

ONLINE DISINHIBITION AND ONLINE AGGRESSION AMONG YOUNG ADULTS: THE MODERATING ROLE OF EMOTIONAL DYSREGULATION

Palwasha Anwar^{*1}, Tuba Khalid², Dr. Samra Javed³, Aqsa Noor⁴, Sana Farooq⁵, Sabeen Sabir⁶

¹Ph.D. Student, Department of Psychology, Institute of Social Sciences, Bursa Uludağ University, Bursa, Turkey

²Clinical Psychologist, MS Clinical Psychology, RICPP, Riphah International University

³Director, Office of Research Innovation and Commercialization, Alghazali University

⁴Clinical Psychologist, Behavior Analyst, SEN Resource Teacher, Beaconhouse School Professional

⁵Clinical Psychologist, Behavior Analyst, SEN Resource Teacher, Riphah International University

⁶Former Visiting Lecturer, Bahria University Islamabad

¹pulwashakhansrk@gmail.com, ²tuba2412@gmail.com, ³samra.javed@agu.edu.pk,
⁴aqsarana591@gmail.com, ⁵sanaahmed6683@gmail.com, ⁶sabeensabir@yahoo.com

¹ORCID: 0009-0004-2492-1979, ²ORCID: 0009-0006-9032-9345

⁴ORCID: 0009-0008-0183-9330, ⁶ORCID: 0009-0003-3004-9171

Corresponding Author: *

Palwasha Anwar

DOI: <https://doi.org/10.5281/zenodo.20701816>

Received	Accepted	Published
18 April 2026	30 May 2026	15 June 2026

ABSTRACT

Online aggression has become an increasingly important concern among young adults due to the widespread use of digital communication platforms. Guided by the Online Disinhibition Effect Theory and the General Aggression Model, the present study examined the relationships among perceived anonymity, online disinhibition, emotional dysregulation, and online aggression among young adults. A cross-sectional survey design was employed, and data were collected from 442 young adults aged 18–30 years. Participants completed measures of perceived anonymity, online disinhibition, emotional dysregulation, and online aggression. Data were analyzed using correlation, regression, mediation, and moderated mediation analyses. Results indicated that perceived anonymity was positively associated with online disinhibition, which in turn significantly predicted online aggression. Online disinhibition mediated the relationship between perceived anonymity and online aggression. Emotional dysregulation significantly moderated the relationship between online disinhibition and online aggression, such that the association was stronger among individuals reporting higher levels of emotional dysregulation. Furthermore, the indirect effect of perceived anonymity on online aggression through online disinhibition increased across levels of emotional dysregulation, supporting a moderated mediation model. The findings suggest that perceived anonymity contributes to online aggression through increased online disinhibition, while emotional dysregulation amplifies this pathway. Interventions aimed at improving emotion regulation may help reduce aggressive online behavior among young adults.

Keywords: online aggression, perceived anonymity, online disinhibition, emotional dysregulation, young adults

INTRODUCTION

Online aggression intentional behaviour conducted through digital platforms to harm, humiliate, or harass another person has become a prominent public health concern among young adults (Kowalski et al., 2014). Unlike face-to-face aggression, online aggression occurs within an architecture that provides anonymity, removes immediate social feedback, and enables hostile acts to reach large audiences rapidly (Suler, 2004). Understanding the psychological mechanisms that translate these environmental conditions into aggressive behaviour is a central question in cyberpsychology.

Suler's (2004) Online Disinhibition Effect describes how online conditions lower behavioural constraints that ordinarily regulate interpersonal conduct. Toxic disinhibition manifested as hostility and norm-violating behaviour is facilitated by perceived anonymity, invisibility, and reduced accountability. Lapidot-Lefler and Barak (2012) experimentally confirmed that anonymity conditions significantly increased hostile communication, while Udris (2014) validated a scale distinguishing benign from toxic disinhibition, finding toxic disinhibition positively associated with cyberbullying perpetration. Perceived anonymity, understood as the subjective belief that one's actions will not be traced to one's offline identity, is the key antecedent of this disinhibitory process (Christopherson, 2007).

