Co-occurring Affective States in Automated Computer Programming Education

Nigel Bosch[1], Sidney D'Mello[1,2]

Departments of Computer Science[1] and Psychology[2], University of Notre Dame
Notre Dame, IN 46556, USA
{pbosch1, sdmello}@nd.edu

Abstract. We investigated the incidence of momentary co-occurrence of affective states in a computerized learning environment. Novice students (N = 99) used a learning environment designed to teach the basics of computer programming. Only 46 of these students reported a sufficient number of co-occurring affective states for statistical modeling. Association rule mining techniques were used to identify patterns of co-occurrence. Two co-occurring pairs of affective states occurred at rates higher than chance: Confusion/Uncertainty + Frustration and Curiosity + Flow/Engagement. The relationship between these states and student interaction patterns and outcomes was investigated. We found that the co-occurrence of Curiosity + Flow/Engagement was related to success and errors when testing code, as well as the use of hints available in the learning environment and performance on the learning task. Implications for learning environments that attend to student affect are discussed.

1 Introduction

The relationship between learning and affect is a topic that has been actively explored over the last decade [1–4]. More recently, with the rising prominence of intelligent tutoring systems (ITSs), there has been a renewed interest in research exploring affective states in the context of learning with ITSs and other advanced educational technologies [5–9]. Most research into affective states in computerized learning systems has assumed that the student is experiencing one affective state at a time (see meta-analysis [10]). We expand this topic by examining co-occurring affective states, or when multiple affective states are experienced at the same time.

Previous research has explored affective state pairs in part by focusing on the transitions between affective states [11–13]. These transitions illustrate the change from one affective state to another and have been linked to learning performance. Co-occurrence is different because multiple affective states are occurring at the same time rather than in sequence. Determining what affective states co-occur and how those co-occurrence patterns are related to learning is important for more effectively tuning intelligent tutoring systems (ITSs) that sense and respond to student affect. For example, should an affect-sensitive ITS respond to confusion, frustration, or both, if these states cooccur? Alternatively, might the somewhat lower accuracies (see [14, 15] for reviews) of state-of-the-art systems be attributed to cooccurrening affect? Should these affect detectors focus on detecting these such *affective blends*?

Outside of computerized learning contexts, there has been some research into co-occurring affective states. As early as 1972, Izard et al. [16] considered the possibility that some emotions previously thought to be individually experienced might have more nuanced manifestations. Polivy [17] found that attempting to induce a specific affective state could instead result in multiple affective states being induced. Barrett [18] examined the correlation between levels of emotions as reported by participants three times per day on a 7-point likert scale. She found strong relationships between sadness, fear, hostility, and guilt.

In the domain of learning, Harley et al. [19] used commercial emotion recognition software to provide measures of the presence of various emotions and determined if there were co-occurring emotions. They found Happiness and Sadness frequently occurred together, as well as Sadness and Disgust. The co-occurrence of Happiness and Sadness is rather surprising and inconsistent with theory given that these emotions have opposite valence profiles (Happiness is positive while Sadness is negative) [20]. Similarly, Sadness and Disgust, though both negative, have opposing activation levels (Sadness is a deactivating state while Disgust is an activating state). These inconcistencies raise the question of whether the cooccurrence relationships uncovered might be attributed to inaccuracies in automated emotion detection, which is a well-known problem in the field of affective computing [15].

In a recent meta-analysis of 24 studies [10], Flow/Engagement was found to occur relatively more frequently than other affective states across studies, with Boredom, Confusion, Curiosity, Happiness, and Frustration occurring frequently in some studies. These studies monitored discrete affect (e.g., Confusion, Frustration, etc) at multiple points in a learning session, but only one affective state was tracked at each time point. The implicit assumption here is that affective states individually occur rather than co-occur. Taking a somewhat different approach, the novel contribution of this paper is in exploring the incidence of co-occurring discrete affective states and their relationships with student interactions during learning with technology.

