Examining the Role of Peer Acknowledgements on Social Annotations: Unraveling the Psychological Underpinnings

Xiaoshan Huang
xiaoshan.huang@mail.mcgill.ca
Department of Educational & Counselling Psychology
McGill University
Montreal, Canada

Haolun Wu
haolun.wu@mail.mcgill.ca
School of Computer Science
McGill University
Montreal, Canada

Xue Liu
xueliu@cs.mcgill.ca
School of Computer Science
McGill University
Montreal, Canada

Susanne Lajoie
susanne.lajoie@mcgill.ca
Department of Educational & Counselling Psychology
McGill University
Montreal, Canada

ABSTRACT

This study explores the impact of peer acknowledgement on learner engagement and implicit psychological attributes in written annotations on an online social reading platform. Participants included 91 undergraduates from a large North American University. Using log file data, we analyzed the relationship between learners' received peer acknowledgement and their subsequent annotation behaviours using cross-lag regression. Higher peer acknowledgements correlate with increased initiation of annotations and responses to peer annotations. By applying text mining techniques and calculating Shapley values to analyze 1,969 social annotation entries, we identified prominent psychological themes within three dimensions (i.e., affect, cognition, and motivation) that foster peer acknowledgment in digital social annotation. These themes include positive affect, openness to learning and discussion, and expression of motivation. The findings assist educators in improving online learning communities and provide guidance to technology developers in designing effective prompts, drawing from both implicit psychological cues and explicit learning behaviours.

CCS CONCEPTS

• Do Not Use This Code → Generate the Correct Terms for Your Paper; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

KEYWORDS

Digital social annotation, peer acknowledgment, Shapley value, text mining, learner behavior, psychological themes

1 INTRODUCTION

Digital social annotation, a platform where readers comment within text margins while reading, has evolved into a dynamic arena for computer-supported collaborative learning (CSCL) [24], facilitating peer interactions [1] and co-construction of knowledge [31]. However, the quality of online discourse varies, affecting learners' interactive experience and learning outcomes [16]. Peer acknowledgement, often conveyed through positive emoticons or supportive words from fellow users, plays a pivotal role in shaping user behaviour, both in social media [17, 25] and online learning environments [37]. However, despite extensive systematic reviews on social annotation (e.g., [18, 58]), limited attention has been devoted to understanding how peer acknowledgement influences learners' subsequent behaviours within social annotation platforms. Moreover, there remains a gap in our understanding of the key factors that establish evaluative criteria for the quality of written content and how this written content influences peers' reactions. Consequently, it is imperative to delve into the examination of linguistic features within written annotations, as they provide valuable insights into the psychological aspects of CSCL. An in-depth exploration of the interplay between these dimensions can inform the development of effective instructional strategies, fostering student engagement with course materials and within the online learning community [20]. This study aims to address the existing gap by exploring the impact of peer acknowledgement on learners' subsequent annotation behaviours and comparing the relative importance of linguistic indicators within four psychological dimensions (i.e., affect, cognition, motivation, and social) in written annotations related to peer acknowledgement. Given that the digital social annotation platform provides rich trace data without interrupting learners' natural learning processes, it is recommended

to apply learning analytics to examine their "actual" learning behaviours [22] within this context. The following sections will delve into the theoretical foundations, draw from previous empirical studies, and detail the human-centred analytical methods employed in this study.

2 BACKGROUND

2.1 Peer Acknowledgement in Digital Social Annotation

Digital social annotation is a productive forum for online learning [27, 31]. It employs a computer-supported communication approach to facilitate interactions among peers within a shared community or individuals with similar interests in specific topics. More precisely, users can highlight valuable information by adding annotations to selected text and respond to others' annotations through written text or by using emoticons to express reactions. This functionality facilitates the co-construction of knowledge [31], nurtures a sense of collaborative learning community [21], and transforms individual learning styles to social ones by promoting flexible asynchronous conversations [4]. In the digital era, social annotation has emerged as a prominent pedagogical approach in higher education, enhancing online interaction without impeding collaborative learning among students. However, the quantitative and qualitative aspects of online discourse can vary, in terms of engagement in annotation activities and the content contributed, respectively.

Furthermore, the quality of interaction, especially as indicated by the responses and reactions of peers, significantly influences online users' behaviours. Individuals may experience feelings of isolation if they perceive a lack of connection with others in virtual spaces [19]. In contrast, the perceived richness of online discussion forums has a significant positive effect on student participation, interaction, and learning [3]. For example, a previous study of a social media platform known for knowledge sharing, identified peer acknowledgement as a significant indicator of the recipient's comment behaviour, positively predicting more frequent and longer comment posting [9]. A recent systematic review of empirical studies in social annotation has found that the majority of previous research has focused on assessing learners' performance by quantifying annotation metrics [18]. However, despite these findings, limited research has explored how peer acknowledgement, in the form of receiving "upvotes" emoticons, impacts learners' subsequent behaviour in digital social annotation contexts. This gap in the literature underscores the need to investigate the influence of peer acknowledgement on learners' behaviours and the linguistic factors that affect peer reactions in such environments.

2.2 Revealing Psychological Signs through Linguistic Features in Computer-Supported Collaborative Learning

Social annotations can significantly benefit students' learning when used effectively [6]. There is compelling evidence supporting the positive impact of annotations on students' memory and learning (e.g., [45, 52]). Moreover, students who exhibit higher levels of critical thinking and connect their personal reflections to the course materials tend to achieve superior outcomes compared to peers who simply replicate provided content or engage in casual conversations to demonstrate social skills [54]. However, some learners may not be skilled enough to identify the key concept of a given content [21], which may lead to reflecting at a surface level. On another occasion, learners with lower working memory capacity may shift their focus towards more superficial conversation when tackling multiple sources [39]. In an autonomous context such as digital social annotation, interactive actions among learners may reflect the quality of a written text; there is an increasing demand to employ computer processing of discourse to highlight crucial exchanges for the benefit of teachers and other stakeholders [46, 48]. Thus, cultivating the written text of annotation will give a fuller picture of what makes a high-quality annotation, as observed and acknowledged by peers.

