Student Scientific Creativity Journal
Volume 4, Nomor 3, Mei 2026
e-ISSN: 2985-3753; p-ISSN: 2985-3761, Hal. 327-338
DOI: https://doi.org/10.55606/sscj-amik.v4i3.6177
Tersedia: https://journal.amikveteran.ac.id/index.php/sscj

Discursive Toxicity and Cyberbullying in TikTok User Comments: A Critical Discourse Analysis Approach

Dio Manik[1]*, Yeni Adventry Tanjung[2], Annisa Ananda Utomo[3], Amelia[4], Muhammad Natsir[5]
[1-5]Department of English Language and Literature Education, Faculty of Languages and Arts, State University of Medan, Indonesia
Author correspondence: dioman3030@gmail.com

Abstract. *The rapid development of social media has created new spaces of interaction that not only expand public participation but also give rise to problematic communication practices, such as toxicity and cyberbullying in netizen comments. This phenomenon has become increasingly relevant in the context of contemporary digital society, where the boundary between individual expression and verbal aggression is often blurred. This study aims to understand how discursive practices in netizen comments shape, reproduce, and construct the meaning of toxicity and cyberbullying from the participants' perspectives. Employing a qualitative approach with a case study design, the research involves 10–15 participants who are active social media users in Indonesia. Data were collected through in-depth interviews, non-participant observation, and document analysis of digital comments, and were analyzed using thematic analysis combined with a Critical Discourse Analysis (CDA) framework. The findings reveal three primary patterns: the normalization of verbal aggression as part of digital communication culture, psychological distress that leads to ambivalent coping strategies, and the dynamics of social identity that reinforce polarization in online interactions. These findings suggest that cyberbullying is not merely an individual act but a discursive practice embedded in power relations and digital social norms. Theoretically, this study enriches discourse analysis by integrating linguistic, social, and subjective experiential dimensions; practically, it offers implications for strengthening digital literacy, content moderation policies, and mental health based interventions. Furthermore, this research opens avenues for broader and more diverse explorations of digital discourse dynamics.*

Keywords: *Critical Discourse Analysis; Cyberbullying; Digital Interaction; Online Social Identity; Toxicity in Social Media.*

1. INTRODUCTION

The rapid development of social media, particularly TikTok, has transformed the landscape of digital communication into a dynamic, fast-paced, and highly participatory space. For many users, especially younger generations, TikTok functions not merely as an entertainment platform but also as a space for self-expression, identity construction, and social interaction. However, beneath this intensive engagement lies a less visible reality: the experience of being subjected to negative comments, mockery, and repeated verbal attacks. From the perspective of affected individuals, the comment section is no longer a space for dialogue but rather an arena that may generate emotional distress, insecurity, and even a crisis of self-confidence. Recent studies on visual and short-video-based social media further demonstrate that repeated exposure to hostile online interactions is closely associated with heightened anxiety, depressive symptoms, and reduced self-esteem, particularly among adolescents and young adults whose identities are strongly shaped through digital validation processes (Marengo et al., 2022; Valkenburg et al., 2022). In this regard, negative comment

Naskah Masuk: 15 Maret 2025; Revisi: 20 April 2026; Diterima: 23 Mei 2026; Terbit: 31 Mei 2026

cultures on TikTok cannot be understood merely as harmless interactional practices, but rather as communicative environments with tangible psychological consequences.

Preliminary observations of several viral TikTok contents indicate that derogatory, sarcastic, and aggressive comments frequently appear in large volumes and recur persistently. Exploratory interviews with active users reveal that some netizens perceive such practices as "normal" or even as part of digital humor culture. In contrast, victims interpret these interactions as forms of cyberbullying that significantly impact their psychological well-being, including anxiety, stress, and reluctance to re-engage in digital spaces. This phenomenon reflects an inherent tension between the normalization of toxicity and the deeply personal, often painful, subjective experiences of those targeted. Contemporary scholarship on participatory digital culture argues that toxic interactions are often legitimized through irony, meme culture, and collective performativity, causing harassment practices to be reframed as entertainment or "just joking" within online communities (Phillips & Milner, 2021). Such normalization mechanisms are particularly significant because they blur the boundaries between humor, aggression, and symbolic violence, thereby making harmful discourse socially acceptable in everyday digital interaction.

