



An architectural design and evaluation of an affective tutoring system for novice programmers

Hua Leong Fwa 

Correspondence: h.l.fwa1@newcastle.ac.uk
Newcastle University, Newcastle, UK

Abstract

Affect is prevalent in learning and it influences students' learning achievement. This paper details the design and evaluation of an Affective Tutoring System (ATS) that tutors student in computer programming. Although most ATSs are purpose built for a specific domain, making adaptation to another domain difficult, this ATS is architected for adaptability and extensibility. This study also addresses a lack of research exploring the theories and methods of integrating affect and learning within the learning process by proposing methods of regulating the negative affect of students. Both quantitative and qualitative techniques were used for evaluation of the effectiveness of the ATS and its usability and acceptance by student participants. The results revealed that the full affective version of the ATS results in more effective tutoring as compared to the version with the affective function disabled and the students are positive on their learning experience with the ATS with the fill in the gap exercises and hints being most highly rated.

Keywords: Affect, Tutoring, Programming, Learning, Architecture

Background

Emotions and cognition are intertwined as evidenced by studies (Dolan & Vuilleumier, 2003; Kort, Reilly, & Picard, 2001; Linnenbrink, 2006; Phelps, 2004; Vuilleumier, 2005) investigating the interactions between emotion and attention, learning and memory. Anatomically, Gray, Braver, and Raichle (2002) also points to the integration of emotional and cognitive processes in prefrontal cortex of the human brain. This anatomic link was played out in real life in the unfortunate events experienced by Elliot, a patient of Damasio (1994). As a result of a tumor cut from his cortex near the brain's frontal lobe, Elliot, despite having a high level intelligence, became incapable of making simple decisions in his daily life and this eventually led to failures in both his career and marriage.

The link between emotions and cognition prompted studies investigating into the relationship between emotions and the learning achievement of students. Pekrun, Goetz, Titz, and Perry (2002) opined that academic emotions (which are linked to academic learning, classroom instruction and achievements) with the exception of test anxiety are largely neglected by educational psychology research. They embarked on a series of qualitative and quantitative studies to uncover the academic emotions experienced by

students and the effects of these emotions on the students' learning and achievements. Their findings revealed that students experience a wide repertoire of emotions in academic settings and both positive emotions e.g. hope, pride and relief and negative emotions such as anxiety and boredom are prevalent in learning. The positive emotions e.g. enjoyment of learning affect achievement positively by strengthening motivation and enhancing flexible learning whereas the negative emotions e.g. anxiety erodes motivation and draws attention away from the task, resulting in shallow learning. The effects that both positive and negative emotions have on learning were also corroborated by other researchers (Ashcraft & Kirk, 2001; Greene & Noice, 1988; Isen, 2000).

Bloom (1984) highlighted the 2 sigma problem which states that students who are tutored on a one to one basis achieve learning gains of 2 sigma over and above those students who are group tutored. He attributed this difference to the process of constant corrective feedback and high engagement of the tutee, maintained through close monitoring and constant encouragement by the tutor. To achieve enhanced learning outcomes, it is crucial that the active engagement and motivation of the student be sustained through this continuous affective feedback loop between the tutor and the student.

In a study by Efklides (2006) on metacognition and affect, she postulated student perceived task difficulty as a factor influencing metacognition in learning. She reasoned that if a very difficult task is assigned to students and they perceived it as beyond their capability, they would develop negative emotions such as frustration and give up on the task. However, if teachers are able to scaffold the students' learning or employ the use of alternative pedagogical intervention techniques e.g. provision of hints to alleviate the negative emotions, they would be able to sustain the engagement of the students.

Intelligent Tutoring Systems are built with the objective of providing learners with the benefit of one to one tutoring automatically and cost effectively. It acts like a personal training assistant that continually assesses one's knowledge through interactions with the system and builds a personalized model of one's acquired knowledge for the provision of tailored instructions or assistance in the form of hints or demonstrations when one seems to require help to move on.

A common criticism of ITS is that they are devoid of emotional awareness and empathy and that limits their tutoring effectiveness (Lepper, Woolverton, Mumme, & Gurtner, 1993). Pivotal to the effectiveness of human tutors in one to one tutoring is the fact that human tutors are able to constantly monitor both the affect and cognitive ability of their tutees and respond with adequate supportive tutoring measures (Goleman, 1995). Analogous to the way human tutors detect and respond to the affect of the their tutees to sustain the engagement of the latter, Affective Tutoring Systems (Picard, 1997) adapt to the tutees' affect autonomously to bring about enhanced learning outcomes. Most Affective Tutoring Systems (ATs) incorporate affect sensing, tutoring strategies and learning progress tracking into a single environment. With the ability to sense the affect of students, ATs can for instance, sense that students are frustrated and offer hints to resolve the impasse.

Although researchers acknowledge the role that affect plays in learning, ATs on the contrary are rarely implemented (Thompson & McGill, 2012). This is often attributed to the fact that the implementation of ATs encompasses cross-disciplinary knowledge spanning domains of education, psychology and computer science. Often, the implemented ATs is highly specialized and purpose-built for a particular tutoring

subject area, making replication or customization for a different subject area difficult, if not impossible.

Related work

AutoTutor is a fully automated computer tutor that helps students learn Newtonian physics and computer literacy topics. It presents problems or questions to the students and engages them in a dialogue to collaboratively build towards a solution (S. D'Mello et al., 2008). D'Mello et al. aim to incorporate learners' affect into existing AutoTutor's pedagogical strategies (e.g. by regulating their negative emotions) and their research efforts culminated in the development of Affective AutoTutor. Affective AutoTutor is one of the few ATSs that detects and responds to students' affective states (D'Mello & Graesser, 2010). Affective AutoTutor uses facial cues, body postures and conversational dialogue feature to infer the affect of students. The tutoring actions are then derived from a set of production rules that dynamically assessed the cognitive and affective states of students to address negative emotions such as frustration and boredom.

Easy with Eve is an ATS that is built by Massey University, New Zealand (Alexander, Sarrafzadeh, & Hill, 2006). It infers the emotions of learners mainly through facial expressions analysis that was developed in-house. In addition, a case based reasoning program was developed to output a weighted set of tutoring actions and facial expressions based on a given sequence of interactions. The recommended facial expression would be expressed through Eve – an animated intelligent agent embodied within the system.

Empathic Companion is an embodied agent type system that detects and responds to the user's affective state (Prendinger, Dohi, Wang, Mayer, & Ishizuka, 2004). It was developed in the context of a web-based job interview scenario with the objective of regulating the user's negative emotions when faced with difficult job interview questions. It employs the use of a decision network for fusing the inputs from a Galvanic Skin Response (GSR) and an Electromyography (EMG) sensor and for translating the sensors' signals into both the emotions and the agent decisions. The results indicated that Empathic Companion does reduce the frustration of its users.

