Han, J.-H., Shubeck, K., Shi, G.-H., Hu, X.-E., Yang, L., Wang, L.-J., Zhao, W., Jiang, Q., & Biswas, G. (2021). Teachable Agent Improves Affect Regulation: Evidence from Betty's Brain. *Educational Technology & Society*, *24* (3), 194–209.

Teachable Agent Improves Affect Regulation: Evidence from Betty's Brain

Jian-Hua Han[1*], Keith Shubeck[2,3], Geng-Hu Shi[2,3], Xiang-En Hu[2,3,4,5], Lei Yang[4], Li-Jia Wang[3,5], Wei Zhao[1], Qiang Jiang[1] and Gautum Biswas[6]

[1]School of Information Science and Technology, Northeast Normal University, China // [2]Department of Psychology, The University of Memphis, USA // [3]Institute for Intelligent Systems, The University of Memphis, USA // [4]School of Psychology, Central China Normal University, China // [5]Electrical and Computer Engineering, The University of Memphis, USA // [6]Institute for Software Integrated Systems, Vanderbilt University, USA // hanjh675@nenu.edu.cn // keithshubeck@gmail.com // genghushi@gmail.com // xiangenhu@gmail.com // yungly@mails.ccnu.edu.cn // wlj54188@gmail.com // zhaow577@nenu.edu.cn // jiangqiang@nenu.edu.cn // gautam.biswas@vanderbilt.edu

[*]Corresponding author

ABSTRACT: Intelligent learning technologies are often applied within the educational industries. While these technologies can be used to create learning experiences tailored to an individual student, they cannot address students' affect accurately and quickly during the learning process. This paper focuses on two core research questions. How do students regulate affect and what are the processes that affect regulation? First, this paper reviews the affect regulation methods and processes in an intelligent learning environment based on affective transition and affect compensation. This process, along with affect analysis, affect regulation, intelligent agents, and an intervention strategy can be used to analyze specific affect regulation methods and improve the affective regulation system. Seventy-two 7th grade students were randomly placed into an experimental condition that used Betty's Brain, an intelligent tutoring system (ITS), or a classroom control. A lag sequence analysis and a multinomial processing tree analysis of video data captured at 25-minute intervals revealed significant differences in affect transitions frequencies between the two groups. Based on the results of the above analyses and after-class interviews, we found that Betty's Brain was able to promote effective affect-regulation strategies to students in the domain of forest ecosystems.

Keywords: Teachable agent, Affect, Regulation, Tutoring, Betty's Brain

1. Introduction

Learning technologies have been widely used in education, which gradually changed the demand for talent and new educational formats (Liu & Lemeire, 2017). On one hand, these technologies benefit education (Popenici & Kerr, 2017). For example, online learning platforms make it possible for students to learn anytime and anywhere (Du et al., 2019). Recommendation algorithms in ITSs can be used to select adaptive content that fits a student's aptitude, characteristics, and learning progress (Wang et al., 2019). However, learning technologies are not without their disadvantages. For example, students may easily find themselves physically isolated in online learning environments, and they may feel helpless when they encounter difficulties (Raufelder et al., 2018). The status quo of learning technologies is that they make learning content easily accessible but they generally do not improve students' affective well-being. Students often suffer from inattention and lose navigation due to non-adaptive media materials, redundant content, and difficult tasks (Burek, 2017; Lim, 2004). Consequently, many students may disengage from the learning content and have unsatisfactory learning gains. Over time, they may feel fatigue and experience negative affect (Fida et al., 2015; Arsenio & Loria, 2014). Therefore, it is important to investigate the role of affect in technology-based learning environments, like ITSs, and the potential solutions for reducing negative affect and their detriments to learning.

Schutz et al. (2007) pointed out that affect influences students' motivation for learning. Research has shown that students who engage in exciting learning activities experience positive affect and have high learning gains (Gross, 1998). Lu (2012) found that learning activities that make students feel happy are important in teaching. Alkhalaf (2018) found that negative affect might lead to poor academic performance. Academic performance can be improved by increasing positive affect and through continuously combating or managing anxiety during learning. By examining students' degree of concentration, patience and learning willingness, Hwang et al. (2020) found that students using an adaptive learning system with affective and cognitive performance analysis mechanisms had significantly lower levels of mathematical anxiety than those who used the conventional learning system.

An ITS is a computer system that aims to provide immediate and individualized instruction or feedback to learners, usually without intervention from a human teacher (Patrut & Spatariu, 2016). ITSs have the potential to help students manage their affect. For example, web cameras and sensors enable ITSs to capture students' facial expressions and other physiological data that can be converted to affect information (i.e., Kołakowska et al., 2020). Then, an ITS can use various algorithms to provide feedback directly or indirectly to learners about how they can regulate their emotions. To explore the impact of an ITS on students' affect during learning, it is important to determine the mechanism of affect regulation and affect transitions when using an ITS.

2. Affect regulation in intelligent tutoring systems

2.1. Affect regulation and recognition

Affect is a kind of inner reaction of cognitive activity. It greatly impacts an individual's behavior. It can also influence an individual's behavior indirectly through affect reinforcement (Zhang, 2008). The affect regulation process can suppress and weaken negative affect, and can also maintain and enhance positive affect (Gross & John, 2003; Thomspon, 1994). For example, e-learning with affect regulation can significantly improve math performance for students with autism spectrum disorder (Chu et al., 2020). The transition from negative affect to positive affect depends on external feedback and internal regulation. Russell (2003) describes affect as consisting of valence (pleasure to displeasure) and arousal (active to inactive). When plotted, valence increases from left to right along the x-axis, and arousal increases moving upwards on the y-axis (Posner et al., 2005). Generally, affective states relevant to learning include boredom, flow, confusion, frustration, surprise, and delight (Craig et al., 2004). Affect occurs between students' cognitive balance and imbalance between boredom, frustration, confusion, and flow (D'Mello & Graesser, 2012). Boredom has negative valence and low arousal. Flow has positive valence and moderate arousal, whereas confusion has negative valence and moderate arousal. Finally, frustration has negative valence with high arousal (Baker et al., 2010). D'Mello and Graesser (2012) developed a model of affective state transitions based on this concept of equilibrium and disequilibrium by observing the main state transitions that occurred in AutoTutor (Nye, Graesser, & Hu, 2014) sessions. For example, flow may transition to confusion, which may transition to frustration or boredom (D'Mello et al., 2007).

Currently, there are three methods that are typically used to detect affect. For example, affect can be detected by using external devices like cameras, recorders, or other sensors that collect student body expressions (e.g., postures and gestures), facial expressions, verbal expressions (e.g., tone and timbre), and physical and psychological information (e.g., heartbeat, blood pressure and skin conductance). Affect can also be tracked through surveys with various affect scales, including questionnaires, self-reports, observations, and interviews. For example, the User Engagement Survey (UES) is used to measure attention, endurance, and participation (Grafsgaard et al., 2012). The third method is through system analysis (Pentel, 2015). This involves analyzing affect based on student interaction logs, accessing paths, frequency of mouse clicks, duration of staying on the page, and interactions with an ITS. Due to the situational and persistent features of affect, scholars can predict the affect of the next moment using the affective characteristics (e.g., intensity and classification) of the previous moment (Yu et al., 2013). An API can match captured images of an individual with the system model and automatically segment the expression into units, then the program can analyze the affect segmentation points to output affect and features (Maheshwari & Nagendhiran, 2017). By using posture estimation (Grafsgaard et al., 2012) and a gesture detection algorithm (Grafsgaard et al., 2012), a depth image regular pattern can be used to analyze the students' interest and concentration in the learning content. Although the analysis of valence and arousal is an effective method for predicting affect, the affect transition framework (D'Mello et al., 2007; D'Mello & Graesser, 2012) provides essential theoretical support for further exploration affect regulation and its effect on learning.

