Northumbria Research Link

Northumbria
University
NEWCASTLE

Online Disinhibition: Conceptualization, Measurement, and Implications for Online Deviant Behavior

Abstract

Purpose – Online disinhibition is one of the key factors leading to the occurrence of cyberaggression, cyberbullying, and various forms of deviant behaviors in the online environment. To understand the composition of online disinhibition, this study aims to conceptualize online disinhibition and develop a measurement instrument for online disinhibition.

Design/methodology/approach – We followed a rigorous procedure to develop and validate the multidimensional instrument of online disinhibition in three phases: item generation, measurement development, and instrument testing.

Findings – We developed a 23-item online disinhibition scale and identified 6 key dimensions: dissociative anonymity, invisibility, asynchronicity, solipsistic introjections, dissociative imagination, and minimization of authority.

Practical implications – The online disinhibition instrument is an accessible and easily administered measure that can be used as a checklist for systems designers and administrators to evaluate the level of online disinhibition among users. It offers systems design information on how to prevent and combat online deviant behaviors on platforms.

Originality/value – This work provides a rich conceptualization of an online disinhibition instrument that can serve as a springboard for future work to understand online deviant behaviors. The newly developed measurement instrument of online disinhibition also adds to the repository of rigorous research scales in this area.

Keywords: Online disinhibition, online deviant behaviors, cyberaggression, cyberdeviance, cyberbullying, multidimensional constructs, instrument development and validation, dark side of technology use

Paper type: Research paper

1. Introduction

With more than 4 billion people connected to the Internet worldwide, information and communications technology (ICT) is increasingly integrated into our daily activities (Statista, 2020). However, the constellation of benefits offered by ICT has recently been overshadowed by numerous accounts of online deviant behaviors, such as cyberbullying, cyberharassment, cyberaggression, online trolling, and online gossip, reported by popular media and the academic literature (e.g., Chan *et al.*, 2019, Lowry *et al.*, 2019, Akhtar and Morrison, 2019, Lee and Yu, 2020). For instance, the Cyberbullying Research Center (2019) conducted a nationwide survey with 4,972 adolescents in the U.S. and found that 37% of the respondents had been cyberbullied and 11.5% had cyberbullied others. The Pew Research Center also reported that 58% of adults experienced social media harassment (Duggan, 2017). Furthermore, collective trolling and other forms of uninhibited crowd behaviors have become a major stream online, whereby people gang up and launch campaigns to disrupt and harass other users. For example, the subreddit /r/KotakuInAction, which has more than 96,000 Reddit subscribers, is the main hub for the #GamerGate movement, which launches systematic trolling campaigns against women (Massanari, 2017).

Previous studies on the Internet and the psychology literature have recognized *online disinhibition* as an important driver of people's positive and negative behaviors in online environments. Specifically, people often say or do things online that they normally would not say or do in the offline world (Joinson, 1998, Suler, 2004, Walther, 1996, Saunders, 2016). Researchers have described online disinhibition as a psychological state in which individuals feel more relaxed and willing to engage in certain behaviors in the online environment. Some individuals relish the relative freedom online to engage in intentional, antisocial, and provocative online deviant behaviors that aim at antagonizing and upsetting other online users (Sanfilippo *et al.*, 2017, Lowry *et al.*, 2016b, Wright *et al.*, 2019).

Although the concept of online disinhibition has been acknowledged in various disciplines for almost two decades, it is still conceptualized and measured in an unsystematic way. Without a consistent definition and conceptualization of online disinhibition, it is difficult to accumulate knowledge in this area. Without a reliable way to measure online disinhibition, it is also difficult to ascertain that the systems we design are healthy online environments that people are willing to interact with and use. Accordingly, this paper has three main research objectives: (1) to identify

the core dimensions that constitute online disinhibition, (2) to develop and validate an online disinhibition measurement instrument, and (3) to conduct an analysis of competing models.

In response to the call for more research on the dark side of technology use in the academic community (Salo *et al.*, 2018, Turel *et al.*, 2019, Venkatraman *et al.*, 2018), we hope that this research will contribute to the development of the information systems (IS) literature. The newly developed and validated measurement instrument of online disinhibition can be added to the repository of rigorous research instruments. Furthermore, systems designers can use the composition of online disinhibition to formulate design tactics that combat online deviant behaviors.

2. Online Disinhibition Research

Based on our literature review, researchers have defined and conceptualized online disinhibition in three main ways (Cheung *et al.*, 2016).

First, online disinhibition is a behavioral concept. This line of studies has focused on different types of disinhibited behaviors in the online environment. For instance, Suler (2004) identified two types of disinhibited online behaviors: benign disinhibition and toxic disinhibition. In the case of benign disinhibition, the online environment motivates individuals to share personal details about themselves and their emotions. These individuals use the Internet as a way to explore their inner selves, and their oversharing is marked by an intrinsic need to better understand existing or new emotions while solving interpersonal issues (e.g., Bareket-Bojmel and Shahar, 2011). In contrast, toxic disinhibition is characteristic of the modern "troll", and is illustrated by displays of rude or crude language, harsh commentary, "hate speech", and even threats that would be extremely rare in a face-to-face setting (e.g., Wachs *et al.*, 2019). Recently, Udris (2014) built on Suler (2004) conceptualization and developed an eleven-item instrument for online disinhibition with two mirror factors.

Second, online disinhibition has been defined as a psychological state. This line of studies has defined online disinhibition as a psychological state of online users (Schouten *et al.*, 2007). It has emphasized that users are less restrained online than offline, and exhibit certain behaviors that they would not normally display in an offline interaction (Wong *et al.*, 2018). Some researchers have attempted to operationalize this construct. For example, Schouten *et al.* (2007) developed a three-item scale that captures the feeling of being less restrained. Wright *et al.* (2019) adapted

Udris (2014) online disinhibition concept to develop a four-item scale and assessed adolescents' perception of disinhibition in their online interaction and engagement.