MetaTutor is an adaptive ITS that is designed to encourage students to employ meta-cognitive self-regulated learning (SRL) in the tutoring context of human circulatory system (Azevedo, Johnson, Chauncey, & Burkett, 2010). MetaTutor's focus is on SRL skills that are fostered by managing learning through monitoring and strategy use. This is achieved in MetaTutor through the use of pedagogical agents which respond with an evaluation of the student's current level of understanding upon request.

Crystal Island is a 3-dimensional narrative-centered learning environment for the tutoring of eighth-grade microbiology (Rowe et al., 2009). Sabourin, Rowe, Mott, and Lester (2013) fed the survey scores and in-game progress of students within Crystal Island into a Dynamic Bayesian Model (DBN) for the identification of students' off task behavior. The students' actions within the environment were logged. The logs were then analyzed to extract the behaviors of students that diverge from the learning task. These serve as the off-task labels for the classification task. The objective of the study is to identify whether students use off-task behaviors to regulate their emotions e.g. to alleviate frustration.

From the above discussion, we can surmise that an ATS first infers the affect of the students through various input modalities and then responds with various tutoring

actions that alleviate or regulate the students' negative affect. The tutoring actions or responses vary though across each of the above implemented ATS and there seems to be no uniform tutoring response even for the same exhibited affect e.g. frustration. There is in fact a critical lack of research in the area of ATS that explores the theories and methods of integrating affect and learning within the learning process.

The effectiveness of the tutoring response is measured by the degree by which the students' negative emotion is alleviated. To achieve this, the students will have to 'learn' to cope with or regulate the negative emotions. The negative emotions direct students to focus inwardly on the emotions and draw attention away from the learning tasks (Hascher, 2010). In Kort, Reilly, and Picard (2001)'s four quadrant spiral learning model, they postulated that external help such as scaffolding can sustain and motivate a learner to cope with and overcome a learning stage characterized by frustration and misconceptions. Within the field of psychology, this process of "coping" refers to efforts to master, reduce or tolerate the demands created by stress (Weiten, Dunn, & Hammer, 2011). The coping strategies can be further segregated into 3 main types – appraisal-focused, emotional-focused and problem-focused.

Appraisal-focused coping strategy involves changing the way one thinks about the problem. One can possibly dispute or challenge one's irrational assumptions. The key is to replace the negative thinking with rational analyses, potentially reducing stressful situations to less threatening ones. Another technique is to de-stress with humor. Humor buffers the effect of stress, allowing people to bounce back from negative situations faster and also helps to promote social interactions and support.

Emotional-focused strategy involves releasing one's pent-up emotions through expression of the emotions. The techniques include managing hostility towards others and being more forgiving. It can also possibly involve the use of relaxation techniques such as mediation. In addition, effort and task difficulty attributions (Weiner, 1972) can also help in mediating the negative emotions experienced by students. Failure to do well in a learning task can be attributed to insufficient effort (a controllable factor) instead of one's intelligence which is less malleable.

Lastly, the problem-focused strategy involves the use of a systematic problem resolution process of seeking help and acquiring new skills for dealing with the problem e.g. time management skills. By exercising self-control through techniques such as operant conditioning (which involves use of rewards and punishments), one can also better cope with the stress.

In this study, appraisal-focused, emotional-focused and problem-focused strategies are utilized for the regulation of negative emotions experienced by the students.

Design approach

Augmenting ITS with the ability to infer the affect of students is the first step in the construction of an ATS. The ensuing challenge is then how do we design the ATS such that it can respond appropriately to the detected affect of students to increase engagement, task persistence and sustain their motivation to learn. The goal is to regulate or ease the negative emotions such as frustration and disengagement occurring in learning so that it does not degenerate into hopelessness. It would be ideal if a recipe that lists the pedagogic strategies to be applied for every tutoring situation that may arise in an ATS exists. This gap in the literature however, has yet to be addressed.

In addition, most implemented ATSs were also not evaluated for their effectiveness and acceptance by the students. Thus, this study adds to the body of knowledge by proposing an architecture for an ATS that tutors in the domain of computer programming and evaluating for its effectiveness and acceptance.

This paper describes the design, implementation and evaluation of an ATS that tutors students in the subject of computer programming. To cater for the generalizability of the ATS for subject areas other than computer programming, the components within the architecture are designed to be loosely coupled so that they can easily be replaced or extended. Equipped with the ability to infer affect, appropriate tutoring response strategies based on psychological principles are also built into the ATS for the regulation of negative affect. A pedagogical based approach is also adopted in this study for the design of the programming exercises and the hints. To close the loop, we conducted qualitative evaluation using focus group discussions to gauge the ATS’s effectiveness and its acceptance by the students.

Proposed architecture

Figure 1 shows the proposed architecture of the ATS. It comprises of the Sensor Subsystem, the Student Subsystem and the Tutoring Subsystem. The functions of each subsystem are detailed in the sections below.

Sensor subsystem

The Sensor Subsystem acquires and processes the raw data from the various sensors e.g. web camera for facial cues, input devices such as keyboards and mouse for key-strokes and mouse clicks and eye trackers for eye tracking patterns. All the sensor logs will be consolidated and processed to derive a list of relevant features. The processed features are then passed to the Affect Inference module for inference of the student’s affective states.

For this study, our considerations for the choice of sensors are their unobtrusiveness and ease of deployment. Previous studies have tapped on input modalities ranging from

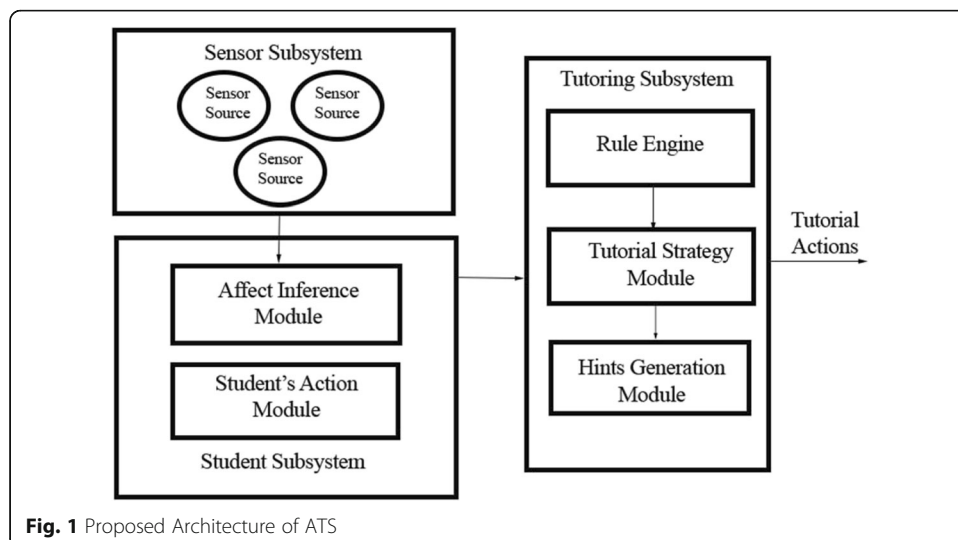


Fig. 1 Proposed Architecture of ATS

facial expressions (Kapoor, Burlison, & Picard, 2007; Kapoor & Picard, 2005; Teeters, El Kaliouby, & Picard, 2006; Yeasin, Bullot, & Sharma, 2006), body postures (Coulson, 2004; Mota & Picard, 2003), physiological (AlZoubi, Calvo, & Stevens, 2009), text (Shaikh, Prendinger, & Ishizuka, 2008) and input devices (Bixler & D'Mello, 2013) for affect sensing. Some of these sensors e.g. physiological sensors require probes to be placed on parts of the subject's body which subjects may find to be obtrusive. Being overly conscious of the presence of these sensors will in turn affect the reliability and validity of the measurements. Most of the physiological sensors also require an adequate level of technical expertise to setup which further constraints their deployment in the classrooms. In contrast, in this study, our proposed array of sensors comprising of web cameras, keyboards and mouse are common commodity devices that are unobtrusive and readily available in most learning environment.

