Research and Practice in Technology Enhanced Learning
Vol. X, No. X (201X) page# start —page# end

AN EXPLORATORY STUDY OF FACTORS INDICATIVE OF AFFECTIVE STATES OF STUDENTS USING SQL-TUTOR

THEA FAYE G. GUIA, MA. MERCEDES T. RODRIGO, MICHELLE MARIE C. DAGAMI, JESSICA O. SUGAY, FRANCIS JAN P. MACAM

Ateneo Laboratory for the Learning Sciences
Department of Information Systems and Computer Science, Ateneo de Manila University
Loyola Heights, Quezon City, Philippines
theafayeguia@yahoo.com, mrodrigo@ateneo.edu, rhyzz_craig_08@yahoo.com, jsugay@ateneo.edu, f_macam@yahoo.com

ANTONIJA MITROVIC

Department of Computer Science and Software Engineering, University of Canterbury
Private Bag 4000, Christchurch, New Zealand
tanja.mitrovic@canterbury.ac.nz

The goal of this study was to model the affective states exhibited by students using SQL-Tutor. Based on current literature, we selected academic affective states of interest and measured their incidence among students during an SQL-Tutor session. We observed that students using SQL-Tutor most often exhibited engaged concentration, confusion and boredom; however, none of these states were correlated with student achievement on the final exam. Using D'Mello's Likelihood metric, *L*, we found that boredom and frustration tended to persist. We then correlated features extracted from SQL-Tutor log files with these two states' *L* values. We found that boredom was negatively correlated with the number of completed/attempted problems, and the number of constraints used. It was positively correlated with the average time needed to complete problems and the average number of attempts. Persistent boredom was negatively correlated with the number of solved problems and positively correlated with the mean time to solve problems and the average number of attempts per solved problem. Frustration was not significantly correlated with any of the factors, but persistent frustration was negatively correlated with the number of constraints used and positively correlated with the average number of attempts per solved problem.

Keywords: SQL-Tutor; observation; achievement; models of affect; boredom.

1. Affect and Learning

Since the 1960s, studies on human affect were left out of the scientific mainstream because affect was considered too vague and too subjective to be studied (Damasio, 2000). Feelings, emotions, moods, motivations, and other constructs that fell under the umbrella of affect were considered too random and not trustworthy; they were seen as the opposite of reason. It was only in the 1990s that more and more researchers in various disciplines including cognitive science, computer science, and engineering began studying emotion in detail. Damasio (2000) and others showed that people who suffered from head injuries or tumors that impaired their emotions continued to possess their rationality but lost their ability to make personal and social decisions. They concluded that reason leverages on emotions and that the absence of emotions hampers judgment at least as much as an excess of emotion does.

In more recent years, learning scientists have taken a keen interest in drawing links between affect and learning. Indeed, researchers have documented a dynamic relationship between affect and learning in which some affective states are antecedents of learning outcomes, while others are consequences of these outcomes (D'Mello et al., 2009). Positive emotions or affective states have been shown to increase learning intensity (Tempelaar et al., 2012), broaden attention (Fredrickson & Branigan, 2005), promote flexibility (Dreisbach & Goschke, 2004), and reduce perserveration (Dreisbach & Goschke, 2004). Unfortunately, positive emotions also have also been shown to lead to increased distractibility (Dreisbach & Goschke, 2004).

There is a growing interest in the literature regarding "negative" affective states, i.e. those with the potential to negatively influence learning and may require intervention. Boredom, for example, is an unpleasant, transient affective state in which the individual feels a pervasive lack of interest in and difficulty concentrating on the current activity (Fisher, 1993). It denotes a detachment from the current situation and a preoccupation with distraction activities such as daydreaming, doodling, texting, or passing notes, just to pass the time (Breidenstein, 2007, Mann & Robinson, 2009). Students experience boredom when they are either under-challenged or over-challenged by a task. In both situations, students focus on the tediousness and meaninglessness of what they are doing as well as on feelings of dissatisfaction and frustration. When students are over-challenged, though, boredom is often accompanied by anger, anxiety, hopelessness, and shame (Acee et al., 2010).

The effects of boredom on learning outcomes are unequivocally negative. Boredom tends to inhibit performance (D'Mello et al., 2009), elaboration, and metacognition (Artino & Jones, 2012). Boredom is related positively to attention problems, negatively to intrinsic motivation, effort, and self-regulation (Pekrun et al., 2010). Boredom has been negatively correlated with the use of adaptive learning strategies (Artino & Jones, 2012). In studies of students using intelligent tutors, boredom has been associated with lower student performance (Lagud et al., 2010; Rodrigo et al., 2010; Craig et al., 2004). Previous study by Baker et al. (2010) showed that across different learning environments, boredom was the only state that tends to lead the students to "game the system", an attempt to advance through the curriculum by taking advantage of the weaknesses and limitations of a computer-based learning systems, without actually learning the material (Baker, 2004). Adding to boredom's insidiousness is that it has been shown to persist regardless of learning environment (Baker et al., 2010). A student who is bored tends to stay bored.

Frustration, on the other hand, is defined as “an interference with the occurrence of an instigated goal-response at its proper time in the behavior sequence” (Dollard et al. in Berkowitz, 1989). When a person is actively striving to reach an objective and is thwarted or blocked, that person experiences frustration. Researchers such as Kapoor et al. (2007) are interested in frustration because it is a potential mediator for disengagement. However, evidence presented by Baker et al. (2010) shows that frustration does not lead to non-learning behaviors such as gaming the system. Frustration, therefore, is not as worrisome an affective state as boredom.

Recognizing the role that affect has in learning, researchers have attempted to identify factors that might be indicative of affective states and build automated detectors of student affective states. In some cases, these models have been incorporated into Intelligent Tutoring Systems (ITSs), enabling these systems to respond to these states in an intelligent manner. Many older ITSs tended to track student cognition and lacked the ability to recognize, respond and adapt to students’ emotional experience. The purpose of this paper is to develop models of affective states exhibited by students using SQL-Tutor. The development of such models and, later, their integration into ITSs broadens the ways in which tutors can interact with and support learners. In specific, we attempt to answer the following research questions:

- To what extent do students exhibit the affective states of interest? Which of these affective states, if any, correlate with learning?
- Which of these affective states, if any, are persistent?
- Which of these persistent affective states, if any, correlate with learning?
- Which SQL-Tutor log data features correlate with these affective states? With these persistent affective states?
- What linear regression model best describes the relationship between these features and these affective states?