At a global level, issues related to cyberbullying and toxic communication in social media have attracted considerable scholarly attention within the fields of digital communication and social psychology. Recent studies highlight how anonymity and platform-specific characteristics intensify aggressive behavior through the online disinhibition effect, while digital discourse simultaneously reproduces social inequalities and identity-based tensions (KhosraviNik, 2020; Matamoros-Fernández & Farkas, 2021). Nevertheless, much of the existing literature remains focused on measuring prevalence or conducting quantitative content analyses, thereby overlooking the deeper exploration of meaning-making processes, lived experiences, and social dynamics embedded in everyday digital interactions, particularly on platforms such as TikTok. Recent discourse-oriented studies consequently emphasize the importance of qualitative and Critical Discourse Analysis approaches for examining how online comments construct identities, negotiate power relations, and normalize exclusionary practices beyond what can be captured through statistical frequency alone (Bouvier & Machin, 2021; KhosraviNik & Unger, 2023). From this perspective, TikTok comment sections should be understood not simply as datasets of interaction, but as discursive arenas in which social meanings and ideological positions are continuously produced and contested.

This gap in the literature underscores the need for a qualitative approach capable of examining toxicity and cyberbullying as complex discursive practices, and Critical Discourse Analysis (CDA) provides a framework for understanding how language in online comments not only reflects reality but also actively constructs and normalizes power relations, social identities, and boundaries of acceptable communication. Accordingly, the analytical focus extends beyond what is said to encompass how and why particular expressions emerge within specific social contexts. Based on this background, the present study aims to analyze the discursive practices of toxicity and cyberbullying in TikTok comment sections through a CDA approach, with the scope limited to interactions within comment sections of selected content that generate significant public engagement. This research is expected to contribute theoretically to the development of digital discourse analysis by elucidating the relationship between language, power, and subjective experience in online environments, and practically to inform the enhancement of digital literacy, the development of more responsive content moderation policies, and the design of interventions aimed at preventing cyberbullying in ways that are sensitive to users' lived experiences.

2. LITERATURE REVIEW

The phenomena of toxicity and cyberbullying in TikTok comment sections cannot be understood merely as individual behavior; rather, they should be viewed as social practices constructed through language, interaction, and power structures within digital spaces. Accordingly, this study is grounded in three main theoretical frameworks: Critical Discourse Analysis (CDA), the Online Disinhibition Effect, and Social Identity Theory, which collectively provide a comprehensive lens to explain the dynamics of meaning, emotion, and social relations in digital interactions.

Critical Discourse Analysis (CDA) in Digital Spaces

Critical Discourse Analysis (CDA) conceptualizes language not merely as a tool of communication but as a social practice that shapes and reproduces power relations, ideology, and social structures. In the context of social media, this perspective has evolved into social media critical discourse studies, emphasizing how digital platforms function as arenas for the production and circulation of discourse (KhosraviNik, 2020, Discourse & Communication).

From this perspective, netizen comments on TikTok are not interpreted as purely spontaneous expressions but as part of discursive practices that construct certain communicative norms, including the normalization of aggressive language. Matamoros-Fernández and Farkas (2021), in Television & New Media, demonstrate that hate speech on

social media is often embedded within broader social structures such as racism, sexism, and social exclusion.

In participants' experiences, for instance, comments involving physical mockery or identity-based insults do not merely cause personal harm but also reproduce implicit social standards regarding what is considered "acceptable" or "unacceptable." Thus, CDA enables the researcher to interpret how language in TikTok comments functions as a mechanism for legitimizing toxicity while simultaneously operating as an invisible form of social control.

Online Disinhibition Effect and the Normalization of Digital Aggression

The online disinhibition effect explains how characteristics of online communication such as anonymity, invisibility, and the absence of immediate consequences encourage individuals to express themselves more freely, including through verbal aggression. In recent scholarship, this phenomenon is increasingly understood not only as a reduction in self-regulation but also as part of an evolving digital communication culture that shapes new norms of interaction (Brown, 2021, Computers in Human Behavior Reports).

