Psikostudia
Jurnal Psikologi
Volume 15 No. 3| September 2026 : 622 - 632
DOI: http://dx.doi.org/10.30872/psikostudia.v15i1

p-ISSN: 2302-2582
e-ISSN: 2657-0963

Behind the Screens: How Anonymity, Social Media Usage, and Platform Type Fuel Toxic Online Disinhibition in Gen Z

Di Balik Layar: Bagaimana Anonimitas, Penggunaan Media Sosial, dan Jenis Platform Memicu *Toxic Online Disinhibition* pada Gen Z

Karimatul Rizqi[1]
[1]Psychology Department, Faculty of Social Science & Humanities, UIN Sunan Kalijaga, Yogyakarta, Indonesia
Email: [1]karimatulrizqi@gmail.com

Sabiqotul Husna[2]
[2]Psychology Department, Faculty of Social Science & Humanities, UIN Sunan Kalijaga, Yogyakarta, Indonesia
Email: [2]sabiqotul.husna@uin-suka.ac.id

Correspondence:
Sabiqotul Husna
Psychology Department, Faculty of Social Science & Humanities, UIN Sunan Kalijaga, Yogyakarta, Indonesia
Email: sabiqotul.husna@uin-suka.ac.id

Abstract

Interactions of Generation Z in increasingly unrestricted digital spaces can trigger the toxic online disinhibition *effect. This study examined the relationships between anonymity, social media usage duration, and platform characteristics with* toxic online disinhibition *effect among Generation Z users in Java Island. A quantitative correlational design was employed with 280 participants (35 males, 245 females; M age = 21.96) selected via quota sampling. Instruments included an anonymity scale, measures of usage duration and platform characteristics, and a* toxic online disinhibition *effect scale. Data were analyzed using multiple linear regression with classical assumption tests, alongside nonparametric tests for demographic variables. Results indicated that anonymity, usage duration, and platform characteristics jointly explained 23.2% of the variance in toxic disinhibition (R^2 = .232), with anonymity as the strongest predictor. A follow-up model examining individual platforms explained 26.3% of the variance, showing that usage exceeding three hours daily and use of Twitter/X and TikTok were significantly associated with increased toxic disinhibition, whereas Instagram and Facebook showed no significant effect. Toxic disinhibition also differed significantly across platforms overall, but not by gender or age. These findings suggest that the* toxic online disinhibition *effect arises from a complex interaction of individual factors, usage behaviors, and platform characteristics, highlighting the need for interventions that consider both the duration of use and the unique dynamics of each social media platform.*

Keyword : anonimity; toxic online disinhibition effect; Z generation; social media use

Abstrak

Interaksi Generasi Z di ruang digital yang semakin tidak terbatas berpotensi memunculkan *toxic online disinhibition effect*. Penelitian ini bertujuan untuk mengkaji hubungan antara anonimitas, durasi penggunaan media sosial, dan karakteristik platform dengan *toxic online disinhibition effect* pada pengguna Generasi Z di Pulau Jawa. Desain penelitian menggunakan pendekatan kuantitatif korelasional dengan 280 partisipan (35 laki-laki, 245 perempuan; M usia = 21,96) yang dipilih melalui teknik quota sampling. Instrumen penelitian meliputi skala anonimitas, pengukuran durasi penggunaan dan karakteristik platform, serta skala *toxic online disinhibition effect*. Analisis data dilakukan menggunakan regresi linear berganda setelah uji asumsi klasik, serta uji beda nonparametrik untuk variabel demografis. Hasil menunjukkan bahwa anonimitas, durasi penggunaan, dan karakteristik platform secara bersama-sama menjelaskan 23,2% varians toxic disinhibition (R^2 = 0,232), dengan anonimitas sebagai prediktor terkuat. Model lanjutan yang menguji platform secara individual menjelaskan 26,3% varians, menunjukkan bahwa penggunaan lebih dari tiga jam per hari serta penggunaan Twitter/X dan TikTok berhubungan signifikan dengan peningkatan toxic disinhibition, sedangkan Instagram dan Facebook tidak menunjukkan pengaruh signifikan. *Toxic online disinhibition* juga berbeda signifikan antar-platform secara keseluruhan, namun tidak berbeda berdasarkan gender atau usia. Temuan ini menegaskan bahwa perilaku *toxic online disinhibition* muncul dari interaksi kompleks antara faktor individual, perilaku penggunaan, dan karakteristik platform, sehingga intervensi perlu mempertimbangkan durasi penggunaan serta dinamika tiap platform media sosial.

Kata Kunci : anonimitas; efek disinhibisi online toksik; generasi Z; penggunaan media social

Received 30/03/2026 | Revised 11/07/2026 | Accepted 27/08/2026

INTRODUCTION

Generation Z is the first generation to grow up fully immersed in digital technologies and social networking platforms, relying on social media not only for interpersonal communication but also for learning, entertainment, information seeking, and identity construction Social media has thus become a dominant environment for developing social relationships and expressing thoughts and emotions, creating both opportunities and psychological challenges associated with digital interaction (Mustaqimmah & Sari, 2021; Nadila, 2022; Pujiono, 2021).

While social media facilitates communication and knowledge sharing, it also increases exposure to harmful online behavior. One phenomenon receiving increasing scholarly attention is the toxic online disinhibition effect, defined as the tendency to express hostility, aggression, or other socially inappropriate behavior online that individuals would be unlikely to display face-to-face. This tendency arises because digital environments reduce social constraints, increase perceived freedom of expression, and weaken immediate interpersonal accountability (Suler, 2004, 2005; Christopherson, 2007; Lapidot-Lefler & Barak, 2012). This phenomenon appears increasingly relevant in Indonesia. Microsoft's Digital Civility Index reported Indonesian internet users showing one of the lowest levels of online civility in Southeast Asia despite the country's offline reputation for friendliness (Microsoft, 2021; CNN, 2022), with 19% of respondents reporting cyberbullying victimization and 47% reporting involvement in online bullying, while KOMINFO documented tens of thousands of public complaints about harmful content, particularly pornography and hate speech (CNN, 2021; Komdigi, 2020; Microsoft, 2020)