The investigation of linguistic features becomes even more critical when considering the four dimensions of psychological factors: affect, cognition, motivation, and social. According to the community of inquiry framework [15], learners' online learning experience is influenced by their perceived emotional feelings, interactions and cohesiveness within the community, which are highly correlated to their cognitive procedures [20]. Studies by Pekrun and colleagues [34] emphasize the role of emotions in learning, highlighting how emotions can significantly impact students' comprehension and retention of information. Additionally, the cognitive dimension encompasses various aspects of memory, cognitive processes, and information processing [2], all of which are deeply intertwined with the way students engage with and make use of annotations and comments in their learning process. Furthermore, Deci and Ryan [43] have extensively explored the motivational aspect, emphasizing the importance of intrinsic motivation in driving effective learning. Finally, the socio-cultural theory [53] underscores the influence of social interactions on cognitive development and learning. It emphasizes the significance of collaborative and social learning environments.

Therefore, the investigation into linguistic features, within the social annotation context, allows us to gain a comprehensive understanding of how these psychological dimensions influence students' learning experiences. This understanding can inform instructional approaches that promote effective learning strategies and foster a deeper engagement with course materials and within the learning community.

2.3 Aim of This Study

In this study, we aim to answer the following two research questions (RQs):

RQ1. *How does peer acknowledgement influence learners' social annotation behaviours?*

RQ2. *What are the predominant linguistic features that affect peer acknowledgement?*

We investigate the two research questions through a series of learning analytics. We hypothesize that peer acknowledgement, indicated by more upvotes received at time t-1, leads to more initiated annotations and responses posted at time t, exceeding the average of peers' contributions. Additionally, we aim to examine the impact of learners' psychological factors on peer acknowledgement by employing text-mining and Shapley value techniques. These

methods help us identify the most significant indicators for peer acknowledgement by comparing emotional, cognitive, motivational, and social perspectives. We will discuss the identified influential predictive factors from these four dimensions in detail.

3 METHODOLOGY

3.1 Participants

This study involved 91 undergraduate students (53 females and 38 males) who were enrolled in the course "Integrating Educational Technology in Classrooms" at a large North American university in Fall 2021. The average age of the participants was 20.67 (SD = 1.25). The course was provided by the Faculty of Education and was open to undergraduate students from various academic disciplines, such as Medicine, Arts, Management, and Engineering. Therefore, the participants represent students from diverse academic backgrounds. Over the semester of twelve consecutive weeks, the students were assigned to complete collaborative reading activities using an asynchronous platform called *Perusall*[1]. The study obtained ethical approval from the university's Research Ethics Board Office.

3.2 Learning Environment, Procedure, and Data Sources

We use *Perusall* as the major learning context in this study. *Perusall* is an interactive and collaborative annotation platform that empowers educators to design collaborative reading sessions utilizing digital textbooks, documents, and online sources. It redefines solitary reading assignments into collaborative social and educational exercises [1]. During the semester, students were directed to annotate class readings weekly through the acts of sharing comments or responses. These annotations created by students were made visible to their classmates (see Figure 1). Students could initiate an annotation by leaving a comment under the annotated content. Additionally, students could respond to annotations generated by their peers. Furthermore, while this participation was not mandatory and would not contribute to the course credit, students could express their support to peers' annotation using the "upvote" feature.

The assigned reading materials on *Perusall* encompassed 29 chapters from books and research articles, including topics related to theories and practical applications of educational technologies. These assigned readings covered a broad spectrum, including comprehending challenges in educational technology, fostering virtual learning through online resources, and engaging in collaborative activities with social technologies.

The data sources analyzed in this study were exclusively retrieved from *Perusall*'s log files. To achieve the research purpose of this study, we specially examined the extracted interactive behaviors of learners using three data sources: (i) learners' interactive behaviors (i.e., number of initiated annotations and responses posted), (ii) peer acknowledgement indicated by received peer upvotes, and (iii) their written content on each annotation.

3.3 Analytic Methods

We leveraged a range of data mining techniques to address our research questions. We used Cross-lag regression [23] to delve into the role of peer acknowledgement in steering learners' future social annotation activities in a sequential time context. Furthermore, we adopted text mining strategies to pull linguistic features from annotations, encapsulating affect, cognition, motivation, and social facets. To pinpoint the crucial features influencing peer acknowledgment in annotations, we applied the Shapley value technique [41]. Subsequent sections elaborate on these methods.

Cross-lag regression. To study the lagged influence of peer acknowledgement on learner annotation behaviors, we employed the Cross-lag regression analysis [23]. This approach is particularly useful in examining the lagged effects of one variable on another over distinct time intervals, and is widely used in various domains including education and psychotherapy [32, 38]. In our context, it helps unravel the complex relationship between peer acknowledgment and the frequency of students initiating annotations and responding to their peers' notes. The Cross-lag regression can be expressed as:

$$Y_t = \beta_0 + \beta_t \cdot X_{t-1} + \epsilon_t, \tag{1}$$

where Y_t is the annotation behavior at time t; X_{t-1} is the peer acknowledgment at the previous time step; β_t denotes the cross-lagged effect; and ϵ_t represents the residual error at time t. The intercept, β_0, signifies the initial state of annotation behavior unaffected by peer acknowledgment. Specifically, time t and t-1 are defined based on the order of 29 assigned readings distributed over 12 weeks. Each week involves the assignment of one to three readings for annotation activities, depending on the overall workload for that specific week. Given the asynchronous nature of social annotation activities, where students often complete annotations before receiving feedback, understanding this lagged relationship becomes vital. Utilizing this method, we intend to uncover the subtle influences of lagged peer acknowledgment on ensuing reading behaviors in social annotation settings.

