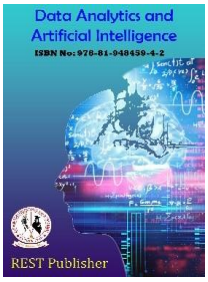
Data Analytics and Artificial Intelligence

Vol: 6(1), 2026

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/

DOI: https://doi.org/10.46632/daai/6/1/7



Non-Linear Patterns in Computer-Mediated Communication Predicting User Satisfaction through Emoji Usage, Message Clarity, and Engagement Metrics

***Vinay Dittakavi**

Distinguished Engineer, PwC

*Corresponding Author Email: vinay.vs.dittakavi@gmail.com

Abstract: *Computer-mediated communication (CMC) is increasingly common across educational, organizational, and social contexts. While it allows real-time interaction and overcomes geographical limitations, it lacks the immediate visual and auditory cues of face-to-face communication. This study examines CMC effectiveness by analyzing key variables such as message length, response time, emoji usage, formality, satisfaction, engagement, and clarity across 100 interactions. Both linear regression and random forest regression were applied to explore predictive patterns among these variables. Descriptive analysis showed considerable variation in communication behaviors, with message lengths ranging from 6 to 496 characters and response times between 65 and 3,562 seconds. Emoji usage averaged 9.06 per message, and satisfaction scores reflected moderate user contentment. Linear regression showed minimal predictive ability ($R^2 = 0.0382$ for training, 0.1037 for testing), whereas random forest regression performed well on training data ($R^2 = 0.8484$) but over fitted, reducing test performance ($R^2 = 0.2451$). Most CMC parameters displayed weak linear correlations, although emoji usage correlated moderately with engagement and clarity. These results indicate that CMC quality is shaped by complex, non-linear interactions rather than simple linear relationships. The study emphasizes the role of emotional expression through emojis and communication style variability, offering insights for improving user satisfaction and message clarity in educational, organizational, and broader digital communication contexts.*

Keywords: *Computer-mediated communication, Emotional expression and emojis, Digital communication effectiveness, Machine learning predictive models, Message clarity and engagement*

1. NTRODUCTION

Technically, while CMC enables real-time interaction and can match the speed and efficiency of face-to-face communication, it lacks the immediate visual and auditory cues that are present in traditional conversational methods. [1]CMC encompasses a range of tools, such as electronic messaging and conferencing systems, sometimes enhanced with audio or video components. Interestingly, this technology presents a paradox: while it enables users to connect with others regardless of physical distance, it can also limit the ability of those nearby like teachers, parents, or supervisors to observe or participate in those interactions, and May similarly restrict access between different online participants.[3] As research into CMC and the Internet is still relatively young, it's only recently that scholars have begun to move beyond simply debating whether these technologies have a positive or negative impact on society.[4] A transaction is entirely facilitated through computer-mediated communication.. The outer edges of the accompanying diagram illustrate various types of computer-based interactions, emphasizing that such communication can only be fully understood or its challenges addressed by considering the interplay of individuals, context, and timing. [5] CMC technologies are continuously advancing, bringing with them growing opportunities for educational use and academic research. As with all new technologies, it is important to evaluate both the strengths and drawbacks of emerging CMC formats within educational contexts.[6] Studies in the field indicate that computer-mediated interaction can positively influence different aspects of second language learning, either directly or indirectly. While a significant body of