Third, online disinhibition has been conceptualized as a composition of Internet attributes. This line of studies has used Internet attributes to capture online disinhibition. For example, anonymity is the most frequently studied Internet attribute in online disinhibition research (e.g., Barak *et al.*, 2008, Görzig and Ólafsson, 2013, Hollenbaugh and Everett, 2013). Suler (2004) summarized six commonly studied Internet attributes of online disinhibition: dissociative anonymity, invisibility, asynchronicity, solipsistic introjections, dissociative imagination, and minimization of status and authority. Barak *et al.* (2008) suggested that these six attributes can interact and supplement each other, resulting in a more complex and amplified effect. This implies that online disinhibition is a multidimensional construct comprising of these six key dimensions. These six attributes have been studied in the context of online social networks (Andalibi *et al.*, 2018, Miller, 2015, Sharon and John, 2018) and online gaming (Gray, 2012, Chen and Wu, 2015). However, no validated measurement instrument has been developed for the six dimensions of the online disinhibition construct.

Overall, our review of the literature showed that there has been little empirical research on how online disinhibition should be measured. Over the last two decades, most online disinhibition studies have been exploratory and descriptive (Dunn, 2012). However, there is no well-established measurement instrument that can be used to measure online disinhibition. Given the increasing severity of online deviant behaviors, it is essential to develop an instrument that identifies the composition of online disinhibition to help designers combat uncivilized online behaviors.

3. Concept and Instrument Development

3.1. Conceptualization

Consistent with Suler (2004) work, we considered online disinhibition to be a composition of six dimensions: (1) dissociative anonymity (DA); (2) invisibility (IV); (3) asynchronicity (AS); (4) solipsistic introjections (SI); (5) dissociative imagination (DI); and (6) minimization of authority (MA). Table 1 presents the definitions of these six dimensions. As mentioned earlier, online disinhibition can be seen as a multidimensional construct (Wu *et al.*, 2017). In this work, we developed measurement items for the six dimensions of online disinhibition and identified their interrelationships and factor structure.

Table 1. Definition of the dimensions of online disinhibition	
Dimensions	**Definition**
Dissociative Anonymity	The degree to which an individual perceives that he/she can hide or change his/her true identity in the online environment.
Invisibility	The degree to which an individual perceives that he/she is not physically seen by others in the online environment.
Asynchronicity	The degree to which an individual perceives that the mode of communication enables delayed responses in the online environment.
Solipsistic Introjection	The degree to which an individual perceives a voice or an image of the other persons in his/her mind in online communication.
Dissociative Imagination	The degree to which an individual perceives the online environment as an imaginary world that has no connection to reality.
Minimization of Authority	The degree to which an individual perceives the absent or diminishing influence of real-life authorities in the online environment.

3.2. Instrument Development

There are various approaches to developing measurement instruments. We followed the classic approach proposed by Moore and Benbasat (1991). Constructing the online disinhibition scale was a longitudinal process (see Figure 1) with three main phases: (1) item generation, (2) measurement development, and (3) instrument testing.

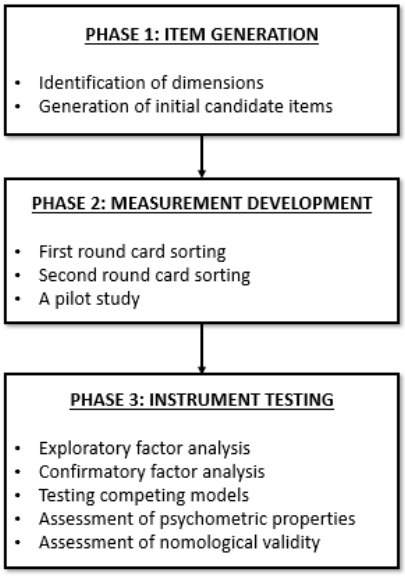

Figure 1. An overview of the instrument development and validation process

3.3. Phase 1: Item Generation

Following a two-step procedure (Belk, 2014), we collected items from the literature (e.g., Barak *et al.*, 2008, Görzig and Ólafsson, 2013, Hollenbaugh and Everett, 2013, Suler, 2004), then derived items from the definitions of the six identified dimensions. We initially developed 50 items to capture the most important aspects of each dimension.

3.4. Phase 2: Measurement Development

We first conducted two rounds of card sorting to ensure the content validity of the initial set of items. We then conducted a pilot test to refine these items.

3.4.1. Card Sorting Exercises. Two rounds of card sorting were conducted. The items were evaluated by a panel of experts who either had a Ph.D. degree or were Ph.D. candidates. Six panel members participated in each round of card sorting. Two measures, Cohen's kappa and item placement ratio were used to assess the reliability of the sorting procedures and the content validity of the scale (Cohen, 1960). In Round 1, Cohen's kappa ranged from 0.62 to 0.83. The overall placement ratio of the items within the target dimension was 84%. Based on the results, we dropped any ambiguous and confusing items. Thirty-eight items were retained for the second round of card sorting. The results of the second round were significantly better than those of the first round. Cohen's kappa ranged from 0.81 to 0.94. The overall placement ratio of the items within the target dimension was 94%, suggesting that the candidate items were satisfactory. In conclusion, the 38 items demonstrated sufficient content validity for the next stage of scale development testing.

3.4.2. Pilot Study. A self-administered questionnaire was distributed to 50 respondents for review and refinement in a pilot study. Hoehle and Venkatesh (2015) suggested that content validity raters should come from the main population of interest. Therefore, we recruited respondents from the Amazon Mechanical Turk (MTurk) system who are regular Internet users. We used this pilot study to purify the wording of the items and obtain initial feedback on the measurement instrument (Lewis *et al.*, 2005). Cronbach's alpha was calculated to assess the validity and reliability of the scale (see Table 2). The scale's reliability met conventional standards for internal consistency (Hair *et al.*, 2006), with a Cronbach's alpha greater than 0.70. In the pilot study, comments pertaining to the survey's functionality, navigability, and clarity were also collected. Specifically,

we asked the respondents to point out unclear items. Based on their feedback, we further refined the wording of the 38 items.