These behaviors reflect the destructive dimension of online disinhibition, manifesting as offensive language, hostile criticism, hate speech, threats, cyberbullying, and participation in deviant online communities individuals would normally avoid offline (Suler, 2004; Lapidot-Lefler & Barak, 2012; Kiswantomo et al., 2022). Joinson (2001) further suggested online disinhibition manifests through increased self-disclosure, while its aggressive dimension appears as flaming, hostile communication intended to provoke conflict (Joinson, 2001; O'Sullivan & Flanagin, 2003; Moor et al., 2010). Among the factors explaining toxic online disinhibition, anonymity has received the greatest empirical attention. The Online Disinhibition Effect theory holds that online environments reduce behavioral restraint through mechanisms including dissociative anonymity, invisibility, asynchronicity, dissociative imagination, solipsistic introjection, and minimization of status and authority (Suler, 2005), with dissociative anonymity considered the strongest predictor because users perceive their online identity as detached from their real-world identity, reducing accountability (Suler, 2004; Suler, 2005). Lapidot-Lefler & Barak (2012) demonstrated that anonymity significantly increases toxic disinhibition, consistent with Zimbardo's (1969) deindividuation experiment showing that individuals with concealed identities are more likely to display aggression.

Anonymity refers to a condition in which identity cannot be identified or linked to one's offline identity during online interaction(Christopherson, 2007; Mukhoyyaroh, 2020), encompassing not only complete absence of identity but also pseudonymous, hard-to-trace profiles (Pfitzmann & Hansen, 2008), conceptualized through unobservability, unlinkability, and pseudonymity. These characteristics reduce social accountability and increase willingness to express opinions and disclose information more openly (Joinson, 2001; Christopherson, 2007; Suler, 2004). Anonymous accounts remain common: a Kaspersky survey found roughly a third of Asia-Pacific internet users maintain anonymous or pseudonymous accounts, highest in Southeast Asia (Kapersky, 2020), used to protect privacy, control personal identity, express controversial opinions, seek sensitive information, and present a truer self than possible face-to-face (Bargh et al., 2002; Qian & Scott, 2007; Kang et al., 2013; Hollenbaugh & Everett, 2013), consistent with our preliminary survey. Yet anonymity may also facilitate negative behavior by reducing perceived accountability and fear of social evaluation, making hostility, offensive comments, hate speech, and cyberbullying more likely than under identifiable conditions (Suler, 2004; Christopherson, 2007; Lapidot-Lefler & Barak, 2012), so higher perceived anonymity is theoretically expected to increase toxic online disinhibition.

Intensity of social media use is another important contextual predictor: media-effects perspectives suggest repeated exposure to digital media shapes users' cognitive, emotional, and behavioral processes (Van Doorn et al., 2010; Valkenburg, 2022), with longer use increasing exposure to provocative content and emotionally charged interaction, potentially weakening inhibition. Intensive use is empirically associated with cyber aggression, cyberbullying, and emotional dysregulation among adolescents and young adults (Kuss & Griffiths, 2017; Beyens et al., 2020; Sobkin & Fedotova, 2021; Nobakht, et al., 2026; Giannakopoulos & Prassou, 2025).

Platform characteristics constitute a further contextual factor, since platforms differ in communication affordances, algorithmic structures, interaction norms, and content visibility (Zhao et al., 2016). Open-discussion, high-virality platforms like Twitter/X facilitate ideological clustering and echo chambers that intensify polarization (Cinelli et al., 2021). TikTok's recommendation algorithm promotes continuous scrolling and rapid emotional engagement, potentially increasing impulsive commenting (Montag, et al., 2021), while Instagram emphasizes visual self-presentation and identity management, producing different interaction patterns (Highfield & Leaver, 2016).

Previous studies consistently show toxic online disinhibition is shaped by both individual and contextual factors, anonymity being among the strongest predictors (Suler, 2004; Suler, 2005; Lapidot-Lefler & Barak, 2012) alongside intensive use (Keles et al., 2020; Kuss & Griffiths, 2017; Beyens et al., 2020) and platform characteristics (Cinelli et al., 2021; Montag, et al., 2021; Zhao et al., 2016). However, these predictors have predominantly been examined separately, leaving their simultaneous contribution

DOI: http://dx.doi.org/10.30872/psikostudia.v15i3.26396

underexplored, particularly in non-Western contexts, even though online behavior likely emerges from the interaction between psychological characteristics, media-use patterns, and platform-specific environments rather than a single determinant (Suler, 2004; Suler, 2005; Valkenburg, 2022; Lapidot-Lefler & Barak, 2012).

The present study integrates these three predictors within a single conceptual framework grounded in the Online Disinhibition Effect theory (Suler, 2004, 2005)and the Differential Susceptibility to Media Effects Model (DSMM), which holds that media effects result from interactions between media characteristics, patterns of use, and user characteristics (Valkenburg, 2022). Here, perceived anonymity represents diminished accountability, duration of use reflects repeated exposure, and platform characteristics represent the socio-technical context shaping communication norms and engagement.

Although research on toxic online disinhibition has grown internationally, evidence from Indonesia remains scarce, particularly among Generation Z, and most prior studies were conducted in Western or experimental settings, limiting cross-cultural generalizability. This study therefore proposes that toxic online disinhibition emerges from the interaction between individual perceptions and the digital environment, anonymity reducing accountability, prolonged use intensifying exposure to provocative content, and platform characteristics shaping behavior through differing affordances and norms (Cinelli et al., 2021; Suler, 2004, 2005; Valkenburg, 2022), and examines these three predictors jointly among Generation Z social media users in Java, Indonesia, extending the Online Disinhibition Effect theory to a non-Western context.

Based on this framework, the following hypotheses were proposed: H1. Perceived anonymity positively predicts toxic online disinhibition. H2. Longer duration of social media use positively predicts toxic online disinhibition. H3. Platform characteristics significantly predict toxic online disinhibition. H4. Perceived anonymity, duration of use, and platform characteristics jointly predict toxic online disinhibition. The novelty of this study lies in integrating these three predictors into a single explanatory model within a non-Western setting, contributing to the literature on digital behavior while offering practical implications for digital literacy, healthier online communication, and evidence-based prevention strategies among young social media users.