Within the TikTok context, the platform's structure characterized by rapid comment exchanges and algorithm-driven virality amplifies the reproduction of negative expressions. Toxic comments often receive significant engagement (e.g., likes or replies), which implicitly validates such behavior within the community. This creates a condition in which toxicity is no longer perceived as deviant but rather as an accepted mode of communication.

From the participants' perspectives, perpetrators often lack awareness of the emotional consequences of their comments, whereas victims experience an internal tension between responding and protecting themselves. This framework thus helps explain the paradox between the perceived freedom of expression and the psychological consequences that follow.

Social Identity Theory and Digital Polarization

Social Identity Theory explains how individuals construct their sense of self based on group membership and how this influences intergroup interactions. In social media contexts, such identities are frequently expressed through fandom affiliations, content preferences, or ideological positions.

Recent studies indicate that digital interactions tend to intensify identity polarization, wherein users construct boundaries between "us" and "them" through language (Carlson & Frazer, 2020, Social Media + Society). In TikTok comment sections, this is manifested in collective attacks directed at individuals perceived as "different" or as deviating from group norms.

Participants' experiences reveal that negative comments are often not isolated acts but rather collective and layered, generating greater social pressure. In this sense, cyberbullying transcends individual behavior and becomes a social practice that reinforces group solidarity through the exclusion of others.

Theoretical Comparison and Research Positioning

These three theoretical approaches offer distinct yet complementary perspectives. The online disinhibition effect primarily addresses the psychological dimensions of individual behavior in digital environments, while Social Identity Theory emphasizes group dynamics and intergroup relations. In contrast, Critical Discourse Analysis provides a broader framework by situating language within the context of power and ideology.

This study adopts CDA as the primary analytical lens, as it enables the integration of psychological and social dimensions within a contextualized analysis of language. Through CDA, toxicity and cyberbullying are understood not merely as behaviors or interactions but as discursive constructions with broader social implications.

Conceptual Framework

Based on these theoretical perspectives, this study conceptualizes TikTok comments as discursive practices that actively construct digital social reality. Language in comments does not simply reflect experience; it produces meaning, shapes identities, and reproduces power relations.

The researcher adopts an interpretive stance that seeks to "listen" to participants' voices as expressions that are not always linear or consistent. Toxicity is understood as a negotiated practice within interaction, while cyberbullying is viewed as an experience shaped by social relations and subjective interpretation.

Accordingly, the analysis is not directed toward measuring frequency or intensity but toward understanding how meaning is constructed, how emotions are negotiated, and how these discursive practices become embedded within broader digital communication cultures.

3. RESEARCH METHOD

This study employs a qualitative research design using Critical Discourse Analysis (CDA) as the primary analytical framework to examine how toxicity and cyberbullying are discursively constructed in TikTok comment sections, reflecting power relations, social identity negotiation, and the normalization of verbal aggression within participatory digital culture. The approach follows Fairclough (2013) and Wodak (2015), complemented by the online disinhibition effect (Suler, 2004) and digital toxicity as discursive violence (Jane, 2018).

Participants consist of 25–30 active TikTok users selected through purposive and snowball sampling, focusing on individuals aged 18–30 in Indonesia who have been involved as perpetrators, victims, or observers of toxic comments. The study examines comment sections of viral TikTok content related to appearance, personal expression, social identity, and lifestyle. Data were collected through semi-structured in-depth interviews (via Zoom or Google Meet), digital document analysis (screenshots of derogatory, sarcastic, or aggressive comments), and non-participant observation within TikTok comment sections to capture interaction dynamics and language patterns in real time. Instruments include interview guidelines, digital field notes, encrypted storage devices, and NVivo software.

Data analysis was conducted using Fairclough's three-dimensional CDA model, encompassing text analysis (linguistic features of comments such as word choice, metaphor, and presupposition), discursive practice analysis (how comments are produced, consumed, and interpreted within TikTok's platform architecture), and social practice analysis (the broader socio-cultural context that normalizes or challenges toxicity). The analysis proceeds through stages of data transcription, open coding, thematic categorization, and critical interpretation. The validity of the data was ensured through triangulation (across interviews, observations, and document analysis), member checking (verifying interpretations with participants), and researcher reflexivity (acknowledging the researcher's position), indicating consistency and credibility of the findings. Ethical considerations were addressed by ensuring informed consent, maintaining participant confidentiality through anonymization (using pseudonyms or codes), avoiding the use of data that could harm or expose participants, securely storing all digital data, and using the data solely for academic purposes, while also considering the ethical use of digital content obtained from public but sensitive online spaces.