Table 2. Results of the pilot test				
Dimension	**Item**	**Cronbach's Alpha**	**Mean**	**S.D.**
Dissociative Anonymity	6	0.92	3.69	1.68
Invisibility	7	0.95	3.93	1.65
Asynchronicity	7	0.74	5.75	0.65
Solipsistic Introjection	6	0.79	5.12	1.02
Dissociative Imagination	5	0.87	2.90	1.43
Minimization of Authority	7	0.91	3.72	1.50

3.5. Phase 3: Instrument Testing

After the measurement instrument had been pretested and refined, we collected new data to re-examine the purified measures. Similar to the pilot test, we collected a sample from MTurk made up of regular Internet users. We believe that using MTurk for data collection was appropriate because samples collected from MTurk are generally more diversified (Matzler *et al.*, 2015) and the data collected are more reliable as those collected from online survey panels (Jia *et al.*, 2017, Lowry *et al.*, 2016a, Hauser and Schwarz, 2016). 430 usable questionnaires were obtained. Of the 430 respondents, 56% were males and 44% were females. Most of the respondents (43%) were aged 29 to 38, followed by 35% aged 17 to 28. The respondents spent more than 3 hours online per day. The measurement instrument was evaluated through both exploratory and confirmatory analyses (Tojib *et al.*, 2008). The two approaches served as a method of triangulation to ensure the development of a rigorous measurement instrument. As suggested by Bentler (1995), we randomly split the data into two data sets: 202 cases for exploratory analysis and 228 cases for confirmatory analysis.

3.5.1. Exploratory Analysis. The 202 cases in the first data set were used for exploratory analysis (exploratory factor analysis [EFA], Cronbach's alpha, and item-total correlation). To assess the validity of the scale, the 38 items were subjected to EFA. To improve the interpretation and obtain some theoretically meaningful factors, principal component analysis with EQUIMAX rotation was applied to these items, and a five-construct solution with a variance explained of 63.10% was obtained. Based on the EFA results (see Appendix A), we found that the items dissociative

anonymity (DA) and invisibility (IV) loaded on the same factor. When we revisited the literature, we found that Hollenbaugh and Everett (2013) proposed a concept called anonymity, which is a higher-order factor made up of dissociative anonymity (labeled discursive anonymity by Hollenbaugh and Everett (2013)) and invisibility (labeled visual anonymity). We also found that AS1 did not load well on its corresponding construct (i.e., asynchronicity). Item-total correlations and Cronbach's alpha were used to evaluate the reliability of the scale. As shown in Table 3, all Cronbach's alpha values exceeded 0.70. However, some items had low item-total correlations (less than 0.50), including AS1, AS2, AS3, and AS4.

Table 3. Psychometric properties

Second-order factor	First-order factor	Item	Item-to-total score correlation
Anonymity ($\alpha = 0.96$)	Dissociative Anonymity	DA1	0.69
		DA2	0.73
		DA3	0.83
		DA4	0.85
		DA5	0.75
		DA6	0.80
	Invisibility	IV1	0.75
		IV2	0.84
		IV3	0.74
		IV4	0.84
		IV5	0.86
		IV6	0.81
		IV7	0.85
	Asynchronicity ($\alpha = 0.77$)	AS1	0.28
		AS2	0.48
		AS3	0.41
		AS4	0.48
		AS5	0.54
		AS6	0.63
		AS7	0.65
	Solipsistic Introjection ($\alpha = 0.84$)	SI1	0.52
		SI2	0.61
		SI3	0.66
		SI4	0.68
		SI5	0.64
		SI6	0.61
	Dissociative Imagination ($\alpha = 0.88$)	DI1	0.75
		DI2	0.76
		DI3	0.84
		DI4	0.58
		DI5	0.63

	Minimization of Authority ($\alpha = 0.91$)	MA1	0.54
		MA2	0.81
		MA3	0.76
		MA4	0.74
		MA5	0.68
		MA6	0.76
		MA7	0.81

3.5.2. Confirmatory Analysis. The 228 cases in the second data set were subjected to confirmatory factor analysis (CFA). The maximum likelihood method was used to detect the unidimensionality of each factor, which indicates the presence of a single trait or construct in a set of items (Anderson and Gerbing, 1988). Following Byrne's (2001) approach, a measurement model with six first-order dimensions was developed. The six dimensions were correlated, with each item having a non-zero loading on its designated factors and a zero loading on other factors. The measurement error terms associated with the items were uncorrelated. Table 4 shows the model fit results for the initial and revised models. As expected, the initial model had a poor model fit. Only χ^2/df was within the recommended threshold, whereas the other fit indices were all below satisfactory levels, suggesting that the initial model was not adequate. Therefore, a respecification of the initial model was performed to identify any ill-fitting parameters and obtain a clearer factor structure. When respecifying the model, standardized factor loadings (Raghunathan *et al.*, 1999) and the modification index (Byrne, 2001) were considered. To avoid over-modification, one item at a time was dropped from the model (Jöreskog and Sörbom, 1988). At the end of the respecification process, 15 items were deleted (two items for dissociative anonymity, two items for invisibility, four items for asynchronicity, three items for solipsistic introjection, one item for dissociative imagination, and three items for minimization of authority). The measurement model with 23 items had satisfactory fit indices ($\chi^2/df = 1.80$; goodness of fit index [GFI] = 0.87; adjusted GFI [AGFI] = 0.83; root mean square residual [RMR] = 0.15; root mean square error of approximation [RMSEA] = 0.06; Tucker–Lewis index [TLI] = 0.94; normed fit index [NFI] = 0.90; comparative fit index [CFI] = 0.95) with a minimized Akaike information criterion (AIC) value (i.e., AIC = 523.56), suggesting that it was the most parsimonious model. The finalized scale is presented in Appendix B.

Table 4. Model fit test results of initial and revised models					
	Threshold	**Initial model (38 items)**	**First revised model (32 items)**	**Second revised model (28 items)**	**Third revised model (23 items)**
χ^2	Smaller is better	1614.51	1017.34	728.21	401.56
d.f.		650.00	449	335	215
p-value	$p > 0.05$	0.00	0.00	0.00	0.00
χ^2/d.f.	$1 < \chi^2/\text{d.f.} < 3$	2.48	2.27	2.17	1.87
GFI	> 0.90	0.72	0.77	0.81	0.87
AGFI	> 0.80	0.68	0.73	0.77	0.83
RMR	< 0.10	0.18	0.16	0.16	0.15
RMSEA	< 0.08	0.08	0.07	0.07	0.06
TLI	> 0.90	0.82	0.87	0.90	0.94
NFI	> 0.90	0.75	0.81	0.85	0.90
CFI	> 0.90	0.83	0.89	0.91	0.95
AIC	Smaller is better	1796.51	1175.34	870.21	523.56
CAIC	Smaller is better	2199.59	1525.26	1184.70	793.75