2.2. Affect regulation methods in intelligent tutoring systems

ITSs have different ways of capturing data relevant to affective states, which can be used to inform future system actions. Some use domain-independent rules (e.g., IF-THEN) and non-independent strategies (e.g., "You have done well"), which are used to achieve affect reinforcement (D'Mello & Graesser, 2013). Some use decision trees and sequential covering algorithms (e.g., AQ, CN2 and PIPPER), which are used to extract dataset rules for learning diagnosis (Quinlan, 1990; D'Mello & Graesser, 2013). Others use probabilistic models, like dynamic decision networks, which can be used to diagnose, evaluate, predict, and determine affect (Conati, 2002). Some are based on affect stratification. For example, the Hidden Markov Model and Baum-Welch algorithm can be used to output state transition probability matrix and vector parameters to evaluate affect (Collins, 1990; Liu & Lemeire, 2017; Thornton & Tamir, 2017). Some use dynamic Bayesian networks to focus on the causes and effects of affect, and probabilistic frameworks to handle high-level uncertainties to identify affect (Conati, 2002). Some use corpora, latent semantic analysis, word vectors and other analytical texts to predict affect response.

Teaching agents in an ITS can respond to and regulate negative affect by providing appropriate tutoring strategies and feedback. D'Mello et al. (2010) observed postures, facial expressions, and dialogue cues to stimulate pedagogical interventions, regulate boredom, frustration, and confusion, and then promoted participation and task persistence in AutoTutor. Wayang Outpost (Arroyo et al., 2014) adopted heuristic strategies for responding to students' affect, including text information and mapping learning behaviors. Their results showed that students can alleviate their boredom and change their behaviors based on digital interventions (Woolf et al., 2009). Although the students in the experimental and control group showed very similar feelings of pleasure, arousal, and dominance, Daradoumis and Arguedas (2020) found that the experimental group was slightly more expressive about their personal satisfaction through an affect pedagogical agent. Based on the theorized model of D'Mello and Graesser (2012), Alexandra et al. (2019) examined three types of affective transitions and their correlations with pretest-to-posttest learning. They found that the presence of boredom indicates a student's knowledge state, but not their learning. In summary, ITSs are mostly used in one-to-one tutoring simulations of human teachers, and they use domain and student models to support students' cognition and affect regulation.

2.3. Affect regulation processes in intelligent tutoring systems

Qin et al. (2014) built an affect compensation structure, using affect recognition, personalized affect regulation, and negative affect compensation. First, multi-modal methods (e.g., facial expressions, language, behavior, and interactive text) use high resolution cameras and wearable sensors to recognize students' positive affect or negative affect, such as frustration. Next, personalized regulation methods are used to analyze the student's characteristics and regulation strategies, and then judge the affect regulation methods. They then use affect compensation (including expert tutoring and peer help) to enhance the system's confidence in the student's optimal affective states. Affect compensation can optimize the affect database, and the affect database can be used for affect recognition and negative affect compensation. Finally, based on historical compensation cases and compensation lists, the systems can be used to alleviate negative affect. According to the affect compensation structure (Qin et al., 2014) and the affect recognition method, the four functional modules and the affect regulation processes can be implemented in an ITS (see Figure 1).

The first module is an affect analysis module. Affect analysis is key to providing intervention strategies. System tracking can determine whether students have studied or not and can also track their affect and transitions. Affect extraction is defined as using self-report and text mining to extract affective valence and arousal. Affect recognition is based on recording and quantifying personal physiological, psychological, and cognitive information to detect affect states. For example, some scholars use cameras and wearable devices to identify student affect in MetaTutor (Harley et al., 2015). The second module is an intelligent agent module. This mainly involves an intelligent agent, like an expert, a teacher, or a peer, who is a virtual animated character that plays a certain role during an interactive session in an ITS. Expert agents have a wealth of knowledge in various disciplines and domains, such as students communicating with virtual doctors and patients, or reasoning about the patient symptoms of island residents in Crystal Island (Taub et al., 2017). Teacher agents track students' knowledge construction processes, for instance, the agents in AutoTutor that judge questions and then give appropriate feedback by leveraging "expectation-misconception tailored dialogue" (D'Mello & Graesser, 2013). Peer agents in Betty's Brain act as learning peers and assistants by using a learn-by-teaching method to build knowledge (Han et al., 2019). The third module is an affect regulation module. Students can regulate their affect by themselves and can also regulate their affect by external feedback and interventions. The external feedback and interventions are mostly based on an analysis of learning characteristics to achieve precise tracing and interventions with dialogues, student logs, and questionnaires in real-time. Self-report data from students suggest they can acquire appropriate affective and cognitive feedback automatically in Betty's Brain (Biswas et al., 2016). Students can adjust their cognitive affect in real-time based on lists and features selection methods, agent dialogue, problem clues, animation prompts and diagnostic reports (Taub et al., 2017). Sometimes, ITSs can present empathetic agents and virtual companions to augment students' awareness of cognitive presence and affect presence, such as in Wayang Outpost (Arroyo et al., 2014). The fourth module is an intervention strategy module. Intervention strategies are the means and methods of affective intervention. Individual moods and cognitive dilemmas are affected by affect, and some scholars use self-explanations and learning-early-warnings to relieve learner confusion about stress analysis in Andes, an ITS for physics (VanLehn et al., 2010). Peer agents can intervene with students in mathematical problem solving in real time, such as SimStudent agent dialogues (Matsuda et al., 2013). Intelligent systems can provide adaptive resources and suggestions based on cognitive impairments or resources property. Taking Wayang Outpost as an example, the system can provide cognitive clues and suggestions in addition to different media materials like video, sound, text, and test (Woolf et al., 2009). A system can provide hints and suggestions and help students solve problems correctly in ASSISTments, given the steps and results of students' questions. (Heffernan & Heffernan, 2014).

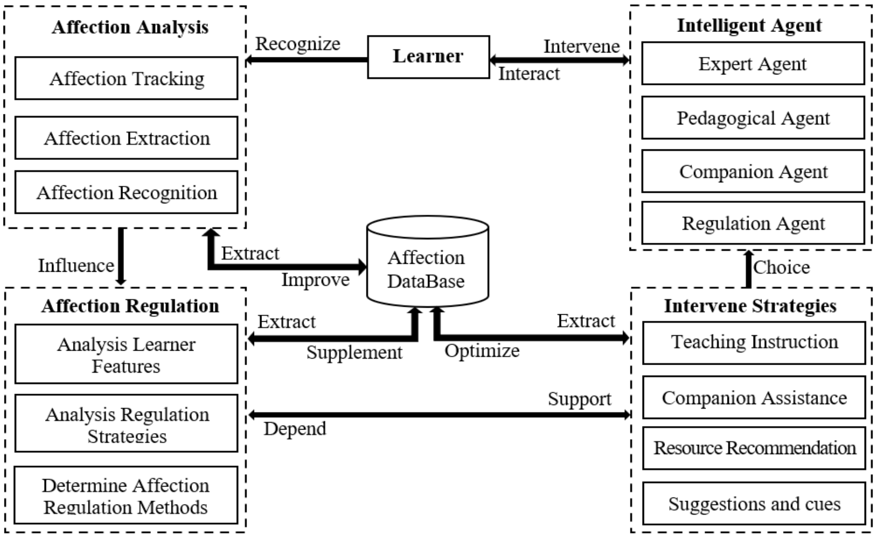

Figure 1. Affective regulation processes in intelligent tutoring environment (adapted from Qin et al., 2014)

In general, based on personal records and known affect, an ITS analyzes resource acceptance manners, preferences, and moods, and uses personalized regulation strategies to increase and decrease positive and negative affect. It can supplement the affect database if the system does not have a student's affect records. Tutoring strategies grounded in ITS cognitive principles and algorithms (i.e., error recognition and correction, student modeling, and natural language dialogue) instruct agents to indirectly address student affect. These strategies are also sent directly to students, which allow computers to act as virtual instructors to impart knowledge and provide adaptive feedback for students. Moreover, based on the frequency of mouse clicks and the path of page access, an ITS analyzes arousal and valence to predict the next affective state, and uses some encouragement, care, praise, and criticism via agents to optimize database adjustment strategies. Additionally, systems can inquire about learner's affect and present them learning tasks and their progress, which can assist students in their learning introspection. In short, ITSs can help students avoid cognitive impasses, errors, and misconceptions, and can also alleviate negative affect. They use process supervision to promote students' reflection and improve their cognition and metacognition.

3. Research design

This paper seeks to identify the affective experiences and effectiveness of using an ITS compared to a non-ITS learning environment. The learning content covered the ecological relationship between wolves, hunters, cows, deer, grass, rainfall, and other concepts about a forest ecosystem. For example, some lessons present these concepts in terms of an increase or decrease in water and food availability and how this affects the animal population.