Text Mining. To address our second research question, we applied text mining techniques to extract linguistic features from student comments using the Linguistic Inquiry and Word Count (LIWC) tool[2] developed by Boyd et al. [5]. This tool applies natural language processing (NLP) to quantify words and phrases corresponding to specific psychological dimensions [50], allowing us to explore the psychological attributes reflected in the students' annotations. It categorizes words into different linguistic and psychological dimensions covering emotions, cognitive processes, social aspects, and linguistic style. Utilizing LIWC for linguistic feature extraction involves the following five steps: text input, tokenization, lookup in the LIWC dictionary, category counting, and analysis [35].

LIWC has facilitated research in various fields, aiding in the analysis of emotional expression and cognitive processes in learning [19, 26, 57]. Using LIWC in our study reveals the nuanced aspects of student interactions in social annotations, enhancing our understanding of affective, cognitive, motivational, and social dynamics.

---

[1]https://www.perusall.com/
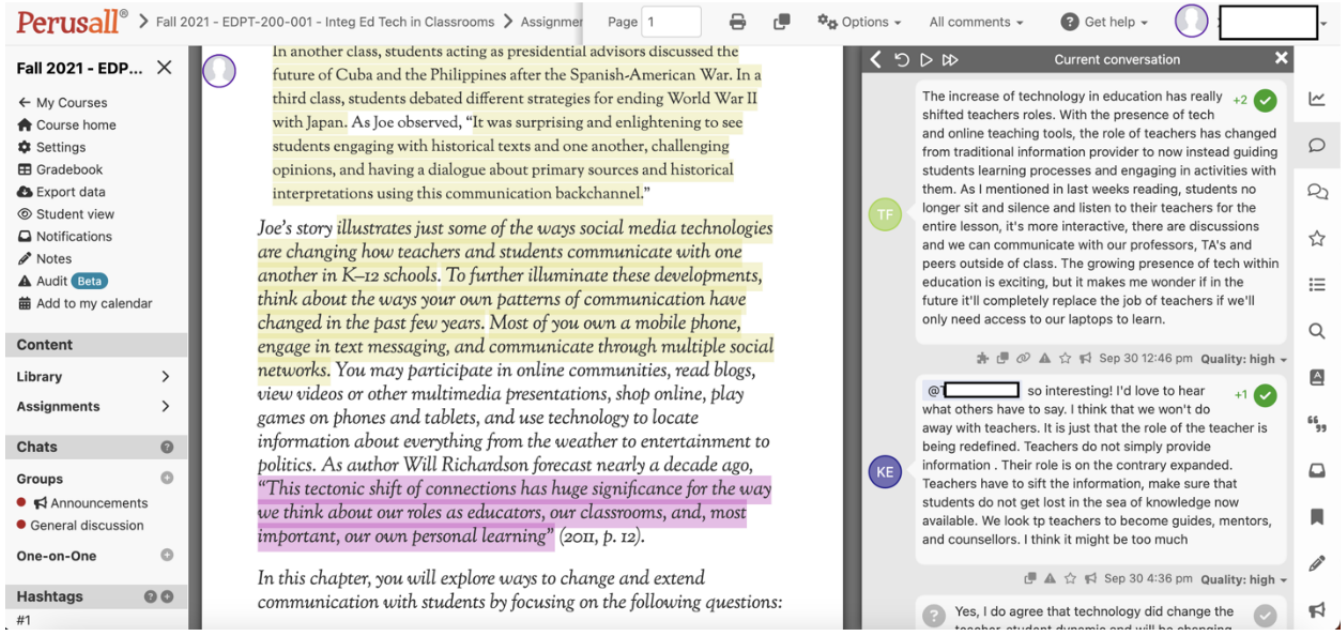
[2]https://www.liwc.app/

Figure 1: The Interface of the Social Annotation Platform Perusall.

Shapley Value. To investigate and filter those linguistic features that have the highest effect on peer acknowledgment, one effective approach is utilizing the Shapley value technique [41] — a concept borrowed from cooperative game theory [14]. The Shapley value assists in distributing the total effect fairly over various contributing factors such as the quality of teaching materials, instructor's teaching style, course difficulty, among others that are present in the diverse features influencing the feedback.

Formally, the Shapley value Φ_i of a player i (in our context, a *feature*) in a cooperative game is defined as:

$$\Phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \Big(M(S \cup \{i\}) - M(S) \Big), \quad (2)$$

where the descriptions of notations are as follows:

- N is the set of all players.
- S is a subset of players not including player i.
- $M(S)$ is the characteristic function/model that outputs a value to each input coalition S.
- $|S|$ and $|N|$ denote the cardinality of set S and N respectively.

By evaluating all possible subsets of the linguistic features and analyzing the marginal contribution of each feature within these subsets, we obtain a detailed understanding of the individual impact of each feature on peer acknowledgement, as indicated by the score of upvotes received. The Shapley values derived from this analysis not only quantify the importance of each feature but also facilitate a hierarchical comprehension of their respective influences, ultimately guiding focus towards the most critical aspects to enhance peer acknowledgement.

4 RESULT

4.1 RQ1. How does peer acknowledgement influence learners' social annotation behaviors?

To investigate our first research question, we carried out two cross-lag regression analyses focusing on the interplay between peer acknowledgment and the distinct dependent variables, namely, the number of annotations initiated and responses given, respectively. Before these analyses, we performed preliminary data cleaning. The anticipated 2,639 data entries — calculated from 91 participants and 29 reading materials — was reduced to 2,205 due to incomplete inputs from 47 participants. Following z-score normalization [11] and outlier exclusion, we finalized 1,989 data entries deemed relatively independent given the diverse reading materials and one-semester time frame.

Our analysis accounts for potential individual-specific effects by implementing a linear mixed-effects approach. To address individual variability within the model, we specifically incorporated a random slope for the grouping variable (i.e., ID), which allows for individual variability in the effect of the predictor variable [7]. In our study, the random slope linear mixed model captures individual differences in how the number of peer acknowledgements affect their annotation behaviours. We show the details in Appendix A.