2. Past Work on Identifying Indicators of Affect

Past research has identified some indicators of student affect. For example, practical predictors of learner engagement already exist in educational systems. Frequent usage of a learning management system for long periods determined engagement and student success (Beer et al., 2010). At a coarse-grained level, engagement was determined by the number of clicks, number of pages visited, time spent on site, number of posts/replies to discussions.

McQuiggan and Lester (2006) identified student self-efficacy while using an online genetics tutorial and problem-solving system by using biofeedback (i.e. heart rate and galvanic skin response) as well as student-tutor interaction logs as inputs. They found that students who have high self-efficacy are those whose heart rates gradually drop as they encounter new questions. On the other hand, students who have low self-efficacy are those whose heart rates spikes dramatically when he/she selects an incorrect answer, without knowing the feedback.

Lagud and Rodrigo's research (2010) with Aplusix, an ITS for Algebra, studied student boredom, confusion and engaged concentration, a subset of flow (Csikszentmihalyi, 1990) using human observations and interaction logs. They found that students who were most bored and confused took the most time and the highest number of steps in solving problems. Those who experienced the most engaged concentration are those who used the least number of steps. Further analysis of the same data set attempted to identify early indicators of boredom (Dagami et al., 2011). Results showed that the precursors of student boredom were frequent usage of "ask" buttons (i.e. *ask for verification*, *ask for solution*, and *ask for score*), less use of number keys, and less use of special buttons (e.g. <=, >= and <>).

Self-assessment surveys and interaction logs, which include mouse movement and control selection rates, within a TC3 performance assessment was utilized to passively classify student mood and performance (Sottilare and Proctor, 2012). Although the study was not able to find reliable predictors of arousal, it found that experience of pleasure and arousal are unstable while experience of dominance is stable over the session. A recent and innovative attempt to advance an ITS's sensitivity and responsiveness to affect was implemented in AutoTutor (D'Mello et al., 2008). Models of student affect were derived based on students' conversational cues, posture, and facial expressions. These models then informed AutoTutor's adaptive affective and cognitive response.

The literature has shown that analyses of overt student actions and physiological signals indicate what it is a student is feeling while working with educational systems. This investigation capitalizes on these earlier successes, attempting to identify log data features that might help detect the affective states of students using the SQL-Tutor.

3. Methods

This section gives an overview of our test bed, SQL-Tutor, target population, and methods for data gathering.

3.1 *SQL-Tutor*

Constraint-Based Modeling (CBM) represents knowledge about a domain as a set of constraints on correct solutions in that domain (Ohlsson & Mitrovic, 2006). Constraints evaluate and judge, rather than infer, and represent both domain and student knowledge (Suraweera et al., 2009). The domain model consists of constraints on basic principles and concepts underlying the domain while the student model is represented by set of constraints that have and have not been violated (Mitrovic & Ohlsson, 1999). An ITS built using this approach is called a constraint-based tutor (CBT).

CBTs can be distinguished from the other types of ITSs by the process of knowledge representation. This kind of tutor tries to overcome the complexity of the formal knowledge representation that ITS researchers have inherited from the field of artificial intelligence. Other ITSs require extremely detailed and accurate models that follow the student step by step and compare each student's step to the incorrect and correct steps generated as possible next moves by the system. The system can then provide immediate feedback on the student's problem-solving step. A CBT, on the other hand, uses constraints to simplify and limit this specificity of modeling process (Mitrovic & Ohlsson, 1999). Every constraint identifies a feature of correct solutions, and at the same time specifying implicitly all the solutions that violate it as incorrect solutions. A CBT evaluates the student's solution against constraints, and if there are no violated constraints, the solution is deemed correct. In the case of constraint violations, the tutor provides feedback.

SQL is currently the most prevailing database language for querying relational databases; it is an essential topic in introductory databases courses in higher education (Dekeyser et al., 2007). Despite its limited set of commands, SQL is not a simple language to master (Renaud & Biljon, 2004). Several researchers have identified a number of common difficulties suffered by students while learning SQL (Dekeyser et al., 2007; Mitrovic, 1998). For this reason, tools have been suggested by various institutions to be able to overcome the identified problems associated with learning SQL. These tools, which are in the form of ITSs, provide a simple environment for students to write and test queries against databases. Some of these are SQLify (de Raadt et al., 2006), Acharya (Bhagat et al., 2002), the SQL automated tutoring system of Kenny and Pahl (2005), and SQL-Tutor (Mitrovic, 1998). We aim to determine the factors that could be detectors of affective states of students through learning indicators from SQL-Tutor.

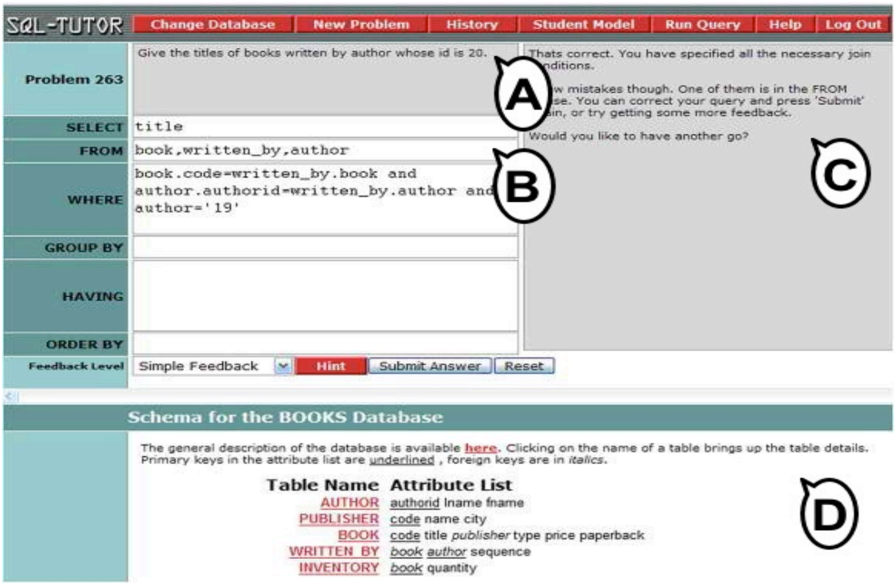

Figure 1. A screenshot of SQL-Tutor main window with its parts: (a) problem text, (b) workspace for composing SQL queries, (c) feedback panel, and (d) the database schema.