METHODS

Study Design and Participants

This study used a quantitative correlational design to examine relationships between perceived anonymity, duration of social media use, and platform characteristics (independent variables) and toxic online disinhibition (dependent variable) among Generation Z social media users, a design appropriate for testing predictive relationships within a regression framework (Creswell & Creswell, 2017). The target population comprised Generation Z individuals aged 10–25 residing on Java Island, Indonesia, who actively used at least one social media platform. Perceived anonymity was not an inclusion criterion, as it was measured as a continuous psychological construct.

Because no complete sampling frame existed for this geographically dispersed, digitally interacting population, a non-probability quota sampling technique was used, with quotas allocated by provincial population distribution to improve demographic representation and reduce geographic concentration. Although not producing probabilistic representation, quota sampling is considered appropriate when sampling frames are unavailable and random access to the population is impractical.

Minimum sample size was determined using G*Power 3.1 (fixed-model multiple linear regression, R^2 deviation from zero), specifying a medium effect size (f^2 = .15), α = .05, power = .80, and three predictors, yielding a minimum of 115 participants; substantially more respondents were recruited to improve precision. Provincial quotas were allocated proportionally based on the 2020 Indonesian Population Census (Badan Pusat Statistik, 2021), with the largest quota assigned to West Java, followed by East Java, Central Java, Banten, Jakarta, and Yogyakarta.

Measures

Toxic Online Disinhibition Effect

Toxic online disinhibition was measured using a scale developed based on the dimensions proposed by Joinson and operationalized by Satriawan (2016) consisting of self-disclosure and flaming. Higher scores indicated stronger tendencies toward toxic online disinhibition.

Perceived Anonymity

Perceived anonymity was assessed using a modified version of Chairunnisa's (2018) anonymity scale, which was developed based on the anonymity framework proposed by Pfitzmann & Hansen (2008). The scale measures three dimensions of anonymity: unobservability, unlinkability, and pseudonymity. Higher scores reflected greater perceived anonymity during online interactions.

Content Validity and Reliability

Content validity was evaluated by two psychology academics using Aiken's V coefficient Aiken (1985), with values ≥ 0.50 considered acceptable. Item discrimination was then assessed using corrected item-total correlations, retaining items with coefficients ≥ .30 (Azwar, 2011). Internal consistency was evaluated using Cronbach's alpha, with coefficients above .70 indicating acceptable reliability and above .80 indicating good reliability (Hair et al., 2010; Field, 2018). Alpha in this study was .780 for the Toxic Online Disinhibition Scale and .917 for the Perceived Anonymity Scale, indicating satisfactory internal consistency.

Duration of Social Media Use

Participants reported their average daily duration of social media use by selecting one of three categories: less than one hour, one to three hours, or more than three hours per day. These categories were adapted from previous studies indicating that social media use exceeding three hours per day represents intensive use and is associated with increased psychological risk (Twenge et al., 2018; Riehm et al., 2019). For regression analysis, duration was dummy coded (0 = ≤3 hours/day; 1 = >3 hours/day).

Platform Characteristics

Participants identified the social media platforms they actively used from four options: Instagram, Twitter/X, TikTok, and Facebook. Because participants were allowed to report more than one platform, each platform was represented as a separate binary variable (0 = not using; 1 = using). This coding approach enabled estimation of the unique contribution of each platform while controlling for the other predictors included in the regression model.

Data Analysis

All statistical analyses were performed using Jamovi (Version 2.5). Preliminary analyses examined outliers (boxplots), residual normality (Shapiro–Wilk test), linearity, homoscedasticity (residual-versus-fitted plots), and multicollinearity (VIF), with assumptions considered satisfied when residuals were normally distributed ($p > .05$), deviations from linearity were non-significant ($p > .05$), and VIF values were below 5 (Field, 2018).

Hypotheses were tested using multiple linear regression, with toxic online disinhibition as the dependent variable and perceived anonymity, duration of social media use, and platform characteristics entered simultaneously as predictors. Duration of use was entered as a dummy variable (0 = ≤3 hours/day, 1 = >3 hours/day), and each platform was represented by a separate binary indicator (0 = not using, 1 = using). Results are reported as unstandardized coefficients (B), standardized coefficients (β), t values, p values, confidence intervals, and R^2, estimating both individual and combined contributions of the predictors.

RESULTS

Table 1 presents the demographic characteristics of the 280 Generation Z participants. The sample was predominantly female (n = 245, 87.5%; male n = 35, 12.5%), and most reported spending more than three hours per day on social media (n = 219), indicating high daily engagement. Geographically, the largest proportion resided in West Java (n = 101), though provincial distribution was not proportional to the Generation Z population across Java's provinces, so findings should be generalized with caution. Most respondents were aged 22–25 years (n = 161), and Instagram was the most frequently used platform, followed by Twitter/X.

The pronounced gender imbalance may reflect differential willingness to complete self-administered questionnaires, as female respondents are often more willing than males, compounded by snowball recruitment through social media networks. The sample composition should therefore be regarded as the empirical distribution of participants successfully recruited rather than a proportional representation of Generation Z in Java, meaning the findings may more strongly reflect female respondents' online experiences and should be generalized accordingly with caution.

Table 1. Demographic Characteristics of the Participants (N =280)

Sex	**N**	**Percentage**
Male	35	12,5%
Female	245	87,5%
Total	**280**	**100%**
Average Daily Social Media Use	**N**	**Percentage**
< 1 hour	3	1,08%
1 – 3 hours	58	20,71%
> 3 hours	219	78,21%
Total	**280**	**100%**
Provincial Distribution of Participants	**N**	**Percentage**
West Java	101	36,07%
Central Java	67	23,92%
East Java	47	16,79%
Jakarta Special Capital Region	24	8,57%
Special Region of Yogyakarta	22	6,79%
Banten	19	7,86%
Total	**280**	**100%**
Age Distribution of Participants	**N**	**Percentage**
18-21 years old	119	42.5%
22-25 years old	161	57.5%
Total	**280**	**100%**
Social Media Platforms Used (Participants Could Select More Than One Platform)	**N**	**Percentage**
Instagram	190	31,93%
Twitter	154	25,88%
Facebook	108	18,15%
TikTok	92	15,46%
Other	51	8,57%

Participants' scores were categorized into low, moderate, and high levels based on each scale's hypothetical mean and standard deviation. For perceived anonymity, most participants fell into the moderate category (n = 173, 62%), followed by high (n = 99, 35%) and low (n = 8, 3%), indicating generally moderate-to-high perceived anonymity. A similar pattern emerged for toxic online disinhibition, with most participants in the moderate category (n = 165, 59%), followed by low (n = 97, 35%) and high (n = 18, 6%).