The result evidenced a significant relationship between received peer acknowledgment at time t-1 and initiated annotation behaviors at time t, with a substantial coefficient of 0.187 (p<.001). This confirms our hypothesis that increased peer acknowledgment positively correlates with a surge in subsequent annotation activities.
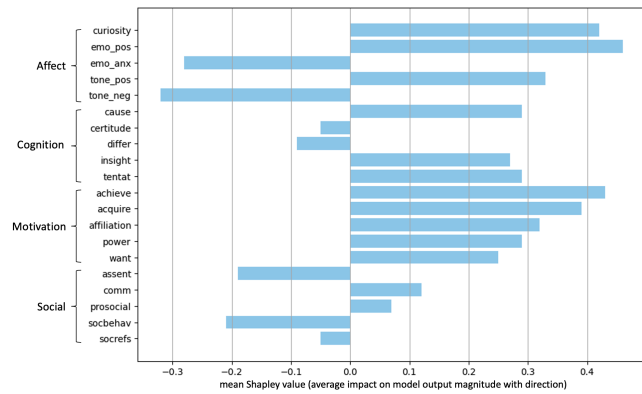
Figure 2: Summary of the Shapley values across all 20 linguistic features, categorized into four dimensions. Each value signifies the average influence of its respective feature on the model's output, illustrating not only the magnitude but also the direction of the impact.

This result underscores the predictive power of lagged peer acknowledgment in determining future annotation behaviors, sustaining a prominent role in fostering engagement in social reading activities. Similarly, we explored the impact of lagged peer acknowledgment on the frequency of response in subsequent social annotation activities, recording a significant coefficient of 0.052 (p=.009). This aligns with our preliminary hypothesis, asserting a positive predictive relationship between peer acknowledgment at time t-1 and the learners' responsive engagement at time t. Furthermore, the reciprocal relationship between each annotation behavior (i.e., numbers of initiated annotations and responses given) at t-1 and peer acknowledgement at t is non-significant, indicating that the previous annotation behaviors by students do not necessarily influence peer acknowledgment in the next annotation activities.

4.2 RQ2. What are the predominant linguistic features that affect peer acknowledgement?

In answering the second question, we conducted a text-mining analysis of learners' written comments. Specifically, we selected twenty linguistic features from the LIWC dictionary covering four dimensions that match best with our learning context, including affect, cognition, motivation, and social (see Table 1). The mean value represents the average linguistic features associated with each selected indicator across all annotation entries, as determined by the LIWC analysis. A higher mean score indicates a greater percentage of words in the annotations expressing the selected indicator. Each dimension contains features (indicators) that can be categorized in that field. For example, written words that reflect a degree of bravado in the sense of certainty such as "*really*", "*actually*", and "*of course*" are categorized into the cognition dimension, while words that reflect motives in higher achievement (e.g., better) are categorized into the motivation dimension [5].

The next step was to identify the most important linguistic features that can predict obtained peer acknowledgment. Initially, we assessed the compatibility of five machine learning models —
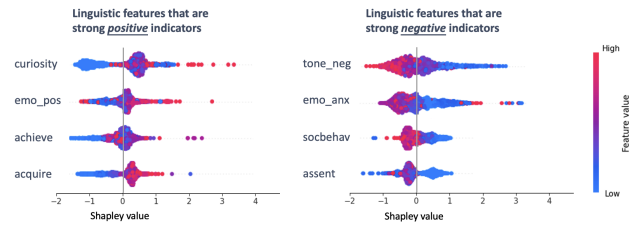
Figure 3: Fine-grained visualization of Shapley values with respect to those selected linguistic features across all possible feature combinations. The left hand side shows the four features that are strong positive indicators, while the right hand side shows the four features that are strong negative indicators. Points of red denote higher feature values, while points of blue represent lower values.

Random Forest [8], Lasso Regression (Lasso) [51], K-Nearest Neighbors (KNN) [47], and ElasticNet [59], Gradient Boosting Regressor (GBR) [13] — with our dataset, using R-squared (R^2) and Mean Squared Error (MSE) as the metrics for evaluating the predictive efficacy of these regression models. The details of model training and overfitting prevention are described in Appendix B. As shown in Table 2, GBR emerged as the best model, exhibiting the highest R^2 and the lowest MSE values on both cross-validation and testing data. Subsequently, we employed the Shapley value technique through the SHAP library[3] with the GBR model (M), as defined in Eq. 2. According to the output Shapley values in Figure 2, most of the selected indicators in the affect (e.g., "*curiosity*", "*emo_pos*"), cognition (e.g., "*insight*", "*tentat*"), and motivation (e.g., "*achievement*", "*acquire*") dimensions exhibit significant positive predictive effects on peer acknowledgement.

In contrast, among the twenty linguistic indicators, only four of them exhibit moderate to strong negative predictive value on peer acknowledgement. The negative indicators include "*tone_neg*" and "*emo_anx*" under the affect dimension, "*socbehav*" and "*assent*" under the social dimension. Two cognitive indicators, "*differ*" (i.e., expressions of discrepancies) and "*certitude*" (i.e., boasting of certainty), and one social indicator "*socref*" (i.e., social reference), also show a negative effect on peer acknowledgement, but these effects are relatively small compared to others.

Figure 3 shows a more fine-grained view of the Shapley values across all feature combinations, specifically focusing on the top four positive indicators and the top four negative indicators. As illustrated, expression of positive affect and motivation are strong predictors of peer acknowledgement, while social indicators and negative emotions have a contradictory impact in prediction. The following section will discuss the implications in detail.

5 DISCUSSION

5.1 Peer Acknowledgement Facilitates Learners' Social Annotation Behaviours

The primary focus of our first research question is to investigate the impact of peer acknowledgement on learners' annotation behaviour.