As mentioned in the introduction, the test bed for this study was SQL-Tutor. Figure 1 shows the problem-solving interface providing the problem text, solution structure, feedback panel and information about the database schema. After logging in, the student is presented with an initial screen that gives information about how to use the system. SQL-Tutor provides help about specific aspects of the system through help menus and tool tips. The student can select a database to work on, and also problems within the selected database. The system can also present system-selected problems appropriate for the student on the basis of his or her knowledge state and learning capabilities. The student would not be interrupted as he or she works on a query. Analysis of the student's solution starts immediately when he or she submits the solution. And if the solution has mistakes, the system will present appropriate feedback.

Students can obtain the descriptions of databases, tables or attributes by selecting appropriate options from the Help menu, or by directly selecting table/attribute names. The tutor assumes that the students have had some preliminary exposure to SQL. However, the Help menu provides descriptions of various clauses and elements of SQL such as function, expressions, predicates and operators. At the moment, the system covers only SELECT statement of SQL, but the same approach could be used with other statements.

3.2 *Population*

For this study, a total of 74 college juniors, aged 18 to 20, in three sections of MIS 21: Introduction to Applications Development class at the Ateneo de Manila University used SQL-Tutor. Because of manpower constraints, though, only 60 of these 74 were observed. Part of the MIS 21 course introduces students to database programming, thus the use of SQL. Students did not have any collegiate-level SQL or database training prior to MIS 21. None of the participants had used SQL-Tutor before. The research team briefed the students on the use of SQL-Tutor for about 10 minutes. No other formal training was provided.

The research team first informed the participants that using SQL-Tutor would help them practice formulating SELECT statements. They then asked the participants to use the tutor for 60 minutes. The students were not given a fixed number of problems to solve. Furthermore, the students did not necessarily solve the same problems as the tutor recommended subsequent problems based on student performance.

3.3 *Affect Observations*

As the students used SQL-Tutor, the research team made them aware that they were going to be observed. Affect was assessed using a quantitative field observation protocol first described in Rodrigo et al. (2007), refined across several studies (e.g. Rodrigo et al., 2008a, 2008b, 2009) and discussed in full detail in Baker, D'Mello, Rodrigo, & Graesser (2010). Quantitative field observations are one of several methods used to collect affect data from subjects. Others include the use of video annotations (D'Mello, Picard & Graesser, 2007), screen replay annotations (De Vicente & Pain, 2002), automatic detection using sensors (D'Mello & Graesser, 2009; McQuiggan & Lester, 2006), and others.

Many of the methods mentioned require equipment and are therefore difficult to scale transfer into actual classroom settings (Rodrigo, et al., 2011). Unless experimenters have multiple video cameras and/or sensors, data will have to be captured one student at a time. Annotation takes considerable time. Video also captures a more limited set of information than live observations. Live observations give observers the opportunity to shift positions to gain a better vantage point, and hence a better approximation, of student affect and behavior. We acknowledge that the technique is not without limitations. Unlike with video, playback is impossible. Furthermore, it is subject to observer fatigue. Nonetheless, it was deemed the best method for this study, given that the researchers had only one class period to the data and no access to sensors.

The observations were carried out by a team of four observers who worked in pairs. The observers were composed of two Masters Students and one Undergraduate Student in Computer Science, and one assistant instructor. The assistant instructor was highly experienced in field observations, having participated in several such studies in the past. The two graduate students and one undergraduate student were in training for this type of fieldwork as it was related to their theses.

The selection of affective states of interest is the subject of much discussion in the literature. Affect is a broad concept that encompasses numerous and diverse human emotional and cognitive experiences. Ekman (1992) enumerates the affective states amusement, anger, contempt, contentment, disgust, embarrassment, excitement, fear, guilt, pride in achievement, relief, sadness/distress, satisfactions, sensory pleasure and shame. Ortony, Core, and Collins (1988) organize these and other states according to the contexts in which they occur—the suffering of a loss, the well-being of others, surprises, and so on. Not all of these states, though, are relevant to learning and the identification of relevant affective states is a subject of continuing research. Craig et al (2004) suggested that the key learning-related affective states are boredom, confusion, eureka, flow, and frustration. D'Mello et al (2005) later amended this list to boredom, confusion, delight, flow, frustration, neutral and surprise. Research since then has suggested that, of this set, boredom and confusion are particularly strongly related to learning (Baker, D'Mello, Rodrigo, & Graesser, 2010). This study makes use of D'Mello's (2005) list of seven academic emotions. Observer training consisted of a pre-observation discussion on the meaning and manifestations of seven these categories of interest:

(1) **Boredom (BOR)** – slouching, resting the chin on his/her palm; statements such as "Can we do something else?" or "This is boring!"

(2) **Confusion (CON)** – scratching his/her head, repeatedly looking at the same interface elements; consulting with a fellow student or the teacher; looking at another student's work to determine what to do next; statements like, "I'm confused!" or "Why didn't it work?"

(3) **Delight (DEL)** – clapping hands; laughing with pleasure; statements such as, "Yes!" or "I got it!"

(4) **Engaged concentration (FLO)** – immersion, focus, and concentration on the system; leaning towards the computer; mouthing solutions; pointing to parts of screen. Engaged concentration can be considered a subset of the construct of flow proposed by Csikszentmihalyi (1990).

(5) **Frustration (FRU)** –banging on the keyboard or mouse; pulling his/her hair; deep sighing; statements such as, "What's going on?"

(6) **Surprise (SUR)** – sudden jerking or gasping; statement such as "Huh?" or "Oh, no!"

(7) **Neutral (N)** – coded when the student did not appear to be displaying any of the other affective states or when the student's affect could not be determined for certain

The observers practiced the coding scheme during an unrelated observation prior to this study. Prior to the observation period, each pair of coders was assigned to ten randomly selected students and the sequence in which they would observe the students was established. When the observation period began, a timed PowerPoint presentation was played to synchronize the observers and tell them which student they should be watching next. Each pair of coders observed the same student for twenty seconds. The observers were standing diagonally behind or in front of the student being observed. The observation was conducted using peripheral vision and surreptitious glances to minimize intrusion with students' natural emotions while still obtaining good affect observations. The observation period lasted for 60 minutes in each class.