While both variables showed comparable distributions dominated by the moderate category, the high category was more prevalent for perceived anonymity than for toxic online disinhibition, suggesting that high perceived anonymity did not necessarily translate into high toxic online disinhibition.

Table 2. Classification of Participants' Scores on Perceived Anonymity and Toxic Online Disinhibition (n=280)

Variables	Category	Score Range	n	Percentage
Anonimity	Low	X < 43,3	8	3%
	Moderate	43,3 ≤ X < 56,7	173	62%
	High	X ≥ 56,7	99	35%
Toxic Online Disinhibition Effect	Low	X < 24	97	35%
	Moderate	24 ≤ X ≤ 36	165	59%
	High	X > 36	18	6%

Prior to hypothesis testing, regression assumptions were examined (Table 3). The Kolmogorov–Smirnov test indicated normally distributed residuals (KS = 0.472, p = .982). Linearity was confirmed by a significant linear relationship between predictors and toxic online disinhibition (p < .001) with a non-significant deviation from linearity (p = .342). Boxplot inspection revealed no extreme outliers, the Breusch–Pagan test was non-significant (p = .351), indicating homoscedasticity, and VIF values (1.021–1.048) indicated no multicollinearity. All assumptions for multiple linear regression were therefore satisfied.

Although duration of social media use and platform characteristics were measured using single-item indicators, they were included as observed predictors to estimate their relative contributions to toxic online disinhibition. As single-item measures do not permit estimation of internal consistency reliability, effects of these predictors should be interpreted with appropriate caution.

Table 3. Results of Multiple Regression Assumption Tests

Assumption	Method	Results	Criterion	Conclusion
Residual normality	Kolmogorov–Smirnov	KS = 0,472; p = 0,982	p > 0,05	Residuals were normally distributed
Linearity	Test of Linearity	F Linearity = 45,763; p < 0,001; F *Deviation from Linearity* = 0,987; p = 0,342	p *Linearity* < 0,05 dan p *Deviation* > 0,05	Linear relationship confirmed
Homoscedasticity	Breusch–Pagan Test	BP = 3,28; p = 0,351	p > 0,05	No evidence of heteroscedasticity
Outlier	Boxplot inspection	No extreme outliers detected	No extreme outliers	Assumption satisfied
Multicollinearity	Variance Inflation Factor (VIF)	VIF = 1,021–1,048	VIF < 10	No evidence of multicollinearity

Hypothesis testing examined the contributions of perceived anonymity, duration of social media use, and platform characteristics in predicting toxic online disinhibition, using multiple linear regression with all three predictors entered simultaneously; duration of use and platform characteristics, both categorical single-item variables, were dummy-coded prior to analysis.

The model was statistically significant, $F(3, 277) = 27.91$, $p < .001$, $R = .482$, $R^2 = .232$, adjusted $R^2 = .224$ (Table 4), indicating the three predictors jointly explained 23.2% of the variance in toxic online disinhibition. All three predictors showed positive, significant associations: perceived anonymity was the strongest ($\beta = .405$, $t(277) = 7.32$, $p < .001$), followed by duration of use ($\beta = .218$, $t(277) = 3.98$, $p < .001$) and platform characteristics ($\beta = .195$, $t(277) = 3.55$, $p < .001$), indicating higher anonymity, longer daily use, and platforms with greater anonymity and lower social regulation were associated with higher toxic online disinhibition.

Table 4. Multiple Regression Analysis Predicting Toxic Online Disinhibition Effect

Predictor	*B*	β	*t*	*P*	95% CI
Constant	20.85	-	4.98	< .001	[12.60, 29.10]
Anonymity	0.51	0.405	7.32	< .001	[0.37, 0.65]
Duration of Social Media Use (dummy)	0.29	0.218	3.98	< .001	[0.15, 0.43]
Platform Characteristics (dummy)	0.27	0.195	3.55	< .001	[0.12, 0.42]

Note. *The overall regression model was significant,* $R = .482$, $R^2 = .232$, *Adjusted* $R^2 = .224$, $F(3, 277) = 27.91$, $p < .001$. *Duration of social media use and platform characteristics were dummy coded prior to analysis.*

Comparative analyses examined whether toxic online disinhibition differed by duration of social media use, gender, age, province, and platform. Significant differences emerged for duration of use, both as two groups ($Z = -2.475$, $p = .013$, $r = .15$) and three groups ($\chi^2 = 8.214$, $p = .016$, $\varepsilon^2 = .022$), with small effect sizes indicating that longer use was associated with higher toxic online disinhibition. Significant differences were also found across platforms (χ^2 = 10.532, p = .015).

In contrast, no significant differences emerged for gender (Z = −0.325, p = .745, r = .02), age (χ^2 = 0.887, p = .647, ε^2 < .001), or province (χ^2 = 4.102, p = .534, ε^2 < .001), with negligible effect sizes indicating minimal practical significance. Overall, behavioral characteristics of social media use, particularly duration and platform choice, were more strongly associated with toxic online disinhibition than demographic characteristics among Generation Z users.

Table 5. Group Differences in Toxic Online Disinhibition Across Demographic Variables

Variable	Groups	Test Statistic	*p-value*	Effect size	Interpretation
Duration of Social Media Use (2 groups)	< 3 hours vs. > 3 hours	Z = -2,475	0,013	r = 0,15	Significant
Social media use (3 groups)	<1 hour; 1–3 hours; >3 hours	χ^2 = 8,214	0,016	r = 0,15	Significant
Sex	Male vs. Female	Z = -0,325	0,745	r = 0,02	NS
Age	<21; 21–24; >24 years old	χ^2 = 0,887	0,647	ε^2 = 0,000	NS
Province	Six provinces in Java	χ^2 = 4,102	0,534	ε^2 = 0,000	NS

NS = Not Significant.