4. Testing Competing Models

We also tested the hypothesized model by analyzing competing models. Five possible alternative models were proposed. Figure 2 shows the representative items from each of these alternatives. In Model 1, the 23 items loaded on one single first-order factor accounting for all of the variance in online disinhibition. This model conceptualized online disinhibition as a first-order construct (Schouten *et al.*, 2007). In Model 2a, the 23 items loaded on six correlated factors following the conceptualization of Suler (2004). In Model 2b, the 23 items loaded on six uncorrelated factors, in which online disinhibition was represented by one of the six factors. To test the multifaceted nature of online disinhibition, Model 3 and Model 4, which were alternative composite latent variable models with different hierarchical structures, were proposed. In Model 3, the 23 items loaded on five correlated factors with one second-order construct based on the EFA results. This model tested the extent to which the correlation between the two first-order factors (dissociative anonymity and invisibility) was strengthened by the second-order factor, anonymity. In Model 4, the 23 items loaded on an alternative hierarchical structure with one second-order factor (anonymity) and four first-order factors (asynchronicity, solipsistic introjection, dissociative imagination, and minimization of authority).

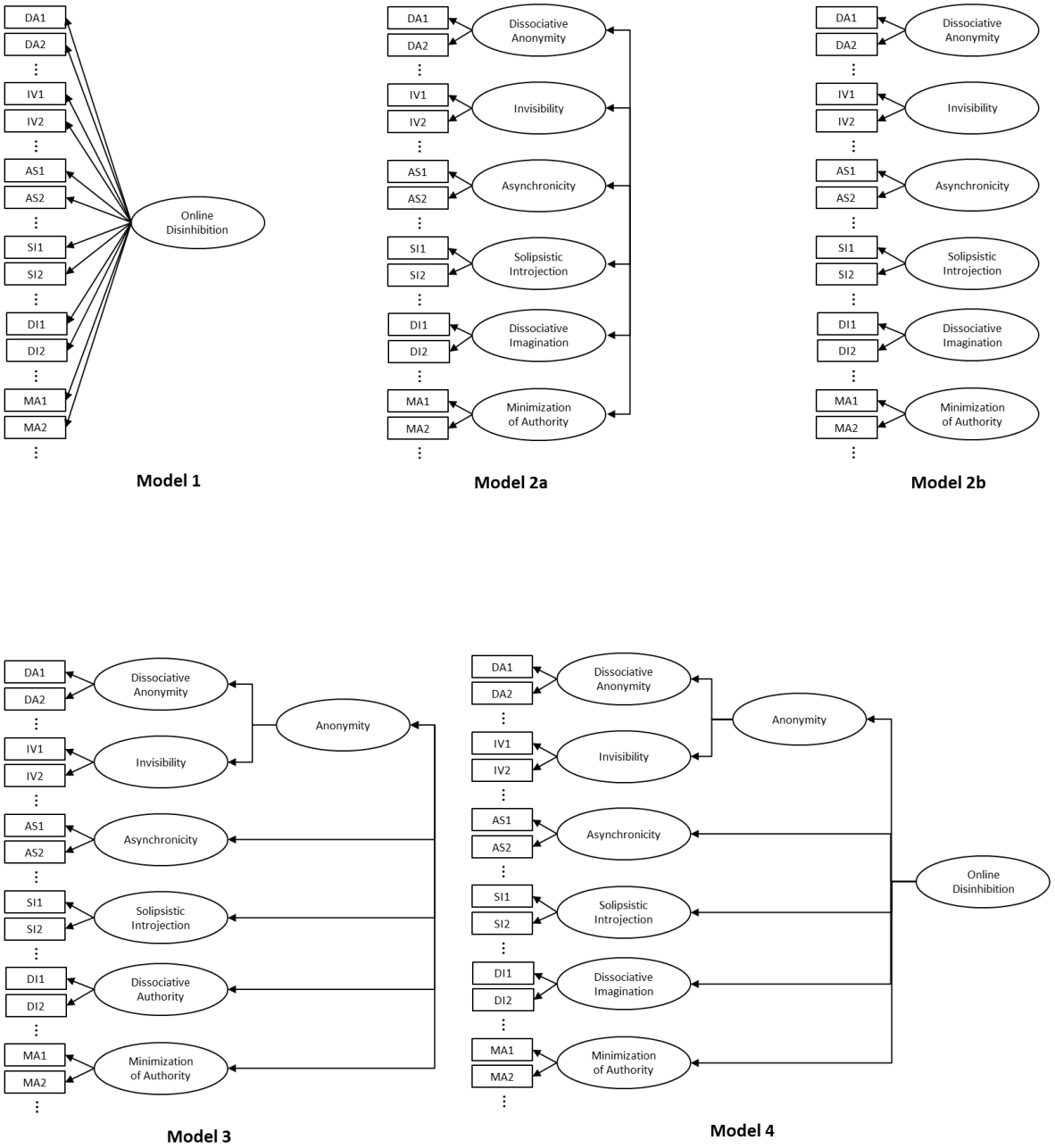

Figure 2. Alternative models of the online disinhibition scale

Table 5. Model fit test results of alternative models							
		Model					
	Threshold	**Null**	**1**	**2a**	**2b**	**3**	**4**
χ^2	Smaller is better	3866.39	1844.53	401.56	788.98	404.22	442.99
d.f.		253.00	230.00	215.00	230.00	219.00	223.00
p-value	$p > 0.05$	0.00	0.000	0.000	0.000	0.000	0.00
χ^2/d.f.	$1 < \chi^2/\text{d.f.} < 3$	15.28	8.02	1.87	3.43	1.85	1.99
GFI	> 0.90	0.26	0.52	0.87	0.79	0.87	0.85
AGFI	> 0.80	0.19	0.43	0.83	0.74	0.83	0.82
RMR	< 0.10	0.98	0.42	0.15	0.71	0.15	0.21
RMSEA	< 0.08	0.25	0.18	0.06	0.10	0.06	0.07
TLI	> 0.90	0.00	0.51	0.94	0.83	0.94	0.93
NFI	> 0.90	0.00	0.52	0.90	0.80	0.90	0.89
CFI	> 0.90	0.00	0.55	0.95	0.85	0.95	0.94
AIC	Smaller is better	3912.39	1936.53	523.56	880.98	518.22	548.29
CAIC	Smaller is better	4014.27	2140.28	793.75	1084.73	770.69	783.05

Table 5 summarizes the results of the model fit tests. The null model assumed that no latent factor was at the basis of the observed items and that the correlations between the items were zero in the population. The null model was included to establish a zero point for the NFI. As expected, poor fit indices were demonstrated in the null model. Models 1 and 2b demonstrated substantial improvements over the null model. However, none of these models reasonably fitted the empirical data. Model 2a showed substantial improvements over Model 1 and Model 2b. Model 2a had a reasonable fit, as indicated by the TLI, NFI, and CFI values, which were within the recommended thresholds. For Models 3 and 4, most of the fit index values were within the recommended thresholds, including AGFI, RMSEA, TLI, and CFI. Only GFI, RMR, and NFI had a marginal fit. Therefore, they were both considered adequate to represent the underlying factor structure of the online disinhibition scale.