Our work used self-reports from students to determine which affective states co-occur, using a retrospective judgment protocol ([21], see below a more detailed description). We contrast previous research of co-occurring affective states by focusing on affective states that are learning-centered, specifically in the domain of computer programming education, and arguably likely to be relavent to ITSs [20]. We investigated the following research questions: 1) What pairs of affective states co-occurred? 2) How do co-occurring affective states relate to student interaction events? and 3) Does one affective state in a co-occurring pair imply the other?

2 Method

Participants. Participants were from a Midwestern United States university. Of 113 students who completed the study, 14 were removed because they reported having prior experience with computer programming and our intended focus was on novices.

Twenty-three students made no secondary affective state judgments at all, and were thus not considered. Additionally, 30 students who made fewer than 10 secondary affect ratings were removed, because they did not provide a usable distributions to analyze co-occurring affective states (see below). The remaining 46 students were 54% female with an average age of 19.2 years (SD = 1.19).

Learning Session. Data was collected using a computerized learning environment, in which students were taught fundamentals of computer programming in the Python language. Students completed 25 minutes of *scaffolded learning*, in which they had access to instructional materials, exercises to solve, and hints.

Fig. 1 shows a screenshot of the learning environment used by students. Numbers overlaid in Fig. 1 indicate the different areas of the learning environment user interface: 1) instructional text, 2) source code editing box, 3) hint display area, and 4) input/output console.

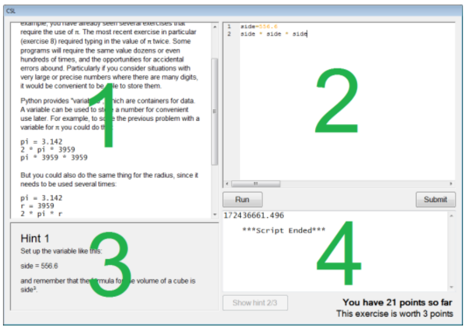

Fig. 1. Screenshot of the learning environment used by students, with key areas numbered.

During the scaffolded learning phase, students could use the hint function only after some time had elapsed since the beginning of the exercise or the last time a hint was used. Hints ranged in scope from expanded details of key concepts to code examples, and finally bottom-out hints containing complete solutions to the current exercise. Hints were specific to the new concepts and code required for each exercise, but were not adapted to students' code or behavior. Instructions for using the interface were given in the introductory exercise. Students were able to test their code with the interactive console, and submit code for automatic correctness checking when they were satisfied with their work. If a submitted solution was correct, the student would automatically be advanced to the next exercise. Otherwise, the learning environment would tell the student their solution was incorrect, and suggest using a hint or trying again. Correctness was determined by comparing the output of the students' code with the output of a correct solution, allowing acceptable variations such as different precision of π in geometry-based solutions. Additionally, solutions to exercises that required reading input from the user were tested by automatically providing differing input values and checking for corresponding correct outputs. There was no limit on number of submission attempts allowed other than the time limit imposed on the entire scaffolding phase of the learning session.

Following the scaffolding phase, students completed a 10-minute *fadeout programming exercise*. The fadeout exercise made use of all major concepts that could be covered in the scaffolding phase. It was designed to be more difficult than novice students would be capable of solving, though they could make progress toward a solution. Students had ten minutes to solve the exercise and did not receive new conceptual information in the instructional material, nor did they have hints available during the fadeout programming exercise.

Affect Judgments. We measured the affective states of students using a retrospective judgment protocol [21], which is a validated offline affect-judgment technique that affords fine-grained affect measurement without any interruptions during the learning session (see review of affect annotation methods [22]). After each student had completed the fadeout phase of the study, they were shown synchronized videos of their own face and on-screen activity and asked to make judgments about what affective states they were experiencing at various points in the learning session. In this manner students were able to make affective state judgments based on a combination of context (as given by screen capture video), facial cues, and memories of the learning session. Fig. 2 shows an illustration of the interface used for retrospective affect judgment.

