Received: 27 November 2018 | Revised: 9 April 2019 | Accepted: 19 May 2019

DOI: 10.1002/cbm.2116

ORIGINAL ARTICLE

WILEY

Understanding the overlap between cyberbullying and cyberhate perpetration: Moderating effects of toxic online disinhibition

Sebastian Wachs[1] | **Michelle F. Wright**[2,3] | **Alexander T. Vazsonyi**[4]

[1]Department of Educational Sciences, Faculty of Human Sciences, University of Potsdam, Potsdam, Germany

[2]Department of Psychology, Pennsylvania State University, State College, PA, USA

[3]Faculty of Social Studies, Masaryk University, Brno, Czech Republic

[4]Department of Family Sciences, University of Kentucky, Lexington, KY, USA

Correspondence
Sebastian Wachs, Department of Educational Sciences, Faculty of Human Sciences, University of Potsdam, Karl-Liebknecht-Str. 24-25, Potsdam 14476, Germany.
Email: wachs@uni-potsdam.de

Abstract

Background: Problem behaviour theory postulates that different forms of norm violations cluster and can be explained by similar antecedents. One such cluster may include cyberbullying and cyberhate perpetration. A potential explanatory mechanism includes toxic online disinhibition, characterised by anonymity, an inability to empathise and to recognise and interpret social cues. The current study to develop a better understanding of the relationship between cyberhate and cyberbullying to inform effective intervention and prevention efforts.

Aims: To test the link between cyberbullying and cyberhate and whether this relationship was moderated by toxic online disinhibition.

Methods: Self-report questionnaires on cyberbullying, cyberhate, and toxic online disinhibition were completed by 1,480 adolescents between 12 and 17 years old (M = 14.21 years; SD = 1.68).

Results: Increases in cyberbullying perpetration and toxic online disinhibition were positively related to cyberhate perpetration. Furthermore, cyberbullies reported more cyberhate perpetration when they reported higher levels of toxic online disinhibition and less frequent cyberhate perpetration when they reported lower levels of toxic online disinhibition.

Conclusion: The current study provides evidence of a possible link between cyberbullying and cyberhate perpetration, moderated by toxic online disinhibition. This suggests that,

 © 2019 John Wiley & Sons, Ltd.

to be effective, prevention and intervention programmes should (i) consider the co-occurrence of varying forms of cyberaggression and (ii) consider potential effects of the online environment on aggressive online behaviour among young people.

1 | INTRODUCTION

The internet has provided a novel medium for the perpetration of new forms of aggression. Cyberbullying, the intentional and repeated infliction of harm via information and communication technologies against a helpless victim (Smith et al., 2008), has been thoroughly investigated over the past two decades. A more recent manifestation of bias based cyberaggression, however, called cyberhate, has not yet received much empirical attention (Lewis, Cogburn, & Williams, 2015). Cyberhate may include denigration through such technologies by harassment, exclusion, and advocacy of violence against specific social groups on the basis of assigned or selected characteristics, such as sexual orientation, race, gender, and religious affiliation (Hawdon, Oksanen, & Räsänen, 2017; Wachs & Wright, 2018).

Currently, little information exists on young people in Germany and their engagement in cyberhate as perpetrators. In one representative study, however, 65% of German adolescents reported having observed cyberhate (Feierabend, Rathgeb, & Reutter, 2018). Some information exists about cyberhate perpetration from other countries. In a study with Finnish adolescents, 6% reported that they have had posted, forwarded, or shared online hate material (Räsänen et al., 2016). With regard to cyberbullying, one representative German study of 9,512 students found that 2.4% of the participants could be classified as cyberbullies (Bergmann & Baier, 2018).

Conceptually, cyberhate has both similar and different characteristics with cyberbullying. Both are carried out with the intention of harming a person or group by using information and communication technologies. Cyberbullying, however, is often described as a repeated activity (Smith et al., 2008), whereas cyberhate, sometimes repetitive, may be carried out as a single act. Whereas cyberbullying can be directed at an individual person, cyberhate is necessarily based on prejudicial views about different social groups (Sponholz, 2018). Sharing conceptual similarities, it can be assumed that there might be some overlap between cyberbullying and cyberhate perpetrators.

Theoretically, consistent with problem behaviour theory, it can be expected that as one form of norm violations increases, the likelihood of another does too (Jessor & Jessor, 1977). Indeed, several studies have shown that different forms of aggression or violence perpetration, including cyberbullying, are associated with traditional bullying perpetration, cyber trolling, physical and cyber dating abuse, and so forth (Olweus, 2012; Wachs, Junger, & Sittichai, 2015; Yahner, Dank, Zweig, & Lachman, 2015; Zezulka & Seigfried-Spellar, 2016). Previous empirical studies have not, however, tested the potential associations between cyberbullying and cyberhate perpetration. Because it can be assumed that cyberbullying and cyberhate are complex behavioural phenomena that are multiply influenced, ranging from individual-level constructs to larger environmental effects, it is also plausible that the relationship might be explained by moderator variables.

