.) in terms of the detector outputs, we triaged the output of each detector so it reports its output as either low, medium and high. The rules are mostly driven by the values: low and high. To implement the triage, we collected logs from 23 students. For each of the nine detectors, we determine the 33<sup>rd</sup> and 66<sup>th</sup> percentile points and used them as thresholds. Thus, for each detector, roughly a third of the 23 students were reported as low, as medium and as high. Because the tasks vary in complexity, different thresholds were calculated for each task.

### 3.2 From Shallow Learning Detection to the ALC Intervention

A series of 6 types of ALC behavioral categories were defined using video analysis of past user's actions on software. Human coders reviewed screen-capture videos and verbal protocols of a pool of 20 students using the meta-cognitive tutor. Following their recommendations and a review of messages transmitted in affective interventions in the literature, the following set of ALC categories was defined:

- *Good Modeling*: The students think about their steps, do not hesitate to go back to the introduction or the situation to look for answers, use the plan feature judiciously in their creation of nodes, and have a minimum of guessing and wrong actions performed on task.

- *Engaged*: The students respond by thinking about the problem rather than guessing, refer back to the instructions or problem situation when they find themselves stuck rather than trying all possible answers. The students take a medium to a high amount of time to complete the task, favoring reflection to quick decisions.

- *Lack of Planning*: The students answer quickly, relying heavily on the feedback given in the interface to get the next steps. While the students sometimes refer to instructions and the situation, they only use the features when they are stuck, not when planning the modeling activity.

- *Help Avoidance*: The students attempt a lot of answers without referring back to the instructions or the problem situation. They rarely make use of the information filled in the plan tab and try to skip the meta-tutor instructions. Instead of using help when they are confused, they spend a lot of time trying to get the interface green or give up rather than thinking about the problem.

- *Gaming*: The students try multiple possible answers within the interface without pausing long enough to think about the problem. They may give up when this random guessing doesn't work. They rarely refer to the instructions or the problem situation and pay little attention to the plan tab or the meta-tutor instructions.

- *Shallow Modeling (default, not recognized as the above mentioned categories)*: The students tend to try several answers on the interface rather than pausing and thinking about the problem. They sometimes refer back to the instructions and problem situation, but not frequently.

**Table 1.** Examples of ALC intervention between-task

Behavior	Example
Good Modeling	You're a Green Master! What was your secret? I know... you make your reading count and thus your brain is getting rewired.
Engaged	Even though it might take a little bit longer, it is worth it to explore the available resources. You are giving your brain a great workout. Look at that green bar! Keep up the good work!
Lack of Planning	Going fast is good, but it doesn't always help you reach your potential... Why don't you stop and think about what you want to model when you are confused. To make more of the bar green, try re-reading the problem description and noting what it asks you to do.
Help Avoidance	It might be worth rereading the problem description and paying more attention to the suggestions presented by the pop-up messages. Our brain needs to engage the material deeply so it can create good connections. That's how we can get more of the bar green!
Gaming	Hmmm! It seems that you need to put quality time into your tasks. Maybe "trial and error" is not always the best strategy. Although you might move quickly through the problem, your brain doesn't get a workout, and it shows in the length of the green bar.
Shallow Modeling (default)	You are getting there! Look at that bar! But remember that to strengthen your brain you have to engage the problem and all its details.

Once these six behaviors were defined, human coders applied them to a sample of 100 tasks and students. The outputs of the detectors on the sample were obtained, and rules were defined to map their values to the behavioral categories.

Using the theories of affect defined in section 2, ALC messages were created for each behavior in order to provide affective support to the learner. A stereotypical message was first created, as illustrated in table 1, for each behavior. The research group then created many synonymous versions of each message, so that the ALC would not repeat itself and thus reduce the student's perception of the ALC as an artificial agent. A separate message was produced for the first and last task performed by the user in the training phase, in order to introduce and wrap-up the ALC interventions.

## 4 Evaluation of the Behavior's Accuracy

Before working with students, we first tested the detectors and behavioral categorizer via test cases. We wrote scenarios of software use that typified each of the six behavioral categories. A member of the research group enacted each scenario, and we confirmed that the detector outputs fell in the anticipated range (low, typical or high) and that the rules assigned the anticipated behavioral classification.

The second part of the validation of ALC behaviors involved pilot subjects and human coders. Seven college students used the AMT system with the ALC turned on. They were asked to speak aloud as they worked. Their voice and screen were recorded as videos. A sample video was made from screen recordings. It included 15 tasks. Three human coders watched each task, paying attention to the depth of modeling shown by the student's actions. Independently of what the software chose, they

chose the ALC intervention that they felt best matched the student's modeling practices. A multi-rater and pairwise kappa was then performed, and showed a sufficient level of inter-reliance with a level of .896.

## 5 Conclusion and Future Work

This article described the development of an affective intervention based on an affective learning companion (ALC) that works with a meta-tutor and a tutor. It described the theories of affect underlying the interventions, and how we defined and implemented the ALC's behavior. The ALC's messages were based on deciding which of six behavioral categories best represented the student's work on the most recently completed task. This categorization was driven by log data. When compared to human coders working with screen captures and verbal reports of students, the detector-driven categorizations agreed with the human coding with a kappa of .896.

The next step in the research is to measure the benefits of this version of the ALC in a two-condition experiment. One group of students will use the system with the ALC turned on during the training phase, and the other will use it without the ALC turned on. We hypothesize that this will cause measurable differences in the depth of students' modeling during the transfer phase.

The forthcoming evaluation will also have students wear physiological sensors while they work so that we can collect calibration data that will be used to supplement the detectors' assessment of the students' affective state. This extra information will be used to help define affective interventions not only between tasks but also while the learner performs on task.