4. RESULT AND DISCUSSION

The findings of this study are presented using a thematic analysis approach, which aims to capture patterns of meaning derived from participants' experiences in interacting within TikTok comment sections. The analysis identified three interrelated main themes: (1) the normalization of toxicity as a form of digital communication culture, (2) the emotional experiences and coping strategies of victims, and (3) the dynamics of identity and social polarization in comment interactions. These themes do not stand independently; rather, they form a complex network of meanings within discursive practices in digital spaces.

The Normalization of Toxicity as Digital Communication Culture

The phenomenon of toxicity in TikTok comments often emerges within the context of viral content that provokes strong public reactions. In such situations, negative comments are not merely sporadic but appear repeatedly and massively, as if they constitute a "normal" pattern of communication.

One participant (P3) stated:

"At first, I was shocked to read harsh comments, but over time, it became normal. In fact, if there are no such comments, it feels strange."

This statement indicates a process of habituation, in which verbal aggression is no longer perceived as deviant but rather as an implicit norm in digital interaction. Observations further reveal that comments involving physical mockery, sarcasm, and personal insults often receive significant engagement, such as likes and replies, which indirectly reinforces the **استمرار** of such practices.

However, this normalization is not entirely free of tension. Another participant (P7) reflected:

"Sometimes I laugh along, but deep down I realize it actually goes too far."

This highlights an ambiguity between participation in toxic practices and individual moral awareness, suggesting that normalization does not necessarily imply full acceptance but rather an ongoing process of negotiation.

Emotional Experiences and Coping Strategies of Victims

Behind comments that may appear casual or humorous lies a deeper emotional experience for those targeted. Several participants who had been victims described feelings of distress, shame, and diminished self-confidence.

Participant (P1) explained:

"I ended up deleting the video not because I was wrong, but because I couldn't handle reading the comments."

This experience demonstrates that cyberbullying extends beyond the digital realm and significantly affects individuals' psychological well-being. In some cases, participants chose to withdraw from digital interactions as a form of self-protection.

Nevertheless, coping strategies were not always passive. Participant (P5) stated:

"Now I'm more selective. If comments become too harsh, I immediately block or ignore them."

This reflects the presence of individual agency in managing one's digital environment. However, a paradox emerges between the desire to remain present in digital spaces and the need to protect mental health, often placing victims in a dilemma.

Identity Dynamics and Social Polarization

Comments on TikTok do not merely reflect individual opinions; they also function as arenas where social identities are negotiated and contested. In many instances, verbal attacks are directed not only at individuals but also at the identities they represent, such as appearance, background, or group affiliation.

Participant (P9) noted:

"Once there is a difference of opinion, people start attacking in groups. It's no longer a discussion it becomes a comment war."

This situation illustrates how comment sections become spaces of polarization, where boundaries between "us" and "them" are constructed through language. Observations indicate that collective forms of commenting intensify social pressure on targeted individuals.

At the same time, polarization can also generate forms of solidarity. Participant (P2) expressed:

"Sometimes there are people who defend you, so you don't feel completely alone."

This suggests that, amid cyberbullying practices, there are also spaces of resistance and social support, although these are often less dominant than the flow of negative comments.

Interconnection Among Themes

The three themes identified demonstrate strong interconnections. The normalization of toxicity creates an environment that enables cyberbullying to flourish, while the emotional experiences of victims reveal the tangible impact of such practices. Meanwhile, identity dynamics and social polarization intensify both the direction and magnitude of aggression within comment interactions.

Overall, the findings do not present a linear pattern but rather a network of experiences characterized by ambiguity, negotiation, and tension. TikTok comment spaces emerge not only as reflections of social reality but also as active sites in which such reality is continuously constructed through recurring and negotiated discursive practices.