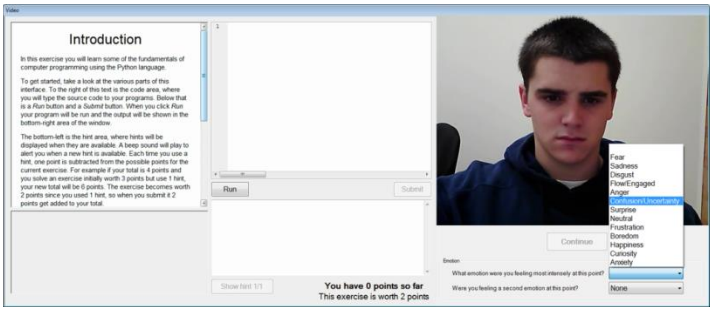

Fig. 2. Retrospective affect judgment interface.

The video streams automatically paused at 100 fixed points for judgments to be made. Points were chosen to correspond to interaction events, such as key presses, running code, showing hints, and other such occurrences. Some periods of idle activity (longer than 30 seconds) were also chosen for judgments. Additionally, students could spontaneously pause the video streams during the retrospective affect judgment process to provide an affect judgment if they wanted to do so.

Students chose affective states from a randomly ordered list comprised of Anger, Anxiety, Boredom, Confusion/Uncertainty (henceforth abbreviated as Confusion), Curiosity, Disgust, Fear, Frustration, Flow/Engagement, Happiness, Sadness, Surprise, and the Neutral state (defined as no apparent emotion). These are largely derived from Pekrun's taxonomy of academic emotions [20]. With each judgment, students were required to choose a primary affective state. Students could also provide a secondary judgment—a co-occurring affective state they were experiencing at that point. Providing a secondary affect judgment was not required.

3 Analysis and Results

For the purposes of this paper, we will focus only on instances where students provided secondary judgments. There were a total of 1,764 affect judgments made that included a secondary affect rating. Table 1 shows the distribution of primary and secondary affect ratings for these judgments sorted by primary ratings.

Table 1. Mean proportions of affective states reported.

Affective State	Primary	Secondary
Confusion	0.301	0.221
Flow/Engagement	0.237	0.150
Frustration	0.167	0.187
Curiosity	0.085	0.099
Boredom	0.056	0.088
Anxiety	0.050	0.116
Neutral	0.038	0.054
Surprise	0.020	0.022
Happiness	0.019	0.017
Anger	0.019	0.023
Sadness	0.004	0.006
Disgust	0.004	0.014
Fear	0.001	0.004

Only Anxiety, Boredom, Confusion, Curiosity, Flow/Engagement, and Frustration ratings were considered further because they were commonly occurring affective states. Neutral was also not considered in co-occurring pairs because it is not conceptually different from no secondary affective rating. Considering only the common affective states resulted in 1,303 pairs of ratings used for analysis.

Question 1: What pairs of affective states co-occurred? The co-occurring affective states were analyzed using an association rule learning metric called Lift (Equation 1) [23].

$$Lift = \frac{\Pr(X \text{ and } Y)}{\Pr(X)\Pr(Y)} \tag{1}$$

Lift is used to compare the observed probability of two co-occurring affective states with the probability of those states co-occurring simply due to random chance. Lift was calculated for each student to ensure that data points would be independent. Values higher than 1 indicates a pair of values is co-occurring more frequently than expected by chance. To identify co-occurring affective states, we performed one-sample t-tests comparing the Lift values of each co-occurring pair with a test value of 1. Table 2 contains the result of this analysis.

Table 2. t-tests of Lift for each affective state pair.

Affective State Pair	t	Mean Lift (SD)
Greater than Chance		
Confusion + Frustration	1.57 (p = .123)	1.14 (0.59)
Curiosity + Flow/Engagement	2.15* (p = .038)	1.33 (0.98)
Less than Chance		
Anxiety + Boredom	-6.72** (p = .000)	0.28 (0.48)
Anxiety + Confusion	-6.04** (p = . 000)	0.55 (0.43)
Anxiety + Curiosity	-12.96** (p = . 000)	0.21 (0.34)
Anxiety + Flow/Engagement	-1.96 (p = .060)	0.76 (0.69)
Anxiety + Frustration	-5.94** (p = . 000)	0.50 (0.49)
Boredom + Confusion	-4.62** (p = . 000)	0.60 (0.47)
Boredom + Curiosity	-3.32** (p = .003)	0.49 (0.80)
Boredom + Flow/Engagement	-4.52** (p = . 000)	0.45 (0.66)
Boredom + Frustration	-3.58** (p = .001)	0.57 (0.66)
Confusion + Curiosity	-5.94** (p = . 000)	0.51 (0.54)
Confusion + Flow/Engagement	-8.46** (p = . 000)	0.58 (0.32)
Curiosity + Frustration	-36.15** (p = . 000)	0.07 (0.17)
Flow/Engagement + Frustration	-19.52** (p = . 000)	0.22 (0.26)

