Implicit and Explicit Emotions in MOOCs

Munira Syed
University of Notre Dame
msyed2@nd.edu

Malolan Chetlur
IBM Research, India
mchetlur@in.ibm.com

Shazia Afzal
IBM Research, India
shaafzal@in.ibm.com

G. Alex Ambrose
University of Notre Dame
gambrose@nd.edu

Nitesh V. Chawla
University of Notre Dame
nchawla@nd.edu

ABSTRACT

Understanding the affect expressed by learners is essential for enriching the learning experience in Massive Open Online Courses (MOOCs). However, online learning environments, especially MOOCs, pose several challenges in understanding the different types of affect experienced by a learner. In this paper, we define two categories of emotions, *explicit* emotions as those collected directly from the student through self-reported surveys, and *implicit* emotions as those inferred unobtrusively during the learning process. We also introduce positivity as a measure to study the valence reported by students chronologically, and use it to derive insights into their emotion patterns and their association with learning outcomes. We show that implicit and explicit emotions expressed by students within the context of a MOOC are independent of each other, however, they correlate better with students' behavior compared to their valence.

Keywords

MOOC, emotions, discussion forum, valence, surveys

1. INTRODUCTION

The exploration of emotions expressed by students in Massive Open Online Courses (MOOCs) has caught the attention of researchers for improving the remote and non-contact learning experience [28, 8, 16, 5]. A few examples of these studies infer emotions of students from their behavior [16], surveys collected during the course [8, 1], clickstream data and discussion forums [28, 5]. The relationship between students' emotions and their behavior, learning outcomes, engagement, and dropout within the MOOC context is established in [1, 25, 21].

Emotions experienced by students during a course impact their behavior and learning outcomes [15, 19]. Detecting the emotion experienced during learning is difficult, and various methods have been employed for this purpose. The methods used to sample emotions mainly fall into three categories as outlined by [27]. The first category consists of methods that take snapshots of students' emotions during the course through survey questionnaires. These methods are intrusive to the learning process and are usually self-reported and subjective in nature. The second category detects emotions during the learning process and includes methods that sample emotions non-intrusively like facial expression detection, conversations, gaze detection, and analysis of text data generated by student interactions within the course [9, 10]. The third category measures emotions after the learning process. The first two categories are relevant to our paper. In [27], the methods in the second category are assumed to counteract the limitations of the methods in the first category. Therefore, in our study we use two categories of emotions to get a more complete view of students' emotional states. In this paper, we measure explicit emotions as the emotions recorded from student's self-reported surveys and Self- Assessment Manikins (SAMs), and implicit emotions as those from the open discussion forum posts of students.

Emotions measured in association with learning seem to be short-lived and last for a few seconds to minutes [15]. Since the emotions were expressed by students in this MOOC at different, non-uniform points in time, one of the challenges of analyzing such a series is the spontaneity of emotions. As the emotions are surveyed after the end of a video or module, we only get a snapshot of the students' emotions during the course [27]. Between two consecutive surveys, a student's emotions can not only change multiple times, but also be conflicting, as students can experience multiple emotions simultaneously [1], which could hinder a chronological analysis of the emotions. However, even if students' emotions are spontaneous and likely to be fraught with missing data, there might be a trend to their emotions over time. An approach that leverages this idea has been proposed in [7], where the positive affect experienced by an individual is averaged over a period of time while the negative reports are ignored. Inspired by this technique, we also calculate the "positivity" of students at each point of the reported emotions and derive a positivity sequence instead of an emotion sequence. This positivity sequence is expected to be more stable over time as compared to the emotion sequence.

We study the implicit and explicit emotions expressed by the MOOC students through the following research questions.

RQ1: Are the explicit and implicit emotions expressed within a MOOC context similar? Can one be used as a proxy for

Munira Syed, Malolan Chetlur, Shazia Afzal, G. Alex Ambrose and Nitesh V. Chawla "Implicit and Explicit Emotions in MOOCs" In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 432 - 437

the other or are both of them equally important for characterizing a student's emotional state?
RQ2: What do the combined (explicit plus implicit) emotional states and positivity sequences characterize about a student's learning?