Copyright@ REST Publisher

research exists on online communication and interaction in distance education, there is a noticeable lack of studies exploring the use of CMC in traditional, face-to-face classroom settings. [7] Although many investigations have focused on student satisfaction and perceptions of CMC in online learning, similar research involving on-campus students remains limited. [8] This analysis introduces a framework for understanding computer-mediated communication (CMC) competence, aiming to structure the growing body of research in this area. Before defining what constitutes CMC competence, certain objective conditions are outlined. Notably, this theory does not aim to rival existing media effects theories. [9]Instead, answering key research questions will help provide a more comprehensive understanding of CMC's overall effectiveness. If emotional content can be conveyed through CMC, identifying the elements that support or obstruct this transfer can enhance communication strategies within organizations. [10] Furthermore, the degree to which emotional information is communicated can guide more informed decision-making. Since the study of CMC in educational contexts is still developing, it is valuable to review recent findings to highlight commonly used methods and focal topics, as well as to pinpoint research gaps that should be addressed in future work. [11] This study offers a meta-analysis that compares communication performance in synchronous computer-mediated communication. [12] While CMC itself is not a recent development, its widespread use today is highlighting the unique environments shaped by digital communication. Social science researchers are showing growing interest in exploring the features of CMC and its impact on individuals, groups, and organizations. [13] Individuals engaged in CMC may feel more comfortable and enriched by a broader range of perspectives, leading to enhanced face-to-face interactions, in contrast to those that are not influenced by CMC. [14] As CMC becomes increasingly integrated into business, academia, and everyday social interactions especially through text messaging on mobile devices there is an indication that both the frequency and variety of communicative cues used in CMC may be expanding.[15] A particularly critical issue involves understanding how the reduced social cues inherent in computer-mediated communication influence the outcomes of the support provided. The absence of facial expressions, tone of voice, and other nonverbal signals may affect learners' perceptions, engagement, and overall effectiveness of the support process. [16] The effectiveness of computer-mediated communication (CMC) in second language learning largely hinges on the opportunities it provides for learners to engage with other speakers, especially native ones. These interactions are crucial not only for language acquisition but also for cultivating the social and pragmatic skills needed for successful communication.[17] The proliferation of computer-mediated communication (CMC) has stimulated extensive theoretical development and research aimed at understanding how it differs from other communication modes. Although much of the early theoretical work emerged from attempts to explain differences in decision-making processes between CMC groups and face-to-face groups, the foundational assumptions of these theories emphasize how the medium's characteristics shape user behavior. These perspectives suggest that the technological environment influences how individuals interact, interpret messages, and construct meaning within digitally mediated contexts. [18] A central function of nonverbal cues is their ability to express emotions and attitudes. They also serve to emphasize, contradict, substitute for, or regulate verbal communication. In computer-mediated communication—particularly in text-based formats such as SMS—emoticons have emerged as key tools for conveying socio-emotional information. Because computer-mediated communication lacks the facial expressions, gestures, and other nonverbal signals present in face-to-face interaction, emoticons were introduced to help recipients interpret the sender's feelings, intentions, or mood through visual cues embedded in text. [19] Communication is a fundamental aspect of human activity, and one of the core contributions of information systems (IS) is their ability to employ technologies that enhance workplace communication. Consequently, numerous computer-mediated communication (CMC) technologies have been developed to improve efficiency by streamlining and supporting interaction. [20]

2. MATERIALS AND METHODS

Materials

Message Length: Message length refers to the total number of characters, words, or sentences in a digital message. It indicates how detailed, explicit, or concise a communication is. Longer messages may convey complex information, while shorter messages often prioritize speed, simplicity, or quick interactions within computer-mediated communication environments.

Response Time (seconds): Response time represents the number of seconds a person takes to reply in a digital communication setting. Faster responses often signal higher engagement or urgency, while slower replies may reflect

multitasking, lower interest, or message complexity. It is a key behavioral indicator used to analyze communication patterns and interaction quality.

Emoji Count: Emoji count refers to the number of emojis included in a message. Emojis contribute emotional tone, clarity, and personality to computer-mediated communication. Higher emoji usage may reflect friendliness or informality, while fewer emojis suggest formality or neutrality. This metric supports research on emotional expression and social cues in digital conversations.

Satisfaction Score: Satisfaction score measures the user's level of contentment following a communication interaction. Typically assessed on a numerical scale, it evaluates message clarity, emotional tone, and overall effectiveness. Higher scores indicate positive communication experiences, whereas lower scores may signal dissatisfaction, misunderstandings, or breakdowns in digital communication dynamics.

Engagement Level: Engagement level indicates how actively and attentively a user participates in digital communication. It includes factors such as messaging frequency, response speed, descriptive expression, and emotional involvement. High engagement reflects strong interest or involvement, whereas low engagement signals reduced attention, passive participation, or diminished motivation during communication.