The General Aggression Model (GAM; Anderson & Bushman, 2002) situates these processes within a broader theoretical architecture: situational inputs such as disinhibitory online conditions alter internal affective and cognitive states, which are then appraised via either automatic-impulsive or controlled-deliberative pathways. Individual differences in the capacity for controlled appraisal moderate whether situational inputs produce overt aggression. Emotional dysregulation defined as difficulty managing emotional responses in contextually appropriate ways (Gratz & Roemer, 2004) directly impairs the controlled-deliberative pathway. Individuals with elevated dysregulation struggle to inhibit impulsive behaviour when emotionally aroused and have limited access to

effective regulatory strategies (Muraven & Baumeister, 2000). Robertson et al. (2012) identified emotion regulation difficulties as consistent predictors of reactive aggression, and Buckels et al. (2014) linked poor impulse control to online trolling behaviour.

Despite this theoretical coherence, prior research has not adequately explained why disinhibitory conditions produce aggression in some individuals but not others. The present study addresses this gap by testing a moderated mediation model in which emotional dysregulation moderates the disinhibition-aggression pathway. A second gap concerns geographic representation: the cyberpsychology literature draws predominantly from Western samples (Kowalski et al., 2014; Willard, 2007), while South Asian contexts including Pakistan, where internet penetration among urban young adults has grown substantially remain underrepresented. The present study tests the proposed model in a Pakistani young adult sample, contributing empirical data from a culturally distinct context.

The following hypotheses are tested:

H1: Perceived anonymity will be positively associated with online disinhibition.

H2: Online disinhibition will be positively associated with online aggression.

H3: Online disinhibition will mediate the relationship between perceived anonymity and online aggression.

H4: Emotional dysregulation will moderate the disinhibition-aggression relationship, such that the association is stronger at higher dysregulation.

H5: The indirect effect of perceived anonymity on online aggression will be significantly stronger at higher levels of emotional dysregulation (moderated mediation).

Method

Research Design

A quantitative cross-sectional survey design was employed to test the moderated mediation model.

Participants

Participants

The final sample consisted of 442 young adults aged 18–30 years. Participants were recruited

through convenience sampling using online survey distribution. Inclusion criteria required participants to be between 18 and 30 years of age and active users of social networking platforms. Individuals who did not complete the survey were excluded from the analyses.

Measures

Perceived Anonymity.

Perceived anonymity was measured using the 12-item scale developed by Lapidot-Leffler and Barak (2012). Responses were recorded on a 5-point Likert scale, with higher scores indicating greater perceived anonymity. Internal consistency in the present study was acceptable ($\alpha = .84$).

Online Disinhibition. Online disinhibition was assessed using the toxic disinhibition subscale of the Online Disinhibition Scale (Udris, 2014). Higher scores reflected greater online disinhibition ($\alpha = .83$).

Emotional Dysregulation. Emotional dysregulation was measured using the DERS-16 (Bjureberg et al., 2016). Higher scores indicated greater emotional dysregulation ($\alpha = .92$).

Online Aggression. Online aggression was assessed using adapted perpetration items from Kowalski et al. (2014). Higher scores reflected greater engagement in aggressive online behavior ($\alpha = .88$).

Ethical Considerations

Participation was voluntary. Participants were informed about the purpose of the study, confidentiality of responses, anonymity of participation, and their right to withdraw at any stage without penalty. Completion of the survey was considered informed consent.

Data Analysis

Data were screened for missing values and outliers prior to analysis. Descriptive statistics, reliability analyses, and Pearson correlations were computed. The hypothesized moderated mediation model was tested using PROCESS Macro Model 14 (Hayes, 2018) with 10,000 bootstrap resamples. Statistical significance was determined using 95% confidence intervals.

Results

Participant Characteristics

The final sample comprised 442 young adults aged 18–30 years. Males constituted 54.5% of the sample ($n = 241$) and females 45.5% ($n = 201$). The largest age group was 22–25 years (42.3%, $n = 187$), followed by 18–21 years (35.3%, $n = 156$) and 26–30 years (22.4%, $n = 99$). Undergraduate students comprised 59.5% of participants ($n = 263$) and postgraduate students 40.5% ($n = 179$). Demographic characteristics are presented in Table 1.

Table 1

Demographic Characteristics of Participants (N = 442)

Variable	n	%
Gender		
Male	241	54.5
Female	201	45.5
Age Group		
18–21 years	156	35.3
22–25 years	187	42.3
26–30 years	99	22.4
Education Level		
Undergraduate	263	59.5
Postgraduate	179	40.5

Preliminary Analyses

Descriptive statistics, internal consistency coefficients (Cronbach's α), and zero-order Pearson correlations among all study variables are presented in Table 2. All four scales demonstrated acceptable to excellent internal consistency (α range = .83–.92). Perceived anonymity was positively and significantly correlated with online disinhibition, $r = .31, p < .01$. Online disinhibition

was positively correlated with online aggression, $r = .34, p < .01$. Emotional dysregulation showed the strongest bivariate association with online aggression, $r = .41, p < .01$. The association between perceived anonymity and emotional dysregulation was non-significant ($r = .12, p > .05$), supporting adequate discriminant separation between the predictor and the moderator.