Observers learned to recognize facial expressions, gestures and utterances of students that implied these emotions. The affective states of interest were not mutually exclusive—confusion and flow, for example, can occur together, as can delight and surprise. Furthermore, a student can exhibit more than one affective state during the 20 second observation period. For tractability, however, the observers only coded the first affective state they saw per observation.

We then computed inter-rater reliability. When inter-rater reliability was low, the observers discussed the points of contention to resolve differences in subjective judging criteria. We conducted another practice round of observations and repeated the reliability checking process until observers reached an acceptable level of agreement.

During the actual data gathering, each pair of observers was assigned to 10 students from each class, for a total of 20 students observed per class or 60 for the three sections. Data from four students was eventually removed because their data was incomplete. The final data set was therefore composed of data from 56 students across three sections.

Eighteen pairs of observations were collected per student. The computed Cohen’s (1960) k during actual observation was 0.94 for affect, which is considered to be a high level of agreement.

3.4 *Learning Indicators*

Learning science researchers used a variety of student-interaction log features as indicators of student learning. The summary in Table 1 reveals some commonalities: time, number of attempted or solved problems and number of errors carry implications about what a student does and does not understand. After studying Table 1, we arrived at a list of learning indicators for SQL-Tutor (Table 2) and distilled these features from the interaction logs.

Table 2. Features distilled from SQL-Tutor.

Feature	Description	Mean	Standard Deviation
SolvedProblems	Number of solved problems	14.36	5.89
AttemptedProblems	Number of attempted problems	16.54	6.58
LearnedConstraints	Number of constraints students learned during the session	12.86	5.28
ConstraintsUsed	Total number of constraints students used	205.32	40.02
SeenMessages	The number of constraints they have seen hints about – as feedback	171.86	74.38
NumOfLogins	Number of sessions or logins	1.86	1.25
TotalTime	Total interaction time in minutes	76.05	28.70
AvgTimeToSolve	Average time per solved problem in minutes	6.22	3.05
TotalAttempts	Total number of attempts (submitted solutions)	81.63	33.09
AvgNumOfAttemptsPer SolvedProb	Average attempts per solved problem	6.50	3.49
Exam Scores	Scores in SQL part of the paper-based exam	33.05	9.81

	(probability of constraint violation %) vs. number of times relevant
CAPIT (Mayo et al., 2000)	- Learning curve: a. Number of errors (constraint violations %) vs. number of solution attempts; or b. Mastery of Constraint C: number of times violated (probability of constraint violation %) vs. number of times relevant; or c. Rate of Mastery of the Target Skill as a Whole: proportion of students with zero errors vs. number of occasions to violate

Finally, to establish ground truth regarding student learning, student test scores were obtained from the participants' performance in SQL proficiency part on their final examination.

4. Results

Table 3 shows the incidence of the seven affective states observed. Engaged concentration was the most common, occurring 57.9% of the time. The second most observed state was confusion, observed 23.9% of the time. Boredom was third most observed with 8.1% of the time. It was followed by states of delight, neutral and frustration with 4.1%, 3.9% and 2.1% incidences respectively. Surprise was never observed.

Table 3. Incidence of affective states and correlation with achievement. None of the relationships were significant.

Affective state	Incidence	Correlation with exam scores
Engaged Concentration	57.90%	0.073
Confusion	23.90%	-0.006
Boredom	8.10%	-0.02
Delight	4.10%	-0.32
Neutral	3.90%	-0.262
Frustration	2.10%	0.15

When correlated with student performance on the SQL portion of the final exams, though, none of the correlations of the affective states with the exam scores were found to be significant (Table 3). Of the seven affective states, the relationship with the highest correlation was frustration with r = 0.15. However, this correlation is still considered to be weak.

We then turned our attention to the transitions among these states and whether some affective states tended to persist. An affective state is said to be persistent if a student is observed to manifest that state during two consecutive observations. We measured persistence using D'Mello's ***L*** (D'Mello et al., 2007). As discussed in Baker, et al. (2010), *L* provides an indication of the probability of a transition above and beyond the base rate of each affective category. *L* explicitly accounts for the base rate of each affective category when assessing how likely a transition is, giving the probability that a transition between two affective states occurs, and given the base frequency of the destination state. *L* has been used repeatedly, by many affect researchers (see Baker, et al., 2010; Forbes-Riley & Litman, 2012; Inventado, et al., 2011; McQuiggan, et al., 2008) to answer the question: Controlling for chance, how likely is it that a user will transition from one state A to another state B?

L is computed as shown in equation 1:

$$L = \frac{\Pr(NEXT \mid PREV) - \Pr(NEXT)}{(1 - \Pr(NEXT))} \tag{4.1}$$

This metric could be formally represented as L[PREV →NEXT], where PREV is the current affective state at time t_1 and NEXT is the next state at time t_{i+1} (D'Mello, 2007) An L value of 1 means that the transition will always occur; a value of 0 means that the transition's likelihood is exactly what it would be given only the base frequency of the destination state. Values above 0 signify that the transition is more likely than it could be expected (i.e. greater than the base frequency of the destination state), and values under 0 signify that the transition is less likely (i.e. less than the base frequency of the destination state).

For a given transition, we calculate a value for L for each student and then calculate the mean and standard error across students. We can then determine if a given transition is significantly more likely than chance (0) using the two-tailed t-test for one sample. The number of degrees of freedom for the two-tailed t-test is the number of students who were ever in the state minus one ($df = N - 1$). Students that never entered the state give no evidence on whether the state is persistent. As a consequence, the number of degrees of freedom varies among affective states within each learning environment.