To the best of our knowledge, this is the first attempt at investigating the effect of explicit and implicit emotion categories within a MOOC context. We find that implicit and explicit emotions expressed by students are indeed different and both are necessary to characterize student emotions. We also see that combined positivity values correlate relatively well with behavior compared to their valence values.

2. RELATED WORK

The comparison of self-reported metrics like emotions and performance in self-regulated learning and other educational contexts has been studied and generally found to be inconsistent with the measured reports [14, 26, 29]. While many of these studies measure the alignment of students' achievement calibration with their actual performance [13, 26, 14], we aim to compare the self-reported emotions of students in MOOCs against the emotions we measure from their behavior in the MOOC, in the form of interactions on the discussion forum. A direct comparison of these methods with ours is infeasible because of the difference in instrumentation and methodology. However, we will compare our general observations with the trends in literature.

We use students' self-reports of emotions along with Self-Assessment Manikin (SAM) as the explicit measures of students' emotions. Self-reports are a very common way of measuring students' emotions because of their subjective nature [11]. Collecting students' emotions through surveys is easy to deploy on a large-scale and is low cost [11], which makes them favourable for use in MOOCs [1]. SAM is a non-verbal assessment technique that allows people to rate their pleasure, represented as valence in our case, on an ordinal scale [4]. SAMs have been used to measure emotion in online learning environments [6, 8].

Among the techniques available for detecting the implicitly expressed emotions of students, analyzing emotions from texts is one of the least invasive ways of detecting students' emotions [17, 22]. Using discussion forums to detect students' emotions in MOOCs is becoming prominent due to its unobtrusiveness and low instrumentation [28]. Many sentiment analysis techniques for detecting valence from text including the word-affect lexicon used in this paper are listed in [18], and education has been noted as one of the applications of sentiment analysis. We use Warriner's [24] word-affect lexicon to calculate the valence values of words in the discussion forum records. The effectiveness of Warriner's word-affect lexicon [24] for sentiment analysis has been demonstrated for detecting sarcasm [20], finding geographical locations associated with happier tweets [12], etc. This automatic method to detect affect from discussion forum data enables a scalable way to glean implicit affect in MOOCs from a large number of forum posts. Sentiment analysis polarity techniques were applied on discussion forum posts in [25]. In [28], a Mechanical Turk is used to obtain confusion ratings among students through simple features like counting the number of question marks to predict the level of confusion in the discussion forum posts. They also use Linguistic Inquiry and Word Count (LIWC) to consider negation words and phrases as an indicator of potential confusion, and clickstream patterns (eg. quiz-quiz-forum) as a feature for detecting confusion. Previous research on using the discussion forum to estimate student retention and performance is complicated due to a vast amount of missing and imbalanced data [3]. We also face challenges to detect implicit emotions in the midst of context-specific terms.

Table 1: 1. Number of students vs. SAM surveys 2. Number of students vs. SAM scores

SAM survey	No. of students
1	4111
2	2815
3	1354
4	906
5	326

SAM score	No. of students
1	3204
2	4355
3	1557
4	295
5	101

3. DATA DESCRIPTION

3.1 Course Description

We use the data from the introductory course on Statistics called "I Heart Stats" for our study. This was a self-paced MOOC on the EdX platform, and the entire course content was released at the start of the course. The course had nine modules, with the ninth module being for the assessment of the overall course. During the course, students were asked to self-report their emotions and valence through emotion surveys and SAM surveys respectively. Initially 24,279 students were enrolled in the course, however, only less than 15,000 students had activity in the first two weeks. Finally, only 1,941 students completed it. Of all the students, 1,629 responded to at least one emotion or SAM survey, and participated in the discussion forum as well. Only these students have been included in the Analysis section of the paper as these are the only students generating both implicit as well as explicit emotions. Note that students completing the course are likely to have longer sequence lengths. Students not interacting with the discussion forum but are still part of the course cannot be included in the analysis leading to an overrepresentation of active users.