The experimental group used Betty's Brain, an ITS developed by a combination of computer science, psychology, and education researchers in the engineering school of Vanderbilt University. The system uses virtual teachers (Mr. Davis) and virtual students (Betty) to intervene and guide students' cognition and affect. The ITS consists of a "Causal Map," "Science Book," "Notes," "Quiz Results," and "Teacher's Guide." The control class used a non-ITS (F_S), which is based on Moodle 2.8 and covers the same domain and content as Betty's Brain, including "Science Book," "Notes," "Quiz Results," and "Teacher's Guide." F_S does not have virtual teachers and students, and participants used Microsoft Word to build causal relationships.

Participants included 72 students in the seventh grade of a middle school in Changchun. Participants consisted of 35 boys and 37 girls. All the students had no experience with using ITSs, and the two classes were taught by the same teachers.

The experimental process was mainly divided into a teaching stage, autonomous learning stage and an after-class interview. In the teaching stage, teachers guided students through content with the theme of "forest ecosystem" and taught them how to use a “Causal Map,” “Science Book,” “Notes,” “Quiz Results,” and “Teacher’s Guide” in 3 minutes. In the autonomous learning stage, students needed to construct a causality diagram in 25 minutes, during which we collected the video data. Afterwards some students needed to complete interviews lasting no longer than 13 minutes.

This study used 46 Mosheng RQES008 HD digital cameras to capture facial expressions with a USB 2.0 interface, and the cameras were assembled and installed on every computer. The coding of the types of affect was based on previous coding schemes used by McDaniel et al. (2007) and Altuwairqi et al. (2021). We referred to the facial expressions in the video data to judge students’ affect. Each coding result was recorded in a table, like Figure 2.

Time	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Stu1	Confusion	Boredom	Confusion	Confusion	Confusion	Boredom	Confusion	Confusion	Confusion	Confusion
Stu2	Confusion	None	Confusion	Flow	Flow	Flow	Flow	Confusion	Confusion	Confusion
Stu3	Confusion	Flow	Confusion	Flow	Confusion	Flow	Flow	Confusion	Boredom	Confusion
Stu4	Boredom	Boredom	Confusion	Flow	Flow	Flow	Confusion	Flow	Flow	Confusion

Figure 2. Affect encoding sample in the control group

As is shown in Figure 2, the first row represents time, which is used to mark 25 encodings. The first column represents students’ identity, for example, stu1 as the first student. Affect for each timepoint were coded as either boredom, flow, confusion, frustration, surprise, delight, or none. After further observation and discussion, because some affective states, such as frustration or surprise were very rare, we only considered “boredom,” “flow,” “confusion,” “delight” and “no affect” in this paper.

4. Results and analysis

Two sets of 25-minute videos during the autonomous learning stage were used for analysis, which coded the following states of: “boredom,” “flow,” “confusion,” “delight” and “no affect.” Our coding process was handled and reviewed by two experimenters in charge. If the number of the matching codes is x, and the number of codes for each person is y, then the quotient (x/y) can be defined as the coding consistency. The two experimenters simultaneously encoded two of the same samples in order. After comparing and contrasting between both results, there were 37 matches in the 50 coded data points. In short, these were recorded every 30 seconds during the autonomous learning stage, and the video coding consistency between the two raters is (37/50) 74%.

4.1. Affective cumulative analysis

To analyze the overall affect distribution, the affective states of each group are summarized below. Taking the autonomous learning stage into account, twenty-five minutes of activities were recorded every 30 seconds. The numbers in the first row represent 50 different recordings, and the data represents the frequency of the corresponding affect. Taking the 21st encoding in the 22nd column in the control group as an example, 1 student showed “boredom,” 11 were in “flow,” 6 were “confused,” 2 showed “delight,” and for 14 of the students we were unable to determine their affect because they were off camera. Accordingly, the cumulative frequency of affect is summarized in Tables 1 and 2.

Table 1. The affective accumulation table of the control group

Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Boredom	1	2	1	1	0	1	1	0	1	1	3	1	4	2	2	2	1	2	2	0	1	0	2	2	1
Flow	10	10	9	13	11	13	11	11	13	13	12	15	13	12	11	14	10	14	14	16	11	13	11	10	9
Confusion	6	4	8	4	7	2	6	7	3	6	4	2	3	6	4	3	7	1	1	2	6	3	5	7	3
Delight	4	1	1	1	1	1	0	0	1	1	1	1	1	1	2	0	2	1	2	0	2	0	0	2	0
None	13	17	15	15	15	17	16	16	16	13	14	15	13	13	15	15	14	16	15	16	14	18	16	13	21
Number	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Boredom	1	0	1	0	1	3	1	2	0	0	0	2	1	1	0	0	0	0	0	0	1	1	1	0	2
Flow	9	12	9	9	11	10	8	6	7	9	7	6	9	8	7	8	7	9	5	6	7	5	7	7	4
Confusion	6	3	4	3	6	3	4	7	7	7	6	3	5	1	3	4	4	1	2	1	0	0	0	0	0
Delight	2	3	2	1	1	0	2	2	2	2	2	1	1	2	1	0	0	2	1	1	0	0	0	0	0
None	16	16	18	21	15	18	19	17	18	16	19	22	18	22	23	22	23	22	26	26	26	28	26	27	28

Table 2. The affective accumulation table of the experimental group

Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Boredom	2	2	0	0	0	0	1	0	1	1	0	0	1	0	1	1	0	3	2	3	2	1	3	0	0
Flow	17	14	12	15	17	22	22	20	23	20	23	24	27	27	26	23	25	23	21	25	27	23	24	26	28
Confusion	3	3	7	5	6	2	4	7	2	6	5	4	1	2	5	6	3	5	8	3	4	6	4	3	2
Delight	1	2	2	2	0	0	1	0	4	2	2	3	1	0	0	2	2	3	0	3	0	1	2	3	0
None	15	17	17	16	15	14	10	11	8	9	8	7	8	9	6	6	8	4	7	4	5	7	5	6	8
Number	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Boredom	2	2	1	0	3	4	3	1	2	4	1	2	3	0	2	1	0	0	1	0	0	0	0	0	0
Flow	27	19	24	25	22	19	24	25	20	18	16	17	18	18	15	14	14	13	10	11	9	9	9	9	9
Confusion	1	5	4	4	1	5	1	2	5	3	4	2	2	1	1	2	1	0	0	0	0	0	0	0	0
Delight	1	4	3	1	2	1	1	3	0	1	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0
None	7	8	6	8	10	9	9	7	11	12	17	17	15	19	19	19	23	25	27	27	29	29	29	29	29

According to the accumulated frequency (as shown in Tables 1 and 2), we observed the following:

- "Flow" and "none" frequently occurred during their learning processes followed with "confusion" in each group, "boredom" and "delight" occurred the least, according to the accumulated frequency.
- Affect changes over time, so it follows that each affect here fluctuates across the recordings. For example, the frequency of "flow" is 9 to 16 in the first 25 recordings in the control group, and "flow" is 4 to 12 in the last 25 recordings.
- Many of the affective states in the groups were coded as "none." This is because the task was not completed, and the camera was disconnected. In this case, it is impossible to determine some affective states. For instance, the frequency of "none" fluctuated from 13 to 21 in the first 25 recordings, and then the frequency increased from 15 to 28 in the last 25 recordings in the control group.

To summarize the overall affect distribution of each group, the cumulative data is represented by a bar graph as shown in Figures 3 and 4.