Clarity Assessment: Clarity assessment evaluates how understandable, accurate, and well-structured a message is in computer-mediated communication. It measures whether recipients can easily interpret the sender's meaning without confusion. High clarity scores denote effective communication, while lower scores indicate ambiguous wording, missing details, or language-related barriers affecting message comprehension.

Optimization Techniques

Linear Regression: A statistical method used a valuable technique to predict quantitative outcomes and has been extensively studied in numerous textbooks over time. Although it may seem less exciting than modern statistical learning methods, it is widely used and very relevant. In addition, it serves as a foundation for more advanced techniques, as many sophisticated statistical learning methods can be seen as extensions or generalizations of linear regression. Therefore, a solid understanding of linear regression is essential before exploring more complex approaches. The fundamental ideas of linear regression are examined in this chapter, along with the least squares method commonly used to build a model. Regression serves two primary purposes. First, it is widely used for forecasting and prediction, often with significant overlap with machine learning applications. With regression analysis, the dependent variable 'y' is predicted based on different values of the independent variables. The variable 'x'. This paper focuses on linear regression and multivariate regression, both of which are well suited for predictive modeling. Regression can take the form of simple linear regression or multiple regressions, which can be a type a regression. Simple linear regression involves a model with a single independent variable to determine its effect on a dependent variable. It is represented by the equation $y = \beta_0 + \beta_1 x + \varepsilon$, which describes the relationship between the variables. In addition, simple regression helps to distinguish the impact of independent variables from the interactions within the dependent variables.

Random Forest Regression: Random forest regression is a powerful supervised machine learning method used for predictive modelling. This method involves training several decision trees on various dataset subsets and their outputs are averaged to improve the prediction accuracy of the method, not only improving performance but also reducing the computational burden associated with training, storing, and predicting with many individual models. Due to their efficiency, random forests are extremely helpful for jobs involving regression, where continuous values are usually predicted. A "forest" of several independently built decision trees is created using the random forest technique, using the ultimate forecast derived by averaging each tree's outputs. By exposing each tree to slightly different data, this approach helps to reduce variance and increase the over fitting, ultimately improving the generalizability of the model.

3. ANALYSIS AND DISCUSSION

The dataset reveals significant variation in communication behaviours and outcomes. Message length and response time vary widely, reflecting different communication styles. Emoji usage and formality scores differ, indicating varying levels of expressiveness and professionalism. Satisfaction, engagement, and clarity ratings point to diverse user experiences, highlighting both highly positive and less effective communication instances.

Copyright@ REST Publisher

TABLE 1. Descriptive Statistics

	Message length	Response time sec	Emoji count	Formality score	satisfaction score	Engagement level	Clarity rating
count	100	100	100	100	100	100	100
mean	257.7	1794.62	9.06	0.534001	2.924182	51.50065	3.166513
std	144.9801	1001.406	6.392578	0.294001	1.181147	31.05125	1.196458
min	6	65	0	0.005062	1.057574	1.083765	1.058179
25%	132.75	940.5	4	0.291829	1.843254	28.00677	2.20199
50%	266	1808.5	8	0.573483	2.989504	52.53995	3.214677
75%	380.25	2692.25	15	0.795057	3.869879	78.02647	4.263922
max	496	3562	19	0.990054	4.962021	99.29648	4.998871

Table1 Detailed statistics indicate significant variation across all communication variables. Message length and response time show wide ranges, suggesting diverse communication styles. Emoji usage and formality scores vary moderately, reflecting mixed expressive trends. Satisfaction, engagement, and clarity ratings show mid-range averages, indicating generally balanced user experiences with considerable individual differences.