3.2 Explicit Emotions

Emotion Surveys: Of all the students, 6,100 submitted 21,448 emotion surveys. During the course, 12 emotion surveys were conducted in which students self-reported their current emotional state. This was optional and students could choose multiple of a list of 15 emotions: *anger, anxiety, boredom, confusion, contentment, disappointment, enjoyment, frustration, hope, hopelessness, isolation, pride, relief, sadness, and shame.* Further details can be found in [1]. The valence values of these emotions were calculated using Warriner's lexicon [24], with a scale of 1 to 9 and 5 being neutral. We shift the scale to [-4, 4] to bring the neutral valence to 0. In the case of multiple emotions being expressed, the associated valence values were averaged to obtain one valence value per survey. Thus, the surveys have positive (0, 4], negative [-4, 0), and neutral {0} valence values.

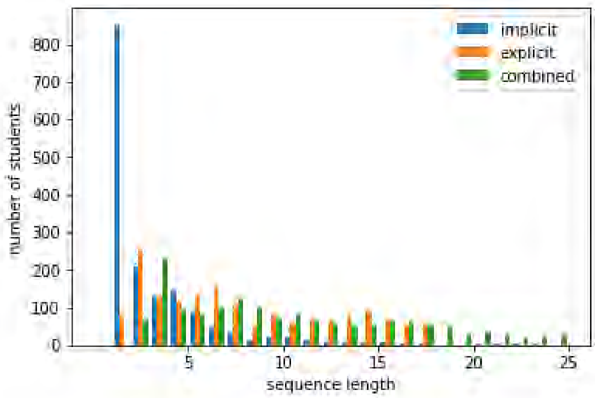

Figure 1: Histogram of implicit, explicit, and combined sequence lengths (until sequence length of 25)

SAM Surveys: A total of 5 SAM surveys, using a 5-point scale, were conducted in this MOOC. The SAM score represented in Table 1 ranges from 1 to 5 with 1 being the least and 5 being the highest state of pleasure. As the distribution of the number of students corresponding to each SAM score is normal, we convert this scale to an interval scale in the range [-4, 4] linearly. In total, 5,363 students have submitted 9,512 SAM surveys with the rest of the details shows in Table 1.

3.3 Implicit Emotions

The discussion forum is a platform that students use to interact with each other, the instructor, and teaching assistant of the MOOC. In total, 1,717 students generated 5,322 discussion forum records. The posts, comments, and replies (i.e. records) on the discussion forum are used to infer the implicit emotions of students.

We use Warriner's word-affect lexicon [24] to calculate the valence values of discussion form records. The tokenized words in tweets are used to calculate the mean valence value of the tweet using Warriner's word-affect lexicon. We use a similar approach to calculate valence values for discussion forum records using the following steps: (i) Tokenize the records to get a list of words, (ii) Remove the stop words from the list, (iii) Make a list v of valence values associated with a word using the lexicon, if present, after re-scaling them between [-4, 4], (iv) Multiply the valence values of words/phrases that follow a negative word with -1 (eg. not, never), and (iv) Return the average valence value of list v.

3.4 Combined Emotions

Throughout the course, students have multiple opportunities, explicit or implicit, to express their emotions. The 12 emotion surveys, 5 SAM surveys, and valence values calculated from discussion forum records were interleaved and ordered chronologically for each student to form a combined sequence of valence values.

A histogram of the number of reports corresponding to the number of students in Figure 3.4 shows that the highest number of students (14%) has a maximum combined sequence length of 3 with the number of students tapering down after that point. The maximum number of reports corresponding to a student is 74, as this student was very active in the discussion forum.

To mitigate the spontaneous nature of emotions, we calculate the positivity of students at each report from the valence sequence values. Thus, if a student reports one negative emotion among a string of positive emotions, the impact of the negative emotion is reduced because of the previously expressed positive emotions. We define positivity as follows.

Positivity: Let $r_1, r_2, ..., r_n$ be the reports made by a student until element n such that: $timestamp(r_{i-1}) < timestamp(r_i)$ for all i. The valences are normalized between [-1, 1], instead of [-4, 4], by dividing them by 4. Let $p_1, p_2, ..., p_m$ be the positive normalized valences where $m <= n$ and $m + 1 > n$. The positivity at the nth element is given by $(p_1 + p_2 + ... + p_m)/n$.

In other words, an element of the positivity sequence is calculated by averaging over only the positive valences in the sequence until that element. Since students have reported more positive than negative valences both explicitly and implicitly, calculating negativity instead of positivity would lead to extremely sparse sequences.