---
[3]https://shap.readthedocs.io/en/latest/

Table 1: Summary and text-mining analysis of twenty linguistic features (indicators) from the LIWC dictionary.

Dimension	Indicator	Description	Sample words	Mean score	Std. deviation
Affect	*curiosity*	Expression of curiosity	wonder, look for, research	0.97	1.02
	emo_pos	Expression of positive emotion	happy, hope, good	0.62	0.76
	emo_anx	Expression of anxiety	worry, nervous, afraid	0.07	0.20
	tone_pos	Positive tone	support, well, new	3.31	1.77
	tone_neg	Negative tone	bad, wrong, too much	0.82	0.90
Cognition	*cause*	Causation	make, because, why	2.94	1.54
	certitude	Boasting of certainty	really, actually, of course	1.05	0.96
	differ	Differentiation	but, not, if	4.01	1.71
	insight	Insight	know, think, make	5.30	2.21
	tentat	Tentative	something, or, any	2.59	1.52
Motivation	*achieve*	Achievement	work, better, best	1.90	1.38
	acquire	Acquire	get, take, getting	0.59	0.67
	affiliation	Affiliation	our, us, help	1.78	1.55
	power	Words with power ideas	order, allow, own	0.85	0.83
	want	Want	hope, want, wish	0.19	0.35
Social	*assent*	Assent	yeah, okay, yes	0.55	0.71
	comm	Acts of communication	tell, thank, said	1.51	1.30
	prosocial	Prosocial behavior	care, help, please	0.60	0.76
	socbehav	Social behavior	say, said, love	4.46	1.99
	socrefs	Social referents	you, we, he/she	6.67	2.86

Table 2: Evaluation and comparison of different machine learning models. The R-squared (R^2) is the higher the better, while the Mean Squared Error (MSE) is the lower the better. The best model is bolded.

Model	cross-validation		testing	
	R-squared (R^2)	Mean Squared Error (MSE)	R-squared (R^2)	Mean Squared Error (MSE)
Random Forest	0.8794	0.2531	0.8682	0.2621
Lasso	0.8217	0.3132	0.8092	0.3207
KNN	0.7911	0.3512	0.7794	0.3603
ElasticNet	0.8123	0.3305	0.8004	0.3410
GBR	**0.9231**	**0.1490**	**0.9123**	**0.1561**

Previous research has consistently shown that active participation in social annotation activities correlates with improved learning outcomes and a greater benefit from insights contributed by peers (e.g., [21, 28, 49]). Furthermore, peer support plays a pivotal role in motivating individuals to engage in autonomous learning environments, where instructors typically take a less active role in guiding discussions [12, 36].

Given the asynchronous nature of online interaction in social annotation, our investigation focuses on the lagged effect of peer acknowledgement rather than the immediate correlation between received peer acknowledgement and social annotation contributions within the same time frame. We specifically explored how peer acknowledgement received in previous activities influences individuals' subsequent annotation behaviours, including initiating new posts and responding to others' contributions. Our findings

provide strong support for the positive impact of peer support on learner engagement in social annotation. In particular, learners who received more peer acknowledgement in previous social annotation activities were more likely to contribute by initiating posts and responding to others. This increased engagement enriches the learning experience and interactions within the community.

Our findings align with the components of social presence outlined in the Community of Inquiry framework [15, 16], suggesting that a sense of genuine connection with others enhances engagement in related activities [42]. Additionally, our results are consistent with positive psychology [29, 30] and teamwork [44] theories, which propose that individuals who feel welcomed in a community, such as a social annotation group, tend to have a higher sense of self-value in their involvement [40]. This, in turn, fosters mutual

trust with peers from the same community, motivating individuals to invest greater effort in contributing to the community.

5.2 Influence of Psychological Features on Peer Acknowledgement in Written Text

Furthermore, we have identified specific psychological themes within the content of written annotations that hold particular significance in eliciting peer acknowledgment. These findings shed light on the key psychological themes that significantly influence peer acknowledgment within digital social annotation environments. Notably, several dimensions of psychological factors emerged as strong predictors of peer acknowledgment. The practical implications for students, teachers, and other practitioners related to each identified theme are also discussed below.

Positive Affect. Our findings reveal that expressions of positive affect, such as curiosity, positive emotion, and a positive tone, exhibit a strong positive impact on peer acknowledgment. Learners who convey positive emotions and maintain an optimistic tone in their annotations are more likely to receive peer acknowledgement. This aligns with previous research highlighting the role of positivity in fostering social interactions and engagement [33, 56]. Conversely, expressions of negative affect, including a negative tone or anxiety, could hinder peer acknowledgment. The finding responds to the call for theoretical buy-in regarding the perceived analytic value by examining the affective components of students' online discussion [55]. Recognizing the impact of positive affect on peer acknowledgment, educators can foster a positive online learning environment. Incorporating elements that cultivate a sense of optimism and camaraderie within the digital community can contribute to a more supportive and collaborative learning space.

Openness for Learning and Discussion. Among the cognitive indicators, annotations characterized by elements such as causation, tentative language, and insight tend to receive more peer acknowledgment. For instance, one sample annotation scored high in *"causation"* in the cognitive process stating, *"I would love to see more of this type of technology integrated into the classroom as it helps students learn, creates new friendships, and teaches them to cooperate"*. By using the logistic conjunction and elaboration on the statement, the annotation is more likely to foster peer acknowledgment. However, expressions of discrepancies and certitude, which reflect a degree of bravado, exhibit a slightly adverse predictive direction. For example, an annotation scored high in *"certitude"* stated, *"100% I definitely agree"*, shows a high degree of certainty but it also presents less effort in the cognitive process and does not lead to further discussion. This suggests that boasting certainty or showing discrepancies in annotations may deter peer acknowledgment, while openness to discussion and providing evidence for critical thinking tend to attract more recognition. The theme offers insights for students; understanding that annotations emphasizing critical thinking and open discussion tend to attract more recognition provides them with a strategic approach. It also provides instructors with guidance on setting evaluation criteria. Encouraging cognitive engagement and discouraging excessive certainty or discrepancies in social annotation can contribute to more meaningful and acknowledged interactions among students.