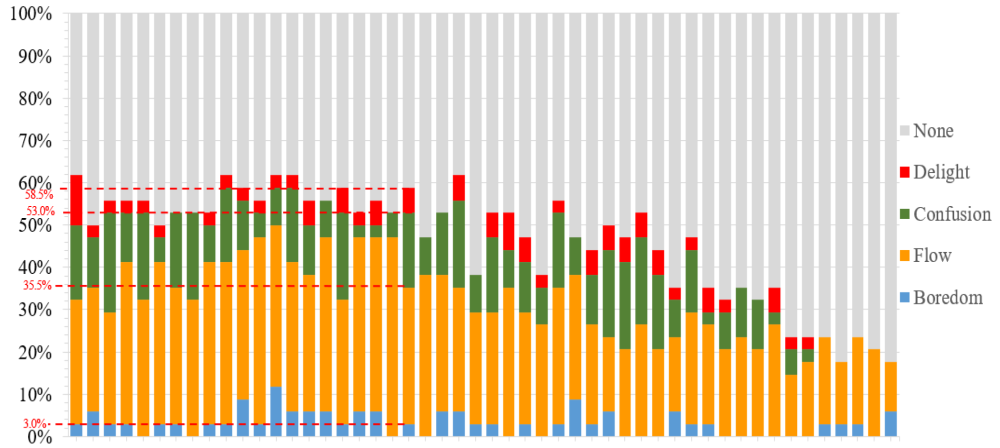

Figure 3. The cumulative affect in the control group

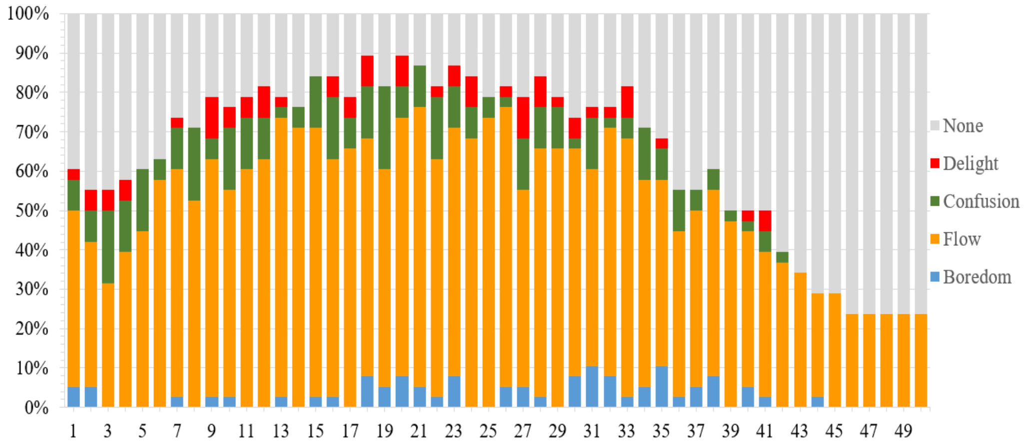

Figure 4. The cumulative affect in the experimental group

Twenty-five minutes of activity were recorded every 30 seconds, and 50 recordings are shown on the x-axis. The y-axis represents the cumulative percentage of each affect. Taking the 21st recording in the x-axis of the control group as an example, 3.0% of students were "bored," 32.5% were in "flow," 17.5% were "confused," 5.5% showed "delight," and 41.5% could not be determined.

Figures 3 and 4 display the proportion of each affect at 50 different video captures throughout the learning sessions. For example, the proportion of "flow" in the experimental group is higher than that of the control group throughout the entire session. In the control group, "flow" increased from time 1 to time 3, peaked at the 20th recording (about 47%), and then slowly declined to 11% by the 50th recording. Comparatively, "flow" in the experimental group fluctuated from approximately 32% to around 45% from time 1 to time 5, then gradually increased to about 74% at the 25th recording. It then gradually declined to roughly 24% by the 50th recording.

In general, there was less "confusion" in the experimental group than the control group. In the control group, the proportion of "confusion" started at 17% and increased to about 21% by the 5th recording. From the 6th to the 45th recording, "confusion" fluctuated between approximately 3% to roughly 21%, and afterwards students did not show "confusion." In the experimental group, the proportion of "confusion" fluctuated between about 2% to around 21% up to the 37th recordings, and students did not display "confusion" from the 43rd to the 50th recording.

In summary, both the cumulative frequency and percentage of "flow" was significantly higher in the experimental group than that in the control group. The cumulative percentage of "confusion" was higher in the control group than that in the experimental group. There was no significant difference between the two conditions for the other affective states.

4.2. Analysis of the difference in each group

Generalized Sequential Querier (GSEQ) can be used to analyze sequence observation data. GSEQ can be used to perform coding and output the frequency of affect transitions (see Table 3). Here, the data consists of the frequency of transitions from the ith affect to the jth affect, denoted as X_{ij}. The variable i represents the affect index of columns, j represents the affect index of rows, N represents the type of affect coded, and the range of changes both i and j is [1, N].

Table 3. Joint frequency table

	Given	Boredom	Flow	Confusion	Delight	None	Totals
Control group	Boredom	13	14	10	4	10	51
	Flow	11	341	59	17	61	489
	Confusion	10	56	74	10	39	189
	Delight	3	18	7	10	15	53
	None	15	54	33	8	774	884
	Totals	52	483	183	49	899	1666
	Given	Boredom	Flow	Confusion	Delight	None	Totals
Experimental group	Boredom	7	21	14	3	11	56
	Flow	21	795	69	29	49	963
	Confusion	11	71	49	7	16	154
	Delight	4	27	10	9	8	58
	None	11	41	9	9	561	631
	Totals	54	955	151	57	645	1862

The frequency of each type of affective transition is different from each group (see Table 3). A total of 1666 transformations were observed in the control group and 1862 changes in the experimental group. Frequent patterns (frequency greater than or equal to 30) include: "*flow/confusion/none→flow/confusion/none*" in the control group, "*flow/confusion/none→flow*," "*flow/confusion→confusion*" and "*none/flow→none*" in the experimental group. Some frequent transition patterns we observed in both conditions are: "*flow/confusion/none →flow*," "*flow/confusion→confusion*," and "*none/flow→none*." The transition patterns of "*confusion→ none*," and "none→confusion" were frequent only in the control group. The frequencies of "*flow/confusion/delight→bored*," "*bored/flow/confusion/delight→flow*," "*bored/flow/delight→confused*," "*flow/none→delight*," and "*bored→no affect*" were all significantly higher in the experimental group than in the control group. The affect transition frequencies of "*bored/none→bored*," "*none→flow*," "*confusion/none→none/confused*," "*bored/confused/delight→delight*," and "*flow/delight→none*" were significantly higher in the control group than in experimental group.

The GSEQ tool calculated the expected frequency of affect transitions shown in Table 4, by using the observed frequencies shown in Table 3 and the M_{ij} formula.

$$M_{ij} = \frac{(X_{i+})*(X_{+j})}{\sum_{i=1}^{i=N}\sum_{j=1}^{j=N}X_{ij}} = \frac{(\sum_{j=1}^{j=N}X_{ij})*(\sum_{i=1}^{i=N}X_{ij})}{\sum_{i=1}^{i=N}\sum_{j=1}^{j=N}X_{ij}} = \frac{X_{i,j=totals}*X_{i=totals,j}}{X_{i=totals,j=totals}} \quad (1)$$

Table 4. Expected frequency table

	Given	Boredom	Flow	Confusion	Delight	None
Control group	Boredom	1.592	14.786	5.602	1.500	27.520
	Flow	15.263	141.769	53.714	14.382	263.872
	Confusion	5.899	54.794	20.761	5.559	101.987
	Delight	1.654	15.366	5.822	1.559	28.600
	None	27.592	256.286	97.102	26.000	477.020
	Given	Boredom	Flow	Confusion	Delight	None
Experimental group	Boredom	1.624	28.722	4.541	1.714	19.398
	Flow	27.928	493.912	78.095	29.480	333.585
	Confusion	4.466	78.985	**12.489**	**4.714**	**53.346**
	Delight	1.682	29.748	4.704	1.776	**20.091**
	None	**18.300**	323.633	**51.171**	19.316	**218.579**

Expected frequency M_{ij} refers to the product of $X_{i,j=totals}$ (sum of the frequencies at which all affective states turn into the j[th] affect) multiplied by $X_{i=totals,j}$ (sum of the frequencies at which the i[th] affective state turns into all affective states) and the quotient of $X_{i=totals,j=totals}$ (sum of all affect transitions). In other words, this formula is used to calculate the expectation of each transition, which is placed in all the transition processes. The expected affect frequency is different from the initial frequency, such as the joint frequency of transition of "*flow*" to "*confusion*" equals 59, and the expected frequency of the transition of "*flow*" to "*confusion*" equals 53.714.

Some of the expected frequencies of affect transitions were significantly different between the two groups. The frequencies of "*bored/flow/delight→bored*," "*bored/confusion/delight/none/flow→flow*," "*flow→confused*," "*bored/flow/delight→delight*" and "*none→none*" in the experimental group are higher than those in the control group. The frequencies of "*confused/none→bored*," "*bored/confused/delight/none→confused*," "*confused/none→delight*," and "*bored/confused/delight/none→none*" in the control group are higher than those in the experimental group.