We see in Table 4 that of all the affective transitions 11 were found to be statistically significant ($p <= .05$) and 2 were marginally significant ($.05 < p <= 0.1$). In the following findings, there are three numbers inside parentheses that are composed of: (1) the D'Mello *L* mean which is also indicated in Table 4, (2) the t-test value *t*, and (3) the significance *p*. Students who are bored tend to stay bored (L = 0.11, t(33) = 2.3, p = 0.03). They do not tend to transition to confusion (L = -0.16, t(33) = -3.02, p = 0.01), frustration (L = -0.01, t(33) = -1.96, p = 0.06), or neutrality (L = -0.04, t(33) = -1.67x10^{16}, p < 0.01). Confused students do not tend to transition to engaged concentration (L = -0.21, t(45) = -2.02, p = 0.05) or neutrality (L = -0.04, t(45) = -1.95x10^{16}, p < 0.01). Frustrated students do not transition to delight (L = -0.04, t(12) = -2.14x10^{16}, p < 0.01), engaged concentration (L = -0.67, t(12) = -2.56, p = 0.03), or neutrality (L = -0.04, t(12) = -2.01x10^{16}, p < 0.01). Rather, they tend to remain frustrated (L = 0.22, t(12) = 2.18, p = 0.05). Neutral students do not tend to transition to frustration (L = -0.02, t(12) = -7.25x10^{16}, p < 0.01). What we concentrated on, however, is the persisting affective states..

We focus the remainder of our analyses and discussion on the two persistent affective states—boredom and frustration. Although neither boredom nor frustration determined student achievement, their persistence merits attention as they may both influence the quality of students' learning experience. Boredom has been shown to lead to problem behaviors such as gaming the system (Baker et al., 2010). Frustration, on the other hand, can be a potential mediator for student disengagement (Kapoor et al., 2007). Also, research by Perkins and Hill (1985) suggests that boredom can be preceded by frustration.

Table 4. Transitions of affective states. Horizontal rows represent the antecedent affective states while the vertical columns represent the subsequent affective states. The first number on each cell is the D'Mello *L* mean, and the number on the parentheses is the standard deviation. Statistically significant transitions are shaded dark gray, and marginally significant transitions are shaded light gray. See affect definitions in Affective Observations part of Methods section.

	BOR	CON	DEL	FLO	FRU	N
BOR	0.11	-0.16	-0.02	0.23	-0.01	-0.04
	(0.27)	(0.31)	(0.08)	(0.81)	(0.04)	(0.00)
CON	0.06	0.07	0.02	-0.21	0.00	-0.04
	(0.29)	(0.34)	(0.13)	(0.71)	(0.06)	(0.00)
DEL	0.06	-0.02	0.06	-0.13	-0.01	-0.03
	(0.32)	(0.43)	(0.23)	(0.93)	(0.05)	(0.05)
FLO	-0.02	0.04	-0.01	0.05	0.00	-0.03
	(0.12)	(0.30)	(0.07)	(0.58)	(0.04)	(0.00)
FRU	0.02	0.18	-0.04	-0.67	0.22	-0.04
	(0.31)	(0.49)	(0.00)	(0.91)	(0.35)	(0.00)

When the features in Table 2 were correlated with all the affective states using IBM SPSS, results showed that only boredom could be predicted using the indicators of learning from SQL-Tutor data (See Table 5). The number of problems solved ($r = -0.31$; $p = 0.02$), number of attempted problems ($r = -0.28$; $p = 0.04$), number of constraints used ($r = -0.24$; $p = 0.07$) are negatively correlated with experience of boredom. Meanwhile, the average time to solve a problem ($r = 0.46$; $p < 0.01$) and average number of attempts per solved problem ($r = 0.50$; $p < 0.01$) are positively correlated with student boredom.

When the features were correlated with the two persistent affective states, we found that the features correlated with the persistent boredom are similar with those of the single instances of boredom. The number of solved problems ($r = -0.28$; $p = 0.04$) and number of attempted problems ($r = -0.26$; $p = 0.06$) were found to be negatively correlated with persistent boredom. Average time to solve a problem ($r = 0.48$; $p < 0.01$) and average number of attempts ($r = 0.57$; $p < 0.01$) per solved problem were found to be positively correlated with persistent boredom. Furthermore, several features were found to predict students' persistent frustration. The number of constraints used ($r = -0.27$; $p = 0.04$) was found to be negatively correlated, while the average number of attempts per solved problem ($r = 0.36$; $p = 0.01$) was found to be positively correlated with persistent frustration.

Table 5. Correlations between indicators of learning from SQL-Tutor data and affective states. The first column titles are the affective states, and the first row titles are the learning indicators. The first number on each cell is the correlation *r*, and the number on the parentheses is the significance *p*. Statistically significant relationships are shaded dark gray, and marginally significant relationships are shaded light gray.

Affect	*Solved Problem*	*Attempted Problem*	*Learned Constraints*	*Constraints Used*	*Seen Messages*	*NumOf Logins*	*Total Time*	*AvgTime ToSolve*	*Total Attempts*	*AvgNumOfAttempts PerSolvedProb*
BOR	-0.31 (0.02)	-0.28 (0.04)	0.06 (0.66)	-0.24 (0.07)	0.11 (0.40)	< 0.01 (0.97)	-0.04 (0.77)	0.46 (< 0.01)	0.08 (0.56)	0.50 (< 0.01)
CON	0.10 (0.46)	0.04 (0.79)	-0.01 (.94)	0.03 (0.82)	0.09 (0.50)	-0.18 (0.19)	-0.03 (0.84)	-0.25 (0.06)	0.07 (0.59)	-0.14 (0.29)
DEL	-0.08 (0.57)	-0.05 (0.71)	0.03 (0.83)	0.03 (0.84)	-0.10 (0.46)	0.09 (0.51)	-0.18 (0.20)	-0.13 (0.33)	-0.11 (0.43)	-0.11 (0.44)
FLO	0.06 (0.64)	0.13 (0.33)	-0.01 (0.96)	0.13 (0.32)	-013 (0.35)	-0.02 (0.91)	0.06 (0.66)	0.06 (0.67)	-0.10 (0.48)	-0.09 (0.49)
FRU	-0.18 (0.18)	-0.16 (0.23)	0.01 (0.92)	-0.17 (0.22)	0.05 (0.74)	-0.10 (0.44)	-0.05 (0.74)	0.14 (0.32)	< 0.01 (0.98)	0.26 (0.05)
BOR-BOR	-0.28 (0.04)	-0.26 (0.06)	0.03 (0.82)	-0.17 (0.20)	0.20 (0.13)	-0.01 (0.91)	< 0.01 (0.97)	0.48 (< 0.01)	0.12 (0.36)	0.57 (< 0.01)

All these features were also correlated with the exam scores in Table 6 . We found that

Table 6. Correlations between indicators of learning from SQL-Tutor data and exam scores. The first column titles are the affective states, and the first row title is the exam score. The first number on each cell is the correlation *r*, and the number on the parentheses is the significance *p*. Marginally significant relationship is shaded light gray.