When interpreting the results, it is important to note that higher-order factors simply explain the covariations between their corresponding first-order factors in a more parsimonious way (Stewart and Segars, 2002). As a result, even when higher-order models (i.e., Model 3 and Model 4) can explain these covariations, the goodness of fit can never surpass that of the original first-order model (i.e., Model 2a). As evidenced by the results in Table 5, the fit indices of Model 2a were slightly better than those of Model 3 and Model 4. However, the target coefficient (T)

suggested that more than 90% of the variation was explained by the two higher-order models, where $T_{Model\ 2a}$ to $T_{Model\ 3}$ = 0.99 and $T_{Model\ 2a}$ to $T_{Model\ 4}$ = 0.91, indicating the existence of higher-order models for the online disinhibition scale. In other words, the factor covariance in Model 2a can be represented in a more economical manner with higher-order models (such as Model 3 and Model 4). Therefore, Model 4 was a more accurate representation of online disinhibition and was accepted in preference to Model 2a. We thus used Model 4 to analyze the reliability and validity of the factors and items.

5. Psychometric Properties

After examining the overall model fit, the psychometric properties of the online disinhibition scale were examined.

5.1. Assessment of Convergent Validity

Convergent validity refers to the extent to which the items in a scale appear to be indicators of a single underlying construct. Convergent validity was evaluated using the following criteria (Fornell and Larcker, 1981, Hair *et al.*, 2006): (1) all factor loadings must exceed 0.70, (2) composite reliability (CR) should exceed 0.70, and (3) average variance extracted (AVE) should exceed 0.5. As shown in Table 6, all factor loadings were greater than 0.70 and all CR and AVE values reached the recommended levels, with CR between 0.84 and 0.96 and AVE ranging from 0.64 to 0.82.

Table 6: Assessment of convergent validity

	Item	Loading	CR	AVE
Dissociative Anonymity	DA3	0.88	0.93	0.78
	DA4	0.93		
	DA5	0.86		
	DA6	0.86		
Invisibility	IV2	0.91	0.96	0.82
	IV3	0.87		
	IV4	0.91		
	IV5	0.91		
	IV7	0.92		
Asynchronicity	AS5	0.71	0.90	0.75
	AS6	0.93		
	AS7	0.94		
Solipsistic Introjection	SI3	0.87	0.84	0.64
	SI4	0.78		
	SI5	0.75		

Dissociative Imagination	DI1	0.84	0.90	0.70
	DI2	0.87		
	DI3	0.86		
	DI4	0.77		
Minimization of Authority	MA3	0.78	0.92	0.75
	MA5	0.89		
	MA6	0.91		
	MA7	0.88		

5.2. Assessment of Discriminant Validity

Discriminant validity refers to the degree to which the measures of distinct constructs differ (Bagozzi and Phillips, 1982). It can be evaluated by measuring AVE (Fornell and Larcker, 1981). That is, each contruct's AVE should be greater than the square of the correlations between the constructs and all other constructs. As shown in Table 7, the square root of the AVE for each factor was higher than the correlations between that factor and other corresponding factors, demonstrating discriminant validity.

Table 7: Assessment of discriminant validity

	AVE	DA	IV	AS	SI	DI	MA
DA	0.78	**0.88**					
IV	0.82	0.83	**0.91**				
AS	0.75	0.18	0.23	**0.87**			
SI	0.64	0.19	0.21	0.21	**0.80**		
DI	0.70	0.27	0.28	-0.11	0.11	**0.84**	
MA	0.75	0.33	0.32	-0.09	0.18	0.49	**0.87**

Note 1: AVE = Average Variance Extracted; DA= Dissociative Anonymity; IV= Invisibility; AS= Asynchronicity; SI=Solipsistic Introjection; DI= Dissociative Imagination; MA= Minimization of Authority

Note 2: Bolded diagonal elements are the square roots of AVE for the constructs

5.3. Assessment of Nomological Validity

Nomological validity refers to the degree to which predictions based on the measured constructs are confirmed in a large theoretical context or network of constructs (Bagozzi, 1981). To assess the nomological validity of online disinhibition, we expected a positive relationship between online disinhibition and online harassment (Fox *et al.*, 2015) based on the literature (Udris, 2014, Wong *et al.*, 2015). Finn (2004) measured online harassment using a four-item scale, including items such as "I harass or bother someone online to make him/her feel worried or threatened," "I embarrass or humiliate someone online to make him/her feel bad," "I insult someone online to make him/her feel mad," and "I make rude or nasty comments to someone

online." A seven-point Likert-type scale (ranging from 1 = strongly disagree to 7 = strongly agree) was used. The CR of online harassment was 0.94, and its AVE was 0.65. The square root of the AVE was greater than all of its inter-construct correlations, demonstrating the reliability and validity of the scale. As expected, there was a strong positive relationship between online disinhibition and online harassment ($\beta = 0.25$, $\rho < 0.001$). The results corroborated prior research on online harassment, thus confirming the nomological validity of the online disinhibition scale.

6. Discussion

Despite its important role in online deviant behaviors, online disinhibition has suffered from both conceptual and operational problems in previous research, thereby limiting its usefulness to researchers and systems designers. To advance this area of research, we identified the core dimensions that constitute online disinhibition. We also developed and validated a measurement instrument for online disinhibition and conducted an analysis of competing models to confirm its dimensionality and multifaceted structure. Specifically, we built on Suler's conceptualization of online disinhibition and identified six key dimensions: (1) dissociative anonymity; (2) invisibility; (3) asynchronicity; (4) solipsistic introjections; (5) dissociative imagination; and (6) minimization of authority. Through a rigorous scale development procedure, we developed a 23-item online disinhibition scale. By analyzing competing models, Model 4 with a hierarchical structure was found to fit the data well and was most theoretically valid, reflecting logical consistency. This work has several implications for researchers and practitioners.