Expression of Motivation. Another influential indicator is the expression of motivation, particularly in terms of learners' desire to achieve, acquire knowledge, seek affiliation, and exercise power. Annotations reflecting high motivation levels are positively associated with peer acknowledgment. One sample annotation received high peer acknowledgement and also expressed high motivation in the learning context stating: *"Three-dimensional dissection is really interesting. I took two anatomy courses last year where we missed the experience of dissecting bodies in real life. Hopefully, I get to experience both dissections in Virtual reality and in real life"*. These words *"hopefully"* and *"get to"* demonstrate a sense of motivation in gaining more experience. This suggests that learners who demonstrate a strong drive and enthusiasm for learning are more likely to attract recognition from their peers. In essence, motivation plays a pivotal role in shaping the dynamics of peer acknowledgment within digital social annotation environments. This insight can motivate students to articulate their enthusiasm, fostering a more engaged and active community within the online learning platform.

Social Aspects: Matter Less. Additionally, the social theme identified from linguistic features does not contribute much to predicting peer acknowledgement, which is aligned with previous study [54]. For example, words such as *"okay"* and *"yeah"*, typically associated with informal content like social conversations and casual dialogues, do not significantly contribute to prompting peer acknowledgment. This implies that learners' engagement in more formal and substantive discussions is more likely to elicit peer acknowledgement in our learning setting.

To encapsulate, we identified three psychological themes that have relative importance in fostering peer acknowledgement, converting dimensions of affect, motivation, and cognition. Specifically, learners who convey positive affect, exhibit high motivation and engage in cognitive processes that encourage critical thinking tend to garner more peer acknowledgment. Conversely, expressions of negative affect and engagement in informal social behaviour are associated with lower peer acknowledgement.

The findings provide insights for learners, teachers, and technology developers by emphasizing the significance of affect, motivation, and cognitive engagement in shaping peer interactions. Teachers can foster the development of a supportive learning community by emphasizing these dimensions through the establishment of evaluation criteria. This encourages students to contribute to a positive and encouraging atmosphere. The identified psychological themes also provide practitioners and technology developers with novel insights for designing effective feedback mechanisms by embedding a learning analytic approach. Building on previous studies that found significant results indicating the impact of learning analytics on learners' online discussions compared to instructor supervision [10], technology developers can design feedback systems that acknowledge and reinforce positive contributions, thereby motivating students to actively participate and contribute meaningfully to discussions. For example, automatic detection algorithms can be implemented to filter students' annotations with the desired themes and provide supportive prompts, such as awards and hints, to those lacking in addressing these themes, along with encouraging reactions.

5.3 Limitation and Future Direction

Despite our findings, our study has several limitations. First, the topics selected for social annotation relied on social sciences within our learning context. Therefore, the generalizability of our findings to other subject-specific contexts may be limited. Future studies could explore participants' learning trajectories in other domains such as engineering or mathematics, which may involve different content designs and workloads for assigned social annotation tasks.

Though valuable for analyzing written language, LIWC has some limitation compared to advanced NLP approaches. For example, its pre-defined dictionary may not cover emerging terms, and it may struggle to capture complex sentences. Still, considering its wide application in psychology [19], sociology [57], and computer science [26] to analyze and understand language content and the efficiency of the analytic process, it is appropriate to use this tool in our social annotation context. To delve deeper into this context, a more profound comprehension of the psychological factors that facilitate peer interaction could be attained by integrating qualitative analysis into future study design. For instance, future research endeavours could encompass human coding of the extracted written texts and the analysis of individuals' reflective diaries or interviews, delving into their experiences with social annotation activities.

6 CONCLUSION

Our research contributes to a deeper understanding of the role of peer acknowledgement in promoting active participation and collaboration among learners in computer-supported collaborative learning environments. Drawing upon established theoretical frameworks and empirical evidence, we shed light on the underlying mechanisms that drive increased engagement and contribution within these learning communities. Our study underscores the significance of peer acknowledgement and its positive impact on individuals' learning behaviors, while also highlighting the intricate interplay of psychological factors that shape peer acknowledgment dynamics within digital social annotation environments. The findings provide empirical evidence for practitioners in education and training using social annotation to promote peer acknowledgement through encouraging positive affect, expression of motivation, and openness for learning and discussion. Furthermore, educational technology developers and programmers can utilize learners' linguistic features to design computer-human interactions that provide adaptive hints to encourage effective interactions in digital social annotations.

REFERENCES

[1] Brittany Adams and Nance S Wilson. 2020. Building community in asynchronous online higher education courses through collaborative annotation. *Journal of Educational Technology Systems* 49, 2 (2020), 250–261.

[2] O Roger Anderson. 1983. A neuromathematical model of human information processing and its application to science content acquisition. *Journal of Research in Science Teaching* 20, 7 (1983), 603–620.

[3] MS Balaji and Diganta Chakrabarti. 2010. Student interactions in online discussion forum: Empirical research from 'media richness theory' perspective. *Journal of interactive online learning* 9, 1 (2010).

[4] C Blyth. 2014. Exploring the affordances of digital social reading for L2 literacy: The case of eComma. *Digital literacies in foreign and second language education* 12 (2014), 201–226.

[5] Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin* (2022), 1–47.

[6] John Bradley and Paul Vetch. 2007. Supporting annotation as a scholarly tool—Experiences from the online Chopin Variorum edition. *Literary and Linguistic Computing* 22, 2 (2007), 225–241.

[7] Markus Brauer and John J Curtin. 2018. Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods* 23, 3 (2018), 389.

[8] Leo Breiman. 2001. Random forests. *Machine learning* 45 (2001), 5–32.