Student subsystem

The Student Subsystem consists of both the Affect Inference and the Student's Action modules. The Affect Inference module will be in charge of inferring the affect of the student. This inference can be achieved through the use of machine learning algorithms that mine through the sensor logs to identify common patterns. The inferred affect will be passed to the Tutorial Strategy module for formulation of an appropriate tutorial strategy that regulates the inferred affect.

The Student's Action module logs the history of student's interactions with the ATS. The interactions that are logged can include the duration of time spent on an exercise, the topic and exercise the student is working on and has completed, date and time of compilations, the number of errors encountered for each compilation attempt, the date and time a hint is provided and the number of hints provided to each student. The interaction logs are in turn passed as input to the Affect Inference module for inference of the student's affect.

Tutoring subsystem

The inferred affect of the student obtained from the Affect Inference module and the history of interactions of the student obtained from the Student's Action module are passed on to the Tutorial Strategy module within the Tutoring Subsystem. A Rete algorithm based rule engine is embedded within the Tutorial Strategy module. The proposed design of the tutoring response strategy rules based upon the coping strategies elaborated in the related work section is shown in Fig. 2. The tutoring strategy rules are encoded within the rule engine. Using a rule engine offers the ease of modification of the strategy rules by non-programmers as the rules are stored in files externalized from the ATS.

One principle of design of the help system provided by the tutor is that it must be contingent on the students' needs (Wood & Wood, 1999). When students make a mistake, it is difficult for the system to ascertain whether it is due to slips, misconceptions or a lack of knowledge (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003). If we offer only hint that points at a generic concept or area to look into, weaker students would most likely be still stuck. On the other hand, if we provide hint that is so detailed that it almost gives the answer away, the more proficient students who slip occasionally