Table 2

Means, Standard Deviations, Reliability Coefficients, and Intercorrelations Among Study Variables (N = 442)

Variable	M	SD	α	1	2	3	4
1. Perceived Anonymity	3.42	0.81	.84	—			
2. Online Disinhibition	3.11	0.74	.83	.31**	—		
3. Emotional Dysregulation	2.65	0.92	.92	.12	.18*	—	
4. Online Aggression	1.98	0.68	.88	.14*	.34**	.41**	—

Note. * $p < .05$. ** $p < .01$. Two-tailed.

Table 3

Simple Linear Regression Analysis Predicting Online Disinhibition

Predictor	B	SE	β	t	p
Perceived Anonymity	0.28	0.04	.31	6.81	< .001

Note. $df = 440$.

Table 4

Simple Linear Regression Analysis Predicting Online Aggression

Predictor	B	SE	β	t	p
Online Disinhibition	0.22	0.04	.34	5.23	< .001

Note. $df = 440$.

Table 5

Indirect Effect of Perceived Anonymity on Online Aggression Through Online Disinhibition

Effect pathway	Estimate	BootSE	95% LLCI	95% ULCI
Perceived Anonymity → Online Disinhibition → Online Aggression	0.06	0.02	0.03	0.10

Note. Bias-corrected bootstrap confidence intervals based on 10,000 resamples. CI excludes zero, confirming significant mediation.

Table 6

Moderation Analysis: Emotional Dysregulation as Moderator of the Disinhibition–Aggression Relationship

Predictor	B	SE	t	p
Online Disinhibition	0.22	0.04	5.23	< .001
Emotional Dysregulation	0.26	0.04	6.44	< .001
Online Disinhibition × Emotional Dysregulation	0.14	0.04	3.32	.001

Note. All predictors mean-centred prior to computation of the interaction term. Outcome variable = Online Aggression.

Table 7

Conditional Indirect Effects of Perceived Anonymity on Online Aggression at Three Levels of Emotional Dysregulation

Level of Dysregulation	Emotional Indirect Effect	BootSE	95% LLCI	95% ULCI
Low (−1 SD)	0.03	0.01	0.01	0.05
Mean (0 SD)	0.06	0.02	0.03	0.10
High (+1 SD)	0.10	0.03	0.05	0.16

Note. Bias-corrected bootstrap confidence intervals based on 10,000 resamples. All CIs exclude zero.

Table 8

Index of Moderated Mediation

Index	BootSE	95% LLCI	95% ULCI
0.04	0.01	0.01	0.07

Note. Bias-corrected bootstrap confidence interval based on 10,000 resamples. CI excludes zero, confirming significant moderated mediation.

Discussion

The present study tested an integrated moderated mediation model of online aggression in a sample of Pakistani young adults. All five hypotheses were supported.

H1: Perceived Anonymity and Online Disinhibition

Perceived anonymity significantly predicted online disinhibition ($\beta = .31$), consistent with Suler's (2004) theoretical account and the experimental findings of Lapidot-Lefler and Barak (2012). The subjective belief that one's actions are unidentifiable attenuates anticipated social consequences, thereby weakening the normative inhibitors of online behaviour (Christopherson, 2007). This relationship held in the Pakistani

sample, providing preliminary cross-cultural support for the anonymity–disinhibition pathway.

H2 and H3: Online Disinhibition as Mediator

Online disinhibition predicted online aggression ($\beta = .34$) and significantly mediated the anonymity–aggression relationship (indirect effect = 0.06, 95% CI [0.03, 0.10]). Perceived anonymity does not produce aggression directly but operates through a psychological mechanism of reduced inhibition, consistent with the GAM's situational-input account (Anderson & Bushman, 2002). Notably, the indirect effect remained significant even at low dysregulation, indicating that structural features of online environments carry independent aggression risk across the user population.