Note. * $p < .05$; ** $p < .01$

Only the Confusion + Frustration and Curiosity + Flow/Engagement affective state pairs occurred at levels above what was expected by chance. Out of these, the Curiosity + Flow/Engagement pair was statistically significant, while the Confusion + Frustration Pair approached significance (p = .123). These pairs were expected in the context of the learning environment, while pairs such as Boredom + Flow/Engagement occurring together do not make theoretical sense. We now examine the two pairs that occurred at higher than chance levels in greater detail.

Question 2: How do co-occurring affective states relate to student interaction events? To investigate this research question we correlated the Lift of the two cooccurring affective state pairs with key events from the learning session. Events analyzed included Key Press (number of times a student pressed keys), Run Error and Run Success (number of times a student tested their code and received syntactic errors or no such errors, respectively), Hints Per Exercise (average number of times hints were used by a student divided by number of exercises completed), and Score. The Score was defined as the number of exercises completed by a student and the number of hints that they did not use. For example, a student who completed five exercises and used three out of fifteen possible hints would have a score of seventeen. These scores were displayed to students during the learning sessions.

Table 4 shows the result of these correlations. Correlations were separately analyzed by phase of the session, because of the different nature of the scaffolding and fadeout phases. Hints were not made available to students in the fadeout phase and thus Score (which was calculated in part from hint usage) was also not available for correlation.

Table 3. Correlations between Lift for co-occurring affective states interaction events.

	Confusion + Frustration		**Curiosity + Flow/Engagement**	
	Scaffolding	**Fadeout**	**Scaffolding**	**Fadeout**
Key Press	-0.040	0.067	**0.208**	-0.114
Run Successes	-0.012	0.199	-0.038	**0.314**
Run Errors	0.031	-0.168	-0.030	**-0.314**
Show Hint	0.002		**-0.203**	
Score	-0.113		**0.226**	

Note. Correlations larger than 0.2 in bold. No correlations were statistically significant. Hints and Score were not available in the Fadeout phase of the study. Run Successes and Run Errors were partial correlations controlling for total Runs.

Due to the small sample size, we focus on the size rather than the significance of the correlations. Confusion + Frustration did not appear to exhibit any meaningful trends (as might be expected from the marginally positive Lift of this co-occurrence), but Curiosity + Flow/Engagement co-occurrence was associated in the expected direction with several variables. Co-occurring Curiosity + Flow/Engagement was tied to a higher proportion of Key Press events, less hint usage, and a better outcome (Score) in the scaffolding phase of the study. Run events were particularly correlated with Curiosity + Flow/Engagement co-occurrence, where students experienced more successful runs and fewer errors than without the co-occurrence.

Question 3: Does one affective state in a co-occurring pair imply the other? The dependence of one affective state on the other in these co-occurring pairs may provide some additional information for interpreting their presence. To examine the dependence we use another association rule learning metric called Confidence (Equation 2).

$$Confidence\ (X \rightarrow Y) = \frac{\Pr(X\ and\ Y)}{\Pr(X)} \quad (2)$$

Confidence is used as an estimate of the probability of an affective state Y occuring, given the presence of another affective state X (to what extent does X imply Y). The Confidences of both possible orderings of the affective states in the two frequently co-occurring pairs were compared to determine if one element of a pair is more likely to imply its co-occurring affective state than the other way around. For example, does Confusion imply the presence of Frustration more than Frustration implies the presence of Confusion? Knowing the answer to this type of question may be helpful for automatic affect detection software that can better identify situations where one affective state may be mistaken for another, as well as situations where a co-occurrance of affective states may be more likely to take place and a simple affective state label may not apply. Table 4 presents the results of comparing the Confidences for the two affective state pairs that occur more often than chance.

