ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2022; 13(3); 232-243

Comparison of Methods of Affect Transition Analysis: An Example of SimInClass Dataset

Şeyma ÇAĞLAR ÖZHAN* Arif ALTUN**

Abstract

Studies of emotional-cognitive sequences are the growing body of research area in educational context. These studies focus on how emotions change during the learning-teaching process due to their dynamic nature. In affect transition studies, the change of emotion, depending on the event and time, is usually analyzed by using (a) lag sequential analysis (LSA), (b) L metric, (c) L* metric, and (d) Yule's Q metric. Yet, various methodological criticisms exist in the literature while utilizing these sequential analysis methods. In this study, it is aimed to comparatively examine lag analysis, L metric, L* metric, and Yule's Q in terms of proportion of invalid values, maximum transition metrics, minimum transition metrics, and analysis results. For this reason, the emotional states of the fifteen prospective teachers were collected and their emotions were labeled every 0.5 seconds with EEG (Electroencephalogram), GSR (Galvanic Skin Response), and Microsoft Kinect in a teacher training simulator (SimInClass). The dataset contained 17570 emotions, and the data were analyzed by utilizing lag analysis, L, L* and Yule's Q. The results showed that LSA had yielded the most proportion of invalid results. In addition, it was observed that the number of invalid values increased as the segment length became shorter in all analysis methods. When the maximum and minimum transition metric values were examined, it was found that as the sequence length increased in L and L* analyses, the value of the metrics approached 1, which is the largest value they can reach. However, it was noted that the lag analysis maximum-minimum transition metrics fluctuate independently from the sequence length. It was concluded that there were differences in the analysis results produced by the four sequential analysis methods with the same functions. It was thought that this situation might be due to the different invalid results produced by the analyses. When the results were compared with the studies in the literature, it was thought that it would be beneficial to pay attention to the nature of the data (emotional or behavioral), the data type (singe modality or multimodal modality), the amount of data (short sequences or long sequences), the environment in which the dataset was created (computer-based or not), and the sampling rate (automated data collection tool or observation) when choosing sequential analysis methods.

Keywords: affect, affect transition, sequential analysis

Introduction

Emotion is an intensely conscious mental response to a specific goal, subjectively experienced and lasting from minutes to hours, causing physiological and behavioral changes (Kleinginna & Kleinginna, 1981). The concept of emotion is sometimes referred to interchangeably with the terms "feeling," "mood," and "arousal" in the literature. The term "affect" is also used in a comparable way as emotion; it is known as a meta concept that covers emotion (Juslin & Slobada, 2013).

Emotions experienced by individuals in an educational context are the product of a cognitive evaluation structured by individuals' goals, teaching requirements, and competencies (Frenzel et al., 2009). According to appraisal theory, by updating any of the elements that construct emotions, individuals initiate the cognitive reappraisal process, and an affect transition occurs (transition from one emotion state to another) (Scherer,1993). In some studies, such as in Han et al., (2021) and Rebolledo-Mendez et al. (2022), affect transitions are examined to understand the effectiveness of intervention methods and the change of emotion in the learning-teaching process.

* Ress. Assist., Bartın University, Faculty of Sciences, Bartın-Turkey, seymacaglar9@gmail.com, ORCID ID: 0000-0002-0106-6285
** Prof. Dr., Hacettepe University, Faculty of Education, Ankara-Turkey, altunar@hacettepe.edu.tr, ORCID ID: 0000-0003-4060-6157

To cite this article:
Çağlar Özhan, Ş., & Altun, A. (2022). Comparison of methods of affect transition analysis: An example of SimInClass dataset. *Journal of Measurement and Evaluation in Education and Psychology*, *13*(3), 232-243. https://doi.org/10.21031/epod.1051716

Received: 31.12.2021
Accepted: 12.09.2022

Instead of using simple ways to examine the affect transition, researchers apply sequential and conditional methods, among which are lag sequential analysis (LSA; Bakemann & Gottman, 1997), L metric (D'Mello & Graesser, 2012), L* metric (Matayoshi & Karumbaiah, 2020), and Yule's Q (Yule, 1900). In these methods, various measures such as likelihood or probability are used regarding the frequency of transition from one state to another (Bosch & Paquette, 2021). These methods are not only used for affect but also used in the discovery of patterns in computer-assisted environments, like exploring the behavioral patterns of learners at different success levels in a mobile learning environment (i.e. Sun et al., 2021), examining discussion orientations in online discussion environments according to cognitive styles (i.e. Wu & Hou, 2015), and examining development trajectories of learners' cognitive behaviors in small private online courses (i.e. Liu et al., 2021).

In studies where affect transition is explored, there are some methodological differences in examining sequential states. While in some studies self-transitions (temporal sequential repetition of an emotional state) are counted in the analysis, in others they are not. Some studies prefer Lag analysis when examining affect transition, while others prefer the L metric, L*, or Yule's Q.

For example, Lajoie et al. (2021) examined the emotional states of individuals according to their performance in a learning process that requires self-regulation. In each step of the self-regulated learning process, data on the mood of individuals and the temporal change of emotion were collected using a camera, and emotions were defined categorically by the software. For six states—happy, surprised, angry, scared, sad, and disgusted—analyses were conducted with 1786 emotional codes for the high-performance groups and 846 emotional codes for the low-performance group. The patterns in the sequences were examined with lag sequence analysis. Sequences do not include self-transitions in states. Lajoie et al.'s (2021) results showed that low-performing individuals set higher goals, experienced more negative emotions, and experienced more emotional transitions than high-performing individuals.