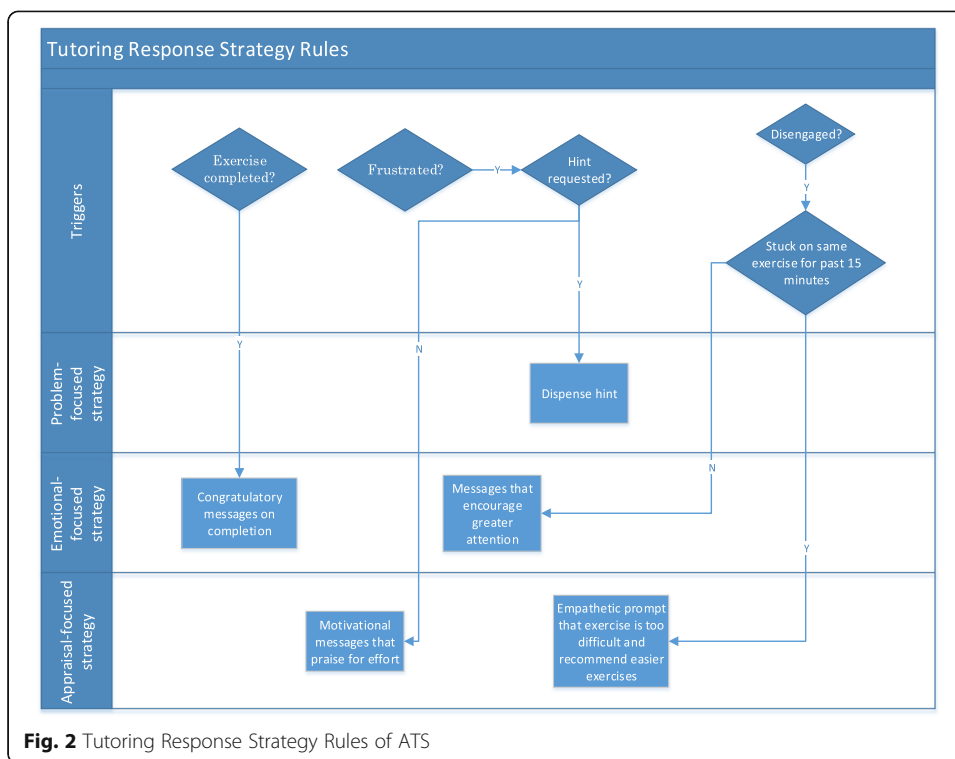


Fig. 2 Tutoring Response Strategy Rules of ATS

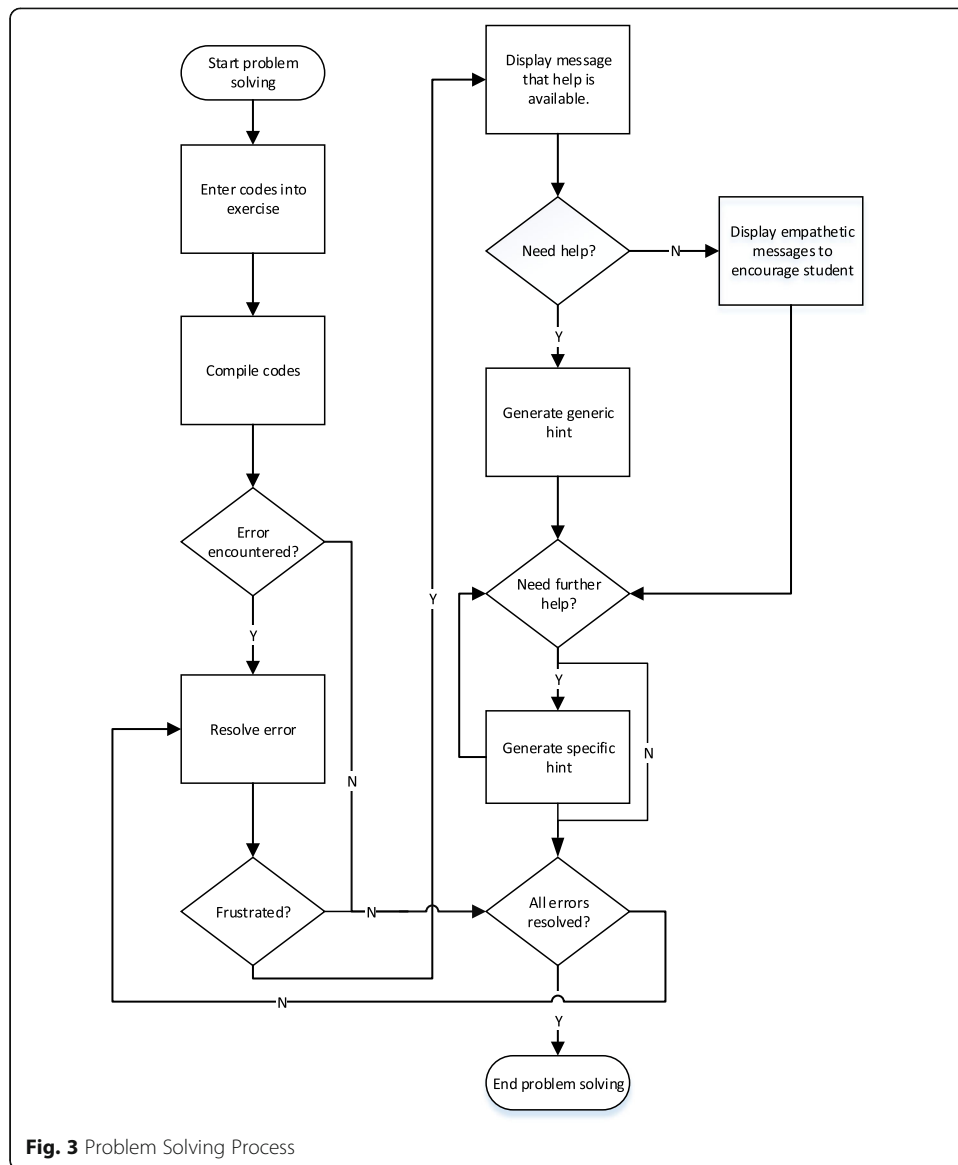
would not be sufficiently challenged and may also be bored or frustrated by the need to go through an elaborate explanation text when they already know the concept. This justifies putting the request for help under the student’s control and devising hints that progressively increase in level of details.

In addition, by opting for the provision of multiple levels of help, students who slip would not have to go through the more detailed explanations that are meant for students with a genuine lack of knowledge. For the students who are genuinely stuck due to knowledge gaps, the help provided would have to be progressively more explicit until the student is able to resolve the problem.

The first instance when students request for hint, a generic hint relating to the topic would be displayed. For subsequent help requests, detailed hints would be provided to the students. The detailed hints are bottom out hints that not only direct the students on the steps to be taken to rectify the error but also provide an explanation on the error.

To illustrate, if the student uses an assignment operator (“=”) instead of relational equals to operator (“==”), instead of just outputting “replace = with == in line 12”, the full explanation text “The assignment operator (=) is used to assign a value to a variable and a relational equals to operator (==) in a==b is used to check whether the variables a and b equal in values. == should be used in place of = in line 12.” will be shown instead. This will help address the knowledge gap of the student as the detailed explanation enhances the student’s understanding on the rationale for the rectification steps.

The flow of the student’s problem solving process is shown in Fig. 3. Students will first key in codes into the missing lines within exercises in the ATS. The students will then compile the codes within the ATS and when errors are encountered, students will try to debug and rectify the errors on their own. Some students may experience frustration after repeated futile attempts to rectify the errors. The regulation of negative affect



such as frustration will be handled by the tutoring response strategy rules within the Tutoring Subsystem.

To illustrate, for the currently configured tutoring response strategy rules, the ATS upon detecting that the students are frustrated, will initiate the display of prompts on screen to remind the students that they can request for hints to resolve the errors. For the version of ATS that was deployed for the trial, these prompts are faded in from the side of the screen and automatically faded out after 8 s. Our design consideration is that the prompts should not be disruptive to the students. In a study by Robison, McQuiggan, and Lester (2009), they concluded that the risk of students reacting negatively to incorrect affective interventions may be high for some affective states and caution should be exercised to avoid unintended effects of inadequate interventions. In consideration of these, we adopted a less disruptive and non-interruptive way of displaying the reminder prompts where students do not have to click on a button to close the prompt. At any instance, the students can also request for hints on their own without the reminder message by clicking

on the request for hint button. If the students did not request for hints for 10 s after the system detected that they are frustrated, the system will fade in an empathetic message encouraging the students to continue resolving the errors on their own. This would be an appraisal-focused response (Weiten et al., 2011) that alleviates the frustration by diverting their attention away from the negative affect itself.

Other emotional-focused and attribution based tutoring responses include the display of congratulatory message when the student successfully complete an exercise, attributing the success to their positive effort. Conversely, when the ATS detects that the student is disengaged and has worked on the designated programming exercise for less than 15 min, prompts reminding the student to devote more attention to the task and the need to invest sufficient effort would be displayed. If the disengaged student has already worked on the designated programming exercise for more than 15 min, the ATS would then recommend that the student work on an exercise with a lower difficulty level. This is an example of a problem-focused response strategy.

Design of Programming Exercises

The Cognitive Load Theory (Sweller, 1988) provides the framework for presentation of instructional information to optimize the cognitive load which will in turn lead to enhanced intellectual performance. It theorizes that if the total cognitive load exceeds the available working memory capacity, learning would not be able to occur. An effective instructional design reduces the cognitive load which puts less strain on the working memory requirements. The freed working memory in turn allows the learner to acquire other more advanced schemas which leads to improved learning when repeated over many cycles.

Computer programming is notoriously difficult to learn and novice learners face a range of difficulties ranging from misconceptions on the programming constructs to failure to translate their intention into logical programming codes. The construction of a computer program requires the programmer to retrieve pre-existing knowledge on the program syntax and semantic in their working memory and then devise a plan for combining these together to fulfil a function. Working memory has in fact been established to be a good predictor of programming skill acquisition in a previous study by Shute (1991).

For a novice learning programming, the cognitive load of having to memorize the syntax and semantics of the various programming constructs may overwhelm their working memory capacity. A technique to reduce the cognitive load employed in this study is the fill in the gap programming model. The fill in the gap programming model is employed for the exercises within this ATS where students are required to fill in missing lines of codes within a provided code template for each exercise. A screenshot of a sample exercise within the ATS is shown in Fig. 4. Using a fill in the gap model not only affords the reduction in the cognitive load of the novice students but also helps the students to overcome the initial inertia of not knowing how to start writing the program – a common issue that is encountered by students learning how to program.

Evaluation methodology

In this study, both qualitative and quantitative evaluations are used for evaluation of the proposed ATS. I conducted both an experimental study and a series of focus group discussions to evaluate the effectiveness and acceptance of the proposed ATS.