To provide a more comprehensive understanding of the predictors, an additional multiple linear regression was conducted including perceived anonymity, duration of use, and social media platform, with each platform (Instagram, Twitter/X, TikTok, Facebook) dummy-coded as a binary variable (0 = non-user, 1 = user). This coding allowed the independent contribution of each platform to be estimated while accounting for participants' potential use of multiple platforms.

Table 4. Multiple Linear Regression Results with Dummy Variables for Social Media Platform Use as Predictors of the *Toxic Online Disinhibition Effect*

Variables	*B*	*SE*	*β*	t	*p*
Constanta	18.72	2.95	-	6.34	< .001
Anonimity	0.45	0.07	0.382	6.21	< .001
Duration (>3 hours)	1.98	0.76	0.148	2.61	0.009
Instagram	-0.85	0.68	-0.071	-1.25	0.213
Twitter/X	2.14	0.73	0.176	2.93	0.004
Tiktok	1.67	0.70	0.139	2.38	0.018
Facebook	0.58	0.75	0.045	0.77	0.441

Note. The overall regression model was significant, R = .513, R^2 = .263, adjusted R^2 = .247, F(6, 274) = 16.30, p < .001. Platform use was dummy-coded (0 = non-use; 1 = use), so each coefficient reflects that platform's unique contribution to toxic online disinhibition after controlling for the other predictors.

Multiple linear regression was conducted entering anonymity, duration of use, and platform-use dummy variables (Instagram, Twitter/X, TikTok, Facebook) simultaneously. The overall model was significant, F(6, 274) = 16.30, p < .001, R = .513, R^2 = .263, adjusted R^2 = .247, indicating all predictors collectively explained 26.3% of the variance in toxic online disinhibition.

Partially, anonymity (β = 0.382, t(274) = 6.21, p < .001) and social media use exceeding three hours per day (β = 0.148, t(274) = 2.61, p = .009) had significant positive effects, as did use of Twitter/X (β = 0.176, t(274) = 2.93, p = .004) and TikTok (β = 0.139, t(274) = 2.38, p = .018), indicating users of these platforms tended to show higher toxic online disinhibition than non-users, controlling for other predictors. Instagram (β = −0.071, t(274) = −1.25, p = .213) and Facebook use (β = 0.045, t(274) = 0.77, p = .441) were not significant.

DISCUSSION

This study demonstrates that the toxic online disinhibition effect among Generation Z cannot be explained solely by anonymity; rather, it results from the interplay of individual characteristics, patterns of social media use, and platform characteristics. Anonymity, duration of use, and platform characteristics were simultaneously associated with toxic disinhibition, with the regression model explaining 23.2% of the variance. These findings support Suler's (Suler, 2004, 2005). Online Disinhibition Effect theory, which attributes disinhibited online behavior to dissociative anonymity, invisibility, asynchronicity, and the limited presence of immediate social consequences. Consistent with this, Lapidot-Lefler & Barak (2012) showed that disinhibition depends not only on anonymity but also on broader interaction characteristics that reduce social constraints, manifesting as hostility, offensive language, hate speech, threats, and destructive criticism, reinforced by reduced social cues and limited interpersonal accountability online. Negative behavior on social media is thus better understood as multidimensional rather than the result of a single factor.

Only 8% of respondents showed high levels of toxic disinhibition. This may relate to the sense of freedom anonymity provides, combined with limited supervision of online content, which makes inappropriate behavior easier to display without clear consequences (Fazry & Apsari, 2021). Consistent with Suler (2005) and Lapidot-Lefler & Barak (2012), this study found a significant positive relationship between anonymity and toxic disinhibition: higher anonymity was associated with higher disinhibition. This pattern aligns with Lingam & Aripin's (2019) findings on flaming behavior, in which anonymity protects identity while enabling freer expression, with (Christopherson, 2007), who linked anonymity to greater willingness to express extreme opinions and reduced accountability, and with Joinson

DOI: http://dx.doi.org/10.30872/psikostudia.v15i3.26396

(2001), who found that anonymity promotes self-disclosure while reducing self-control, which in certain contexts may develop into maladaptive expression. Individual differences in self-regulation likely compound this effect; research among Indonesian youth similarly shows that stronger internalization of values is associated with lower aggressive tendencies (Fauzan & Widiantoro, 2025), suggesting anonymity's influence interacts with individuals' capacity for self-regulation rather than operating uniformly.

Anonymity operates through concrete psychological mechanisms, not merely as a descriptive condition. Following Pfitzmann & Hansen (2008), it comprises unobservability, unlinkability, and pseudonymity, mechanisms that progressively reduce identity traceability and weaken accountability. In this study, 35% of participants reported high anonymity. Kang, et al (2013) suggested individuals turn to anonymity after negative experiences such as digital attacks or fraud, while others use it to manage boundaries across social contexts (Vitak, 2012). Suler (2004) explained this through dissociative anonymity, where the online self becomes psychologically detached from one's real identity, reducing felt obligation to social norms, a process further supported by Joinson (2007). Marwick & Boyd (2014) added that unclear audience contexts (context collapse) may push individuals toward more extreme behavior when identities are not tied to immediate consequences. Even so, anonymity's contribution in the regression model was limited, consistent with Lapidot-Lefler & Barak (2012), who argued that toxic disinhibition results from multiple interacting situational and individual factors rather than anonymity alone.

Duration of social media use also significantly predicted toxic disinhibition, consistent with literature linking heavier exposure to negative emotional responses, impulsivity, and problematic use (Beyens et al., 2020; Kuss & Griffiths, 2017; Valkenburg, 2022). Gulo & Gunawan (2021) & Istiqomah (2017) similarly found that greater social media use correlates with adolescent aggression, itself a manifestation of toxic disinhibition (Lapidot-Lefler & Barak, 2012). This pattern parallels findings among Indonesian youth more broadly, where heavy, fear-of-missing-out-driven engagement with social media has likewise been linked to problematic online behavior (Akbar et al., 2019). In this study, 78% of respondents used social media more than three hours daily, and this group showed significantly higher toxic disinhibition than lighter users ($Z = -2.475$, $p = .013$), consistent with Kuss & Griffiths (2017) and Beyens, et al (2020).