Table 5. Summary table of adjusted residuals of affective transformation

	Given	Boredom	Flow	Confusion	Delight	None
Control group	Boredom	**9.330**	-0.246	**2.000**	**2.104**	**-4.999**
	Flow	-1.319	**23.624**	0.910	0.834	**-21.899**
	Confusion	1.822	0.205	**13.153**	**2.031**	**-9.763**
	Delight	1.080	0.811	0.526	**6.974**	**-3.809**
	None	**-3.555**	**-21.887**	**-10.064**	**-5.230**	**29.250**
	Given	Boredom	Flow	Confusion	Delight	None
Experimental group	Boredom	**4.347**	**-2.096**	**4.701**	1.013	**-2.395**
	Flow	-1.915	**27.936**	-1.545	-0.129	**-27.737**
	Confusion	**3.276**	-1.344	**11.253**	1.116	**-6.604**
	Delight	1.843	-0.733	**2.588**	**5.595**	**-3.390**
	None	**-2.130**	**-27.685**	**-7.564**	**-2.932**	**35.234**

Note. $|Z_{ij}|>1.96$.

This paper uses the Z_{ij} formula to calculate the adjusted residual value given the data shown in Table 4 and the joint frequencies in Table 5.

$$Z_{ij} = \frac{X_{ij}-M_{ij}}{\sqrt{M_{ij}*\left(1-\frac{M_{ij}}{X_{+j}}\right)*\left(1-\frac{M_{ij}}{X_{i+}}\right)}} = \frac{X_{ij}-M_{ij}}{\sqrt{M_{ij}*\left(1-\frac{M_{ij}}{\sum_{i=1}^{i=N}X_{ij}}\right)*\left(1-\frac{M_{ij}}{\sum_{j=1}^{j=N}X_{ij}}\right)}} = \frac{X_{ij}-M_{ij}}{\sqrt{M_{ij}*\left(1-\frac{M_{ij}}{X_{i=totals,j}}\right)*\left(1-\frac{M_{ij}}{X_{i,j=totals}}\right)}} \quad (2)$$

Z_{ij} is used to calculate the difference between the observation and the expectation. We use the formula (Haberman, 1979) to execute and compute the adjusted residual value. The product of probability of neither belonging to $X_{i,j=totals}$ nor belonging to $X_{i=totals,j}$ is used as the weight of M_{ij}. The difference between the actual value and the expected value is used as the dividend, and the root of the expected value including weight is used as the divisor. The quotient of the two is called the adjusted residual value. The adjusted residuals are similar to Z-scores; Z_{ij} is normally distributed, and the Z-test can be used to test the statistical significance. According to the standard normal distribution in the Z-value table, Z_{ij} is substituted into the normal distribution to find the corresponding probability *P*-value. Also, $|Z_{ij}|>1.96$ (95% confidence interval) is selected to indicate a significant change in affect, which is marked in bold font.

According to Table 5, the significant $|Z_{ij}|$ is marked on the affect conversion graph, and the arrows point to the next affect of the transition, and thicker lines indicate more significance of the affect transitions. The conversion relationship is drawn, as shown in Figures 5 and 6.

We observed the following affective conversions.

- There are repeating or recurrent patterns of affect conversions, which include: "*boredom→boredom/confusion/none*," "*flow→flow/none*," "*confusion→confusion/none*," "*delight→delight/none*," and "*none→boredom/flow/confusion/delight/none*." The "*boredom→boredom*," "*confusion→confusion*," "*confusion→none*," "*none→confusion*," "*delight→delight*," "*delight→none*," "*none→delight*," "*boredom→none*," "*none→boredom*," and "*confused→delight*" transitions are more significant than rest of the affect transition in both the control group and experimental group.
- There are different transition patterns between the two groups. For example, "*boredom/confusion→delight*" is significant in the control group, but not in the experimental group. "*Delight→confusion*" and "*confusion→boredom*" are significant patterns in the experimental group, but not in the control group. Additionally, the transition of "*boredom→delight*" and "*confused→delight*" is significant in the control group, but not the experimental group. The transition of "*delight→confused*" and "*confused→boredom*" are significant in the experimental group, but not the control group.

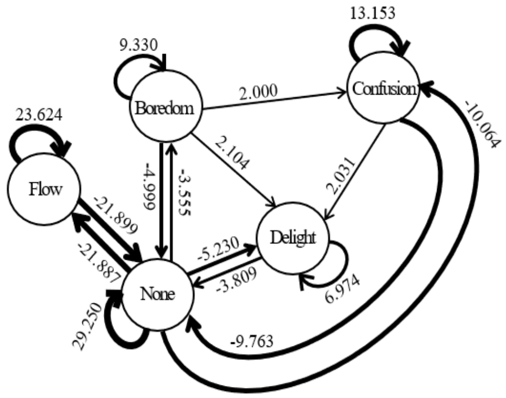

Figure 5. Affective conversion in the control group

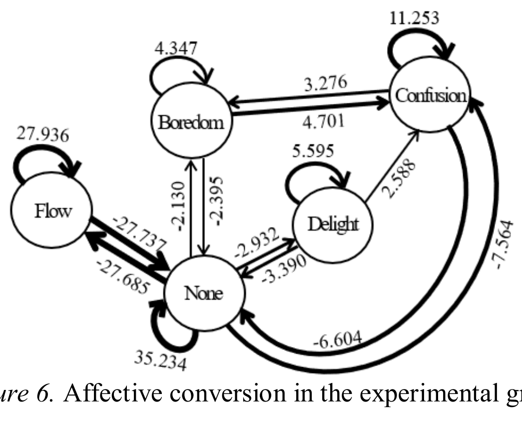

Figure 6. Affective conversion in the experimental group

In summary, based on the data analysis above, we can see the specific difference and significance in each transition in each group. However, we cannot compare the specific Z-value across different groups, and we can only compare the Z-value in the inner group.

4.3. Analysis of the difference between the two groups of affect

This paper analyzes the significance of the difference between the two groups, by using a multinomial processing tree (MPT) to analyze the frequency of transition (see the Table 3) in a general processing tree (GPT) software. The results are shown in Table 6.

In Table 6, the parameters (PA) in the first column represents every transition in each group, for example, the CAA pattern refers to the A→A in the control group, and EAA refers to the A→A in the experimental group. The star sign (*) in the first column in the table refers to the parameter(s) that are restricted as constant. Additionally, in the second column, EV is the estimated value of the parameter. SD refers to standard deviation in the third column. For the confidence intervals (CI), in the fourth column, LO represents the lower limit of the confidence interval and UP represents the upper limit of the confidence interval in the fifth column. "Sig of Difference" refers to the statistical significance of the difference in artificial processing. "LO_SI" shows the lower limit of the significance of the difference, and the last column shows the upper limit of the significance of the difference (UP_SI).

By using the confidence interval of the parameter estimation values in the groups, we can directly assess the difference of the significance of the difference of the parameter. The LO_SI of difference value equals the upper limit of the confidence interval in the experimental group. The negative sign corresponds to the lower limit of the confidence interval in the control group. Similarly, the UP_SI of difference value equals the upper limit of the confidence interval in the control group minus corresponds to the lower limit of the confidence interval in the experimental group. The difference is significant if the one of the LO_SI value and UP_SI value is less than 0, otherwise it cannot be intuitively judged.