Baker et al. (2010) aimed to examine the persistence and occurrence rate of six emotional states, namely boredom, delight, confusion, concentration, frustration, and surprise, in three computer-assisted learning environments. Since the persistence of the emotion is also examined, self-transitions are included in the sequences. Analyses were performed on sequences of 706 and 3640 emotional codes using the L metric for two different implementations. They found that boredom had persistence in all three environments and was associated with poorer learning, while the affect state with the least persistence was frustration.

Botelho et al. (2018) aimed to explore the dynamics of students' affect in a computer-based learning environment. The affect states of the students were determined as sensor-free by using the behaviors in the system, and the dataset with 48276 observations was evaluated with the L metric analysis. Five different states, namely concentration, boredom, confusion, frustration, and neutral, took part in the participants' sequences. The results showed that the transitions from engagement to boredom and from confusion to engagement were significant.

Karumbaiah et al. (2021) aimed to examine methodological differences in the use of L metrics in affect transition studies. They emphasized that the number of affect states was a minimum of five (See. Botelho et al., 2018) and a maximum of thirteen (See. Bosch & D'Mello, 2017). The studies were carried out in computer-based tutoring systems and game-based learning environments, and in some of them, self-transition was excluded, while in others, it was included. In order to see the results of these differences, they examined ten different datasets used in previous studies and provided a corrected method for the use of L metrics. These investigations yielded that excluding the self-transition violated the assumption of independence causing more patterns to appear while providing no information on the persistence of affect states. They concluded that it is possible to obtain information about the permanence of the states even though the emotional transitions were diluted with the inclusion of self-transitions.

Bosch and Paquette (2021) examined different transition analyses from a methodological perspective. They compared the findings of sequential analysis methods by using two simulation datasets obtained in a computer-based environment. For this purpose, it was tested using simulated datasets containing 10,000 data. The comparisons were made by performing analyses with two and four variables using the dataset obtained from the simulation. They also analyzed another dataset with three variables: self-transitions were removed entirely, and the states showed a balanced distribution. They sequentially examined 99 students' participation in practice and learning activities in a computer-based learning environment to compare the data produced by the simulation with data obtained from the natural environment. The sequence lengths of the participants ranged from 87 to 1087. In addition, analyses were applied to sequences with sequence lengths ranging from 5 to 50 extracted from these long sequences. The results showed that invalid values were higher for all sequential analysis metrics in short sequences than in the long sequences. In addition, invalid values increased as the number of variables increased in all sequential analyses. The maximum average transition value was calculated to determine whether all metrics deviated from zero. The results showed that short sequences produced artificially meaningful results in other transition metrics except for Yule's Q metric. To keep the invalid metrics under control, it has been stated that sequential analyses can be performed in sequences with at least ten observations for sequences with two variables; at least 20 for sequences with four variables; and, at least 35 observations for sequences with seven variables. Long sequences should be used to avoid false positive results.

In summary, transition metrics and related analysis methods are frequently preferred in studies exploring cognitive, behavioral, and affective patterns. However, differences in the applications of these analytical methods cause methodological concerns. In re-examining studies using L-metric analyses, Karumbaiah et al. (2021) found that excluding the states of repeating states in sequences and self-transition will result in discovering false patterns. Bosch and Paquette (2021), who examined the transition metrics and analyzed them comparatively, found that transition analysis methods produced different results according to the sequence length and number of affective states. However, Karumbaiah et al. (2021) focused only on the L metric, and Bosch and Paquette (2021) compared the analysis methods on the simulated data.

However, as Bosch and Paquette (2021) state, when examining behavior in online environments, there is no such thing as pressing a button many times in succession, so there is no repetitive data. When it comes to affect, the persistence state is high, and at the same time, the sampling rate is very high because of automatic emotion-recognition systems. These situations cause the formation of datasets with long sequences and repetitive data. In addition, due to the ever-changing and multidimensional nature of emotion, the number of states may be high in the data collected in natural environments; however, it may not be possible to distribute them in a certain balance. In this study, based on the findings and limitations in the literature, the goal is to perform and compare an analysis of transition metrics in data collected in the natural environment, where the seven affective states are included; the states are not evenly distributed, and self-transitions are included in the sequences. It is questionable, though, which sequential analysis method will work in the datasets with this feature and whether they will produce similar results.

From this point of view, the first aim of this study is to comparatively examine invalid values and maximum–minimum transition values produced by different methods, according to sequence length, by applying sequential analysis methods to the affect datasets formed by making sense of physical and physiological data in a virtual classroom simulation (SiminClass). The second aim is to compare the subsequent results by using the sequential analysis method.

As can be seen in the equations presented in the Methods section, the ratio is calculated with transition metrics. For this reason, when the denominator is equal to 0, the affected transition metrics cannot be calculated, and they produce an invalid value. That value is not calculated in all metrics for a sequence because their formulas are different. In this study, the proportion of invalid values in all values was examined among the analysis methods. Because the excess of invalid values in a sequence will reduce the probability of the occurrence of an expected value, the excess of these values reduces the statistical power (Bosch & Paquette, 2021). For this reason, it is important for the accuracy of the analysis to examine the status of invalid values according to the sequence length.

Another situation to be examined is the maximum–minimum values reached by the metrics. As mentioned in the Methods section, there are values in a certain range that each metric can tolerate. These values were also examined in order to understand how the trends of the metrics changed according to the sequence length. Each metric produced for each student's emotional transitions was evaluated within itself, and maximum and minimum values were determined. In this context, the details of participants, the dataset, data collection tools, and data analysis are presented in the Methods section.