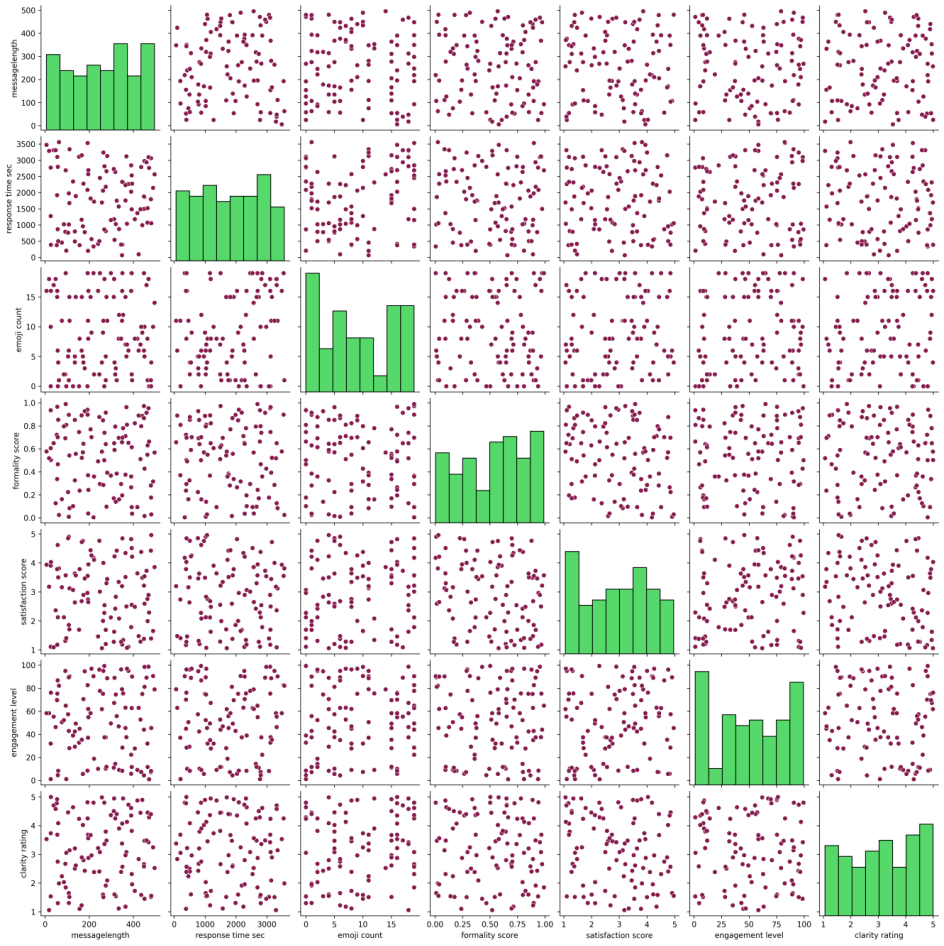


FIGURE 1. Scatter Plot of the various computer-mediated communications

Figure 1 shows a scatter plot matrix illustrating the relationships between various computer-mediated communication variables. These plots reveal diverse patterns; most variables show scattered distributions, indicating weak or no linear

Copyright@ REST Publisher

correlations. The histograms on the diagonal illustrate the data distribution of each variable and reveal the variability in message length, emoji usage, engagement, and clarity.