Table 4. Comparisons of Confidence for affective state pairs. Standard deviations are in parentheses.

Primary → Secondary	Mean (SD)	t
Confusion → Frustration	0.419 (0.225)	-5.71*
Frustration → Confusion	0.672 (0.263)	-5.71*
Curiosity → Flow/Engagement	0.495 (0.359)	4.53*
Flow/Engagement → Curiosity	0.259 (0.190)	4.53*

Note. * $p < .001$

As can be seen in Table 4, the affective states in a co-occurrence pair do not imply each other equally. That is, given a co-occurence pair with Confusion as one affective state, the probability of the other state in the pair being Frustration is .419. On the other hand, for a pair with Frustration, the probability of the other state being Confusion is .672 (significantly higher than .419, $p < .001$). Similarly Curiosity and Flow/Engagement do not imply each other at the same probability. Curiosity implies the presence of Flow/Engagement at a significantly higher level than Flow/Engagement implies Curiosity ($p < .001$).

4 General Discussion

In exploring potentially co-occurring affective states we discovered several salient patterns. Perhaps most noteable was the sparsity of secondary affect ratings. Though the phenomenon of co-occurrence does occur at levels higher than expected by chance in some situations, we found it to be rare in general. Only 46 of 99 students reported 10 or more instances of co-occurring affective states. We found Curiosity + Flow/Engagement co-occurrences to be associated with fewer errors encountered and more success in the learning task, as expected, which may prove useful in intelligent tutoring systems if they can encourage the co-occurrence of these affective states.

Because affect detection and recognition can be an integral part of an intelligent tutoring system, particularly as a means of determining when to intervene and when a student is learning effectively, affect detection may be well served to include detection of co-occurrence in affective states. Furthermore, it may be useful to explore the possibility that frequently co-occurring affective states are likely to be confused for each other, both by humans and computers. Adding more possible labels tends to make classification tasks more difficult, but as the field of affective computing evolves and more advanced multimodal techniques are developed, this may become a practical possibility. We found differences between the Confidence levels of co-occurring affective states that may be helpful for affect recognition systems as well, since these Confidence levels could inform affect recognition systems about the probability of other affective states in a potential co-occurence.

The present study is not without limitations. In particular, the sample size is limited (N = 46) after removing students with few or no secondary affect judgments. Therefore future work might be better served by obtaining affect judgments on a Likert scale for each affective state. This process is more time consuming than simply selecting two affective states from a list, but can be improved by using information gained in the current study to limit choices to affective states that frequently occur.

An additional limitation of this study is the potential for results being specific to the nature of the learning environment. Some correlations between interaction events differed noticeably between the scaffolding and fadeout phases of the study, which may indicate that the informative relationships between co-occurring affective states and student interaction events differ in important ways between learning environments and instruction formats. Future work should include plausible variations in learning materials and instruction formats to further explore the potential relationships between those factors and co-occurring affective states.

Though a seemingly infrequent phenomenon, co-occurring affect states do exist and have some connections to the learning process. Understanding more about the complex nature of affective states in learning environments can lead to better affective awareness in intelligent tutoring systems. Affective awareness can in turn improve the efficacy of teaching computer programming in a world where computers play the role of teacher more and more frequently.

References.