Method

Participants

Fifteen prospective teachers who took classroom management and teaching practice courses participated in the study. Prospective teachers were healthy individuals between the ages of 20-22, using their right hand dominantly. Before the application, the participants were informed through a consent form. Ethical permission for this study was obtained from the university.

Preparation of Dataset

The dataset in this study includes the emotional states of fifteen prospective teachers who teach in a virtual classroom simulation (SimInClass) during this process. In SimInClass, the participants plan the lesson, arrange the classroom seating, perform the teaching activities, and review the teaching report. Each session lasted approximately 6 to 10 minutes for each participant in the simulation. Simultaneously with the teaching process, physical and physiological data were collected from the participants. The participants' facial expressions were collected via Microsoft Kinect as physical data. As physiological data, the brain signals of the participants were obtained by EEG (Electroencephalogram), and galvanic skin responses were received by GSR. The data received from multiple modalities were interpreted by an emotion recognition system and labeled by the system to reflect a single dominantly experienced emotional state. Null values were observed when the physical and physiological sensors could not collect data. These values are not included in the dataset. A total of 17570 emotion codes were included in the dataset, in which labels were made according to seven basic emotional states. The minimum sequence length among the participants was 516, and the maximum sequence length was 1708. Details on data collection tools and emotion recognition are presented in the sub-headings.

Data Sources

In the study, physiological data were collected with GSR and EEG, and physical data were collected with Microsoft Kinect. With GSR, temperature and electrical changes in the skin with sweat and nerves are measured. Galvanic skin response occurs due to the difference in the electrical properties of the skin, which occurs in the interaction of individuals with stimuli (phasic) and the absence of any stimulation (tonic). This measure does not give a direct idea of emotions. It is related to the level of arousal of the sympathetic nervous system. It causes an activation similar to excitement, fear, anger, and surprise.

Electrical activation of the brain can be measured with EEG. It is a non-invasive measurement method that can obtain information about cognitive, affective, and motor functions by examining the activation created by the brain regarding a stimulus on time-base and frequency-base (Bayazıt, 2018). Microsoft Kinect makes sense of facial expressions with the facial mapping library.

Since it is essential to use multiple modalities in emotion recognition (Sebe et al., 2005) GSR, EEG, and facial expression data were interpreted with an emotion recognition module. The sampling rate of this module was approximately 0.5 sec. In the module, the data produced from GSR was combined with the data from facial expression by determining the arousal dimension of emotion, and the data produced from EEG determined the positive-negative dimension of emotion. As a result, the dominant one out of seven emotional states was determined by the value created. Three different (CNNA, CNNV, CNNF) InceptionResnetV2 Convolutional Neural Network (CNN) models were used to construct the multimodal emotion recognition model.

The model was tested on public datasets such as SEED and LUMED-LUMED2 datasets created by the developers of the emotion recognition module. Certain emotion categories were found to be confused with others when the emotion recognition module identified emotions using only facial expressions. Emotions were determined with an accuracy rate of 74.2% when facial expression, EEG, and GSR were used together (see. Cimtay et al., 2020).

Data Analysis Method

The dataset was analyzed with four different sequential analysis methods and the findings were compared. The statistical significance of the sequential relationships between each emotion was examined via the L metric (D'Mello et al., 2007), L* metric (Matayoshi & Karumbaiah, 2020), lag sequential analysis-LSA (Bakeman & Gottman, 1997), and Yule's Q.

L metric is a transition likelihood metric that is used to determine affect transition and affect-behavior transition (D'Mello et al., 2007). The L metric indicates the probability and direction of the transition from the base frequency to the destination state. "Values above 0 signify that the transition is more likely than it could be expected to be given only the base frequency of the destination state, and values under 0 signify that the transition is less likely than it could be expected to be given only the base frequency of the destination state" (Baker et al., 2007). In this study, as suggested in the literature (D'Mello et al., 2007), D'Mello's L metric was calculated separately for each participant. One-sample t-test was applied to each participant's L metrics to test the transitions' significance. Benjamini and Yekutieli's (2001) post-hoc control method was applied to control false discoveries, as Matayoshi and Karumbaiah (2021) suggested. Transitions with an adjusted p-value less than .05 are considered statistically significant. The L metric formula is as follows:

$$L = \frac{P(Ynext|Xprev) - P(Ynext)}{1 - P(Ynext)}$$

L* is a transition metric modified from D'Mello's L metric. L* is a method of shifting the chance value of D'Mello's L from zero to a positive value to eliminate the undesirable effects of excluding self-transition (Matayoshi & Karumbaiah, 2020). L* is calculated using the formula for D'Mello's L metric. The only difference between the two methods is that when calculating the transition from A to B in L*, all transitions where A is the destination state are removed from the transition frequency matrix.

Lag Sequential Analysis (LSA) is an analysis method that is frequently used in the computer-based environment to examine the behavioral and affective interaction of users according to sequential and conditional probabilities (i.e Sun et al., 2017; Yang et al., 2018; Yang et al., 2018). In this study, as Bakeman and Gottman (1997) suggested, a coding scheme listing the emotions of prospective teachers in chronological order was prepared. Then, a transitional frequency matrix was created by calculating the frequency of each emotion category following another emotion category. In order to test the significance between the transitions, Z value was calculated using the transitional frequency. Z value represents the deviation of the probabilities of each transition from the expected values. It is accepted that Z values greater than +1.96 reach a significant level ($p < .05$). The Z value formula is as follows:

$$Z = \frac{f_{rc} - f_r p_c}{\sqrt{f_r p_c (1-p_c)(1-p_r)}}$$

Yule's Q is defined as a simple transformation of the odd ratio (OR) (Yule, 1900). Like the Pearson correlation coefficient, it takes a value between -1 and 1. Values between 0.3 and 0.7 indicate a moderate correlation between transitions, and values greater than 0.7 indicate a high correlation. In sequential analysis studies, it is mostly given with the Z value (i.e. Pohl et al., 2016). The Q metric formula is as follows:

$$Q = \frac{OR-1}{OR+1}$$

Results

1. Invalid Values and Maximum-Minimum Transition Values

In order to compare the results produced by the transition metrics in the real environment, the invalid values (i.e. "not a number values") and maximum-minimum transition values produced by the analyses were examined.