Fig. 4 Fill in the gap programming exercise

For the experimental study, the effectiveness of the ATS for the students' learning was evaluated. As the proposed ATS is to be used as a Java programming tutor, to evaluate its effectiveness, I seek to measure the learning achievement of the students during the trial. Specifically, the learning achievement is defined as the number of exercises completed, the duration of time taken to complete each exercise, the number of exercises attempted and the number of compilation required to complete each exercise. These can be derived from the students' action logs that are captured within the Student's Action Module elaborated in the section on Proposed Architecture. The intuition is that if the ATS is effective, the students would be able to attempt and complete more programming exercises and complete them in a shorter amount of time with lesser compilations. Our hypothesis for this study is thus the full affective version of the ATS results in more effective tutoring as compared to the version with the affective function disabled.

This measurement of learning achievement through tracking the number of programming exercises completed and the time taken to complete them was also adopted in the study by Corbett and Anderson (2001). In his study, he further concluded that test achievement was strongly related to the set of problems the students successfully solved and the solution paths that the students follow had little or no impact (Corbett & Anderson, 2001), thus validating our approach to focus on the outcomes and not the process of problem resolution for measurement of learning achievement.

Experimental study

In an experimental study, a researcher actively manipulates which group receives the designated treatment in a randomized control trial and then records the outcome.

A total of 39 students participated in the experimental design for evaluation of the effectiveness of the ATS on their learning. The 39 participants were randomly assigned to the control ($n = 21$) and test group ($n = 18$). All the participants were first briefed and then requested to sign consent forms indicating their consent to participate in the trial and for collection and publication of data captured in the trial. They were also briefed on the basic functions of the ATS and that they will be required to complete as many Java exercises as possible in the ATS for an hour in the trial. Each of them was

assigned a user id and password to login to the ATS. Participants in the control group worked on a version of the ATS without the affect related functions while participants in the test group worked on the full affective version.

The student participants were 18 years old on average and were undertaking an Information Technology diploma in Nanyang Polytechnic, Singapore. They had undertaken a basic course in Java for a semester before the trial. In this trial, no pre-test was administered as the student participants were matched in their programming experiences.

Focus group discussion

Focus group discussion is a group interviewing technique relying on interaction within the group based on topics that are selected by the researcher (Morgan, 1996). The participants within the focus group are usually selected by the researcher. In a focus group, the interaction among the members of the group allows the researcher to extract the attitudes, feelings, beliefs and experiences of the members which may not be feasible with the use of other research methods.

In this research, the developed ATS will likely be deployed in a classroom to complement a human tutor in the teaching of the designated subject area. The use of focus group discussions facilitates the retrieval of a range of students' perceptions on the usability and acceptance of the system by the students. In addition, it allows for the extraction of genuine feelings and opinions of the students, though on the condition that the environment is comfortable and conducive for them to voice their views. In a focus group discussion, it is envisaged that a group's capability would be more than the sum of its parts, thus generating more opinions than what can be obtained from the sum of opinions from each individual interview.

In our context, a total of 4 focus group discussions were conducted with the objective of evaluating the effectiveness of the ATS. Twenty-eight students (different from the students involved in the experimental trial) in total were recruited for the focus groups with 6, 7, 8 and 7 students for each of the 4 focus groups.

The participants were asked to trial the ATS for an average of 30 min before the start of the focus group discussion to familiarize them with the functions of the ATS. The focus group discussions were next held in classrooms with the participants seated in a circle. Each focus group discussion lasted for an average of about 20 min. The participants were asked to self-introduce themselves at the start of the session. All the focus group sessions were conducted by a moderator and were tape-recorded. After each session, the tape recordings were then transcribed. The analysis and consolidation of the focus group discussions transcripts using the content analysis methodology is summarized in Fig. 5. Two moderators with an average teaching experience of 7 years each conducted the focus group discussion sessions.

The set of guiding questions for the focus group discussions are

1. How was your learning experience with the ATS?
2. Why do you feel that way about the ATS?
3. What do you like about the ATS?
4. What do you think can be improved for the ATS?

5. What do you think of the feature where the system senses your frustration and engagement and respond accordingly e.g. by providing you with hints if it senses that you are frustrated?
6. What else would you like to say about the tutoring experience or about the system?

These questions serve as a guide to the moderator and the moderator is free to rephrase and re-sequence the questions when he or she conducts the discussion session.

During the conduct of the session, the moderator tries to ensure that every participant has an equal opportunity of expressing his or her view on the topic and that the session is not dominated by any individual participant. After the session, the moderator thanks each participant for their participation and assures the participants that their views are valued and will contribute towards the enhancement of the ATS which will in turn benefit future users of this ATS.

Results

Experimental study

The descriptive statistics for the number of exercises completed, time taken to complete exercise, number of compilations and number of exercises attempted for the non-affective (control) and affective (test) groups are shown in Table 1.

The time taken to complete each exercise was compiled for each exercise completed by individual students, resulting in a total of 65 records for the non-affective group and 60 for the affective group. The number of compilations was recorded for each compilation attempt by each student, resulting in a total of 101 and 118 compilation attempts by the students in the non-affective and affective group respectively.

As the data from the experimental trial did not fulfill the assumption of normality, the Mann-Whitney U test was used to determine whether the students are able to achieve more with the use of the full ATS functionality. The Mann-Whitney test results showed that the time taken to complete each exercise ($U = 1406$, $n_1 = 60$, $n_2 = 65$, $z = -2.69$, $p < 0.05$ two-tailed) and the number of exercises attempted ($U = 83$, $n_1 = 18$, $n_2 = 21$, $p < 0.05$ two-tailed) were significantly different for the non-affective and affective group. On the other hand, the number of exercises completed ($U = 174$, $n_1 = 18$, $n_2 = 21$, $p < 0.05$ two-tailed) and the number of compilations ($U = 5907$, $n_1 = 101$, $n_2 = 118$, $z = -0.112$, $p < 0.05$ two-tailed) were not significantly different for the non-affective and affective groups.

Focus group discussions

Top rated functions

The top rated functions of the ATS that most of the participants found most beneficial to their learning were the hint function and the fill in the gap programming exercises. The students reflected that the programming exercises are beneficial to their learning as on the one hand, it allowed them to apply what they have learnt from the lessons and on the other, to assess the adequacy of their knowledge for the various topics. They also opined that filling in missing lines of codes in the exercises is less daunting than writing an entire program on their own. For some of the students, they recalled that frequently when writing a computer program from scratch, they do not know how to start. This is exacerbated by the cognitive load of having to interpret long lists of errors,