H4 and H5: Emotional Dysregulation as Moderator

Emotional dysregulation significantly moderated the disinhibition-aggression relationship and produced a significant index of moderated mediation. The indirect effect at high dysregulation (0.10) was more than three times that at low dysregulation (0.03), a practically meaningful difference. These findings support the GAM's distinction between controlled-deliberative and automatic-impulsive appraisal pathways (Anderson & Bushman, 2002): when regulatory capacity is compromised, disinhibitory conditions more reliably activate impulsive aggressive responding. This is consistent with the Self-Regulatory Resource Model (Muraven & Baumeister, 2000) and with Robertson et al.'s (2012) identification of emotion regulation difficulties as predictors of reactive aggression.

Theoretical and Practical Implications

Theoretically, these findings demonstrate the utility of integrating the Online Disinhibition Effect with the GAM: neither framework independently accounts for who is most at risk from disinhibitory conditions. Practically, the moderation finding identifies emotional dysregulation as an intervention target. Emotion-focused approaches with established efficacy including dialectical behaviour therapy skills training and cognitive-behavioural strategies for affect regulation (Gratz & Roemer, 2004) may reduce online aggression risk among high-dysregulation young adults. Universities should consider embedding such skills in student welfare programmes. Platform designers should note that perceived anonymity carries indirect aggression risk and that design choices enhancing accountability and identity continuity may attenuate disinhibitory effects. Policymakers should support regulatory frameworks that combine platform accountability standards with educational interventions targeting emotional competencies.

Limitations

Three limitations should be noted. First, the cross-sectional design precludes causal inference;

longitudinal or experimental designs are needed to establish temporal precedence. Second, reliance on self-report raises common method variance concerns; future research should incorporate behavioural or platform-level measures. Third, convenience sampling from university students limits generalisability to non-student and non-urban populations.

Future Directions

Future research should employ experimental paradigms in which anonymity is manipulated and participants are pre-classified on dysregulation, providing stronger causal tests. Ecological momentary assessment designs capturing real-time regulatory processes and online behaviour would also strengthen causal inference. Examining which specific DERS-16 dimensions carry the greatest moderating effect would inform targeted intervention design.

Conclusion

Online disinhibition mediates the relationship between perceived anonymity and online aggression, and this indirect pathway is substantially amplified among individuals with elevated emotional dysregulation. These findings extend cyberpsychology theory by integrating the Online Disinhibition Effect with the GAM and identify emotional dysregulation as a clinically and practically relevant moderator of online aggression risk in a Pakistani young adult sample.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.
- Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, 53(1), 27-51. <https://doi.org/10.1146/annurev.psych.53.100901.135231>

- Bjureberg, J., Ljótsson, B., Tull, M. T., Hedman, E., Sahlin, H., Lundh, L. G., Bjärehed, J., DiLillo, D., Messman-Moore, T., Gumpert, C. H., & Gratz, K. L. (2016). Development and validation of a brief version of the Difficulties in Emotion Regulation Scale: The DERS-16. *Journal of Psychopathology and Behavioral Assessment*, 38(2), 284–296. <https://doi.org/10.1007/s10862-015-9514-x>
- Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, 67, 97–102. <https://doi.org/10.1016/j.paid.2014.01.016>
- Christopherson, K. M. (2007). The positive and negative implications of anonymity in internet social interactions. *Computers in Human Behavior*, 23(6), 3038–3056. <https://doi.org/10.1016/j.chb.2006.09.001>
- Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, 26(1), 41–54. <https://doi.org/10.1023/B:JOBA.0000007455.08539.94>
- Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, 50(1), 1–22. <https://doi.org/10.1080/00273171.2014.962683>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis* (2nd ed.). Guilford Press.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4), 1073–1137. <https://doi.org/10.1037/a0035618>
- Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, 28(2), 434–443. <https://doi.org/10.1016/j.chb.2011.10.014>
- Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126(2), 247–259. <https://doi.org/10.1037/0033-2909.126.2.247>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891. <https://doi.org/10.3758/BRM.40.3.879>
- Robertson, T., Daffern, M., & Bucks, R. S. (2012). Emotion regulation and aggression. *Aggression and Violent Behavior*, 17(1), 72–82. <https://doi.org/10.1016/j.avb.2011.09.006>
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>
- Udris, R. (2014). Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale. *Computers in Human Behavior*, 41, 253–261. <https://doi.org/10.1016/j.chb.2014.09.036>

Willard, N. E. (2007). *Cyberbullying and cyberthreats: Responding to the challenge of online social aggression, threats, and distress*. Research Press.