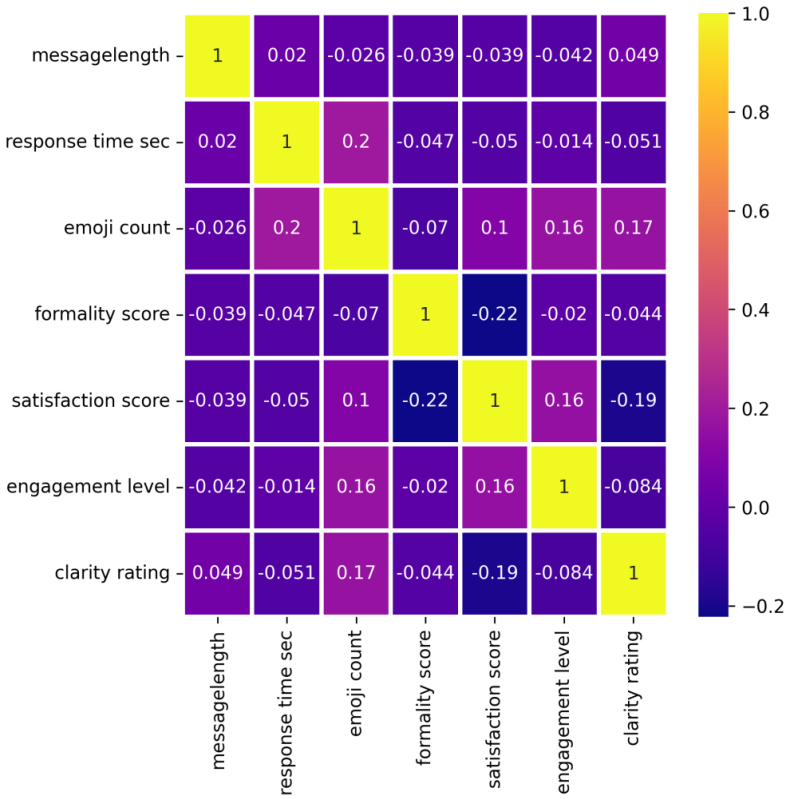


FIGURE 2. Correlation Heat Map on Process Parameters and Effects

Figure 2 provides a heat map of the correlations between process parameters and communication outcomes. The analysis generally shows weak correlations between the variables. Specifically, emoji count shows moderate positive relationships with engagement and clarity, while the formality score shows a slight negative correlation with satisfaction. Overall, the results indicate minimal linear dependence between most CMC parameters.

Linear Regression

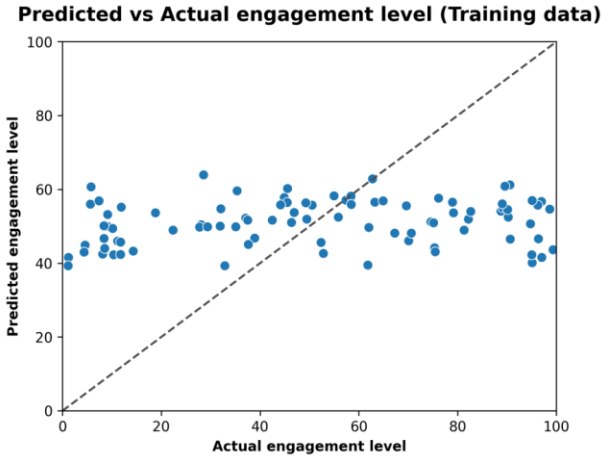


FIGURE 3. Linear Regression (Training data)

Copyright@ REST Publisher

Figure3 the linear regression model shows poor predictive performance on both the training and testing datasets. During training, the R^2 value is 0.0382, indicating that almost no variance is explained, and its RMSE value is 30.28. During testing, the R^2 value improves slightly to 0.1037, and the RMSE value is 29.25.

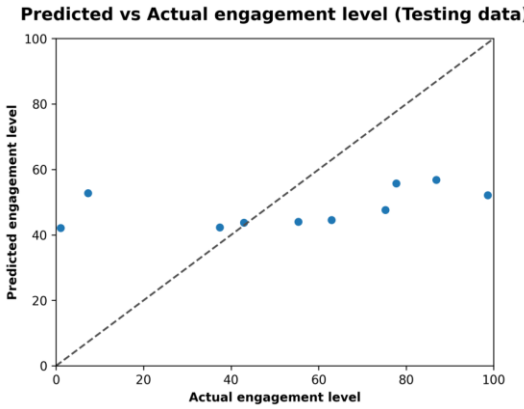


FIGURE 4. Linear Regression (Testing data)

Figure 4 presents the linear regression results for both the training and testing datasets. This model performs poorly; the training (R^2) value is 0.0382 and the testing (R^2) value is 0.1037, indicating that very little variance is explained. Error metrics such as RMSE and MAE are also high, suggesting that this model struggles to accurately predict values in both datasets.

TABLE 2. Performance Metrics of Linear Regression (Training Data and Testing Data)

Data	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Train	Linear Regression	0.0382	0.0382	916.9899	30.2818	26.0770	55.6453	0.9082	27.6052
Test	Linear Regression	0.1037	0.1459	855.8142	29.2543	24.8135	46.5591	1.3745	24.7720

Table 2 summarizes the performance of the linear regression model on the training and testing datasets. The model exhibits very low R^2 values (0.0382 for training, 0.1037 for testing), indicating that its explanatory power is weak. High error metrics, including RMSE and MAE, reflect significant inaccuracies in the predictions, suggesting that the model did not adequately capture the underlying data patterns.