Affect	*Solved Problem*	*Attempted Problem*	*Learned Constraints*	*Constraints Used*	*Seen Messages*	*NumOf Logins*	*Total Time*	*AvgTime ToSolve*	*Total Attempts*	*AvgNumOfAttempts PerSolvedProb*
Exam Scores	-0.01 (0.96)	< -0.00 (0.98)	-0.11 (0.44)	0.26 (0.06)	0.01 (0.92)	-0.12 (0.40)	0.07 (0.59)	0.07 (0.62)	0.03 (0.85)	0.02 (0.86)

We produced two linear regression models, one for boredom and one for frustration, using the learning indicators from SQL-Tutor as features. For each affective state in turn, we first selected all the features as the independent variables and identified the affective state as the dependent variable. We then reviewed the results to see which features contributed significantly to the model and which did not. We then performed backwards elimination to remove features that did not significantly contribute to the models. That is, we removed the non-significant features and performed regression on the remaining features. We did this repeatedly until all features that remained in the model were significant. Of the two models, only the one for boredom was statistically significant. The generated final model is:

$$\text{BOREDOM} = -0.002*\mathit{SeenMessages} + -0.002*\mathit{TotalTime} + 0.031*\mathit{AvgTimeToSolve} + 0.007*\mathit{TotalAttempts} + 0.068 \quad (4.2)$$

The model shows that boredom (r=0.647; p<0.001) can be predicted by the amount of feedback the student receives, total interaction time, average time per solved problem, and total attempts. And to test the generalizability of our model we computed its Bayesian Information Criterion for Linear Regression (BiC') (Raftery, 2003). The BiC' is used to assess the tradeoff between model fit and the number of parameters (which can spuriously increase model fit). Values of BiC' values between -6 and -10 correspond to p-values of 0.05, implying that the model has a significantly better fit than chance. BIC' values of less than -10 have a corresponding p-value of 0.006. As shown in Table 8, the model for boredom had a BiC' of -14.27 implying that it had a much better fit than chance.

5. Discussion and Conclusion

In this study, we attempted to identify factors that might be indicative of the affective states of students who used SQL-Tutor. We used two sources of data for the analysis: human observations and learning indicators taken from the tutor's log files. From the observations, we found that students using SQL-Tutor most often exhibited engaged concentration, confusion, and boredom. None of the seven affective states had significant correlation with the student achievement. However, boredom and frustration were found to persist and were nonetheless interesting because other studies indicate that these states may affect the student learning experience. Our findings suggest that while students using SQL-Tutor express boredom and frustration, infrequently (8.10% and 2.10% of the time respectively), these affective states tend to persist when experienced. The results also show that boredom alone was negatively correlated with number of solved problems, number of attempted problems, and total number of constraints used. It was positively correlated with average time to solve a problem and average attempts per solved problem. Persistent boredom was negatively correlated with number of problems solved and positively correlated with average time per solved problem. Frustration alone was not significantly correlated with any of the factors, but persistent frustration was negatively correlated with number of constraints used and positively correlated with average number of attempts per problem.

That boredom is indicated by fewer problems solved and attempted, and fewer constraints used follows intuition. These students are not progressing as steadily or successfully as their peers, perhaps because the material is too difficult or too easy. Time is an indication of boredom. The longer it takes for students to solve an SQL problem correctly, possibly because the problem is more difficult, the more boredom is experienced. Similarly, disengagement may lead students to go through several unsuccessful attempts to solve a problem, which results into more submitted solutions. Lastly, that higher average time and average number of attempts per solved problem with a lower number of solved problems indicates an even less desirable affective state: persistent boredom. A model of boredom may therefore be worth incorporating into the tutor to inform the tutor of students' negative affective states and to cue meaningful interventions. The form which these interventions take is a topic for further study but may include lowering or raising levels of difficulty, or changing the learning task.

The features that characterize the other persistent affective state, frustration, follow intuition: persistently frustrated students do not use as many constraints and tend to use more steps on average to solve a problem. Using fewer constraints in solving a problem might lead to incorrect solution. Frustration could have kept the student from thinking further for the correct solution and thus use fewer constraints in the process. Also, the fact that they need more steps implies that previous steps were unsuccessful, hence their attempts to solve the problem were frustrated. As mentioned in the introduction, though, frustration is not necessarily negative. Frustration is a natural consequence of learning new material and the experience of frustration does not necessarily lead to non-learning behaviors (Baker et al., 2010). A detector for frustration might still be worth incorporating in intelligent tutors, if only to mitigate extreme or persistent experiences of the emotion that may lead to disengagement from the subject matter.

The work presented in this paper can be continued in a number of ways. First, we examined the data at a coarse grain size—the session level, as opposed to the transaction level. This means that the features, even if applied, will require at least one session's worth of data in order to indicate boredom. If the researcher's goal were to develop a system that detects boredom (or any other affective state) in real time, the current learning indicators is not sensitive enough to do so.

Second, part of the task of detection refinement is the feature space engineering. The current feature space used for building the models presented in this paper was based on features from previous studies (Lagud & Rodrigo, 2010; Matsuda et al., 2007; Tabanao et al., 2008; Suraweera & Mitrovic, 2002; Holland et al., 2009; Mayo et al., 2000). Researchers can consider expanding the feature set, to include other features that could be indicative of student affect.

The development of student affect detectors is the first step towards affect-sensitive intelligent tutors. Researchers hope that the inclusion of affect sensitivity to tutoring provides another criterion for system adaptation. Intelligent tutors can use these models to fire interventions such as providing remedial lessons or raising or lowering the difficulty level of the material. Alternatively, they can notify teachers that students may be struggling, cuing human intervention. Ultimately, these interventions should raise student achievement or improve students' overall learning experience.

Acknowledgments

We thank the MIS 21 students who participated as subjects during the experiment. We thank Mr. Alipio Gabriel, Mr. John Paul Contillo, and the Ateneo Laboratory for the Learning Sciences for their assistance. We thank Mr. Hubert Ursua and the technical and secretarial staff of the Ateneo de Manila University's Department of Information Systems and Computer Science. Finally, we thank the Department of Science and Technology for the grant entitled "Development of an Educational Data Mining Workbench."