6.1. Research and Practical Implications

Our work has several implications for research and practice. First, we conceptually defined and rigorously operationalized online disinhibition. The findings provide a better understanding of the multidimensionality of the online disinhibition construct, contributing a new 23-item measurement instrument of the construct to the repository of rigorous research scales. Our comprehensive conceptualization of online disinhibition contributes to the IS literature on the dark side of technology use. Specifically, research on online deviant behaviors could use our conceptualization and instrument to investigate how the online environment allows users to engage in deviant behaviors, such as cyberbullying, cyberaggression, and online gossip.

Second, adopting a parsimonious approach, most studies have not sought to specifically explain how Internet attributes contribute to the online disinhibition construct. For instance, most studies have only examined the roles of anonymity and online deviant behaviors (Barlett *et al.*, 2016, Munger, 2017). We believe that our fine-grained measurement instrument can help researchers accurately understand Internet attributes and provide a better understanding of how Internet attributes other than anonymity affect online deviant behaviors.

Third, our conceptualization of and instrument for online disinhibition can be used as a springboard for future research. Researchers could replicate our work for some specific form of online deviant behavior or test the stability of the measure over time (Johns, 2006). For example, researchers could test the conceptualization and instrument of online deviant behaviors in different online platforms. The interrelationships between the dimensions of online disinhibition may vary depending on the characteristics of the platforms, resulting in different effects on various forms of deviant behaviors.

Finally, our findings have important implications for practitioners. The 23-item online disinhibition instrument is an accessible and easily administered measure. The items can serve as a checklist for practitioners, such as systems designers and administrators, to evaluate the level of online disinhibition among users. They can use the checklist to adjust systems design to prevent and combat online deviant behaviors on their platforms.

6.2. Limitations and Directions for Future Research

Although the newly developed measurement instrument of online disinhibition underwent rigorous development and validation procedures, this study is not without limitations. First, the study only measured online disinhibition in the general online environment. An obvious extension of this research would be to conduct replication studies in different online environments (e.g., e-mail, social networking sites, online forums). We expect this to give systems designers better insights into the characteristics of the system and the level of online disinhibition. In addition, this study only examined the association between online disinhibition and online harassment. A more systematic research framework should be used to guide the study of online disinhibition and online deviant behaviors. We believe that conceptualizing and operationalizing this important concept is the first step toward theory building in research on the dark side of technology use.

7. Conclusion

Online deviant behaviors have attracted increasing attention from both academics and practitioners. This work contributes to IS research by conceptualizing online disinhibition and by developing and validating a new instrument. The measurement instrument of online disinhibition has the potential to contribute to research on a wide range of online deviant behaviors. Systems designers can also use this measurement instrument to adjust their systems design tactics to combat online harassment and provide a healthy online atmosphere for their users.

References

Akhtar, S. and Morrison, C. M. (2019). "The prevalence and impact of online trolling of UK members of parliament," *Computers in Human Behavior,* Vol. 99, No., pp. 322-327.

Andalibi, N., Haimson, O. L., Choudhury, M. D. and Forte, A. (2018). "Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media," *ACM Transactions on Computer-Human Interaction (TOCHI),* Vol. 25, No. 5, pp. 1-35.

Anderson, J. C. and Gerbing, D. W. (1988). "Structural equation modeling in practice: A review and recommended two-step approach," *Psychological Bulletin,* Vol. 103, No. 3, pp. 411-423.

Bagozzi, R. P. (1981). "An examination of the validity of two models of attitude," *Multivariate Behavioral Research,* Vol. 16, No. 3, pp. 323-359.

Bagozzi, R. P. and Phillips, L. W. (1982). "Representing and testing organizational theories: A holistic construal," *Administrative Science Quarterly,* Vol. 27, No. 3, pp. 459-489.

Barak, A., Boniel-Nissim, M. and Suler, J. (2008). "Fostering empowerment in online support groups," *Computers in Human Behavior,* Vol. 24, No. 5, pp. 1867-1883.

Bareket-Bojmel, L. and Shahar, G. (2011). "Emotional and interpersonal consequences of self-disclosure in a lived, online interaction," *Journal of Social and Clinical Psychology,* Vol. 30, No. 7, pp. 732-759.

Barlett, C. P., Gentile, D. A. and Chew, C. (2016). "Predicting cyberbullying from anonymity," *Psychology of Popular Media Culture,* Vol. 5, No. 2, pp. 171-180.

Belk, R. (2014). "You are what you can access: Sharing and collaborative consumption online," *Journal of Business Research,* Vol. 67, No. 8, pp. 1595-1600.

Bentler, P. M. (1995), *EQS Structural Equations Program Manual, Multivariate*, Encino, CA: Software, Inc.

Byrne, B. (2001), *Structural equation modeling with AMOS: Basic concepts, applications, and programming* Mahwah, NJ: Lawrence Erlbaum.

Chan, T. C. H., Cheung, C. M. K. and Wong, R. Y. M. (2019). "Cyberbullying on social networking sites: The crime opportunity and affordance perspectives," *Journal of Management Information Systems*, Vol. 36, No. 2, pp. 574-609.

Chen, V. H. H. and Wu, Y. (2015). "Group identification as a mediator of the effect of players' anonymity on cheating in online games," *Behaviour & Information Technology,* Vol. 34, No. 7, pp. 658-667.

Cheung, C. M. K., Wong, R. Y. M. and Chan, T. K. H. (2016), "Online disinhibition: Conceptualization, measurement, and relation to aggressive behaviors," in *Thirty Seventh International Conference on Information Systems*, Dublin.

Cohen, J. (1960). "A coefficient of agreement for nominal scales.," *Educational and Psychosocial Measurement*, Vol. 20, No., pp. 37-46.

Cyberbullying Research Center (2019), "2019 Cyberbullying Data," available at: https://cyberbullying.org/2019-cyberbullying-data (accessed 28 August, 2020).