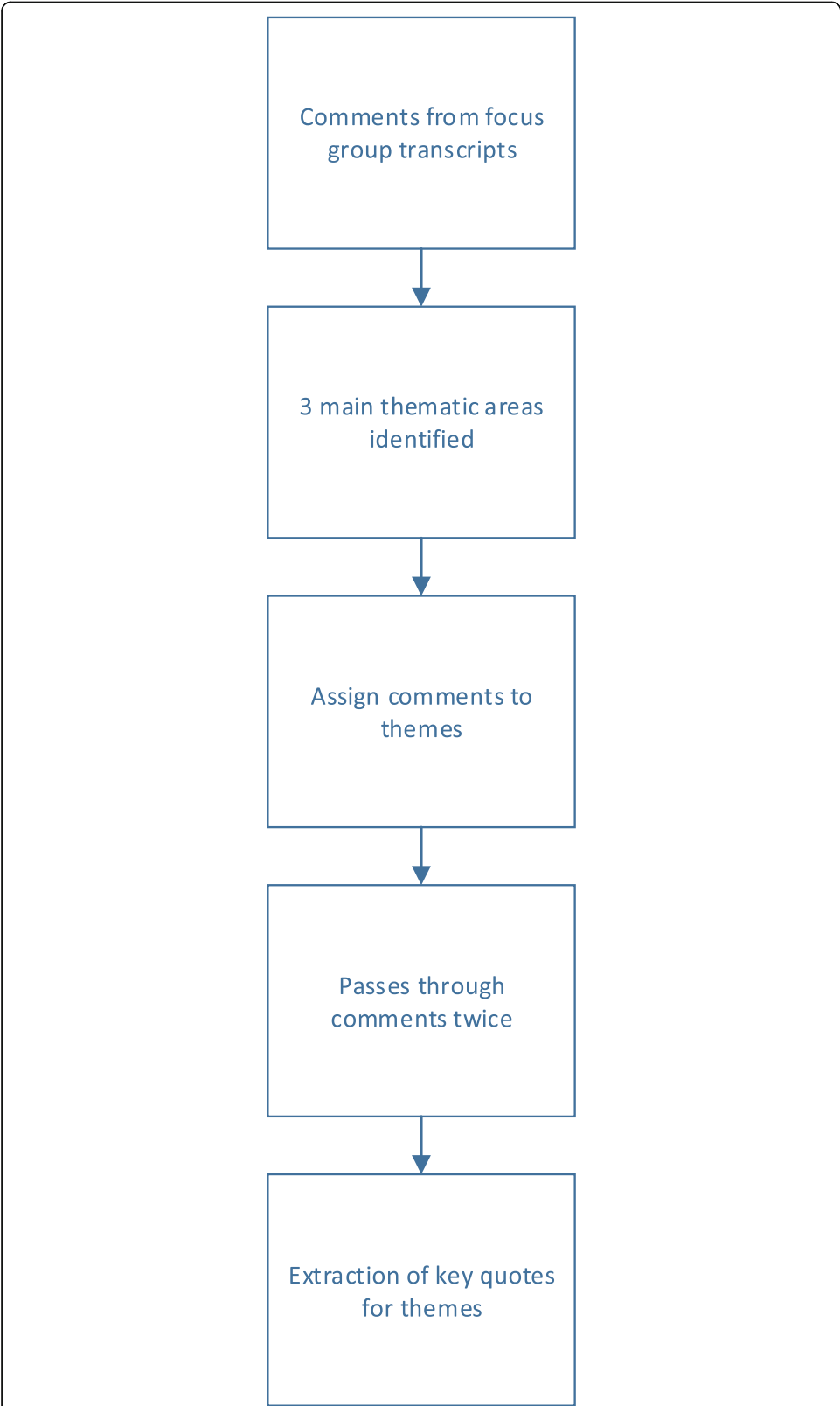


Fig. 5 Analysis of focus group discussion transcripts

Table 1 Descriptive Statistics for determinants of learning achievement

Groups	No of exercises completed		Time to complete exercise (seconds)		No of compilations		No of exercises attempted	
	Median	<i>n</i>	Median	<i>n</i>	Median	<i>n</i>	Median	<i>n</i>
Non-affective	3	21	831	65	7	101	5	21
Affective	3.5	18	470.5	60	13	118	7.5	18

which makes it difficult for them to identify the exact cause and location of the errors. They felt that the filling in of missing lines design of the exercises within the ATS substantially lessened the cognitive load and made it easier for them to pick up the relevant programming concepts. They also found that the hints are especially useful for students who are weaker in Java programming as the bottom-out hints tell them exactly how to correct the related errors. As the student aptly puts it, “The hint gives the answer e.g. put in the static keyword in line 12... the hint is the top feature and it can remind me of stuff that I missed out”.

In the trial, most of the students used the hint function although some of the students also reflected that they did not understand the hints. They expressed that the hints could be in a form that is easier to understand. The hints are formulated such that they not only offer the student steps on how to fix the relevant error but also a detailed explanation on the syntax or logical error. This appears to put some of the students off due to the length of the hint text. In the trials, the duration of time in which the students kept the hint message box opened was tracked and it was found that most of the students closed the hint message box within a few seconds, suggesting that they might be scanning the hints quickly without spending much effort to read and understand the accompanying explanation text. Some mechanisms may have to be built into the hint module to encourage the students to spend more time reading and contemplating about the provided hint and explanation to further enhance their understanding.

Usability

On the usability of the ATS, the students felt that the user interface was generally well designed. Some of the students reflected that they needed some time to get around the ATS and to familiarize themselves with the various functions of the tutoring system. This was despite the fact that they were given a brief introduction to the various functions of the ATS before the trial. These students felt that the user interface of the ATS can be more intuitive especially if it was to be used for their self-directed learning. One suggestion put forth by the students was that a pre-recorded video tutorial that introduces the students to the functions of the ATS would help.

The majority of the students across the focus groups also commented favorably on the use of the tutoring system for their learning. This can be concluded from the students' quotes: “It is more interactive as compared to programming on an Integrated Development Environment (IDE)”, “I wished I had this to learn in Year 1”, “This is better than Code Academy as the hints in Code Academy are not helpful”. From the quotes, the students supported their positive reviews of the tutoring system by comparing against existing alternatives that they had previously used for their learning e.g. Code Academy and Integrated Development Environment. Further clarifications with

the students also revealed that their perceived system interactivity features are the fade in congratulatory messages and reminder prompts.

Acceptance

The students were briefed at the beginning of the trial that the ATS will be equipped with web cameras, key and mouse click loggers that will sense their frustration and engagement through the capturing of their facial expression, keystrokes, mouse clicks, head postures and contextual logs. Most of the students in the focus group discussions, when told that web cameras will be installed to monitor their facial expressions and head postures were surprised. However, when the moderator explained that this was necessary to sense their frustration and engagement so that the ATS can respond with empathetically sensitive tutorial response, the students were able to accept the presence of the web cameras. Three out of the 28 students however, expressed that they are still uneasy about the presence of the recording web cameras. As one student says it: "It's like being watched... Why do you need to watch me?"

Some students were also skeptical of the ability of the ATS to sense their emotions. These were the words of one of the student - "If it senses well, then it's good but what happens if it didn't sense well Some people can code even when they are frustrated the question is how does it sense". It thus seemed that some students were not confident that the ATS can sense their emotions most probably because they do not understand how this emotion sensing mechanism works. They were also concerned with the consequences on their learning if the tutoring system should sense their emotions wrongly.