Figure 1 shows the proportion of invalid values for L, L*, and Z. In an affect dataset with sequence lengths ranging from 516 to 1708, it is seen that the proportion of invalid values decreases as the number of observations increases. In addition, the proportion of invalid values is higher in LSA (Z values) than in the L and L* metrics.

Figure 1

Invalid Values of Transition Metrics

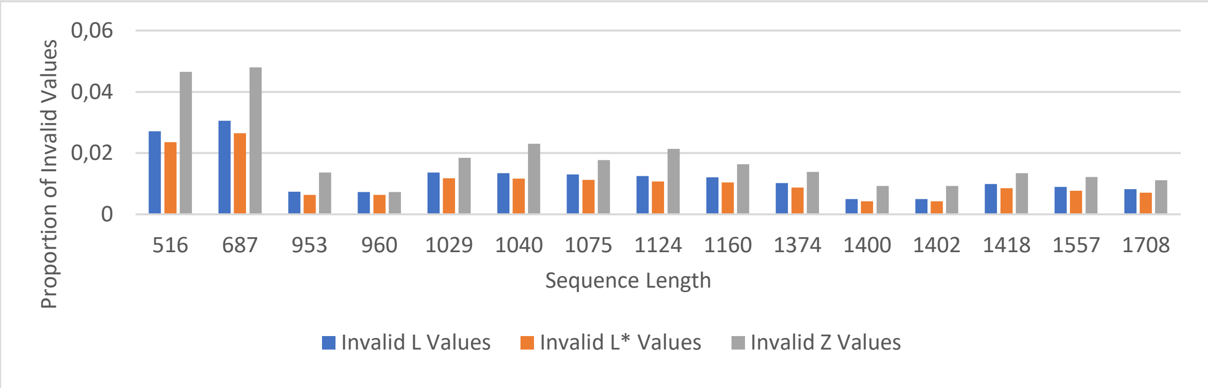
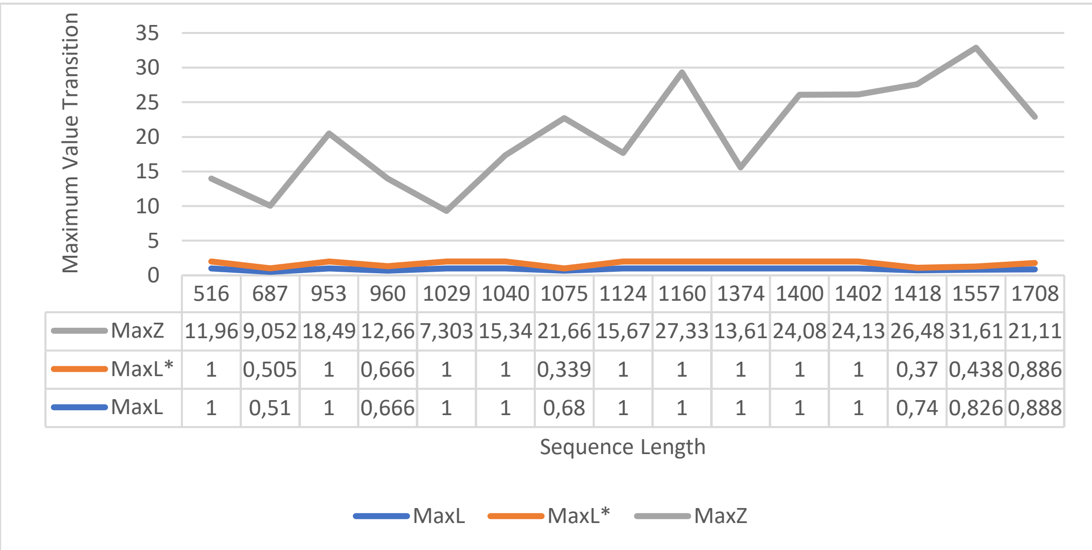