Duggan, M. (2017), "Online harassment," available at: http://www.pewinternet.org/2017/07/11/online-harassment-2017/ (accessed 11 August, 2020).

Dunn, K. (2012). "A qualitative investigation into the online counselling relationship: To meet or not to meet, that is the question," *Counselling and Psychotherapy Research*, Vol. 12, No. 4, pp. 316-326.

Finn, J. (2004). "A survey of online harassment at a university campus," *Journal of Interpersonal Violence*, Vol. 19, No. 4, pp. 468-483.

Fornell, C. and Larcker, D. F. (1981). "Evaluating structural equation models with unobservable variables and measurement error," *Journal of Marketing Research*, Vol. 18, No. 3, pp. 39-50.

Fox, J., Cruz, C. and Lee, J. Y. (2015). "Perpetuating online sexism offline: Anonymity, interactivity, and the effects of sexist hashtags on social media," *Computers in Human Behavior*, Vol. 52, No., pp. 436-442.

Görzig, A. and Ólafsson, K. (2013). "What makes a bully a cyberbully? Unravelling the characteristics of cyberbullies across twenty-five European countries," *Journal of Children and Media*, Vol. 7, No. 1, pp. 9-27.

Gray, K. L. (2012). "Intersecting oppressions and online communities: Examining the experiences of women of color in Xbox Live," *Information, Communication & Society*, Vol. 15, No. 3, pp. 411-428.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. and Tatham, R. L. (2006), *Multivariate data analysis*, Upper Saddle River, NJ: Pearson Prentice Hall.

Hauser, D. J. and Schwarz, N. (2016). "Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants," *Behavior Research Methods*, Vol. 48, No. 1, pp. 400-407.

Hoehle, H. and Venkatesh, V. (2015). "Mobile application usability: Conceptualization and instrument development," *MIS Quarterly*, Vol. 39, No. 2, pp. 435-472.

Hollenbaugh, E. E. and Everett, M. K. (2013). "The effects of anonymity on self-disclosure in blogs: An application of the online disinhibition effect," *Journal of Computer-Mediated Communication*, Vol. 18, No. 3, pp. 283-302.

Jia, R., Steelman, Z. R. and Reich, B. H. (2017). "Using mechanical turk data in IS research: risks, rewards, and recommendations," *Communications of the Association for Information Systems*, Vol. 41, No. 1, pp. 301-318.

Johns, G. (2006). "The essential impact of context on organizational behavior," *Academy of Management Review*, Vol. 31, No. 2, pp. 386-408.

Joinson, A. N. (1998), "Causes and implications of disinhibited information on the Internet," *Psychology and the Internet*, Gackenbach, J., ed.^eds., p.^pp. 43-60, San Diego, CA: Academic Press.

Jöreskog, K. and Sörbom, D. (1988), *LISREL 7: A guide to the program and its application*, Chicago: SPSS Inc.

Lee, C. H. and Yu, H. (2020). "The impact of language on retweeting during acute natural disasters: uncertainty reduction and language expectancy perspectives," *Industrial Management & Data Systems*, Vol. 120, No. 8, pp. 1501-1519.

Lewis, B. R., Templeton, G. F. and Byrd, T. A. (2005). "A methodology for construct development in MIS research," *European Journal of Information Systems*, Vol. 14, No. 4, pp. 388-400.

Lowry, P. B., D'Arcy, J., Hammer, B. and Moody, G. D. (2016a). ""Cargo Cult" science in traditional organization and information systems survey research: A case for using nontraditional methods of data collection, including Mechanical Turk and online panels," *The Journal of Strategic Information Systems*, Vol. 25, No. 3, pp. 232-240.

Lowry, P. B., Zhang, J., Moody, G. D., Chatterjee, S., Wang, C. and Wu, T. (2019). "An integrative theory addressing cyberharassment in the light of technology-based opportunism," *Journal of Management Information Systems*, Vol. 36, No. 4, pp. 1142-1178.

Lowry, P. B., Zhang, J., Wang, C. and Siponen, M. (2016b). "Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model," *Information Systems Research*, Vol. 27, No. 4, pp. 962-986.

Massanari, A. (2017). "# Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures," *New Media & Society*, Vol. 19, No. 3, pp. 329-346.

Matzler, K., Veider, V. and Kathan, W. (2015). "Adapting to the sharing economy," *MIT Sloan Management Review*, Vol. 56, No. 2, pp. 71-77.

Miller, B. (2015). ""Dude, Where's Your Face?" Self-Presentation, Self-Description, and Partner Preferences on a Social Networking Application for Men Who Have Sex with Men: A Content Analysis," *Sexuality & Culture*, Vol. 19, No. 4, pp. 637-658.

Moore, G. C. and Benbasat, I. (1991). "Development of an instrument to measure the perceptions of adopting an information technology innovation," *Information Systems Research*, Vol. 2, No. 3, pp. 192-222.

Munger, K. (2017). "Tweetment effects on the tweeted: Experimentally reducing racist harassment," *Political Behavior*, Vol. 39, No. 3, pp. 629-649.

Raghunathan, B., Raghunathan, T. and Tu, Q. (1999). "Dimensionality of the strategic grid framework: the construct and its measurement," *Information Systems Research*, Vol. 10, No. 4, pp. 343-355.

Salo, J., Mäntymäki, M. and Islam, A. N. (2018). "The dark side of social media–and Fifty Shades of Grey introduction to the special issue: the dark side of social media," *Internet Research*, Vol. 28, No. 5, pp. 1166-1168.

Sanfilippo, M., Yang, S. and Fichman, P. (2017). "Trolling here, there, and everywhere: Perceptions of trolling behaviors in context," *Journal of the Association for Information Science and Technology*, Vol. 68, No. 10, pp. 2313-2327.

Saunders, K. C. (2016). "A double-edged sword: Social media as a tool of online disinhibition regarding American sign language and Deaf cultural experience marginalization, and as a tool of cultural and linguistic exposure," *Social Media+ Society*, Vol. 2, No. 1, pp. 1-9.

Schouten, A. P., Valkenburg, P. M. and Peter, J. (2007). "Precursors and underlying processes of adolescents' online self-disclosure: Developing and testing an "Internet-attribute-perception" model," *Media Psychology*, Vol. 10, No. 2, pp. 292-315.