Random Forest Regression

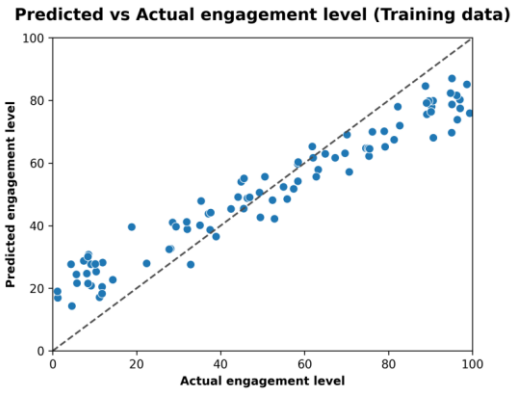


FIGURE 5. Random Forest Regression (Training data)

Copyright@ REST Publisher

Figure 5 shows the Random Forest Regression results. This model fits the training data well ($R^2 = 0.8484$) with a low RMSE (12.02), but its performance decreases on the test data ($R^2 = 0.2451$, RMSE = 26.85), indicating over fitting. While the training errors are low, the test errors are relatively high, reflecting limited generalization.

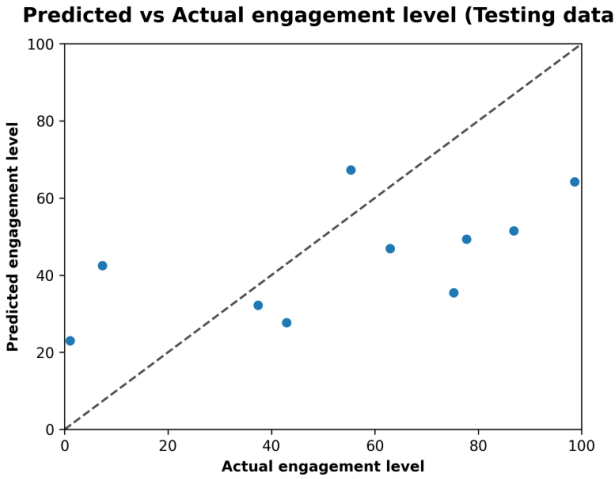


FIGURE 6. Random Forest Regression (Testing data)

Figure 6 shows the performance of the Random Forest Regression model on the training and test datasets. The model achieves high training accuracy ($R^2 = 0.8484$) with low errors, but the test performance drops significantly ($R^2 = 0.2451$), indicating over fitting. The test RMSE and MAE values are very high, which shows that the model struggles to generalize to unseen data.

TABLE 3. Performance Metrics of Random Forest Regression (Training Data and Testing Data)

Data	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Train	Random Forest Regression	0.8484	0.8484	144.5345	12.0223	10.1926	25.3025	0.3323	9.6191
Test	Random Forest Regression	0.2451	0.3614	720.7523	26.8468	24.3502	39.7754	1.0249	25.1546

Table 3 presents the performance metrics of the Random Forest Regression model. This model fits the training data well with low RMSE and MAE (R^2 = 0.8484), but the test performance is poor (R^2 = 0.2451), showing higher errors. This indicates over fitting; the model effectively learns the training patterns but struggles to generalize to unseen data.

4. CONCLUSIONS

This study of computer-mediated communication (CMC) variables highlights the complexity of digital interactions. Analysis of 100 exchanges shows substantial variability, reflecting diverse individual strategies. Linear regression performed poorly, indicating that simple linear models cannot capture the multifaceted factors influencing communication outcomes. Random forest regression excelled on training data but over fitted, underscoring the difficulty of creating generalizable predictive models for human communication. Notably, emoji usage showed moderate positive correlations with engagement and clarity, supporting their role in compensating for absent nonverbal cues. Encouraging appropriate emoji use may enhance emotional expression and overall communication quality in educational and organizational settings. Weak links between formality and satisfaction suggest that professionalism alone does not increase user satisfaction; informal, emotionally expressive messages may yield better outcomes. This challenges traditional assumptions about professional communication in digital contexts. Future research should focus on nonlinear interactions and advanced machine learning methods, alongside qualitative studies of user perceptions, to better understand and optimize communication effectiveness in CMC environments.