Figure 2a

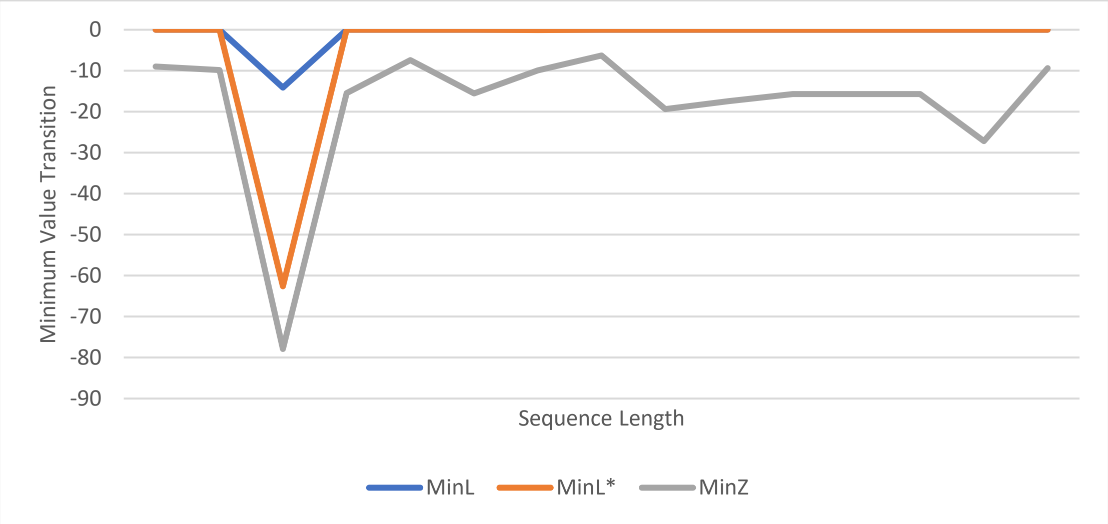
Maximum Transition Values for L, L and Z values*



In order to see how the trends of the metrics differ in an affect dataset collected in the real environment, maximum transition values (Figure 2a) and minimum transition values (Figure 2b) obtained from the analyses were examined. The results show that L and L* get closer to the maximum value of 1 as the sequence length increases. In addition, it was observed that the minimum value of L and L* approached 0 for almost all transitions (except for one transition). Since the Z value in LSA can be between –∞ and ∞, it can reach no absolute maximum value. However, it was not observed that the maximum and minimum Z values produced from LSA increased proportionally with the sequence length. It is noteworthy that the maximum and minimum values of the Z value fluctuate independently of the sequence length.

Figure 2b

Minimum Transition Values for L, L and Z values*



2. Comparison of Affect Transition Analysis Results

In order to examine whether the affect transition analyses produce similar results in the SiminClass dataset, all analyses were applied to the dataset separately, and the results were examined comparatively. As seen in Table 1, the cyclical transition for happiness was statistically significant according to the L metric (Mean L = 0.64, SE = 0.03, t=19.6, p=0.00), LSA (ptr=0.89, z=103.53, p<.05) and Yule's Q (Q=0.97). Cyclical transition of sadness was also statistically significant according to the L metric (Mean L = 0.4, SE = 0.06, t=5.63, p=0.00), LSA (ptr=0.76, z=90.33, p<.05), and Yule's Q (Q=0.96). Cyclical transition for neutral was statistically significant according to the L metric (Mean L = 0.45, SE = 0.04, t=10.22, p=0.00), LSA (ptr =0.79, z=87.05, p<.05) and Yule's Q (Q=0.93). In addition, cyclical transition of disgust was statistically significant according to the L metric (Mean L = 0.29, SE = 0.07, t=4.2, p=0.001), LSA (ptr =0.55, z=67.81, p<.05) and Yule's Q (Q=0.96).

As seen in Table 1, the cyclical transition for anger was statistically significant according to LSA (ptr =0.17, z=15.9, p<.05) and Yule's Q (Q=0.99). According to LSA and Yule's Q, cyclical transition for surprise (ptr =0.33, z=39.81, p<.05; Q=0.97) and for fear (ptr =0.36, z=44.87, p<.05; Q=0.95) were statistically significant. In addition to cyclical transitions, the transition from fear to surprise (ptr =0.03; z=3.88, p<.05; Q=0.49), from surprise to fear (ptr =0.086; z=5.55, p<.05; Q=0.64), and from disgust to anger (ptr =0.004; z=2.85, p<.05; Q=0.61) are statistically significant for LSA and Yule's Q analyses.

Since self-transition is not included in L*, these findings would not inherently be common to L* as with other analyses. According to this analysis, the transition from happiness to disgust (Mean L*= 0.22, SE = 0.07, t=3.11, p<0.05) and from happiness to fear were meaningful (Mean L*= 0.26, SE = 0.08, t=3.12, p<0.05). Also, the transition from neutral to sadness (Mean L*= 0.18, SE = 0.05, t=3.15, p<0.05) and from neutral to disgust were also statistically significant (Mean L*= 0.18, SE = 0.64, t=2.85, p<0.05).