4. ANALYSIS

4.1 Calculated Valences

Section 3.3 lists the steps to calculate the valence values of the discussion forum records. To validate these valence values, 440 samples of the discussion forum records were manually annotated by three human raters in which each rater chooses one, two, or none of the 15 emotion choices that students had for their emotion surveys. The fourth rater is the calculated valence. We use Fleiss' Kappa [2] to calculate the inter-rater agreement by converting the valence scores to positive, negative, or zero valence. The inter-rater agreement of the three human raters is 0.457 (moderate agreement), whereas the inter-rater agreement of the four raters including the calculated valences is 0.218 (fair agreement) [23]. While the agreement including the calculated valences is lower, it is adequate, and so we use the calculated valence of these discussion forum records as the implicit valence values.

4.2 Implicit vs. Explicit features (RQ1)

Both implicit and explicit sequences are instances of irregular time-series data. However, since emotion data is spontaneous and might change multiple times between consecutive reports [15], averaging, downsampling, interpolating or duplicating valence values in an emotion sequence might misrepresent the true emotional trajectory of the student.

4.2.1 Feature vectors description

Since the valence sequences are not uniform in length, we create fixed length feature vectors for analysis. The features are used in Sections 4.2.2 and 4.2.3 with their description given: (i) *pos*: ratio of the number of positive valences to the total length of the sequence (ii) *neg*: ratio of the number of negative valences to the total length of the sequence (iii) *neu*: ratio of the number of neutral valences to the total length of the sequence (iv) *trans*: ratio of the number

Table 2: Corr. between implicit & explicit features

Features	Pearson's r	Spearman's rho
pos	0.0401	0.0696**
neg	0.0413*	0.102***
neu	0.0150	0.0380
seq_len	0.346***	0.422***
trans	0.125***	0.162***
neg_pos	0.113***	0.165***
pos_neg	0.0805**	0.127***
range	0.243***	0.257***

*:p-val.<0.1, **:p-val.<0.05, ***:p-val.<0.0001

of transition of valences from positive to negative or vice versa in the sequence to the sequence length (v) *pos_neg*: ratio of the number of transition of valences from positive to negative to the sequence length (vi) *neg_pos*: ratio of the number of transition of valences from negative to positive to the sequence length (vii) *range*: calculated by subtracting the minimum valence value from the maximum valence value expressed (To normalize the value the resulting range is divided by 8, as the valence values lie in the range [-4, 4].) (viii) *seq_len*: length of the valence sequence (integral value).

4.2.2 Correlation

In Table 2, we see that *pos*, *neg*, and *neu*, as defined in Section 4.2.1, between implicit and explicit emotions of students are not correlated with each other. This shows that both types of sequences are somewhat independent of each other and might show different insights into students' affect. There are relatively few neutral discussion forum records which is why its correlation with completion is not significant. That is why transitions from neutral to positive and negative valences, and vice-versa have been left out of the features list. The sequence lengths seem to be mildly correlated showing that students reporting more emotions in the emotion surveys were also more likely to submit more records in the discussion forum. This correlation is expected since the number of students with larger sequence lengths decreases as seen from Figure 3.4.

4.2.3 Clustering of Feature Vectors

We cluster the 7-dimensional feature vector to identify groups of similar students using K-Means. To visualize the clusters created, we decompose the 7-dimensional feature vectors of students' implicit and explicit emotion sequences to a 2-dimensional space using Principal Component Analysis (PCA) separately. The PCA decomposition in Figure 4.2.3 shows very separable clusters in the 2-dimensional space. The explicit clusters have significantly different ratios of course completion: orange: 37.2%, purple: 25.5%, olive: 51.9%. Similarly, the completion ratios of the implicit clusters are: red: 34.5%, blue 32.6%:, green: 60.3%, with the green cluster having significantly more students completing the course than the other two.