References

Acee, T. W., Kim, H., Kim, H. J., Kim, J. I., Chu, H. N. R., Kim, M., Cho, Y., et al. (2010). Academic boredom in under- and over-challenging situations. *Contemporary Educational Psychology, 35*(1), 17–27. Elsevier Inc. doi:10.1016/j.cedpsych.2009.08.002.

Artino, A. R. & Jones, K. D. (2012). Exploring the complex relations between achievement emotions and self-regulated learning behaviors in online learning. *The Internet and Higher Education, 15*(3), 170–175. Elsevier B.V. doi:10.1016/j.iheduc.2012.01.006.

Baker, R. S. J. d., Corbett, A. T., Koedinger, K. R. & Wagner, A. Z. (2004). Off-task behavior in the Cognitive Tutor Classroom: When students 'game the system'. *Proceedings of ACM CHI: 2004: Computer-Human Interaction*. 383–390.

Baker, R. S. J. d., D'Mello, S. K., Rodrigo, M. M. T. & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies* (Vol. 68, pp. 223–241).

Beer, C., Clark, K. & Jones, D. (2010). Indicators of engagement. Curriculum, Technology and Transformation for an Unknown Future. *Proceedings of ASCILITE Sydney 2010*. (pp. 75–86).

Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin*, *106*(1), 59–73.

Bhagat, S., Bhagat, L., Kavalan, J. & Sasikumar, M. (2002). Acharya: An intelligent tutoring environment for learning SQL. *Proc. of the Vidyakash 2002 International Conference on Online Learning*, Mumbai, India.

Breidenstein, G. (2007). The meaning of boredom in school lessons. Participant observation in the seventh and eight form. *Ethnography and Education, 2*(1), 93–108. doi:10.1080/17457820601159133.

Cohen, J. A. (1960). Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20. 37–46.

Craig, S. D., Graesser, A. C., Sullins, J. & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media* (Vol. 29, No. 3, pp. 241–250).

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper & Row, New York.

D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., Graesser, A. (2005). Integrating affect sensors in an intelligent tutoring system. In “Affective Interactions: The Computer in the Affective Loop Workshop” In conjunction with International conference on Intelligent User Interfaces, 7–13.

D'Mello, S. K. & Graesser, A. (2009). Automatic detection of learners' emotions from gross body language. *Applied Artificial Intelligence*, *23* (2), 123–150.

D’Mello, S. K., Jackson, T., Craig, S., Morgan B., Chipman, P., White, H., Person, N., Kort, B., el Kaliouby, R., Picard, R. & Graesser, A. (2008). AutoTutor detects and responds to learners affective and cognitive states. *Workshop on Emotional and Cognitive Issues at the International Conference of Intelligent Tutoring Systems*, 31–43.

D’Mello, S. K., Pearson, N,, & Lehman, B. (2009). Antecedent-consequent relationships and cyclical patterns between affective states and problem solving outcomes. In V. Dimitrova, R. Mizoguchi, B. duBoulay, & A. Graesser (Eds.), *Proceedings of 14th International Conference on Artificial Intelligence in Education*. (pp 57–64). Amsterdam: IOS Press.

D'Mello, S. K., Picard, R. W. & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, *22* (4), 53–61.

D’Mello, S. K., Taylor, R. S. & Graesser, A. (2007). Monitoring affective trajectories during complex learning. *Proc. 9th Annual Meeting of the Cognitive Science Study*, 203–208.

Dagami, M. M. C., Guo, B., Pating, M. B., Guia, T. F. G., Leonor, W. B. S. & Rodrigo, M. M. T. (2011). Determining the precursors of boredom. *Philippine Computing Journal, 6* (1), 43–66.

Damasio, A. R. (2000). A second chance for emotion. In *Cognitive Neuroscience of Emotion*, R. D. Lane and L. Nadel (Eds.). New York, NY: Oxford University Press.

de Raadt, M., Dekeyser, S. & Lee, T. Y. (2006). Do students SQLify? Improving learning outcomes with peer review and enhanced computer assisted assessment of querying skills. *Proc. of the 6th Baltic Sea conference on Computing education research: Koli Calling 2006*, 101–108. doi=10.1145/1315803.1315821.

De Vicente, A. & Pain, H. (2002). Informing the detection of the students' motivational state: An empirical study. *Proc. 6th International Conference on Intelligent Tutoring Systems*, 933–943.

Dekeyser, S., de Raadt, M. & Lee, T. Y. (2007). Computer assisted assessment of SQL query skills. *Proc. 18th Conf. Australasian Database*, 53–62.

Dreisbach, G. & Goschke, T. (2004). How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. *30*(2), 343–353. doi:10.1037/0278-7393.30.2.343.

Ekman, P. (1999). Basic Emotions. In T. Dalgleish and T. Power (Eds.) *The Handbook of Cognition and Emotion*, 45–60. Sussex, U.K.: John Wiley & Sons, Ltd.

Fisher, C. D. (1993). Boredom at work: A neglected concept. *Human Relations* (Vol. 46, No. 3, pp. 395–417).

Forbes-Riley, K. & Litman, D. (2012). Adapting to multiple affective states in spoken dialogue. Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL). Seoul, South Korea, 5-6 July 2012.

Fredrickson, B. L. & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion, 19*(3), 313–332. doi:10.1080/02699930441000238.

Holland, J., Mitrovic, A. & Martin, B. (2009). J-LATTE: A constraint-based tutor for Java. *Proc. 17th Int. Conf. Computers in Education*, 142–146.

Inventado, P. S., Legaspi, R., Suarez, M., & Numao, M. (2011). Investigating transitions in affect and activities for online learning interventions. In T. Hirashima et al. (Eds.), *Proceedings of the 19th International Conference on Computers in Education*. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education.

Kapoor, A., Burleson, W. & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies* (Vol. 65, pp. 724–736).

Kenny, C. & Pahl, C. (2005). Automated tutoring for a database skills training environment. In *Proceedings of the 36th SIGCSE technical symposium on Computer science education* (SIGCSE '05). ACM, New York, NY, USA, 58-62. *37*(1), 58–62. doi: 10.1145/1047344.1047377 http://doi.acm.org/10.1145/1047344.1047377.