Table 1

Comparasion of Affect Transition Analysis

		Happiness	***Sadness***	***Neutral***	***Anger***	***Disgust***	***Surprise***	***Fear***
Happiness	**D'Mello L**	**0.64*** (0.03)	0.19 (0.06)	-0.72 (0.9)	-	0.24 (0.07)	-	0.27 (0.08)
	L*	-	**0.17*** (0.06)	-3.08 (3.24)	0.13 (0.09)	**0.22*** (0.07)	-	**0.26*** (0.08)
	LSA	**0.89***	0.007	0.07	0.00	0.02	0.003	0.012
	Yule's Q	**0.97***	-0.95	-0.87	-0.7	-0.49	0.51	-0.33
Sadness	**D'Mello L**	0.02 (0.01)	**0.4*** (0.06)	-0.82 (0.91)	-	0.06 (0.02)	-	0.11 (0.06)
	L*	0.02 (0.01)	-	-0.99 (1.08)	0.05 (0.04)	0.06 (0.02)	-	0.11 (0.06)
	LSA	0.02	**0.76***	0.18	0.001	0.03	0.001	0.08
	Yule's Q	-0.94	**0.96***	-0.43	0.009	-0.3	-0.73	-0.43
Neutral	**D'Mello L**	0.11 (0.03)	0.2 (0.05)	**0.45*** **(0.04)**	0.95 (0.07)	0.18 (0.06)	-	0.16 (0.06)
	L*	-0.02 (0.08)	**0.18*** (0.05)	-	0.09 (0.07)	**0.18*** (0.64)	-	0.16 (0.06)
	LSA	0.07	0.09	**0.79***	0.001	0.02	0.01	0.011
	Yule's Q	-0.88	-.045	**0.93***	-0.009	-0.55	0.1	-0.3
Anger	**D'Mello L**	-0.002 (0.003)	0.03 (0.02)	-0.934 (0.9)	0.09 (0.07)	0.02 (0.016)	-	-0.0001 (0.0001)
	L*	-0.002 (0.003)	0.03 (0.02)	-0.99 (0.9)	-	0.02 (0.01)	-	-0.0001 (0.0001)
	LSA	0.09	0.09	0.45	**0.27***	0.09	0.00	0.00
	Yule's Q	-0.31	0.009	-0.14	**0.99***	0.1	-1	-1
Disgust	**D'Mello L**	0.17 0.01	0.02 0.008	-0.91 0.91	0.15 (0.09)	**0.29*** (0.07)	-	0.03 (0.02)
	L*	0.01 (0.01)	0.019 (0.009)	-0.97 (1.00)	0.15 (0.009)	-	-	0.03 (0.02)
	LSA	0.18	0.09	0.13	**0.004***	**0.55***	0.004	0.024
	Yule's Q	-0.51	-0.31	-0.53	**0.61***	**0.96***	-0.22	0.06
Suprise	**D'Mello L**	-0.00 (0.004)	0.007 (0.007)	-0.92 (0.91)	-	0.002 (0.002)	-	0.018 (0.009)
	L*	-0.0003 (0.004)	0.007 (0.008)	-0.92 (0.94)	-0.00 (0.00)	0.002 (0.002)	-0.00 (0.00)	0.018 (0.009)
	LSA	0.15	0.02	0.40	0.00	0.01	**0.33***	**0.09***
	Yule's Q	-0.47	-0.85	0.08	-1	-0.73	**0.97***	**0.64***
Fear	**D'Mello L**	0.006 (0.005)	0.057 (0.032)	-0.91 (0.91)	-.0003 (0.000)	.022 (0.01)	-	0.16 (0.05)
	L*	0.004 (0.005)	0.057 (0.033)	-0.94 (0.97)	-.0003 (0.000)	.022 (0.01)	-	-
	LSA	0.29	0.07	0.20	0.00	0.05	**0.03***	**0.36***
	Yule's Q	-0.31	-0.36	-0.33	-1	0.03	**0.49***	**0.95***

* Statistically significant transitions are shown in bold.

**Mean L and Mean Standard Error for D'Mello L method, Mean L* and Mean Standard Error for L* method, and Z value for LSA are given in the table.

Discussion and Conclusion

In this study, it was aimed to comparatively examine the invalid values and maximum-minimum transition values produced by different sequential analysis methods in the affect dataset, which was formed by making sense of physical and physiological data in a virtual classroom simulation. In addition, the results produced by the sequential analysis methods were examined comparatively and based on the theoretical framework. In this context, an implementation was carried out with fifteen prospective teachers in a classroom simulation (SiminClass) to create a dataset. Participants' affect states were recognized every 0.5 seconds using EEG, GSR, and facial expressions to include seven basic emotional states. The sequential analysis methods L, L*, LSA, and Yule's Q were applied to the created dataset.

After applying the sequential analysis methods, the invalid values produced by L, L*, and LSA were compared. The results showed that the proportion of invalid values of LSA is higher than L and L*. Moreover, as the sequences get shorter, invalid values increase in all analysis methods. As a matter of fact, in the study of Bosch and Paquette (2021), it was found that LSA produced more invalid values in the log dataset containing behavioral data compared to other sequential analysis methods, and short sequences produced more invalid values than long sequences. Due to invalid values, the number of transitions that show a meaningful pattern in a dataset may decrease, and the transition parameters may be found less than they should be. This issue increases the chance of finding outliers (Bosch & Paquette, 2021). This may lead to type 1 or type 2 error. Therefore, analysis can be made using L or L* methods in data collected in the real environment, where there are seven states and the distribution ratio of the states is random. In addition, it is predicted that the sequences being as long as possible will increase the statistical power of the results obtained from the sequential analysis method to be used.

When the maximum and minimum values obtained from the analyses were examined, considering the limits of L and L* $[-\infty,1]$, it was seen that the maximum value could be approached as the sequence length reaches a certain saturation in both analyses. It has been observed that the minimum value curve converges at zero for L and L*. This indicates that at least one transition between states never occurred. Since it was a dataset created in the real environment, the transitions between the states were random; therefore, this result was expected (Matayoshi & Karumbaiah, 2020). There is no maximum or minimum value that the Z value produced with LSA can take. However, the findings show that the maximum and minimum values of LSA fluctuate independently of the number of observations.

On the contrary, Bosch and Paquette (2021) found that as the number of observations increased, the maximum value of LSA increased, and the minimum value decreased. This difference may be due to the characteristics of the datasets. Bosch and Paquette (2021) performed LSA with sequence lengths ranging from 5 to 50 and log data in a computer-based environment. This finding could not be reached in the dataset with seven affect states and a sampling rate of 0.5 seconds. This shows that the nature of the data (emotional or behavioral), the data type (singe modality or multimodal modality), the amount of data (short sequences or long sequences), the environment in which the dataset was created (computer-based or not), and the sampling rate due to the characteristics of the data collection tool (automated data collection tool or observation) are essential. In sequential analyses, the analysis method to be used can be decided by paying attention to these dataset features.