Discussion

The results of the experimental study show that the time taken to complete each exercise and the number of exercises attempted by the students differs significantly between the non-affective and affective groups. The number of exercises completed and the number of compilations were however not significantly different between the two groups. This implies that the use of the full affective version of the ATS enhances the efficiency (solving exercises in a shorter time) and persistence (expending effort in attempting more exercises) of the students.

It is surprising though that the number of exercises completed was not significantly different. Closer examination of the data further revealed that the students using the full affective version were able to solve the easier exercises in a shorter time but they were unable to resolve the more difficult exercises covering more advanced concepts such as arrays and functions. One possible explanation for this is that these difficult exercises were scheduled towards the end of the 1 h session and students were encouraged to solve the exercises in sequence from the fundamental to the more advanced level. It is plausible the students ran out of time for the resolution of the more difficult exercises which were attempted only towards the end of the session.

The number of compilations was not significantly different for the affective and non-affective groups. It is possible that more sophisticated measures that examines the compilation errors and the compilation behaviour of the students in resolving the errors between compilation attempts may be required (Jadud, 2006). As this version of

ATS did not capture the details of the compilation errors encountered by the students, this was not further investigated.

It is thus reasonable to support the hypothesis that the full affective version of the ATS results in more effective tutoring as compared to the version with the affective function disabled as the affective ATS does enhance the efficiency and persistence of the students in solving the programming exercises.

From the focus group discussions, the students were generally positive on their learning experience with the ATS. They were willing to use the ATS as a complement to face-to-face lessons conducted by their tutors and also as a form of e-learning. The students compared the proposed tutoring system with existing alternatives such as Code Academy and IDE and prefer the interactivity and tutoring aid provided by the proposed tutoring system. This also suggests that our affect based interventions in the form of empathetic messages and reminder prompts are rated positively by the students due to their interactivity effects. Thus, usability of the system is generally not an area of concern for the students.

The top rated functions that the students found most useful for their learning were the fill in the gap programming exercises and the hint function. The fill in the gap programming exercises lowers the cognitive load as students only have to fill in the missing lines of codes while the hint helps the students to resolve compilation errors that they are unable to solve on their own. However, the students felt that the explanation provided in the hint messages could be further enhanced to be more comprehensible and possibly shorter to entice them to read it (Anderson, Corbett, Koedinger, & Pelletier, 1995). The students further suggested that audio messages might help. Other than audio, animations and videos may also help in encouraging the students to spend more effort on comprehending the provided hints. Further research may be initiated in the future to investigate whether this would lengthen the reading time of the hints by the students and whether this would subsequently result in deep enhanced learning.

Some of the students expressed their discomfort with their actions being monitored with sensors such as web cameras even though they understood that the recorded logs are essential for the ATS to infer their emotions for the optimization of their learning. We are of the opinion that other than disclosing the purpose of recording the actions of the students within the ATS, more may have to be done to address the privacy concerns of some of these students. One possible solution will be to assure these students that the captured logs would not be used for purposes other than to enhance their learning. The data access and retention policy will need to be made known to the students as well so that they know who is using the data, the purpose it is used for and how the data will be stored and for how long. Another way to circumvent the privacy concerns of students would be to build in the option of allowing the students to opt out from being monitored by sensors for affect inference.

The students also raised concerns on the accuracy of affect sensing and the ramifications should the affect sensing be incorrect. To address this, we would suggest disclosing to the students that the tutoring response strategies formulated in response to the sensed affective states of the students as elaborated in the Tutoring subsystem section are mainly fail-safe. Thus, even if the students' affect is wrongly diagnosed, there would not be dire consequences on their learning.

S. K. D'Mello, Lehman, and Graesser (2011) proposed the use of affective messages, feedback with accompanying affective expression and dialogue moves for tutorial interventions. They evaluated the efficacy of Affective AutoTutor with an experimental study to evaluate its efficacy in promoting learning and engagement. The affective version of MetaTutor (VanLehn et al., 2014) also used motivational messages that addresses specific affective states of students. The evaluation study that was conducted however, did not use the affect sensors and detectors to determine the optimal moment for displaying the motivational messages and this possibly constraints the effectiveness of the affective interventions. In comparison, our study proposes pedagogical driven intervention measures which are activated at the moment when the relevant affective states of frustration and boredom are exhibited by the students. In addition, both qualitative and quantitative evaluation instruments are used in our study for evaluating the effectiveness and acceptance of the ATS by the students.

Conclusion

Incorporating empathy into an intelligent tutoring system enhances its tutoring effectiveness. In this study, we detailed the design, implementation and evaluation of an ATS. From the literature, we surmise that ATSs are seldom implemented and in addition, for the implemented ATSs, most tutored in the domain of mathematics and science and not for the domain of computer programming.

To cater for the generalizability and subsequent extension of the ATS for use in other tutoring domains, the various components of the ATS are designed to be loosely coupled so that individual component can be extended or modified without affecting other components. This study also addresses the gap in the literature where appropriate tutoring responses must be formulated in response to the inferred affect of the students to optimize their learning.

Santos (2016) reviewed 26 systems that leverages affect in an educational context and concluded that not many of the systems provide affective interventions. In addition, the few works (Grawemeyer et al., 2015; Santos, Saneiro, Salmeron-Majadas, & Boticario, 2014) that evaluated the effects of affective interventions in an educational context adopted a Wizard of Oz approach in contrast to our study which integrated affect sensing and intervention and evaluated the effects of the affective interventions in a tutoring system. Santos (2016) further acknowledged that devising autonomously the appropriate tutoring response to specific students' affective states (including when to intervene and what affective support to provide) is a challenging task and is an open research issue. This underlines the importance of our study in contributing to the body of knowledge for autonomous affect driven tutorial intervention.

The proposed ATS is evaluated for its effectiveness, usability and acceptance by the students using both the quantitative technique of experimental study and the qualitative technique of focus group discussions. The results for the experimental study support the hypothesis that the full affective version of the ATS results in more effective tutoring as compared to the version with the affective function disabled. The analysis of the transcripts for the focus group discussion further revealed that the students are generally positive on their learning experience with the ATS. The top rated functions from the students are the fill in the gap programming exercise and the hint function which offers explanation on top of the rectification steps.

This research also highlights privacy concerns with regards to the monitoring of students' behaviors and actions via sensors in an ATS. One suggestion is for options to be provided for students in an ATS to opt in or out of the monitoring but opting out will also mean that the empathetic tutoring response will be disabled as well. This privacy and usability conflict is an issue that will have to be resolved for widespread deployment of ATSs in classrooms and would be a potential area for future research.

Author's contribution

The author designed and conducted the experimental trials and focus group discussions, analyzed the results and wrote up the findings in this research paper. The author read and approved the final manuscript.

Funding

This work has been fully supported by Nanyang Polytechnic Capability Development Fund 8th Grant Call.

Availability of data and materials

The raw data in the form of audio recordings and sensor logs is not available for sharing with the audience of this publication as participants did not consent to the sharing of their raw data.

Ethics approval

Ethical approval was granted for this project on 30th Sep 2014 by the Newcastle University's IRB.