[9] Gordon Burtch, Qinglai He, Yili Hong, and Dokyun Lee. 2022. How do peer awards motivate creative content? Experimental evidence from Reddit. *Management Science* 68, 5 (2022), 3488–3506.

[10] Juan Pedro Cerro Martinez, Montse Guitert Catasus, and Teresa Romeu Fontanillas. 2020. Impact of using learning analytics in asynchronous online discussions in higher education. *International Journal of Educational Technology in Higher Education* 17, 1 (2020), 1–18.

[11] Jay L Devore. 2015. *Probability and Statistics for Engineering and the Sciences.* Cengage Learning.

[12] Maureen Feeley and Julian Parris. 2012. An Assessment of the PeerWise Student-Contributed Question System's Impact on Learning Outcomes: Evidence from a Large Enrollment Political Science Course. *Available at SSRN 2144375* (2012).

[13] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.

[14] Drew Fudenberg and Jean Tirole. 1991. *Game theory.* MIT press.

[15] D Randy Garrison and J Ben Arbaugh. 2007. Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and higher education* 10, 3 (2007), 157–172.

[16] D Randy Garrison and Martha Cleveland-Innes. 2005. Facilitating cognitive presence in online learning: Interaction is not enough. *The American journal of distance education* 19, 3 (2005), 133–148.

[17] Jing Ge and Ulrike Gretzel. 2018. Emoji rhetoric: a social media influencer perspective. *Journal of marketing management* 34, 15-16 (2018), 1272–1295.

[18] Hajar Ghadirian, Keyvan Salehi, and Ahmad Fauzi Mohd Ayub. 2018. Social annotation tools in higher education: a preliminary systematic review. *International Journal of Learning Technology* 13, 2 (2018), 130–162.

[19] Xiaoshan Huang, Stephanie Beck, Lingyun Huang, and Susanne P Lajoie. 2023. Emotion and Emotion Regulation Matter: A Case Study on Teachers' Online Teaching Experience During COVID-19. (2023).

[20] Xiaoshan Huang, Alejandra Ruiz-Segura, Chengyi Tan, Tingting Wang, Robin Sharma, and Susanne P Lajoie. 2023. Social presence in technology-rich learning environments: how real we are feeling connected and how does it matter for learning? *Information and Learning Sciences* (2023).

[21] Tristan E Johnson, Thomas N Archibald, and Gershon Tenenbaum. 2010. Individual and team annotation effects on students' reading comprehension, critical thinking, and meta-cognitive skills. *Computers in human behavior* 26, 6 (2010), 1496–1507.

[22] Jelena Jovanovic, Dragan Gasevic, Shane Dawson, Abelardo Pardo, and Negin Mirriahi. 2017. Learning analytics to unveil learning strategies in a flipped classroom. *Internet and Higher Education* 33 (2017), 74–85.

[23] David A Kenny. 1975. Cross-lagged panel correlation: A test for spuriousness. *Psychological bulletin* 82, 6 (1975), 887.

[24] Akrivi Krouska, Christos Troussas, and Maria Virvou. 2018. Social annotation tools in digital learning: A literature review. In *2018 9th international conference on information, intelligence, systems and applications (IISA).* IEEE, 1–4.

[25] Chei Sian Lee, Hamzah Osop, Dion Hoe-Lian Goh, and Gani Kelni. 2017. Making sense of comments on YouTube educational videos: A self-directed learning perspective. *Online information review* 41, 5 (2017), 611–625.

[26] Shan Li, Juan Zheng, Xiaoshan Huang, Tingting Wang, and Susanne P Lajoie. 2023. Detection of Goal Setting and Planning in Self-regulated Learning Using Machine Learning and Think-aloud Protocols. (2023).

[27] Amanda Licastro. 2019. The past, present, and future of social annotation. In *Digital reading and writing in composition studies.* Routledge, 87–104.

[28] Chia-Ching Lin and Chin-Chung Tsai. 2012. Participatory learning through behavioral and cognitive engagements in an online collective information searching activity. *International Journal of Computer-Supported Collaborative Learning* 7 (2012), 543–566.

[29] Shane J Lopez. 2011. *The encyclopedia of positive psychology.* John Wiley & Sons.

[30] Shane J Lopez and Charles R Snyder. 2009. *The Oxford handbook of positive psychology.* Oxford University Press.

[31] Esteban Morales, Jeremiah H Kalir, Alice Fleerackers, and Juan Pablo Alperin. 2022. Using social annotation to construct knowledge with others: A case study across undergraduate courses. *F1000Research* 11 (2022).

[32] Haruka Notsu, Shigeru Iwakabe, and Nathan C Thoma. 2023. Enhancing working alliance through positive emotional experience: A cross-lag analysis. *Psychotherapy Research* 33, 3 (2023), 328–341.

[33] Shigehiro Oishi, Janetta Lun, and Gary D Sherman. 2007. Residential mobility, self-concept, and positive affect in social interactions. *Journal of personality and social psychology* 93, 1 (2007), 131.

[34] Reinhard Pekrun, Thomas Goetz, Wolfram Titz, and Raymond P Perry. 2002. Positive emotions in education. (2002).

[35] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71, 2001 (2001), 2001.

[36] Jeffrey Perrin. 2014. Features of engaging and empowering experiential learning programs for college students. *Journal of University Teaching & Learning Practice* 11, 2 (2014), 2.

[37] Krystle Phirangee and Jim Hewitt. 2016. Loving this dialogue!!!!: Expressing emotion through the strategic manipulation of limited non-verbal cues in online learning environments. In *Emotions, technology, and education.* Elsevier, 69–85.

[38] Robert C Pianta, Margaret Burchinal, Faiza M Jamil, Terri Sabol, Kevin Grimm, Bridget K Hamre, Jason Downer, Jennifer LoCasale-Crouch, and Carollee Howes. 2014. A cross-lag analysis of longitudinal associations between preschool teachers' instructional support identification skills and observed behavior. *Early Childhood Research Quarterly* 29, 2 (2014), 144–154.