When the analysis results are examined, the cyclical transition of happiness, sadness, neutral, and disgust is statistically significant according to the results of L, LSA, and Yule's Q. However, anger, surprise, and fear's cyclical transition were statistically significant only according to LSA and Yule's Q metric (cyclical transitions were not calculated since self-transitions were not examined in L*). When the results produced by the analyses are examined, the cyclical transition of happiness, sadness, neutral, and disgust is statistically significant according to the results of L, LSA, and Yule's Q. However, anger, surprise, and fear's cyclical transition were statistically significant only according to LSA and Yule's Q metric (cyclical transitions were not calculated since self-transitions were not examined in L*).

When these results were evaluated in the theoretical framework, it was an expected result that the cyclical transition of more intense and long-lasting emotions such as sadness would be found to be statistically significant. The statistical significance of the surprise cyclical transition did not match the theoretical framework. As surprise is the shortest of all emotions, it lasts for a few seconds and disappears (Ekman, 2021). For this reason, the statistical significance of the surprise cyclical transition did not match the theoretical framework.

The statistical significance of transitions between different emotions differed between analysis methods. According to LSA, the mutual pattern between fear and surprise is statistically significant, and there is a moderate relationship between these transitions according to Yule's Q. According to LSA, the pattern of transition from disgust to anger was also statistically significant, and the Yule's Q metric shows a moderate relationship between these two states. However, there was no statistically significant transition between these states according to D'Mello's L and L*. This situation among the results of the analysis may have caused these contradictory results in the dataset, as more invalid values were produced in the LSA and the rare occurrence of emotions such as anger and fear.

Since state persistence is ignored in L* sequences, the results it naturally produces are different from the others. This metric can be used in studies that focus only on transitions between states where state persistence is not essential (Karumbaiah et al., 2019).

Studies examining emotional-cognitive sequences in educational contexts are a growing body of research area. However, it has been observed that different results were obtained for the same dataset in different methods of affect transition analysis serving the same purpose. To increase statistical power in studies examining affect transition, it is important to carefully examine the structure of the dataset and the purpose of the research. In this way, the appropriate affect transition analysis method can be selected. The data type, amount of data, the environment in which the data was created, and the characteristics of the data collection tool can be the determinants of this decision process. Also, in similar studies, it would be helpful to have the sequences as long as possible to avoid potential errors. On the other hand, should a transition frequency, which is very rare in long sequences, be excluded? Should the number of observations be determined according to the number of states in such studies? It is one of the topics recommended to be examined in future studies. In addition, this study was conducted in a classroom simulation with a limited number of participants and an automatic emotion identification system with little margin of error. The research can be conducted with more participants in different contexts and data collection tools in future studies.

Declarations

Conflict of Interest: No potential conflict of interest was reported by the authors.

Ethical Approval: This study was approved by the Ethics Boards and Commissions of Hacettepe University (date 22.05.2020, document number GO20/459).

Acknowledgments: This work was supported by the Newton-Katip Çelebi fund, delivered by the Scientific and Technological Research Council of Turkey (TÜBİTAK) through the project titled 'Investigating the Effects of Computer-Based Affective Recommendation System on Teacher Trainees Cognitive-Emotional Development' (Grant No:117R036).

Author Contribution: Caglar-Ozhan S. and Altun A. contributed to the design and implementation of the research to the analysis of the results. All authors discussed the results and commented on the manuscript.

Studies included in the current reliability generalization meta-analysis are marked with an asterisk (*).

References

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511527685

Baker, R. S., 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. https://doi.org/10.1016/j.ijhcs.2009.12.003

Baker, R. S., Rodrigo, M. M. T., & Xolocotzin, U. E. (2007). The dynamics of affective transitions in simulation problem-solving environments. In A. Paiva, R. Prada & W. Picard (Eds.), *International conference on affective computing and intelligent interaction* (pp. 666-677). Springer. https://doi.org/10.1007/978-3-540-74889-2_58

Bayazıt, T. (2018). Event Related Potentials (ERP). *Journal of Medical Clinics, 1*(1), 59-65. https://dergipark.org.tr/tr/pub/atk/issue/38771/451155

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics, 29*(4),1165-1188. https://doi.org/10.1214/aos/1013699998

Bosch, N., & D'Mello, S. (2017). The affective experience of novice computer programmers. *International Journal of Artificial Intelligence in Education, 27*(1), 181-206. https://doi.org/10.1007/s40593-015-0069-5

Bosch, N., & Paquette, L. (2021). What's next? Sequence length and impossible loops in state transition measurement. *Journal of Educational Data Mining, 13*(1), 1-23. https://eric.ed.gov/?id=EJ1320638

Botelho, A. F., Baker, R., Ocumpaugh, J., & Heffernan, N. (2018). Studying affect dynamics and chronometry using sensor-free detectors. In E. Boyer & M. Yudelson (Eds.), *Proceedings of the 11th international conference on educational data mining* (pp. 157–166). EDM. https://files.eric.ed.gov/fulltext/ED593106.pdf

Cimtay, Y., Ekmekcioglu, E., & Caglar-Ozhan, S. (2020). Cross-subject multimodal emotion recognition based on hybrid fusion. *IEEE Access*, *8*, 168865-168878. https://doi.org/10.1109/ACCESS.2020.3023871

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145-157. https://doi.org/10.1016/j.learninstruc.2011.10.001