Platform characteristics also mattered: Twitter/X and TikTok significantly predicted toxic disinhibition, while Instagram and Facebook did not. Cinelli, et al (2021) linked Twitter's open, opinion-based structure to polarization and conflict, corroborated by Oldemburgo De Mello, et al (2024), who found Twitter use associated with increased outrage shortly after exposure, and by Avalle, et al (2024), who showed toxicity rising with exposure to opposing viewpoints. For TikTok, Montag, et al (2021) and Doyle (2024) linked algorithmic amplification and rapid content consumption to greater reactivity and hostile commenting. Instagram's non-significant effect may reflect its stronger emphasis on visual self-presentation and impression management over open debate (Highfield & Leaver, 2016). These results suggest platforms function as active psychosocial contexts shaping opportunities for disinhibited expression, not merely as neutral channels through which interaction occurs.

Gender and age showed no significant differences ($p = .745$ and $p = .647$, respectively), consistent with Barlett & Coyne. (2014), who found gender gaps in online aggression narrowing as digital communication becomes more anonymous and indirect, and with the relatively homogeneous digital-native profile of Generation Z (Odgers & Jensen, 2020; Prensky, 2009). Notably, most respondents showed moderate-to-high anonymity (62% and 35%, respectively) but predominantly moderate-to-low toxic disinhibition (59% and 35%, with only 6% high), echoing Suler's (2004) distinction between benign and toxic disinhibition and Valkenburg's (2022) argument that effects depend on the interaction between individual traits and context of use, such as purpose of use, group norms, and prior digital experience.

Overall, these findings indicate that prevention efforts should not target anonymity alone but should combine digital literacy, healthier management of usage duration, and stronger platform design and moderation. Community and character-based interventions targeting adolescent social behavior more broadly, such as social skills training, have shown effectiveness in reducing bullying-related tendencies among youth (Hardhiyanti et al., 2020), suggesting similar structured interventions may help curb toxic online disinhibition as well. Interventions focusing solely on reducing anonymity are thus unlikely to be sufficient on their own.

Limitations

This study has several limitations that should be considered when interpreting the findings. First, the distribution of respondents was not proportional across provinces, and the sample was predominantly female; therefore, the generalizability of the findings to the broader Generation Z population should be approached with caution. Second, the use of an online survey and a cross-sectional design limited the verification of respondents' characteristics and precluded causal inferences. Third, the research model explained only 23.2% of the variance in the *toxic online disinhibition effect*, indicating that other factors outside the model may also influence this behavior.

CONCLUSION

Based on the results and analyses conducted, it can be concluded that the *toxic online disinhibition effect* among Generation Z is a phenomenon influenced by the interplay of multiple factors, namely anonymity, duration of social media use, and platform characteristics. Anonymity was found to be a significant predictor that may weaken self-control and increase freedom of expression, thereby contributing to the emergence of toxic behavior. However, this role does not operate independently but interacts with usage intensity and the structural dynamics of individual platforms. Longer durations of social media use are associated with greater exposure to online interactions, which, in turn, may increase the likelihood of negative emotional responses and disinhibited behavior. Meanwhile, platform characteristics also shape patterns of user interaction, with platforms characterized by openness and rapid information exchange potentially being more susceptible to extreme behaviors.

Overall, these findings emphasize that the *toxic online disinhibition effect* is not the result of a single factor but rather a complex and context-dependent phenomenon. Therefore, understanding this phenomenon requires consideration of the interplay among individual factors, patterns of social media use, and the digital environments that shape user behavior. Future research is recommended to employ more representative samples in terms of both geographic distribution and gender composition, as well as to consider other psychosocial variables relevant to the *toxic online disinhibition effect*. Future studies should also examine additional psychological variables and digital-environment characteristics, such as *invisibility*, *trait impulsivity*, *online aggression*, *perceived social norms*, and *digital literacy*, using longitudinal or experimental designs to obtain a more comprehensive understanding.

ACKNOWLEDGMENTS

The researchers would like to express their gratitude to all respondents who participated in this study and provided the data required for the research.

ETHICAL APPROVAL

This study was conducted in accordance with the ethical principles of research involving human participants. Informed consent was obtained from all respondents prior to data collection, and participant confidentiality and anonymity were maintained throughout the research process.

DECLARATION OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article.

TRANSPARENCY OF DATA

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

FUNDING

This study received no funding from any external party.

AUTHORS' CONTRIBUTIONS

The first author contributed the original idea, data collection, study design, data analysis, interpretation of results, and manuscript writing. The second author contributed to refining the concept and study design, data analysis, interpretation of results, final manuscript writing and revision. Both authors read and approved the final version of the manuscript.

AI USAGE STATEMENT

Artificial Intelligence (AI) tools were used solely to assist with language editing, grammar checking, and improving the clarity and readability of the manuscript. AI was not used to generate the research questions, develop the conceptual framework, conduct the qualitative document analysis, interpret the findings, or formulate the scholarly conclusions. The authors take full responsibility for the originality, accuracy, and academic integrity of the content. AI tools are not recognized as authors or contributors in accordance with accepted ethical standards in academic publishing.