Sharon, T. and John, N. A. (2018). "Unpacking (the) secret: Anonymous social media and the impossibility of networked anonymity," *New Media & Society*, Vol. 20, No. 11, pp. 4177-4194.

Statista (2020), "Global digital population as of July 2020," available at: https://www.statista.com/statistics/617136/digital-population-worldwide/ (accessed 28 August, 2020).

Stewart, K. A. and Segars, A. H. (2002). "An empirical examination of the concern for information privacy instrument," *Information Systems Research*, Vol. 13, No. 1, pp. 36-49.

Suler, J. (2004). "The online disinhibition effect," *Cyberpsychology & Behavior*, Vol. 7, No. 3, pp. 321-326.

Tojib, D. R., Sugianto, L.-F. and Sendjaya, S. (2008). "User satisfaction with business-to-employee portals: conceptualization and scale development," *European Journal of Information Systems*, Vol. 17, No. 6, pp. 649-667.

Turel, O., Matt, C., Trenz, M., Cheung, C. M., D'Arcy, J., Qahri-Saremi, H. and Tarafdar, M. (2019). "Panel report: The dark side of the digitization of the individual," *Internet Research*, Vol. 29, No. 2, pp. 274-288.

Udris, R. (2014). "Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale," *Computers in Human Behavior*, Vol. 41, No., pp. 253-261.

Venkatraman, S., MK Cheung, C., Lee, Z. W., D. Davis, F. and Venkatesh, V. (2018). "The "Darth" side of technology use: An inductively derived typology of cyberdeviance," *Journal of Management Information Systems*, Vol. 35, No. 4, pp. 1060-1091.

Wachs, S., Wright, M. F. and Vazsonyi, A. T. (2019). "Understanding the overlap between cyberbullying and cyberhate perpetration: Moderating effects of toxic online disinhibition," *Criminal Behaviour and Mental Health*, Vol. 29, No. 3, pp. 179-188.

Walther, J. B. (1996). "Computer-mediated communication impersonal, interpersonal, and hyperpersonal interaction," *Communication Research*, Vol. 23, No. 1, pp. 3-43.

Wong, R. Y. M., Cheung, C. M. K. and Xiao, B. (2018). "Does gender matter in cyberbullying perpetration? An empirical investigation," *Computers in Human Behavior*, Vol. 79, No., pp. 247-257.

Wong, R. Y. M., Cheung, C. M. K., Xiao, S. B. and Chan, T. K. H. (2015), "The instigating, impelling, and inhibiting forces in cyberbullying perpetration across gender," in *Pacific Asia Conference on Information Systems (PACIS)*, Singapore.

Wright, M. F., Harper, B. D. and Wachs, S. (2019). "The associations between cyberbullying and callous-unemotional traits among adolescents: The moderating effect of online disinhibition," *Personality and Individual Differences*, Vol. 140, No., pp. 41-45.

Wu, S., Lin, T.-C. and Shih, J.-F. (2017). "Examining the antecedents of online disinhibition," *Information Technology & People*, Vol. 30, No. 1, pp. 189-209.

Appendix A. Factor loading of the measurement items

Rotated Component Matrix					
	Component				
	1	2	3	4	5
DA1	.682	.137	.158	.074	.013
DA2	.705	.239	.164	.011	.008
DA3	.788	.188	.169	.130	.040
DA4	.794	.263	.103	.084	.063
DA5	.711	.301	.062	.176	.075
DA6	.791	.193	.082	.123	.046
IV1	.757	.227	.011	.129	.044
IV2	.834	.173	.179	.145	.031
IV3	.707	.223	.191	.257	.139
IV4	.811	.124	.149	.105	.090
IV5	.839	.142	.120	.072	.076
IV6	.777	.194	-.123	.138	.057
IV7	.809	.213	.168	.134	.048
AS1	.037	.267	-.241	.191	.261
AS2	.081	.173	-.095	.158	.516
AS3	-.055	-.110	.346	.020	.716
AS4	-.051	-.055	.286	.090	.728
AS5	.113	-.019	-.336	-.010	.678
AS6	.028	.065	-.402	.091	.702
AS7	.109	.010	-.409	.052	.725
SI1	.233	.070	.147	.632	-.093
SI2	.153	.084	-.001	.701	.059
SI3	.060	.137	.048	.786	-.056
SI4	-.006	.091	.009	.787	.137
SI5	.021	.059	-.002	.750	.173
SI6	-.020	.051	.063	.721	.219
DI1	.140	.141	.781	.219	-.136
DI2	.078	.119	.808	.095	-.089
DI3	.064	.168	.858	.037	-.051
DI4	.120	.200	.632	.002	.033
DI5	.138	.222	.680	-.022	-.164
MA1	.080	.622	.200	.020	-.111
MA2	.179	.822	.153	.122	-.028
MA3	.180	.740	.284	.130	.045
MA4	.071	.813	.214	.022	-.040
MA5	.194	.723	.008	.177	.199
MA6	.241	.787	.082	.155	.050
MA7	.188	.803	.153	.150	.138
Extraction Method: Principal Component Analysis. Rotation Method: Equamax with Kaiser Normalization. Rotation converged in 6 iterations.					

Appendix B. The finalized online disinhibition scale (23 items)

Item	Statement
Dissociative Anonymity	I feel I am anonymous in the online environment.
	I believe that my personal identity remains unknown to others in the online environment.
	I feel that I can hide my identity online.
	My actions are not identifiable in the online environment.
Invisibility	I feel like invisible in the online environment.
	I feel that online environment averts others from seeing any of me.
	I am invisible online.
	Others do not see me in the online environment.
	My actions are invisible in the online environment.
Asynchronicity	I do not need to reply others immediately in the online environment.
	I can delay my feedback to others in the online environment.
	I can postpone replying others in the online environment.
Solipsistic Introjection	I assign a character to that person I am communicating with online.
	I interpret others' messages with my expectation during online communication.
	I perceive how that person's intended to talk about in the online environment.
Dissociative Imagination	I feel like people in the online space are just imaginary with no connection to reality.
	The online environment is an imaginary world.
	The online environment has no connection to reality.
	The online environment is separated from the offline world.
Minimization of Authority	I am away from real life authorities in the online environment.
	I feel less fear of offline authorities in the online environment.
	I feel free from real life authorities in the online environment.
	I feel that offline authorities are absent in the online environment.