4.3 Combined sequence features (RQ2)

From the previous subsection, we saw that implicit and explicit sequences are not identical and should both be incorporated into a student's valence trajectory. So we use both

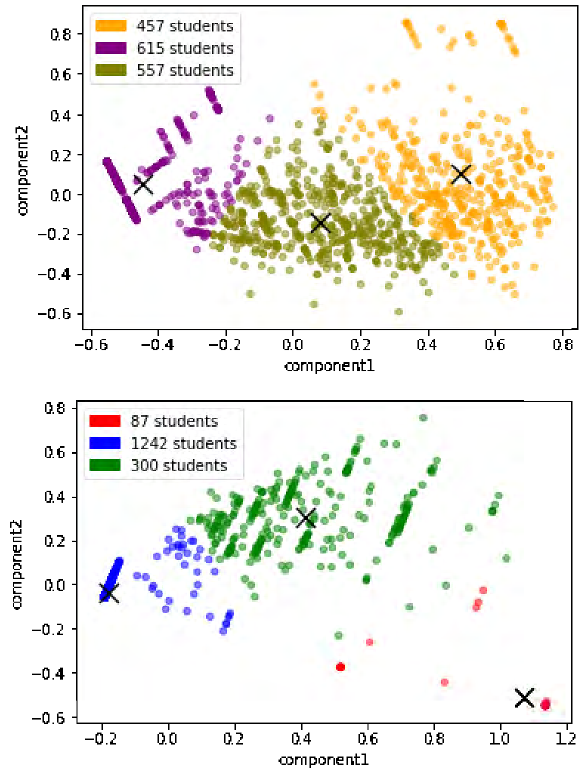

Figure 2: PCA decomposition of explicit (top) and implicit (bottom) seq. clusters ('x': cluster centers)

implicit and explicit sources of emotions ordered by time to generate a combined valence sequence for students. The features from Section 4.2.1 are used in the analysis below.

4.3.1 Correlation of features with completion

We generate the 7-dimensional feature vector from the combined valence sequence for each as defined in Section 4.2.1 and show the correlation of each dimension with completion in Table 3. Completion is defined by a student reaching module 8 [1]. We see that *seq_len* has the highest correlation with completion possibly because sequence length could act as proxy for the amount of time students spent in the course. A similar reasoning might hold for *trans*. The *pos*, *neg*, or *neu* features do not seem to be correlated with completion. However, *neg_pos* seems to be better correlated with

Table 3: Corr. of combined vectors with completion

Feature	Pearson's r	Spearman's rho
pos	-0.0549***	-0.110***
neg	0.0807***	0.156***
neu	-0.0382**	0.0828***
neg_pos	0.215***	0.300***
pos_neg	0.112***	0.201***
trans	0.186***	0.223***
seq_len	0.523***	0.460***
range	0.390***	0.392***

**: p-val. <0.05, ***: p-val. <0.0001

Table 4: Corr. of features with quiz performance

Features	average	minimum	maximum
range	-0.0804*	-0.181***	0.0735*
seq_len	-0.232***	0.0681*	-0.405***

*: p-val.<0.1, **: p-val.<0.05, ***: p-val.<0.0001

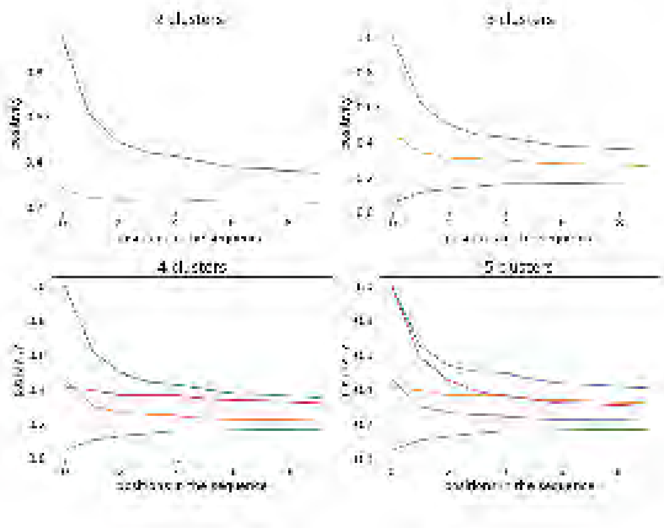

Figure 3: Positivity Clustering of Combined Seqs.

completion than *pos_neg*. This supports our intuition that students transitioning from a negative to positive emotional state are more likely to stay in the course, compared to the other way round. The feature *range* is better correlated with completion than *trans* which indicates that higher intensity of changes in emotions is more likely to result in completion.