Lagud, M. C. V. & Rodrigo, M. M. T. (2010). The affective and learning profiles of students while using an intelligent tutor for Algebra. *ITS 2010* (Vol. 1, pp. 255–263).

Mann, S. & Robinson, A. (2009). Boredom in the lecture theatre: An investigation into the contributors, moderators and outcomes of boredom amongst university students. *British Educational Research Journal, 35*(2), 243–258. doi:10.1080/01411920802042911.

Matsuda, N., Cohen, W. W., Sewall, J., Lacerda, G. & Koedinger, K. R. (2007). Predicting students' performance with SimStudent: Learning cognitive skills from observation. In R. Luckin, K. R. Koedinger & J. Greer (Eds.). *Proc. Int. Conf. Artificial Intelligence in Education*, 467–476. Amsterdam, Netherlands. IOS Press.

Mayo, M., Mitrovic, A. & McKenzie, J. (2000). CAPIT: An intelligent tutoring system for capitalization and punctuation. *International Workshop on Advanced Learning Technologies*, 151–154.

McQuiggan, S. W. & Lester, J. C. (2006). Diagnosing self-efficacy in intelligent tutoring systems: An empirical study. In M. Ikeda, K. Ashley & T. W. Chan (Eds.), ITS *2006*. LNCS 4053, 565–574.

McQuiggan, S. W., Robinson, J. L., & Lester, J. C. (2008). Affective transitions in narrative-centered learning environments. In B. Woolf, et al. (Eds.), Intelligent Tutoring Systems 2008, Lecture Notes in Computer Science 5091, 490-499. Springer-Verlag Berlin Heidelberg.

Mitrovic, A. (1998). Learning SQL with a computerized tutor. *Proc. 29th SIGCSE Technical Symposium on Computer Science Education*, 307–311.

Mitrovic, A. & Ohlsson, S. (1999). Evaluation of a constraint-based knowledge representation for individualized instruction. *Int. J. Artificial Intelligence in Education*, 10(3-4), 238–256.

Ohlsson, S. & Mitrovic, A. (2006). Constraint-based knowledge representation for individualized instruction, *Computer Science and Information Systems. 3*(1), 1–22.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge: The University of Cambridge.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H. & Perry, R. P. (2010). Boredom in achievement settings: Exploring control – value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*(3), 531–549. doi:10.1037/a0019243.

Perkins, R. E. & Hill, A. B. (1985). Cognitive and affective aspects of boredom. *British Journal of Psychology* (Vol. 76, No. 2, pp. 221–232).

Raftery, A. E. (2003). Bayesian model selection in social research, *Sociological Methodology 25*, 111–163.

Renaud, K. & Biljon, J. v. (2004). Teaching SQL – Which pedagogical horse for this course? *Key Technologies for Data Management*. *3112*, 244–256.

Rodrigo, M.M.T. & Baker, R.S.J.d. (2009). Coarse-grained detection of student frustration in an introductory programming course. *Proceedings of ICER 2009: the International Computing Education Workshop*. pp. 75–80. doi: 10.1145/1584322.1584332

Rodrigo, M. M. T., Baker, R. S. J. d., Agapito, J., Nabos, J., Repalam, M. C., Reyes, S. S. Jr., & San Pedro, M. O. C. (2011). The effects of an embodied conversational agent on student affective dynamics while using an intelligent tutoring system. *IEEE Transactions on Affective Computing*, *2*(4), 18-37.

Rodrigo, M. M. T., Baker, R. S. J. d., D'Mello, S., Gonzalez, M. C. T., Lagud, M. C. V., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., Santillano, J. Q., Sugay, J. O., Tep, S., & Viehland, N. J. B., (2008a). Comparing learners' affect while using an intelligent tutoring system and a simulation problem solving game. In Woolf, B. P., Aimeur, E., Nkambou, R. & Lajoie, S. P. (Eds.) *Proc.* i*ntelligent tutoring systems*, (40–49). Montreal, Canada.

Rodrigo, M. M. T., Baker, R. S. J. d., Lagud, M. C. V., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., Santillano, J. Q., Sevilla, L. R. S., Sugay, J. O., Tep, S. & Viehland, N. J. B. (2007). Affect and usage choices in simulation problem-solving environments. In R. Luckin, K. R. Koedinger & J. Greer (Eds.), *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work*. Amsterdam, The Netherlands: IOS Press, 145–152.

Rodrigo, M.M.T., Rebolledo-Mendez, G., Baker, R.S.J.d., du Boulay, B., Sugay, J.O., Lim, S.A.L., Espejo-Lahoz, M.B., Luckin, R. (2008b). The effects of motivational modeling on affect in an intelligent tutoring system. *Proceedings of International Conference on Computers in Education*, 57-64.

Rodrigo, M. M. T., Baker, R. S. J. d. & Nabos, J. Q. (2010). The relationships between sequences of affective states and learner achievement. *Proc. 18th International Conference on Computers in Education*, 55–60.

Sottilare, R. A. & Proctor, M. (2012). Passively classifying student mood and performance within intelligent tutors. *Educational Technology & Society*, *15*(2), 101 – 114.

Suraweera, P. & Mitrovic, A. (2002). KERMIT: A constraint-based tutor for database modeling. In S. A. Cerri, G. Gouarderes & F. Paraguacu (Eds.), *ITS 2002*, LNCS 2363 (pp. 377–387).

Suraweera, P., Mitrovic, A., Martin B., Holland, J., Milik, N., Zakharov, K. & McGuigan, N. (2009). Using ontologies to author constraint-based intelligent tutoring systems. In D. Dicheva et al. (Eds.), *Semantic Web Technologies for e-Learning, IOS Press, 77-95.*

Tabanao, E. S., Rodrigo, M. M. T. & Jadud, M. C. (2008). Identifying at-risk novice Java programmers through the analysis of online protocols. *Philippine Computing Society Congress*. (UP Diliman, Quezon City, February 23-24, 2008).

Tempelaar, D. T., Niculescu, A., Rienties, B., Gijselaers, W. H. & Giesbers, B. (2012). How achievement emotions impact students' decisions for online learning, and what precedes those emotions. *The Internet and Higher Education, 15*(3), 161–169. Elsevier B. V. doi:10.1016/j.iheduc.2011.10.003.