D'Mello, S., Taylor, R., & Graesser, A. (2007). Monitoring affective trajectories during complex learning. In McNamara, D., Trafton, J. (Eds.), *Proceedings of 29th annual cognitive science society* (pp. 203–208). Cognitive Science Society. https://doi.org/10.1007/978-1-4419-1428-6_849

Ekman (2021, September 7). *What is surprise?* Paul Ekman. https://www.paulekman.com/universal-emotions/what-is-surprise/

Frenzel, A. C., Goetz, T., Stephens, E. J., & Jacob, B. (2009). Antecedents and effects of teachers' emotional experiences: An integrated perspective and empirical test. In P. A. Schutz & M. Zembylas (Eds.), *Advances in teacher emotion research: The impact on teachers' lives* (pp. 129-152). Springer. https://doi.org/10.1007/978-1-4419-0564-2_7

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. https://www.jstor.org/stable/27032865

Juslin, P. N., & Sloboda, J. A. (2013). Music and emotion. In D. Deutsch (Ed.), *The psychology of music* (pp. 583-645). Academic Press. https://doi.org/10.1016/B978-0-12-381460-9.00015-8

Karumbaiah, S., Baker, R. S., & Ocumpaugh, J. (2019). The case of self-transitions in affective dynamics. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *International conference on artificial intelligence in education* (pp. 172-181). Springer. https://doi.org/10.1007/978-3-030-23204-7_15

Karumbaiah, S., Baker, R. B., Ocumpaugh, J., & Andres, A. (2021). A re-analysis and synthesis of data on affect dynamics in learning. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2021.3086118

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion, 5*(4), 345-379. https://doi.org/10.1007/BF00992553

Lajoie, S. P., Zheng, J., Li, S., Jarrell, A., & Gube, M. (2021). Examining the interplay of affect and self regulation in the context of clinical reasoning. *Learning and Instruction*, *72,* 101219. https://doi.org/10.1016/j.learninstruc.2019.101219

Liu, Z., Zhang, N., Liu, S., & Liu, S. (2021). Development trajectory of student cognitive behaviors in a SPOC forum: An integrated approach combining epistemic network analysis and lag sequential analysis. In L. Lee, L. Wang, Y. Kato & S. Sato (Eds.), *2021 International symposium on educational technology (ISET)* (pp. 26-30). IEEE. https://doi.org/10.1109/ISET52350.2021.00016.

Matayoshi, J., & Karumbaiah, S. (2020). Adjusting the L statistic when self-transitions are excluded in affect dynamics. *Journal of Educational Data Mining, 12*(4), 1-23. https://eric.ed.gov/?id=EJ1298368

Matayoshi, J., & Karumbaiah, S. (2021). Using marginal models to adjust for statistical bias in the analysis of state transitions. In M. Scheffel, N. Dowell, S. Joksimovic & G. Siemens (Eds.), *LAK21: 11th International learning analytics and knowledge conference* (pp. 449-455). Association for Computing Machinery. https://doi.org/10.1145/3448139.3448182

Rebolledo-Mendez, G., Huerta-Pacheco, N. S., Baker, R. S., & du Boulay, B. (2022). Meta-affective behaviour within an intelligent tutoring system for mathematics. *International Journal of Artificial Intelligence in Education, 32*(1), 174-195. https://doi.org/10.1007/s40593-021-00247-1

Pohl, M., Wallner, G., & Kriglstein, S. (2016). Using lag-sequential analysis for understanding interaction sequences in visualizations. *International Journal of Human-Computer Studies, 96,* 54-66. https://doi.org/10.1016/j.ijhcs.2016.07.006

Scherer, K. R. (1993). Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition and Emotion, 7*(3), 325–355. https://doi.org/10.1080/02699939308409192

Sebe, N., Cohen, I., & Huang, T. S. (2005). Multimodal emotion recognition. In C. Chen & P. Wang (Eds.), *Handbook of pattern recognition and computer vision* (pp. 387-409). World Scientific. https://doi.org/10.1142/1802

Sun, J. C. Y., Kuo, C. Y., Hou, H. T., & Lin, Y. Y. (2017). Exploring learners' sequential behavioral patterns, flow experience, and learning performance in an anti-phishing educational game. *Journal of Educational Technology & Society, 20*(1), 10-20. https://www.proquest.com/scholarly-journals/exploring-learners-sequential-behavioral-patterns/docview/2147743221/se-2

Sun, Z., Lin, C. H., Lv, K., & Song, J. (2021). Knowledge-construction behaviors in a mobile learning environment: A lag-sequential analysis of group differences. *Educational Technology Research and Development, 69*(2), 533-551. https://doi.org/10.1007/s11423-021-09938-x

Wu, S. Y., & Hou, H. T. (2015). How cognitive styles affect the learning behaviors of online problem-solving based discussion activity: A lag sequential analysis. *Journal of Educational Computing Research*, *52*(2), 277-298. https://eric.ed.gov/?id=EJ1076314

Yang, X., Song, S., Zhao, X., & Yu, S. (2018). Understanding user behavioral patterns in open knowledge communities. *Interactive Learning Environments, 26*(2), 245-255. https://doi.org/10.1080/10494820.2017.1303518

Yang, X., Li, J., & Xing, B. (2018). Behavioral patterns of knowledge construction in online cooperative translation activities. *The Internet and Higher Education, 36*, 13-21. https://doi.org/10.1016/j.iheduc.2017.08.003

Yule, G. U. (1900). On the association of attributes in statistics: With illustrations from the material of the childhood society. *Philosophical Transactions of the Royal Society of London. 66*(194), 252–261. https://doi.org/10.1098/rspl.1899.0067