4.3.2 Correlation of features with Quiz Performance

The performance score of students for a quiz is normalized between 0 and 1. The average, minimum, and maximum performance score of the quizzes (total 4) that students have attempted is used as the y-variable for correlation. The features that are significantly correlated with these statistics using Pearson's correlation are in Table 4. While the negative correlation with *seq_len* is unsurprising given that harder quizzes are towards the end of the course, the positive correlation with *range* suggests that student who experience extreme emotions tend to perform better.

4.4 Positivity clustering (RQ2)

We compare fixed length positivity sequences by clustering the first 10 elements of 767 students who have a sequence length of at least 10. We see that k=3 is the highest number that shows no overlap of cluster centers. While there is no significant difference between the clusters for quiz performance, the difference between clusters in terms of quiz participation using ANOVA is significant at p-value < 0.05. Specifically, in the k=3 chart in Figure 4.4, there are more students in the most positive (green) cluster that do not submit a single quiz (29.3%) than the other two clusters (20%). A possible explanation is that students had trouble with the quizzes and the ones who did not attempt them were more likely to be happier. All three cluster centers converge towards a narrow range of positivity, suggesting that students tend towards the same positivity in the course even though they started out differently.

5. DISCUSSION AND FUTURE WORK

Similar to the studies [14, 26, 29], we found that the self-reported emotions did not reflect the implicitly measured emotions. Clustering students by their emotion sequence had different ratios of students that completed the course in each cluster. This observation is similar to what [14] found about different learning strategies and activity of students. To investigate whether the temporally proximal self-report was correlated with the outcome completion, we measured the correlation of the last reported valence and the final positivity in the students' sequences with completion. However, similar to [29], we found no correlation. This suggests that the proximity of students' emotions to the outcome completion does not have a bearing on completion.

Through RQ1, we show that both the implicit and explicit emotion sequences are independent of each other and contribute different emotional information. Through RQ2, we showed that students tend to converge towards the same positivity even though they start out differently, indicating that they end up feeling the same way. This might be because of external factors that remained constant for all the students, e.g., how the course was conducted, possibly explaining the lack of correlation with the course outcomes. We see significant differences between these clusters in quiz participation but not in other learning outcomes. This may be because students who did not attempt the quizzes did not struggle through the course and remained relatively happy. Our results show that there is potential for identifying different groups of students that participate in a MOOC. Table 2 shows that the explicit and implicit sequences are associated with behavior, but not valence. One of the possible reasons is that students who participate more in the discussion forum tend to submit more surveys as well but the two types of sequences do not corroborate each other in valence. From Table 3, we also observe that students who feel negatively about the course and then transition to a positive emotional state are more likely to stay in the course. We found that the range of valence that students experience is more indicative of their course completion and quiz performance possibly because the students who struggle through the course report higher valence values after achieving their course objectives, resulting in their highly varied emotions.

A limitation of our work is our sentiment analysis technique that uses a bag-of-words model with the discussion forum records only and does not consider other implicit measures of emotions. In this work, we have only relied on a single word-affect lexicon. However, we can make the calculated valence values more stable by triangulating the valences with other lexicons. We would also like to improve granularity and quantify the extra information conveyed by either type of emotion sequence. Even so, as most emotion research in MOOC relies on only one category of emotions, we conclude that it might be advantageous for researchers in this area to supplement their current method with a method from the other category of emotions. It is important to continue exploring emotions in MOOCs in pursuit of goals such as the personalization of MOOCs, improving the emotional well-being of students, and the design of MOOCs.

6. ACKNOWLEDGEMENTS

We thank John Dillon for his contributions to this project. This work was supported in part by the National Science Foundation (NSF) Grant IIS-1447795.

7. REFERENCES

[1] S. Afzal, B. Sengupta, M. Syed, et al. The ABC of MOOCs: Affect and its inter-play with behavior and cognition. In *ACII*, pages 279–284. IEEE, 2017.

[2] M. Banerjee, M. Capozzoli, L. McSweeney, et al. Beyond kappa: A review of interrater agreement measures. *Canadian journal of statistics*, 27(1):3–23, 1999.