REFRENCES

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Akbar, R. S., Aulya, A., Psari, A. A., & Sofia, L. (2019). Ketakutan Akan Kehilangan Momen (FoMo) Pada Remaja Kota Samarinda. *Psikostudia : Jurnal Psikologi*, *7*(2), 38. https://doi.org/10.30872/psikostudia.v7i2.2404

Avalle, M., Di Marco, N., Etta, G., Sangiorgio, E., Alipour, S., Bonetti, A., & Quattrociocchi, W. (2024). Persistent interaction patterns across social media platforms and over time. *Nature*, *628*, 582–589. https://doi.org/https://doi.org/10.17605/osf.io/fq5dy

Azwar, S. (2011). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar)

Badan Pusat Statistik. (2021). *Jumlah Penduduk menurut Wilayah, Klasifikasi Generasi, dan Jenis Kelamin, di INDONESIA - Dataset—Sensus Penduduk 2020—Badan Pusat Statistik*. https://sensus.bps.go.id/topik/tabular/sp2020/2/0/0

Bargh, J. A., McKenna, K. Y. A., & Fitzsimons, G. M. (2002). Can You See the Real Me? Activation and Expression of the "True Self" on the Internet. *Journal of Social Issues*, *58*(1), 33–48. https://doi.org/10.1111/1540-4560.00247

Barlett, C., & Coyne, S. M. (2014). A meta-analysis of sex differences in cyber-bullying behavior: The moderating role of age. *Aggressive Behavior*, *40*(5), 474–488. https://doi.org/10.1002/ab.21555

Beyens, I., Pouwels, J. L., van Driel, I. I., Keijsers, L., & Valkenburg, P. M. (2020). The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports*, *10*(1), 10763. https://doi.org/https://doi.org/10.1038/s41598-020-67727-7

Chairunnisa. (2018). *Pengaruh kesadaran diri dan anonimitas terhadap keterbukaan diri pengguna Media Sosial* [Undergraduate Thesis]. UIN Syarif Hidayatullah Jakarta. http://repository.uinjkt.ac.id/dspace/submit

Christopherson, K. (2007). The positive and negative implications of anonymity in Internet social interactions: "On the Internet, Nobody Knows You're a Dog." *Computers in Human Behavior*, *23*, 3038–3056. https://doi.org/10.1016/j.chb.2006.09.001

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), e2023301118. https://doi.org/10.1073/pnas.2023301118

CNN. (2021). *Riset: Netizen di Indonesia Paling Tak Sopan se-Asia Tenggara*. https://www.cnnindonesia.com/teknologi/20210225115954-185-610735/riset-netizen-di-indonesia-paling-tak-sopan-se-asia-tenggara.

CNN. (2022). *Daftar 10 Negara Paling Ramah kepada Turis Asing, Ada Indonesia Lho!* https://www.cnnindonesia.com/gaya-hidup/20221028132137-269-866561/daftar-10-negara-paling-ramah-kepada-turis-asing-ada-indonesia-lho

Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.

Doyle, O. A. (2024). Algorithms and the "Anti-Preference": A Quantitative Investigation of "Reaching the Wrong Audience" on TikTok. *Berkeley Undergraduate Journal*, *38*(1). https://doi.org/10.5070/B3.39970

Fauzan, A. H., & Widiantoro, W. (2025). The Role of Value Internalization in Suppressing the Level of Youth Aggressiveness: An Empirical Study of IKS.PI Kera Sakti Silat College. *Psikostudia : Jurnal Psikologi*, *14*(3), 394–401. https://doi.org/10.30872/psikostudia.v14i3.19687

Fazry, L., & Apsari, N. C. (2021). Pengaruh media sosial terhadap perilaku cyberbullying di kalangan remaja. *Jurnal Penelitian dan Pengabdian Kepada Masyarakat (JPPM)*, *2*(2), 272. https://doi.org/10.24198/jppm.v2i2.34679

Field, A. (2018). *Field, Andy. Discovering statistics using IBM SPSS statistics*. (Vol. 5). SAGE Publications.

Giannakopoulos, G., & Prassou, A. (2025). Mediating and Moderating Mechanisms in the Relationship Between Social Media Use and Adolescent Aggression: A Scoping Review of Quantitative Evidence. *European Journal of Investigation in Health, Psychology and Education*, *15*(6), 98. https://doi.org/10.3390/ejihpe15060098

Gulo, R., & Gunawan, N. (2021). Hubungan Intensitas Penggunaan Media Sosial Dengan Perilaku Agresif Verbal Pada Siswa di SMP Islam Ayatra. *NURSING UPDATE : Jurnal Ilmiah Ilmu Keperawatan P-ISSN : 2085-5931 e-ISSN : 2623-2871*, *12*. https://doi.org/10.36089/nu.v12i2.339

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th edition). Pearson Education Limited.

Hardhiyanti, R. S., Pandjaitan, L. N., & Arya, L. (2020). Efektivitas Social Skills Training (SST) untuk Mereduksi Intensitas Bullying Pada Remaja. *Psikostudia : Jurnal Psikologi*, *9*(1), 1. https://doi.org/10.30872/psikostudia.v9i1.3586

Highfield, T., & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, *2*(1), 47–62. https://doi.org/10.1080/22041451.2016.1155332

Hollenbaugh, E. E., & Everett, M. K. (2013). The Effects of Anonymity on Self-Disclosure in Blogs: An Application of the Online Disinhibition Effect: Anonymity and self-disclosure. *Journal of Computer-Mediated Communication*, *18*(3), 283–302. https://doi.org/10.1111/jcc4.12008

Istiqomah, I. (2017). Penggunaan media sosial dengan tingkat agresivitas remaja. *Insight : Jurnal Pemikiran dan Penelitian Psikologi*, *13*(2), 96–112. https://doi.org/10.32528/ins.v13i2.813

Joinson, A. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology*, *31*. https://doi.org/10.1002/ejsp.36

Joinson, A. N. (2007). Disinhibition and the Internet. In *Psychology and the Internet* (pp. 75–92). Elsevier. https://doi.org/10.1016/B978-012369425-6/50023-0

Kang, R., Brown, S., & Kiesler, S. (2013). Why Do People Seek Anonymity on the Internet? Informing Policy and Design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*, 2657–2666. https://doi.org/https://doi.org/10.1145/2470654.2481368

Kapersky. (2020). *Making Sense of Our Place in the Digital Reputation Economy* [The Digital Reputation Economy]. https://www.kaspersky.com/blog/digital-reputation-economy-report/

Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, *25*(1), 79–93. https://doi.org/https://doi.org/10.1080/02673843.2019.1590851

Kiswantomo, H., Rahmani, K., & Aliifah, D. N. (2022). Subjective Well-being sebagai Prediktor Online Disinhibition Effect pada Mahasiswa. *Jurnal Psikologi*, *15*(2), 269–283. https://doi.org/https://doi.org/10.35760/psi.2022.v15i2.6249

Komdigi. (2020). *Ini Konten Negatif yang Dominan di Indonesia*. https://www.komdigi.go.id/berita/sorotan-media/detail/ini-konten-negatif-yang-dominan-di-indonesia