**Acknowledgements.** This material is based upon work supported by the National Science Foundation under Grant No. 0910221.

## References

1. Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., et al.: Repairing disengagement with non-invasive interventions. *Frontiers in Artificial Intelligence and Applications*, vol. 158, p. 195 (2007)
2. Arroyo, I., Woolf, B.P., Cooper, D.G., Burleson, W., Muldner, K.: The Impact of Animated Pedagogical Agents on Girls' and Boys' Emotions, Attitudes, Behaviors and Learning. In: *Proceedings of the 2011 IEEE 11th International Conference on Advanced Learning Technologies*. Proceedings from ICALT 2011, Washington, DC, USA (2011)
3. Baker, R.S.J.d., et al.: Adapting to when students game an intelligent tutoring system. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) *ITS 2006*. LNCS, vol. 4053, pp. 392–401. Springer, Heidelberg (2006)
4. Baker, R.S.J.d., Gowda, S.M., Corbett, A.T.: Towards predicting future transfer of learning. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) *AIED 2011*. LNCS, vol. 6738, pp. 23–30. Springer, Heidelberg (2011)



5. Baker, R.S.J.d., Gowda, S.M., Corbett, A.T., Ocumpaugh, J.: Towards automatically detecting whether student learning is shallow. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 444–453. Springer, Heidelberg (2012)
6. Baylor, A.L., Kim, Y.: Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education* 15(2), 95–115 (2005)
7. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12(2), 293–327 (2005)
8. Bursleson, W., Picard, R.W.: Gender-Specific Approaches to Developing Emotionally Intelligent Learning Companions. *IEEE Intelligent Systems* 22(4), 62–69 (2007), doi:10.1109/MIS.2007.69
9. D’Mello, S.K., Lehman, B., Person, N.: Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education* 20(4), 361–389 (2010), doi:10.3233/JAI-2010-012
10. Dweck, C.: *Self-Theories: Their role in motivation, personality and development*. Psychology Press, Philadelphia (2000)
11. Girard, S., Zhang, L., Hidalgo-Pontet, Y., VanLehn, K., Bursleson, W., Chavez-Echeagary, M.E., Gonzalez-Sanchez, J.: Using HCI task modeling techniques to measure how deeply students model. In: Chad Lane, H., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 766–769. Springer, Heidelberg (2013)
12. Gulz, A.: Benefits of Virtual Characters in Computer Based Learning Environments: Claims and Evidences. *International Journal of Artificial Intelligence in Education* 14(3), 313–334 (2004)
13. Gulz, A., Haake, M., Silvervarg, A.: Extending a teachable agent with a social conversation module – effects on student experiences and learning. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS, vol. 6738, pp. 106–114. Springer, Heidelberg (2011)
14. Hayashi, Y.: On pedagogical effects of learner-support agents in collaborative interaction. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 22–32. Springer, Heidelberg (2012)
15. Katz, S., Connelly, J., Wilson, C.: Out of the lab and into the classroom: An evaluation of reflective dialogue in Andes. In: *Proceeding of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, pp. 425–432 (2007)
16. Kim, Y., Baylor, A., Shen, E.: Pedagogical agents as learning companions: the impact of agent emotion and gender. *Journal of Computer Assisted Learning* 23(3), 220–234 (2007)
17. Lehman, B., D’Mello, S., Graesser, A.: Interventions to regulate confusion during Learning. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 576–578. Springer, Heidelberg (2012)
18. Muldner, K., Bursleson, W., Van de Sande, B., VanLehn, K.: An analysis of students’ gaming behaviors in an intelligent tutoring system: predictors and impacts. *User Modeling and User-Adapted Interaction* 21(1-2), 99m–135m (2011), doi:10.1007/s11257-010-9086-0
19. Rodrigo, M.M.T., Baker, R.S.J.d., Agapito, J., Nabo, J., Repalam, M.C., Reyes, S.S., San Pedro, M.O.C.Z.: The Effects of an Interactive Software Agent on Student Affective Dynamics while Using an Intelligent Tutoring System. *IEEE Transactions on Affective Computing* 3, 224–236 (2012), doi:http://doi.ieeecomputersociety.org/10.1109/T-AFFC.2011.41

20. Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., Collins, H.: The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies* 66(2), 98–112 (2008), doi:10.1016/j.ijhcs.2007.09.003
21. Weiner, B.: An attributional theory of achievement motivation and emotion. *Psychological Review* 92(4), 548 (1985)
22. Walonoski, J.A., Heffernan, N.T.: Prevention of off-task gaming behavior in intelligent tutoring systems. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) *ITS 2006*. LNCS, vol. 4053, pp. 722–724. Springer, Heidelberg (2006)
23. Woolf, B.P., Arroyo, I., Muldner, K., Burleson, W., Cooper, D.G., Dolan, R., Christopherson, R.M.: The Effect of Motivational Learning Companions on Low Achieving Students and Students with Disabilities. In: Aleven, V., Kay, J., Mostow, J. (eds.) *ITS 2010, Part I*. LNCS, vol. 6094, pp. 327–337. Springer, Heidelberg (2010)
24. Zhang, L., Burleson, W., Chavez-Echeagaray, M.E., Girard, S., Gonzalez-Sanchez, J., Hidalgo-Pontet, Y., VanLehn, K.: Evaluation of a meta-tutor for constructing models of dynamic systems. In: Chad Lane, H., Yacef, K., Mostow, J., Pavlik, P. (eds.) *AIED 2013*. LNCS (LNAI), vol. 7926, pp. 666–669. Springer, Heidelberg (2013)