Table 6. Analysis in model multinomial processing tree

Control Group					Experimental Group						
			95% CI					95% CI		Sig of Difference	
PA	EV	SD	LO	UP	PA	EV	SD	LO	UP	LO_SI	UP_SI
CA*	0.031	constant			EA*	0.030	constant				
CAA	0.255	0.061	0.135	0.375	EAA	0.125	0.044	0.038	0.212	0.076	0.336
CAB	0.275	0.062	0.152	0.397	EAB	0.375	0.065	0.248	0.502	0.350	0.149
CAC	0.196	0.056	0.087	0.305	EAC	0.250	0.058	0.137	0.363	0.276	0.168
CAH	0.078	0.038	0.005	0.152	EAH	0.054	0.030	-0.005	0.113	0.108	0.158
CB*	0.294	constant			EB*	0.517	constant				
CBA	0.022	0.007	0.009	0.036	EBA	0.022	0.005	0.013	0.031	0.022	0.023
CBB	**0.697**	**0.021**	**0.657**	**0.738**	**EBB**	**0.826**	**0.012**	**0.802**	**0.850**	**0.193**	**-0.064**
CBC	**0.121**	**0.015**	**0.092**	**0.150**	**EBC**	**0.072**	**0.008**	**0.055**	**0.088**	**-0.004**	**0.094**
CBH	0.035	0.008	0.019	0.051	EBH	0.030	0.006	0.019	0.041	0.022	0.032
CC*	0.113	constant			EC*	0.083	constant				
CCA	0.053	0.016	0.021	0.085	ECA	0.071	0.021	0.031	0.112	0.091	0.054
CCB	**0.296**	**0.033**	**0.231**	**0.361**	**ECB**	**0.461**	**0.040**	**0.382**	**0.540**	**0.309**	**-0.021**
CCC	0.392	0.036	0.322	0.461	ECC	0.318	0.038	0.245	0.392	0.070	0.216
CCH	0.053	0.016	0.021	0.085	ECH	0.045	0.017	0.013	0.078	0.057	0.072
CG*	0.472	constant									
CGA	0.017	0.004	0.008	0.025	EGA	0.017	0.005	0.007	0.028	0.019	0.018
CGB	0.061	0.008	0.045	0.077	EGB	0.065	0.010	0.046	0.084	0.039	0.031
CGC	**0.037**	**0.006**	**0.025**	**0.050**	**EGC**	**0.014**	**0.005**	**0.005**	**0.024**	**-0.001**	**0.045**
CGH	0.009	0.003	0.003	0.015	EGH	0.014	0.005	0.005	0.024	0.021	0.010
CH*	0.032	constant			EH*	0.031	constant				
CHA	0.057	0.032	-0.006	0.119	EHA	0.069	0.033	0.004	0.134	0.140	0.115
CHB	0.340	0.065	0.212	0.467	EHB	0.466	0.066	0.337	0.594	0.382	0.130
CHC	0.132	0.047	0.041	0.223	EHC	0.172	0.050	0.075	0.270	0.229	0.148
CHH	0.189	0.054	0.083	0.294	EHH	0.155	0.048	0.062	0.248	0.165	0.232

Note. A refers to boredom, B refers to flow, C refers to confusion, H refers to delighted, G refers to no affect.

According to the MPT model (as shown in Table 6), some affect transitions are significantly different between the two groups, and the significant transitions are bolded in the table. For instance, the B→B transition in the experimental group (estimated value = 0.826, *SD* = 0.012, lower limit of the confidence interval = 0.802) is significantly higher than the B→B in the control group (estimated value = 0.697, *SD* = 0.021, upper limit of confidence interval = 0.738); the C→B in the experimental group (estimated value = 0.461, *SD* = 0.040, lower limit of confidence interval= 0.382) is significantly higher than the C→B in control group (estimated value = 0.296, SD = 0.033, upper limit of confidence interval = 0.361). The B→C (estimated value = 0.072, *SD* = 0.008, upper confidence interval= 0.088) in the experimental group is significantly smaller than B→C in the control group (estimated value = 0.121, *SD* = 0.015, lower confidence interval = 0.092). The G→C in the experimental group (estimated value =0.014, *SD* = 0.005, upper confidence interval = 0.024) is significantly smaller than in the control group (estimated value = 0.037, *SD* = 0.006, lower confidence interval = 0.025). The results of a chi-square test indicate that there are significant differences in the affective transitions between both groups ($X^2[9]$ = 0.01988).

According to the affect transitions of the two groups in Table 3, this paper calculates the difference between combined frequency of affect changes in two groups, denoted as X'_{ij}, which represents the difference of frequency between the ith affect and the jth affect transition, see Table 7.

Table 7. The combined frequency of the difference between the two groups

Given	Boredom	Flow	Confusion	Delight	None	Totals
Boredom	-6	7	4	-1	1	5
Flow	10	454	10	12	-12	474
Confusion	1	15	-25	-3	-23	-35
Delight	1	9	3	-1	-7	5
None	-4	-13	-24	1	-213	-253
Totals	2	472	-32	8	-254	196

According to the value of X'_{ij}, the combined frequency is different from the initial joint frequency. Some transitions in the experimental group were less than those in the control group, such as, "*none→none*," "*confusion→confusion*," "*none→confusion*," "*confusion→none*," and "*none→flow*." Some affective changes in the experimental group are more frequent than those in the control group, for instance, "*flow→flow*," "*confusion→flow*," and "*flow→delight*."

According to the frequency of transformation in Table 7, this article calculates the expected transformation of the difference between the experimental group and the control group, denoted as M'_{ij}, which represents the expected frequency of transformation from the ith affect to the jth affect.

$$M'_{ij} = \frac{(X'_{i+})*(X'_{+j})}{\sum_{i=1}^{i=N}\sum_{j=1}^{j=N} X'_{ij}} = \frac{(\sum_{j=1}^{j=N} X'_{ij})*(\sum_{i=1}^{i=N} X'_{ij})}{\sum_{i=1}^{i=N}\sum_{j=1}^{j=N} X'_{ij}} = \frac{X'_{i,j=totals} * X'_{i=totals,j}}{X'_{i=totals,j=totals}} \quad (3)$$

Table 8. The expected frequency of the difference between the two groups

Given	Boredom	Flow	Confusion	Delight	None
Boredom	**0.051**	**12.041**	-0.816	**0.204**	-6.480
Flow	**4.837**	**1141.469**	-77.388	**19.347**	-614.265
Confusion	-0.357	-84.286	**5.714**	-1.429	**45.357**
Delight	**0.051**	**12.041**	-0.816	**0.204**	-6.480
None	-2.582	-609.265	**41.306**	-10.327	**327.867**

According to the expected frequency (as shown in Table 8), the difference frequency between the two groups is also different from the combined frequency (as shown in Table 7). Some transitions in the experimental group are less than those in the control group including: "*flow→none*," "*none→flow*," "*confusion→flow*," "*flow→confusion*," "*none→delight*," "*boredom→confusion*," "*boredom→none*," "*delight→none*," "*none→boredom*," "*confusion→delight*," "*delight→confusion*," and "*confusion→boredom*." Some changes in the experimental group are more likely than that in the control group including: "*flow→flow*," "*none→none*," "*confusion→none*," "*none→confusion*," "*flow→delight*," "*delight→flow*," "*boredom→flow*," "*confusion→confusion*," "*flow→boredom*," "*boredom→delight*," "*delight→delight*," "*delight→boredom*," and "*boredom→boredom*."

Z'_{ij} is used to calculate the difference between observations and expectations. The product of the probabilities of neither belonging $X'_{i,j=totals}$ nor belonging $X'_{i=totals,j}$ is used as the weight of M'_{ij}. The difference between the initial value and the expected value is used as a dividend, and weighted expected value is used as a divisor, and a quotient of the two is called the adjusted residual value.

According to frequency (Table 7) and expectation (Table 8), this paper calculates the adjusted residual value, which is expressed as Z'_{ij} (see Table 9).

$$Z'_{ij} = \frac{X'_{ij} - M'_{ij}}{\sqrt{M'_{ij} * \left(1 - \frac{M'_{ij}}{X'_{+j}}\right) * \left(1 - \frac{M'_{ij}}{X'_{i+}}\right)}} = \frac{X'_{ij} - M'_{ij}}{\sqrt{M'_{ij} * \left(1 - \frac{M'_{ij}}{\sum_{i=1}^{i=N} X'_{ij}}\right) * \left(1 - \frac{M'_{ij}}{\sum_{j=1}^{j=N} X'_{ij}}\right)}} = \frac{X'_{ij} - M'_{ij}}{\sqrt{M'_{ij} * \left(1 - \frac{M'_{ij}}{X'_{i=totals,j}}\right) * \left(1 - \frac{M'_{ij}}{X'_{i,j=totals}}\right)}} \tag{4}$$

Table 9. The residual of the difference between the two groups

Given	Boredom	Flow	Confusion	Delight	None
Boredom	**-27.277**	-1.240	**5.007**	**-2.757**	**1.964**
Flow	**1.981**	**-14.398**	**7.734**	-1.432	**13.466**
Confusion	**2.103**	**8.395**	**-10.973**	-1.237	**-6.170**
Delight	**4.278**	-0.748	**3.967**	**-2.757**	-0.137
None	-0.586	**13.450**	**-6.225**	**2.378**	**-13.025**

Note. $|Z'_{ij}|>1.96$.