Copyright@ REST Publisher

REFERENCES

[1]. Kiesler, Sara, Jane Siegel, and Timothy W. McGuire. "Social psychological aspects of computer-mediated communication." American psychologist 39, no. 10 (1984): 1123.

[2]. Derks, Daantje, Agneta H. Fischer, and Arjan ER Bos. "The role of emotion in computer- mediated communication: A review." Computers in human behavior 24, no. 3 (2008): 766-785.

[3]. Jones, Rodney H. "The problem of context in computer-mediated communication." Discourse and technology: Multimodal discourse analysis (2004): 20-33.

[4]. Spears, Russell, Martin Lea, Rolf Arne Corneliussen, Tom Postmes, and Wouter Ter Haar. "Computer-mediated communication as a channel for social resistance: The strategic side of SIDE." Small group research 33, no. 5 (2002): 555-574.

[5]. Hesse, Bradford W., Carol M. Werner, and Irwin Altman. "Temporal aspects of computer-mediated communication." Computers in Human Behavior 4, no. 2 (1988): 147-165.

[6]. Luppicini, Rocci. "Review of computer mediated communication research for education." Instructional science 35 (2007): 141-185.

[7]. Smith, Bryan. "The use of communication strategies in computer-mediated communication." System 31, no. 1 (2003): 29-53.

[8]. An, Yun-Jo, and Theodore Frick. "Student perceptions of asynchronous computer-mediated communication in face-to-face courses." Journal of Computer-Mediated Communication 11, no. 2 (2006): 485-499.

[9]. Spitzberg, Brian H. "Preliminary development of a model and measure of computer-mediated communication (CMC) competence." Journal of Computer-Mediated Communication 11, no. 2 (2006): 629-666.

[10]. Harris, Ranida B., and David Paradice. "An investigation of the computer-mediated communication of emotions." Journal of Applied Sciences Research 3, no. 12 (2007): 2081-2090.

[11]. Hrastinski, Stefan, and Christina Keller. "Computer-mediated communication in education: A review of recent research." Educational Media International 44, no. 1 (2007): 61-77.

[12]. Ziegler, Nicole. "Synchronous computer-mediated communication and interaction: A meta-analysis." Studies in Second Language Acquisition 38, no. 3 (2016): 553-586.

[13]. Riva, Giuseppe, and Carlo Galimberti. "Computer-mediated communication: Identity and social interaction in an electronic environment." Genetic Social and General Psychology Monographs 124, no. 4 (1998): 434-464.

[14]. Dietz-Uhler, Beth, and Cathy Bishop-Clark. "The use of computer-mediated communication to enhance subsequent face-to-face discussions." Computers in Human Behavior 17, no. 3 (2001): 269-283.

[15]. Riordan, Monica A., and Roger J. Kreuz. "Cues in computer-mediated communication: A corpus analysis." Computers in Human Behavior 26, no. 6 (2010): 1806-1817.

[16]. Rains, Stephen A., Steven R. Brunner, Chelsie Akers, Corey A. Pavlich, and Selin Goktas. "Computer-mediated communication (CMC) and social support: Testing the effects of using CMC on support outcomes." Journal of Social and Personal Relationships 34, no. 8 (2017): 1186-1205.

[17]. González-Lloret, Marta. "Conversation analysis of computer-mediated communication." Calico Journal 28, no. 2 (2011): 308-325.

[18]. Walther, Joseph B. "Relational aspects of computer-mediated communication: Experimental observations over time." Organization science 6, no. 2 (1995): 186-203.

[19]. Spears, Russell, and Martin Lea. "Panacea or panopticon? The hidden power in computer-mediated communication." Communication research 21, no. 4 (1994): 427-459.

[20]. Xiaojuan Ou, Carol, Choon Ling Sia, and Chun Kit Hui. "Computer-mediated communication and social networking tools at work." Information Technology & People 26, no. 2 (2013): 172-190.

Copyright@ REST Publisher