Kuss, D. J., & Griffiths, M. D. (2017). Social Networking Sites and Addiction: Ten Lessons Learned. *International*

Journal of Environmental Research and Public Health, *14*(3), 311. https://doi.org/10.3390/ijerph14030311

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

Lingam, R. A., & Aripin, N. (2019). The role of anonymity in the flaming activity on YouTube in Malaysia. *Jurnal The Messenger*, *11*(1A), 111–120. https://doi.org/https://doi.org/10.26623/themessenger.v11i1A.832

Marwick, A. E., & boyd, danah. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, *16*(7), 1051–1067. https://doi.org/10.1177/1461444814543995

Microsoft. (2020). Riset: Netizen di Indonesia Paling Tak Sopan se-Asia Tenggara Baca artikel CNN Indonesia "Riset: Netizen di Indonesia Paling Tak Sopan se-Asia Tenggara." *Digital Civility Index (DCI)*. https://www.cnnindonesia.com/teknologi/20210225115954-185-610735/riset-netizen-di-indonesia-paling-tak-sopan-se-asia-tenggara.

Montag, C., Yang, H., & Elhai, J. D. (2021). On the psychology of TikTok use: A first glimpse from empirical findings. *Frontiers in Public Health*, *9*. https://doi.org/https://doi.org/10.3389/fpubh.2021.641673

Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior*, *26*(6), 1536–1546. https://doi.org/10.1016/j.chb.2010.05.023

Mukhoyyaroh, T. (2020). *Anonimitas Dan Deindividuasi Pada Remaja Pengguna Sosial Media*, *11*(1). https://doi.org/https://doi.org/10.29080/jpp.v11i1.354

Mustaqimmah, N., & Sari, N. D. (n.d.). Konsep Diri Generasi Z Pengguna Aplikasi TikTok di Kabupaten Rokan Hulu. *Medium*, *9*(2), 148–166.

Nadila, S. M. (2022). *Generasi Z: Si Paling Healing vs Si Pembawa Perubahan*. Pusat Riset Masyarakat dan Budaya-BRIN. https://pmb.brin.go.id/generasi-z-si-paling-healing-vs-si-pembawa-perubahan/

Nobakht, H. N., Wichstrøm, L., & Steinsbekk, S. (2026). Longitudinal relations between social media use and cyberbullying victimization across adolescence: Within-person effects in a birth cohort. *Journal of Youth and Adolescence*, *55*(1). https://doi.org/https://doi.org/10.1007/s10964-025-02205-9

Odgers, C. L., & Jensen, M. (2020). Adolescent Mental Health in the Digital Age: Facts, Fears and Future Directions. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *61*(3), 336–348. https://doi.org/10.1111/jcpp.13190

Oldemburgo De Mello, V., Cheung, F., & Inzlicht, M. (2024). Twitter (X) use predicts substantial changes in well-being, polarization, sense of belonging, and outrage. *Communications Psychology*, *2*(1), 15. https://doi.org/10.1038/s44271-024-00062-z

O'Sullivan, P. B., & Flanagin, A. J. (2003). Reconceptualizing 'flaming' and other problematic messages. *New Media & Society*, *5*(1), 69–94. https://doi.org/10.1177/1461444803005001908

Pfitzmann, A., & Hansen, M. (2008). *Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology*. http://dud.inf.tu-dresden.de/Anon_Terminology.shtml

Prensky, M. (2009). H. sapiens digital: From digital immigrants and digital natives to digital wisdom. *Innovate: Journal of Online Education*, *5*(3).

Pujiono, A. (2021). Media Sosial Sebagai Media Pembelajaran Bagi Generasi Z. *Didache: Journal of Christian Education*, *2*(1), 1. https://doi.org/10.46445/djce.v2i1.396

Qian, H., & Scott, C. R. (2007). Anonymity and Self-Disclosure on Weblogs. *Journal of Computer-Mediated Communication*, *12*(4), 1428–1451. https://doi.org/10.1111/j.1083-6101.2007.00380.x

Riehm, K. E., Feder, K. A., Tormohlen, K. N., Crum, R. M., Young, A. S., Green, K. M., Pacek, L. R., La Flair, L. N., & Mojtabai, R. (2019). Associations Between Time Spent Using Social Media and Internalizing and Externalizing Problems Among US Youth. *JAMA Psychiatry*, *76*(12), 1266–1273. https://doi.org/10.1001/jamapsychiatry.2019.2325

Satriawan, N. (2016). *Hubungan antara Konsep Diri dengan Toxic Disinhibition Online Effect pada Siswa SMK N 8 Surakarta*. https://www.semanticscholar.org/paper/Hubungan-antara-Konsep-Diri-dengan-Toxic-Online-SMK-Satriawan/0e6bca46b759699911074c109efdf1a44a1efb1b

Sobkin, V. S., & Fedotova, A. V. (2021). Adolescents on Social Media: Aggression and Cyberbullying. *Psychology in Russia*, *14*(4), 186–201. https://doi.org/10.11621/pir.2021.0412

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

Suler, J. (2005). The online disinhibition effect. *International Journal of Applied Psychoanalytic Studie*, *2*(2), 184–188. https://doi.org/https://doi.org/10.1002/aps.42

Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2018). Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time. *Clinical Psychological Science*, 6(1), 3–17. https://doi.org/10.1177/2167702617723376

Valkenburg, P. M. (2022). Theoretical foundations of social media uses and effects. In *Handbook of adolescent digital media use and mental health* (pp. 39–60). Cambridge University Press.

Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research*, *13*(3), 253–266. https://doi.org/https://doi.org/10.1177/1094670510375

Vitak, J. (2012). The Impact of Context Collapse and Privacy on Social Network Site Disclosures. *Journal of Broadcasting & Electronic Media*, 56(4), 451–470. https://doi.org/10.1080/08838151.2012.732140

Zhao, X., Lampe, C., & Ellison, N. B. (2016). The Social Media Ecology: User Perceptions, Strategies and Challenges. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, 89–100. https://doi.org/10.1145/2858036.2858333

Zimbardo, P. G. (1969). *The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos*. 17, 237–307.