Discussion

This study identifies three main findings: (1) the normalization of toxicity as part of digital communication culture, (2) the emotional experiences of victims accompanied by coping strategies, and (3) identity dynamics that reinforce social polarization in TikTok comment interactions. These findings indicate that cyberbullying cannot be understood merely

as individual behavior; rather, it constitutes a complex discursive practice shaped by the interplay of language, platform structures, and social context.

The Normalization of Toxicity as a Discursive Practice

The finding on the normalization of toxicity suggests that verbal aggression has undergone a process of legitimization in digital spaces. From the perspective of Critical Discourse Analysis (CDA), this phenomenon can be understood as the reproduction of ideology through language, in which aggressive expressions are no longer perceived as deviant but rather as part of accepted communicative norms. KhosraviNik (2020) argues that social media functions as a space where discursive practices not only reflect social reality but also actively shape it through repetition and circulation of discourse.

These findings are consistent with Matamoros-Fernández and Farkas (2021), who demonstrate that hate speech on social media is often normalized through platform mechanisms and user interactions. However, this study extends prior research by revealing a dimension of moral ambiguity among participants, wherein individuals simultaneously engage in toxic practices while remaining aware of their problematic nature. This suggests that normalization is not a linear process but rather the result of continuous negotiation between individual awareness and social pressure.

Emotional Experiences and Agency in Digital Spaces

The second finding highlights that victims' emotional experiences are not only personal but also socially situated. Within the framework of the online disinhibition effect, verbal aggression in digital environments is often facilitated by anonymity and social distance (Brown, 2021). However, this study demonstrates that its impact extends beyond the digital realm, with tangible consequences for individuals' psychological well-being.

Notably, the study also reveals forms of agency among participants, such as blocking accounts or ignoring negative comments. This indicates that individuals are not entirely passive in the face of cyberbullying, but actively negotiate their positions within digital spaces. This finding enriches previous studies that tend to portray victims primarily as vulnerable subjects, by offering a perspective that acknowledges their capacity to manage and reconstruct their experiences.

Nevertheless, a tension emerges between the desire to remain present in digital spaces and the need for self-protection. This tension reflects a broader paradox within contemporary digital culture, where social participation simultaneously constitutes a source of potential vulnerability.

Social Identity and Polarization in Digital Interaction

The identity dynamics identified in this study demonstrate that TikTok comment sections function as arenas for the construction and contestation of social identity. From the perspective of Social Identity Theory, language is used to establish boundaries between "ingroup" and "outgroup," thereby reinforcing polarization (Carlson & Frazer, 2020).

These findings align with existing literature suggesting that social media intensifies social fragmentation. However, this study contributes further by showing that polarization does not only produce conflict but can also generate forms of solidarity. In some instances, support from other users serves as a form of resistance against cyberbullying. This indicates that digital spaces are inherently ambivalent simultaneously operating as sites of symbolic violence and as potential spaces for social support.

Integrating Findings within the Critical Discourse Analysis Framework

The three identified themes reinforce the position of CDA as the primary analytical framework, as it enables the integration of language, power, and social context. Toxicity and cyberbullying emerge not merely as communicative phenomena but as discursive practices that reproduce norms, shape identities, and regulate social relations within digital environments.

Conceptually, this study proposes that cyberbullying on TikTok can be understood as the outcome of the interaction between the normalization of aggressive language, group identity dynamics, and platform structures that facilitate visibility and virality. Thus, the phenomenon cannot be reduced to issues of individual ethics alone but must be situated within a broader digital communication ecosystem.

Researcher Reflexivity

In the process of interpretation, the researcher's position as part of a digital society inevitably influences the reading of the data. Familiarity with social media culture allows for a more empathetic understanding of participants' experiences, yet it also requires critical reflexivity to avoid reproducing the same normalization observed in the data.

Moreover, the socio-cultural background of participants primarily young users shapes how toxicity is perceived, often as something ambiguous, positioned between entertainment and symbolic violence. This underscores that meaning is not fixed but continuously negotiated within specific social contexts.