As shown in Table 9, the significant Z'_{ij} is marked on the affect conversion graph, with the arrows pointing to the next affect of the transition. Thicker lines indicate more significance of the affect transitions. The conversion relationship is drawn, as shown in Figure 7.

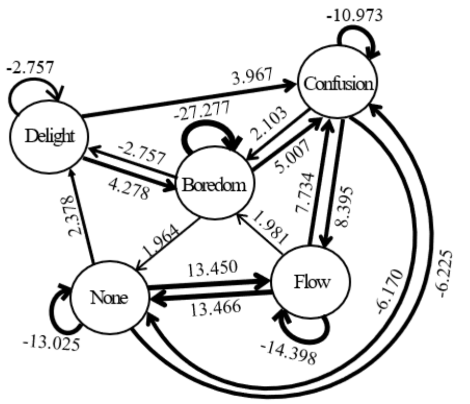

Figure 7. Affective conversion diagram of the groups

According to the residuals in Figure 7, some patterns with significant differences between the groups were observed, which include: "*none→flow*," "*flow→none*," "*confusion→flow*," "*flow→confusion*," "*boredom→confusion*," "*boredom→boredom*," "*flow→flow*," "*confused→confused*," and "*none→none*."

In summary, there are significant differences between the two groups in affective transitions, especially when negative affect transformed into positive affect, such as "*confusion/none→flow*" and "*none/boredom→delight*." There are significant differences in negative affect changes. Positive affect transformed into positive affect, such as "*flow→flow*" and "*delight→delight*." Additionally, there are positive changes into negative, such as "*flow/delight→bored/confused*." There are also negative affect changes into negative, such as "*confusion→boredom*" and "*boredom→confusion*." Finally, there are also significant differences for the transitions of "*boredom/flow/confusion/none→none*."

5. Conclusion

This paper first summarized the affect regulation methods supported by teachable agents and the affect regulation processes in ITSs. Four ITS functions that can be used to detect or help regulate affect were described. To supervise and adjust negative affect, ITSs use intelligent algorithms and technologies to analyze learning data (e.g., cognition and mood), determine learning affect, and then provide reasonable and flexible strategies (e.g., refined learning materials) or rigid strategies (e.g., simple rehearsing).

According to the accumulation analysis, students in the experimental group were prone to "*confusion*" and "*boredom*," but they spent more time in a "*flow*" state. It should be noted that while "*confusion*" is typically thought of a negative affective state, research has shown that it can be beneficial to learning when it does not lead to "*frustration*," "*boredom*," and disengagement (D'Mello, et al., 2014). With the system's help, students adjusted these negative affective states to "*delight*" and "*flow*." The use of scaffolds (such as prompts, tests, responses, and notes) often showed that students were surprised about their results. The "*flow*" state was more common in the experimental group than the control group, which suggests a higher degree of concentration in the experimental group. Therefore, the affect of the experimental group was more positive than in the control group.

Lag sequence analysis was used to analyze the different affect transitions in each group (see Table 3). The quantity of the affect transitions in the two groups is different and the frequency of standardized emotional transitions of two groups are also different (Table 4). The adjusted residual value of each affect transition in each group is standardized and the size indicates differing scales of affect transition in the groups. Significant differences were observed for each type of affect transition.

To further explore the differences in emotional transitions between the two groups, we used the MPT method to analyze the differences in affect transitions between the two groups. The results revealed significant differences in the transitions of the two groups in individual affect transition types. For example, the transitions from "*flow*" to "*flow*" (i.e., staying in a "*flow*" state) and "*confusion*" to "*flow*" (i.e., resolving some "*confusion*") in the experimental group are significantly higher than the same transitions in the control group. The transition from "*flow*" to "*confusion*" (i.e., reach an impasse) and "*none*" to "*confusion*" in the experimental group are significantly lower than the same in the control group. There are not only internal differences in each group, but also significant differences between the two groups, which we observed from our lag sequence analysis (see Tables 7 and 9). For example, in the control condition, the likelihood of students remaining in a "*bored*" state ("*bored* —>*bored*") is stronger (i.e., more significant) than in the experimental group. Comparatively, students in the experimental condition remained in a confused state less frequently than in the control condition.

We suggest two reasons for the occurrence of positive affect regulation, based on our observations of the video data and after-class interviews. First, learning with an ITS is engaging and has game-like features. With Betty's Brain, students learned about biological relationships and exercised thinking strategies to solve a task. A second reason why positive affect regulation occurred may be in part due to the virtual characters that help students find content related to the task at hand. This is consistent with the experimental conclusion of Segedy et al. (2014), who also used Betty's Brain for their study. They found that the ITS provides students with the necessary support in a timely manner, so that students can apply cognitive and metacognitive strategies to solve "cause-and-effect" problems. They also concluded that the system helps promote students' deep learning and guides them to use suitable strategies to solve problems. Some students in the experimental group began to use more optimized logic to complete tasks. This indicates that they consciously took advantage of the system's cognitive and metacognitive scaffolds to assess causality. These help students better regulate their affect and enhance their learning effectiveness. As one participant said, "I study causality very seriously. I always hope to teach students the correct knowledge. Therefore, I am confident that I can complete this task." This sentiment is consistent with the Kobylińska and Karwowska (2015) research on using automated affect regulation to influence students' negative affective experience. Students in the experimental condition appeared to be attending to the tasks at hand, given their time spent on task and mouse click frequency within the interface. While students were attending to the tasks, the ITS helped students become aware of their affect through dialogue. Students could then report their affective state to the teacher agent. Accurately grasping affect perception, evaluation, and expression requires understanding affect and affective knowledge, controlling affect and affective intelligence development, and thereby enhancing students' ability to recognize, regulate and manage affect.

This study took into account the affect transitions both within and between the two groups. However, this study has some limitations. Differences in the learning level and cognitive development of students varies by region, so the conclusions of this experiment may not be universal. It is also necessary to combine iterative experiments to avoid time shortage and contingency problems. It is possible to gain more fruitful results after multiple rounds of repeated experiments. We encoded students' affect by human observation, so the results may have some bias, so it is necessary to adopt artificial intelligent technologies to analyze the specific information automatically. A large amount of data is needed to further explore the deep relationship between affect regulation and cognition. Likewise, future studies can capture video and audio data more frequently, which would strengthen the reliability of the results. It is also necessary to collect both cognitive and metacognitive data at the same time. This allows for an exploration of the in-depth relationship between affect and metacognition by using a multi-branch tree analysis. Hwang et al. (2020) reported that changes in affect result in performance changes. Therefore, future studies of affect regulation in ITSs would benefit from tracking both affect changes and performance changes over time.

Acknowledgement

This research was partially sponsored by the National Science Foundation under the award The Learner Data Institute (award #1934745). The opinions, findings, and results are solely the authors' and do not reflect those of the funding agencies. This study was also supported by National Social Science Foundation of China for education programs (Project Approval # BCA190076). This study was also supported by the National Natural Science Foundation of China (Project Approval # 62077012).

References

Altuwairqi, K., Jarraya, S. K., Allinjawi, A., & Hammami, M. (2021). A New emotion-based affective model to detect student's engagement. *Journal of King Saud University - Computer and Information Sciences, 33*(1), 99-109.

Alexandra, J. M., Ocumpaugh, J., Baker, R. S., Slater, S., Paquette, L., Jiang, Y., Karumbaiah, S., Bosch, N., Munshi, A., Moore, A., & Biswas, G. (2019). Affect sequences and learning in Betty's Brain. In D. Azcona & R. Chung (Eds.), *Proceedings of 9th International Conference on Learning Analytics & Knowledge (LAK 19)* (pp. 383-390). New York, NY: Association for Computing Machinery. doi:10.1145/3303772.3303807

Alkhalaf, A. M. (2018). Positive and negative affect, anxiety, and academic achievement among medical students in Saudi Arabia. *International Journal of Emergency Mental Health and Human Resilience*, *20*(2), 397-401. doi:10.4172/1522-4821.1000397

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A Multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, *24*(4), 387-426.

Arsenio, W. F., & Loria, S. (2014). Coping with negative emotions: Connections with adolescents' academic performance and stress. *The Journal of Genetic Psychology*, *175*(1), 76-90.