1. Arroyo, I., Cooper, D.G., Burleson, W., Woolf, B.P., Muldner, K., Christopherson, R.: Emotion Sensors Go To School. AIED. pp. 17–24 (2009).
2. Baker, R.S.J. d, Moore, G.R., Wagner, A.Z., Kalka, J., Salvi, A., Karabinos, M., Ashe, C.A., Yaron, D.: The Dynamics between Student Affect and Behavior Occurring Outside of Educational Software. In: D’Mello, S., Graesser, A., Schuller, B., and Martin, J.-C. (eds.) Affective Computing and Intelligent Interaction. pp. 14–24. Springer, Berlin Heidelberg (2011).
3. Baker, R.S.J. d., D’Mello, S.K., Rodrigo, M.M.T., Graesser, A.C.: 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. 68, 223–241 (2010).
4. Craig, S., Graesser, A., Sullins, J., Gholson, B.: Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. Journal of Educational Media. 29, 241–250 (2004).
5. Andallaza, T.C.S., Rodrigo, M.M.T., Lagud, M.C.V., Jimenez, R.J.M., Sugay, J.O.: Modeling the Affective States of Students Using an Intelligent Tutoring System for Algebra. Proceedings of The Third International Workshop on Empathic Computing (IWEC 2012). p. 4 (2012).
6. D’Mello, S., Craig, S.D., Gholson, B., Franklin, S., Picard, R., Graesser, A.C.: Integrating affect sensors in an intelligent tutoring system. Affective Interactions: The Computer in the Affective Loop Workshop at. pp. 7–13 (2005).
7. Graesser, A., D’Mello, S., Chipman, P., King, B., McDaniel, B.: Exploring relationships between affect and learning with AutoTutor. Proceedings of Artificial Intelligence in Education (2007).
8. Litman, D.J., Forbes-Riley, K.: Predicting student emotions in computer-human tutoring dialogues. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. p. 351 (2004).

9. Robison, J., McQuiggan, S., Lester, J.: Evaluating the consequences of affective feedback in intelligent tutoring systems. 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. pp. 1–6 (2009).
10. D'Mello, S.: A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. Journal of Educational Psychology. 105, 1082–1099 (2013).
11. Baker, R.S.J. d, Rodrigo, M.M.T., Xolocotzin, U.E.: The dynamics of affective transitions in simulation problem-solving environments. In: Paiva, A.C.R., Prada, R., and Picard, R.W. (eds.) Affective Computing and Intelligent Interaction. pp. 666–677. Springer, Berlin Heidelberg (2007).
12. Bosch, N., D'Mello, S.: Sequential Patterns of Affective States of Novice Programmers. The First Workshop on AI-supported Education for Computer Science (AIEDCS 2013). pp. 1–10 (2013).
13. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learning and Instruction. 22, 145–157 (2012).
14. D'Mello, S., Kory, J.: Consistent but modest: a meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. Proceedings of the 14th ACM international conference on Multimodal interaction. pp. 31–38. ACM, New York, NY, USA (2012).
15. Calvo, R.A., D'Mello, S.: Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. IEEE Transactions on Affective Computing. 1, 18 –37 (2010).
16. Izard, C.E., Bartlett, E.S.: Patterns of emotions: A new analysis of anxiety and depression. Academic Press, Oxford, England (1972).
17. Polivy, J.: On the induction of emotion in the laboratory: Discrete moods or multiple affect states? Journal of Personality and Social Psychology. 41, 803–817 (1981).
18. Barrett, L.F.: Discrete Emotions or Dimensions? The Role of Valence Focus and Arousal Focus. Cognition & Emotion. 12, 579–599 (1998).
19. Harley, J.M., Bouchet, F., Azevedo, R.: Measuring Learners' Co-Occurring Emotional Responses during Their Interaction with a Pedagogical Agent in MetaTutor. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., and Panourgia, K. (eds.) Intelligent Tutoring Systems. pp. 40–45. Springer Berlin Heidelberg (2012).
20. Pekrun, R., Stephens, E.J.: Academic emotions. In: Harris, K.R., Graham, S., Urdan, T., Graham, S., Royer, J.M., and Zeidner, M. (eds.) APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors. pp. 3–31. American Psychological Association, Washington, DC, US (2012).
21. Rosenberg, E.L., Ekman, P.: Coherence between expressive and experiential systems in emotion. Cognition & Emotion. 8, 201–229 (1994).
22. Porayska-Pomsta, K., Mavrikis, M., D'Mello, S., Conati, C., Baker, R.S.J. d.: Knowledge Elicitation Methods for Affect Modelling in Education. International Journal of Artificial Intelligence in Education. 22, 107–140 (2013).
23. Tan, P.-N., Kumar, V., Srivastava, J.: Selecting the Right Interestingness Measure for Association Patterns. Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 32–41. ACM, New York, NY, USA (2002).