Competing interests

The author declares that he has no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 November 2017 Accepted: 18 July 2018

Published online: 24 October 2018

References

- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73(3), 277–320.
- Alexander, S., Sarrafzadeh, A., & Hill, S. (2006). *Easy with Eve: a functional affective tutoring system*, Paper presented at the workshop on motivational and affective issues in ITS. 8th international conference on ITS (pp. 5–12).
- AlZoubi, O., Calvo, R. A., & Stevens, R. H. (2009). *Classification of EEG for affect recognition: an adaptive approach AI 2009: Advances in Artificial Intelligence*, (pp. 52–61). Springer.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, 4(2), 167–207.
- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, 130(2), 224. <https://doi.org/10.1037/0096-3445.130.2.224>
- Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). *Self-regulated learning with MetaTutor: Advancing the science of learning with MetaCognitive tools New science of learning*, (pp. 225–247). Springer.
- Bixler, R., & D'Mello, S. (2013). Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. Paper presented at the 18th international conference on intelligent user interfaces (IUI'13).
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16.
- Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, 28(2), 117–139.
- D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H.,... Picard, R., (2008). AutoTutor detects and responds to learners affective and cognitive states. Paper presented at the Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems.
- D'Mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modelling and User-Adapted Interaction*, 20(2), 147–187.
- D'Mello, S. K., Lehman, B., & Graesser, A. (2011). *A motivationally supportive affect-sensitive autotutor New perspectives on affect and learning technologies*, (pp. 113–126). Springer.
- Damasio, A. (1994). *Descartes' Error: Emotion, reason, and the human brain*, (vol. 178). New York: Grosset/Putnam.
- Dolan, R. J., & Vuilleumier, P. (2003). Amygdala automaticity in emotional processing. *Annals of the New York Academy of Sciences*, 985(1), 348–355.
- Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1(1), 3–14.
- Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam Books.
- Grawemeyer, B., Mavrikis, M., Holmes, W., Hansen, A., Loibl, K., & Gutiérrez-Santos, S. (2015). The impact of feedback on students' affective states. Paper presented at the CEUR Workshop Proceedings.

- Gray, J. R., Braver, T. S., & Raichle, M. E. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences*, 99(6), 4115–4120.
- Greene, T. R., & Noice, H. (1988). Influence of positive affect upon creative thinking and problem solving in children. *Psychological Reports*, 63(3), 895–898. <https://doi.org/10.2466/pr0.1988.63.3.895>.
- Hascher, T. (2010). Learning and emotion: Perspectives for theory and research. *European Educational Research Journal*, 9(1), 13–28.
- Isen, A. M. (2000). Some perspectives on positive affect and self-regulation. *Psychological Inquiry*, 11(3), 184–187.
- Jadud, M. C. (2006). Methods and tools for exploring novice compilation behaviour. Paper presented at the Proceedings of the second international workshop on Computing education research.
- Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8), 724–736.
- Kapoor, A., & Picard, R. W. (2005). Multimodal affect recognition in learning environments. Paper presented at the proceedings of the 13th annual ACM international conference on multimedia.
- Kort, B., Reilly, R., & Picard, R. W. (2001). *An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion*. Paper presented at the advanced learning technologies. IEEE International Conference on.
- Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In *Computers as cognitive tools*, (vol. 1993, pp. 75–105).
- Linnenbrink, E. A. (2006). Emotion research in education: Theoretical and methodological perspectives on the integration of affect, motivation, and cognition. *Educational Psychology Review*, 18(4), 307–314.
- Morgan, D. L. (1996). *Focus groups as qualitative research*, (vol. 16). Sage publications.
- Mota, S., & Picard, R. W. (2003). Automated posture analysis for detecting learner's interest level. Paper presented at the conference on computer vision and pattern recognition workshop (CVPRW'03).
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91–105.
- Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, 14(2), 198–202.
- Picard, R. W. (1997). *Affective computing*, (vol. 252). MIT press Cambridge.
- Prendinger, H., Dohi, H., Wang, H., Mayer, S., & Ishizuka, M. (2004). *Empathic embodied interfaces: Addressing users' affective state Affective Dialogue Systems*, (pp. 53–64). Springer.
- Robison, J., McQuiggan, S., & Lester, J. (2009). *Evaluating the consequences of affective feedback in intelligent tutoring systems*, Paper presented at the affective computing and intelligent interaction and workshops, 2009. ACII 2009 (). 3rd International Conference on.
- Rowe, J., Mott, B., McQuiggan, S., Robison, J., Lee, S., & Lester, J. (2009). *Crystal island: A narrative-centered learning environment for eighth grade microbiology*, Paper presented at the workshop on intelligent educational games at the 14th international conference on artificial intelligence in education (). Brighton, UK.
- Sabourin, J. L., Rowe, J. P., Mott, B. W., & Lester, J. C. (2013). Considering alternate futures to classify off-task behavior as emotion self-regulation: A supervised learning approach. *Journal of Educational Data Mining*, 5(1), 9–38.
- Santos, O. C. (2016). *Emotions and personality in adaptive e-learning systems: an affective computing perspective Emotions and Personality in Personalized Services*, (pp. 263–285). Springer.
- Santos, O. C., Saneiro, M., Salmeron-Majadas, S., & Botcario, J. G. (2014). A methodological approach to eliciting affective educational recommendations. Paper presented at the advanced learning technologies (ICALT), 2014 IEEE 14th International Conference on.
- Shaikh, M. A. M., Prendinger, H., & Ishizuka, M. (2008). Sentiment assessment of text by analyzing linguistic features and contextual valence assignment. *Applied Artificial Intelligence*, 22(6), 558–601.
- Shute, V. J. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research*, 7(1), 1–24.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Teeters, A., El Kaliouby, R., & Picard, R. (2006). Self-Cam: feedback from what would be your social partner. Paper presented at the ACM SIGGRAPH 2006 research posters.
- Thompson, N., & McGill, T. J. (2012). Affective tutoring systems: Enhancing e-learning with the emotional awareness of a human tutor. *International Journal of Information and Communication Technology Education*, 8(4), 75–89.
- VanLehn, K., Burleson, W., Girard, S., Chavez-Echeagaray, M. E., Gonzalez-Sanchez, J., Hidalgo-Pontet, Y., & Zhang, L. (2014). The affective meta-tutoring project: Lessons learned. Paper presented at the International Conference on Intelligent Tutoring Systems.
- Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, 9(12), 585–594.
- Weiner, B. (1972). Attribution theory, achievement motivation, and the educational process. *Review of Educational Research*, 42(2), 203–215.
- Weiten, W., Dunn, D., & Hammer, E. (2011). *Psychology applied to modern life: Adjustment in the 21st century*, (10th ed.,). Cengage Learning.
- Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education*, 33(2), 153–169.
- Yeasin, M., Bullot, B., & Sharma, R. (2006). Recognition of facial expressions and measurement of levels of interest from video. *IEEE Transactions on Multimedia*, 8(3), 500–508.