[3] Y. Bergner, D. Kerr, and D. E. Pritchard. Methodological challenges in the analysis of MOOC data for exploring the relationship between discussion forum views and learning outcomes. *International Educational Data Mining Society*, 2015.

[4] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[5] J. C. Cheng. An exploratory study of emotional affordance of a massive open online course. *European Journal of Open, Distance and e-learning*, 17(1):43–55, 2014.

[6] K.-H. Cheng, H.-T. Hou, and S.-Y. Wu. Exploring students' emotional responses and participation in an online peer assessment activity: A case study. *Interactive Learning Environments*, 22(3):271–287, 2014.

[7] E. Diener, R. J. Larsen, S. Levine, et al. Intensity and frequency: dimensions underlying positive and negative affect. *Journal of personality and social psychology*, 48(5):1253, 1985.

[8] J. Dillon, N. Bosch, M. Chetlur, et al. Student emotion, co-occurrence, and dropout in a MOOC context. In *EDM*, pages 353–357, 2016.

[9] S. D'Mello, A. Olney, C. Williams, et al. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies*, 70(5):377–398, 2012.

[10] S. K. D'mello, S. D. Craig, A. Witherspoon, et al. Automatic detection of learner's affect from conversational cues. *User modeling and user-adapted interaction*, 18(1-2):45–80, 2008.

[11] J. A. Fredricks and W. McColskey. The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of research on student engagement*, pages 763–782. Springer, 2012.

[12] L. Gallegos, K. Lerman, A. Huang, et al. Geography of emotion: Where in a city are people happier? In *WWW*, pages 569–574. ACM, 2016.

[13] L. S. Garavalia and M. E. Gredler. An exploratory study of academic goal setting, achievement calibration and self-regulated learning. *Journal of instructional psychology*, 29(4):221–231, 2002.

[14] D. Gaševic, J. Jovanovic, A. Pardo, et al. Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, 4(2):113–128, 2017.

[15] A. Graesser and S. K. D'Mello. Theoretical perspectives on affect and deep learning. In *New perspectives on affect and learning technologies*, pages 11–21. Springer, 2011.

[16] D. Leony, P. J. M. Merino, J. A. R. Valiente, et al. Detection and evaluation of emotions in massive open online courses. *J. UCS*, 21(5):638–655, 2015.

[17] Z. Liu, W. Zhang, J. Sun, et al. Emotion and associated topic detection for course comments in a mooc platform. In *2016 International Conference on Educational Innovation through Technology (EITT)*, pages 15–19. IEEE, 2016.

[18] S. M. Mohammad. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier, 2016.

[19] R. Pekrun, T. Goetz, W. Titz, et al. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, 37(2):91–105, 2002.

[20] A. Rajadesingan, R. Zafarani, and H. Liu. Sarcasm detection on twitter: A behavioral modeling approach. In *WSDM*, pages 97–106. ACM, 2015.

[21] A. Ramesh, D. Goldwasser, B. Huang, et al. Understanding mooc discussion forums using seeded LDA. In *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 28–33, 2014.

[22] P. Rodriguez, A. Ortigosa, and R. M. Carro. Extracting emotions from texts in e-learning environments. In *2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*, pages 887–892. IEEE, 2012.

[23] A. J. Viera and J. M. Garrett. Understanding interobserver agreement: the kappa statistic. *Fam med*, 37(5):360–363, 2005.

[24] A. B. Warriner, V. Kuperman, and M. Brysbaert. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207, 2013.

[25] M. Wen, D. Yang, and C. Rose. Sentiment analysis in MOOC discussion forums: What does it tell us? In *EDM*. Citeseer, 2014.

[26] P. H. Winne and D. Jamieson-Noel. Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27(4):551–572, 2002.

[27] M. Wosnitza and S. Volet. Origin, direction and impact of emotions in social online learning. *Learning and instruction*, 15(5):449–464, 2005.

[28] D. Yang, M. Wen, I. Howley, et al. Exploring the effect of confusion in discussion forums of massive open online courses. In *L@S*, pages 121–130. ACM, 2015.

[29] M. Zhou and P. H. Winne. Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, 22(6):413–419, 2012.