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*, *68*(4), 223-241.

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: Betty's Brain. *International Journal of Artificial Intelligence in Education*, *26*(1), 350-364.

Burek, B. L. (2017). *Pilot study investigating the impacts of behavioural inattention and meta-attention on post-secondary students' online information seeking for academic purposes* (Unpublished master degree). University of Toronto, Canada. Retrieved from http://hdl.handle.net/1807/79045

Chu, H. C., Tsai, W. W. J., Liao, M. J., Chen, Y. M., & Chen, J. Y. (2020). Supporting e-learning with emotion regulation for students with Autism Spectrum Disorder. *Educational Technology & Society*, *23*(4), 124-146.

Collins, R. (1990). Stratification, emotional energy, and transient emotions. In T. D. Kemper (Ed.), *Research agendas in the sociology of emotions*, (pp. 27-57). Albany, NY: State University of New York Press. Retrieved from https://psycnet.apa.org/record/1990-97864-002

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, *16*(7-8), 555-575.

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 an AutoTutor. *Journal of Educational Media*, *29*(3), 241-250.

Daradoumis, T., & Arguedas, M. (2020). Cultivating students' reflective learning in metacognitive activities through an affective pedagogical agent. *Educational Technology & Society*, *23*(2), 19-31.

D'Mello, S., & Graesser, A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(4), 1-39.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145-157.

D'Mello, S., Taylor, R. S., & Graesser, A. (2007). Monitoring affective trajectories during complex learning. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society* (29) (pp. 203-208). Austin, TX: Cognitive Science Society. Retrieved from https://escholarship.org/uc/item/6p18v65q

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153-170.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Pekins, L., & Graesser, A. (2010). A Time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In V. Aleven, J. Kay & J. Mostow (Eds.), *Intelligent Tutoring Systems. ITS 2010. Lecture Notes in Computer Science* (vol. 6094, pp. 245-254). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-13388-6_29

Du, X., Zhang, M., Shelton, B. E., & Hung, J. L. (2019). Learning anytime, anywhere: A Spatiotemporal analysis for online learning. *Interactive Learning Environments,* (8), 1-15. doi:10.1080/10494820.2019.1633546

Fida, R., Paciello, M., Tramontano, C., Fontaine, R. G., Barbaranelli,C., & Farnese, M. L. (2015). An Integrative approach to understanding counterproductive work behavior: The Roles of stressors, negative emotions, and moral disengagement. *Journal of Business Ethics*, *130*(1), 131-144.

Grafsgaard, J. F., Fulton, R. M., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2012). Multimodal analysis of the implicit affective channel in computer-mediated textual communication. In L. P. Morency, D. Bohus, H. Aghajan, A. Nijholt, J. Cassell, & J. Epps (Eds.), *Proceedings of the14th ACM International Conference on Multimodal Interaction* (pp. 145-152). New York, NY: Association for Computing Machinery. doi:10.1145/2388676.2388708

Gross, J. J. (1998). The Emerging field of emotion regulation: An Integrative review. *Review of General Psychology*, *2*(3), 271-299.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348-362.

Haberman, S. J. (1979). *Analysis of qualitative data* (new developments, Vol 2, 1st ed.). Chicago, IL: Academic Press.

Han, J., Zhao, W., Jiang, Q., Dong, Y., & Zhang, N. (2019). STEM intelligent learning environment and cognitive science research: An Interview with Prof. Gautam Biswas. *Open Education Research*, *25*(2), 4-11.

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A Multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, *48*, 615-625.

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, *24*(4), 470-497.

Hwang, G. J., Sung, H. Y., Chang, S. C., & Huang, X. C. (2020). A Fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors. *Computers and Education: Artificial Intelligence*, *1*, 100003. doi:10.1016/j.caeai.2020.100003

Kobylińska, D., & Karwowska, D. (2015). How automatic activation of emotion regulation influences experiencing negative emotions. *Frontiers in Psychology*, *6*, 1-4.

Kołakowska, A., Szwoch, W., Szwoch, M. (2020). A Review of emotion recognition methods based on data acquired via smartphone sensors. *Sensors*, *20*(21), 6367-6409. Retrieved from doi:10.3390/s20216367

Lim, C. P. (2004). Engaging learners in online learning environments. *TechTrends*, *48*(4), 16-23.

Liu, T., & Lemeire, J. (2017). Efficient and effective learning of HMMs based on identification of hidden states. *Mathematical Problems in Engineering*, 1-26.

Lu, J. (2012). The Research on the psychology of affective instruction. *Journal of Psychological Science*, *35*(3), 522-529.

Maheshwari, K., Nagendhiran, N. (2017). Facial recognition enabled smart doors using Microsoft face API. *International Journal of Engineering Trends and Applications (IJETA)*, *4*(3), 1-4.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Cohen, W. W., Stylianides, G. J., & Koedinger, K. R. (2013). Cognitive anatomy of tutor learning: Lessons learned with SimStudent. *Journal of Educational Psychology*, 105(4), 1152-1163.

McDaniel, B., D'Mello, S., King, B. G., Chipman, P., Tapp,K. M., & Graesser, A.(2007). Facial features for affective state detection in learning environments. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 467-472). Austin, TX: Cognitive Science Society.

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A Review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education, 24*(2014), 427-469.

Patrut, B., & Spatariu, R. P. (2016). Implementation of artificial emotions and moods in a pedagogical agent. In *Emotions, technology, design, and learning* (pp. 63-86). Academic Press. doi:10.1016/B978-0-12-801856-9.00004-9

Pentel, A. (2015). Patterns of confusion: Using mouse logs to predict user's emotional state. In M. Kravcik, O. C. Santos, J. G. Boticario, M. Bielikova, T. Horvath (Eds.), *Proceedings of the 5th Preface International Workshop on Personalization Approaches in Learning Environments (PALE 2015)* (pp. 40-45). Dublin, Ireland: CEUR workshop proceedings. Retrieved from http://ceur-ws.org/Vol-1388/PALE2015-complete.pdf

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research & Practice in Technology Enhanced Learning*, *12*(22), 1-13.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The Circumplex model of affect: An Integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715-734.

Qin, J., Zheng, Q., & Li, H. (2014). A Study of learner-oriented negative emotion compensation in e-learning. *Journal of Educational Technology & Society*, *17*(4), 420-431.

Quinlan, J. R. (1990). Decision trees and decision-making. *IEEE Transactions on Systems, Man, and Cybernetics*, *20*(2), 339-346.

Raufelder, D., Regner, N., & Wood, M. A. (2018). Test anxiety and learned helplessness is moderated by student perceptions of teacher motivational support. *Educational Psychology*, *38*(1), 54-74.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145-172.

Schutz, P. A., Pekrun, R., & Phye, G. D. (2007). *Emotion in education* (Vol. 10). doi:10.1016/B978-0-12-372545-5.X5000-X

Segedy, J. R., Biswas, G., & Sulcer, B. (2014). A Model-based behavior analysis approach for open-ended environments. *Educational Technology & Society*, *17*(1), 272-282.

Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in game performance during gameplay with Crystal Island. *Computers in Human Behavior*, *76*, 641-655.

Thompson, R. A. (1994). Emotion regulation: A Theme in search of definition. *Monographs of the Society for Research in Child Development*, *59*(2-3), 25-52.

Thornton, M. A., & Tamir, D. I. (2017). Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences*, *114*(23), 5982-5987.

VanLehn, K., Van De Sande, B., Shelby, R., & Gershman, S. (2010). The Andes physics tutoring system: An Experiment in freedom. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 421-443). Berlin Heidelberg: Springer.

Wang, S., Wu, H., Kim, J. H., & Andersen, E. (2019). Adaptive learning material recommendation in online language education. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Proceedings of 20th International Conference on Artificial Intelligence in Education* (pp. 298-302). Cham, Switzerland: Springer.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognizing and responding to student affect. *International Journal of Learning Technology*, *4*(3-4), 129-164.

Yu, L. C., Wu, J. L., Chang, P. C., & Chu, H. S. (2013). Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge-Based Systems*, *41*, 89-97.

Zhang, T. (2008). *Research on learning emotions and emotion models under the network teaching environment* (Unpublished doctor dissertation). Capital Normal University, China. Retrieved from https://kns.cnki.net/kcms/detail/detail.aspx?FileName=2008144384.nh&DbName=CMFD2009