5. CONCLUSION

This study demonstrates that the practices of toxicity and cyberbullying in TikTok comment sections cannot be understood merely as individual behavior, but rather as discursive practices normalized within digital communication culture, characterized by three main patterns: the legitimization of verbal aggression, the complex emotional experiences of victims, and identity dynamics that reinforce social polarization; from these findings emerges the understanding that toxicity is often negotiated as a form of "normal" interaction, while simultaneously generating moral tension at the individual level, whereas victims are not entirely passive but develop coping strategies and forms of agency, and at the same time cyberbullying functions as a social mechanism for constructing group solidarity through the exclusion of others, thus, at a conceptual level, this study contributes by positioning the phenomenon as an ambivalent discursive practice situated between normalization, resistance, and the negotiation of meaning; practically, these findings highlight the importance of educationally oriented moderation policies, the integration of critical digital literacy within educational curricula, and the strengthening of social interventions that support users' mental health, particularly among young people, although this study is limited by its focus on a single platform, the relative homogeneity of participants, and constraints in the depth of exploration, future research is therefore encouraged to expand contexts, employ alternative methodological approaches such as digital ethnography or multimodal analysis, and further investigate aspects of resistance and the transformation of communication norms, ultimately emphasizing that understanding cyberbullying requires a critical reading of language as a dynamic social practice within the evolving digital ecosystem.

REFERENCES

Citron, D. K. (2014). *Hate crimes in cyberspace*. Harvard University Press.

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the International AAAI Conference on Web and Social Media, 11*(1), 512–515.

Fairclough, N. (2013). *Critical discourse analysis: The critical study of language* (2nd ed.). Routledge.

Goffman, E. (1959). *The presentation of self in everyday life*. Anchor Books.

Jane, E. A. (2015). Flaming? What flaming? The pitfalls and potentials of researching online hostility. *Ethnography, 16*(1), 65–87. https://doi.org/10.1177/1466138114529087

Jane, E. A. (2018). Systemic misogyny exposed: Translating rapeglish from the manosphere with a random rape threat generator. *International Journal of Cultural Studies, 21*(6), 661–680. https://doi.org/10.1177/1367877917734042

KhosraviNik, M. (2020). Social media critical discourse studies (SMCDS). In S. Zhu (Ed.), *The Routledge handbook of critical discourse studies* (pp. 582–596). Routledge.

KhosraviNik, M., & Unger, J. W. (2023). Critical discourse studies and social media: Power, resistance and critique in changing media ecologies. *Critical Discourse Studies, 20*(1), 1–15.

Marengo, D., Settanni, M., & Longobardi, C. (2022). Seeking likes: Longitudinal associations between Instagram use, self-esteem, and depressive symptoms in adolescence. *Journal of Adolescence, 94*(1), 1–10.

Marwick, A. E., & Boyd, D. (2011). To see and be seen: Celebrity practice on Twitter. *Convergence, 17*(2), 139–158. https://doi.org/10.1177/1354856510394539

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media, 22*(2), 205–224. https://doi.org/10.1177/1527476420982230

Patchin, J. W., & Hinduja, S. (2015). Measuring cyberbullying: Implications for research. *Aggression and Violent Behavior, 23*, 69–74. https://doi.org/10.1016/j.avb.2015.05.013

Phillips, W., & Milner, R. M. (2021). *You are here: A field guide for navigating polarized speech, conspiracy theories, and our polluted media landscape*. MIT Press.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*(3), 321–326. https://doi.org/10.1089/1094931041291295

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

Valkenburg, P. M., Meier, A., & Beyens, I. (2022). Social media use and its impact on adolescent mental health: An umbrella review of the evidence. *Current Opinion in Psychology, 44*, 58–68.

Van Dijk, T. A. (2015). Critical discourse analysis. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed., pp. 466–485). Wiley.

Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research. *Sociology, 51*(6), 1149–1168. https://doi.org/10.1177/0038038517708140

Wodak, R. (2015). Critical discourse analysis, discourse-historical approach. In K. Tracy, C. Ilie, & T. Sandel (Eds.), *The international encyclopedia of language and social interaction* (pp. 1–14). Wiley-Blackwell.

Wodak, R., & Meyer, M. (2016). *Methods of critical discourse studies* (3rd ed.). Sage Publications.

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web*. Continuum.