[39] Jan L Plass, Dorothy M Chun, Richard E Mayer, and Detlev Leutner. 2003. Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior* 19, 2 (2003), 221–243.

[40] Harry T Reis and Shelly L Gable. 2003. Toward a positive psychology of relationships. (2003).

[41] Alvin E Roth. 1988. *The Shapley value: essays in honor of Lloyd S. Shapley.* Cambridge University Press.

[42] Alfred P Rovai. 2007. Facilitating online discussions effectively. *The Internet and Higher Education* 10, 1 (2007), 77–88.

[43] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.

[44] Eduardo Salas, Dana E Sims, and C Shawn Burke. 2005. Is there a "big five" in teamwork? *Small group research* 36, 5 (2005), 555–599.

[45] Rustam Shadiev, Wu-Yuin Hwang, Yueh-Min Huang, and Tzu-Yu Liu. 2015. The impact of supported and annotated mobile learning on achievement and cognitive load. *Journal of Educational Technology & Society* 18, 4 (2015), 53–69.

[46] Tuya Siqin, Jan Van Aalst, and Samuel Kai Wah Chu. 2015. Fixed group and opportunistic collaboration in a CSCL environment. *International Journal of Computer-Supported Collaborative Learning* 10 (2015), 161–181.

[47] Yunsheng Song, Jiye Liang, Jing Lu, and Xingwang Zhao. 2017. An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing* 251 (2017), 26–34.

[48] Gerry Stahl. 2015. A decade of CSCL. *International Journal of Computer-Supported Collaborative Learning* 10 (2015), 337–344.

[49] Anna Sun and Xiufang Chen. 2016. Online education and its effective practice: A research review. *Journal of Information Technology Education* 15 (2016).

[50] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[51] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58, 1 (1996), 267–288.

[52] Jeroen JG Van Merrienboer and Paul Ayres. 2005. Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development* 53, 3 (2005), 5–13.

[53] L S Vygotsky and Michael Cole. 1978. *Mind in Society: Development of Higher Psychological Processes.* Harvard University Press.

[54] Leah E Wickersham and Kim E Dooley. 2006. A content analysis of critical thinking skills as an indicator of quality of online discussion in virtual learning communities. *Quarterly Review of Distance Education* 7, 2 (2006), 185.

[55] Alyssa Wise, Yuting Zhao, and Simone Hausknecht. 2014. Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics* 1, 2 (2014), 48–71.

[56] Carolyn M Youssef-Morgan and Kristi M Bockorny. 2013. Engagement in the context of positive psychology. *Employee engagement in theory and practice* (2013), 50–70.

[57] Juan Zheng, Shan Li, Xiaoshan Huang, Tingting Wang, and Susanne P Lajoie. 2023. Do Thinking Styles Change with Task Complexity in Problem-solving? (2023).

[58] Xinran Zhu, Bodong Chen, Rukmini Manasa Avadhanam, Hong Shui, and Raymond Zhuo Zhang. 2020. Reading and connecting: using social annotation in online classes. *Information and Learning Sciences* 121, 5/6 (2020), 261–271.

[59] Hui Zou and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67, 2 (2005), 301–320.

A EXPLANATION OF MIXED-EFFECTS LINEAR REGRESSION MODEL

In our study, we integrated a linear mixed-effects model in the cross-lag regression analysis to investigate the influence of peer acknowledgment on learners' social annotation behaviors. The model accounts for individual-specific effects, which is crucial in our context where the effect of peer acknowledgment may vary across participants: $Y_{ij} = \beta_0 + \beta_1 \cdot X_{ij} + u_{1j} \cdot X_{ij} + \epsilon_{ij}$.

- Y_{ij}: Dependent variable representing the social annotation behavior (either the number of annotations initiated or responses given) for the i-th observation of the j-th participant.
- X_{ij}: Independent variable indicating peer acknowledgment for the i-th observation of the j-th participant.
- β_0: Fixed intercept of the model.
- β_1: Fixed slope, representing the overall effect of peer acknowledgment on social annotation behavior.
- u_{1j}: Random slope for the predictor X, specific to each participant j. This component allows the model to capture individual differences in the relationship between peer acknowledgment and social annotation behaviors.
- ϵ_{ij}: Residual error for the i-th observation of the j-th participant.

In summary, the model incorporates a random slope for the grouping variable (participant ID), denoted as u_{1j} in the formula. This approach acknowledges that the effect of peer acknowledgment on annotation behaviors may not be the same for all participants, thus providing a more nuanced understanding of the underlying relationships. This mixed-effects linear regression model with a random slope component is pivotal in our analysis, as it aptly handles the variability across individual participants, enhancing the robustness and interpretability of our findings.

B MODEL TRAINING DETAILS

Training/Test Split. The dataset, consisting of 1989 data entries, was divided into a training set (70%) and a testing set (30%). This division resulted in approximately 1392 data entries for training and about 597 entries for testing.

Cross-Validation. We employed k-fold cross-validation with k set to 5 on the training data. This method divides the training data into five subsets. The model is trained on four subsets and validated on the remaining one, with this process being repeated five times, each with a different subset as validation data.

Overfitting Prevention. Several strategies were implemented to mitigate overfitting. These included applying regularization techniques in models such as Lasso and ElasticNet, and limiting the complexity of the Random Forest and Gradient Boosting Regressor (GBR) models by controlling the maximum depth of trees. Performance metrics were consistently monitored across both cross-validation and testing phases.

Comparison of Cross-Validation and Testing Results. Overfitting is often indicated when a model performs significantly better on training data compared to unseen test data. In our case, the cross-validation and testing results for each model are quite close. For instance, the GBR, our best model, shows R-squared values of 0.9231 in cross-validation and 0.9123 in testing. This closeness in performance metrics across both phases suggests good generalization and is a positive sign against overfitting.