One potential moderating variable that might explain the relationship between cyberbullying and cyberhate is toxic online disinhibition. It describes the loosening of social restrictions and inhibitions that are usually present in face-to-face interactions because it can make individuals feel part of an anonymous mass, and thus he/she does not sense the need to take responsibility for his or her actions (Suler, 2004). In the online context, an individual might feel more comfortable engaging in aggression because of the visual, emotional, and physical distance from his/her victims (Udris, 2014). These circumstances can also manifest as a limited ability to empathise with others and to recognise social cues (Voggeser, Singh, & Göritz, 2018). Cyberbullies might be even more sensitive to toxic online disinhibition because they have shown a lack of empathy and an inability to recognise social cues (Zych, Baldry, Farrington, &

Llorent, 2019). In fact, some past research supports the link between toxic online disinhibition and cyber aggressions (Udris, 2014; Voggeser et al., 2018; Wachs & Wright, 2018, 2019; Wright, Harper, & Wachs, 2019). There is also some evidence that toxic online disinhibition moderates the association between cyberaggression and callous-unemotional traits (Wright et al., 2019) and the relationship between being a cyberhate bystander and perpetrator, after controlling for cyberhate victimisation (Wachs & Wright, 2018).

In sum, there appears some preliminary evidence that suggests a link between cyberbullying and cyberhate perpetration, potentially moderated by toxic online disinhibition. Findings from the current study might help to provide novel insights about correlates of cyberhate perpetration among young people and inform more effective intervention and prevention programmes to tackle online aggression among young people. Specific study hypotheses included that

1. Cyberbullying perpetration would be positively associated with cyberhate perpetration.
2. Toxic online disinhibition would moderate the link between cyberbullying and cyberhate perpetration, where greater disinhibition would potentiate the relationship between the two variables.

2 | METHODS

2.1 | Ethics

All materials and procedures received ethical approval by the educational authority and data protection officer of the federal state of Bremen, Germany.

2.2 | Procedure

Twenty schools were randomly selected from a list of 167 schools from the federal states of Bremen, Brandenburg, and Berlin in Germany. Nine school principals did not reply to the recruitment email, four expressed some interest in the study but had other commitments preventing them from participating; five school principals from Bremen, one from Brandenburg, and one from Berlin agreed to participate. As the participants were minors, parents had to sign a written consent form allowing them to participate. A letter and parental permission slip were distributed to participants through the teachers. Classes were chosen randomly by the schools. Of the 1,788 parental permission slips sent home, 1,480 parental permission slips were returned. Trained research assistants administered an online survey in the school's computer lab during one regular school hour.

Participants were told that their participation was anonymous, that their participation was optional, that they could choose not to answer questions, and that participation could be ended at any time without giving a reason and with no consequence. To prevent distress and further harm by participating in this study, participants were told and given written information about whom they could talk to if they believed they needed support following the completion of the survey.

2.3 | Measures

2.3.1 | Cyberhate perpetration

For cyberhate perpetration, participants were asked: "How often did it happen in the past 12 months that you have posted hateful or degrading writings or speech online, which inappropriately attacks certain groups of people or individuals based on their sex, religious affiliation, race, or sexual orientation?" This item was adopted from past research (Hawdon, Oksanen, & Räsänen, 2015). Participants responded on a five-point Likert scale, ranging from 0 (*never*) to 4 (*very frequently*).

2.3.2 | Cyberbullying perpetration

To measure cyberbullying perpetration, participants were asked: "How many times have you cyberbullied others in the last 12 months?" (Olweus, 2012). Participants responded on a five-point ordinal scale: 0 (*never*) to 4 (*several times a week*).

2.3.3 | Toxic online disinhibition

The extent to which adolescents believed that the online environment disinhibited their online behaviour was measured by one scale consisting of four items (Udris, 2014), with items such as, "I don't mind writing insulting things about others online, because it's anonymous" or "It is easy to write insulting things online because there are no repercussions," and response options ranging from 0 (*definitely do not believe*) to 4 (*definitely do believe*). A mean score was computed by averaging all items. Cronbach's alpha was .79.

2.3.4 | Control variables

For demographic characteristics, participants reported their grade, sex (girl/boy), and migration background by indicating whether German was mainly spoken at home (yes/no).

Socio-economic status (SES) was measured using the Family Affluence Scale (FAS; Boyce, Torsheim, Currie, & Zambon, 2006; Hartley, Levin, & Currie, 2016). Participants answered six items regarding, for example, family's car, children having their own bedroom, the number of bathrooms, the number of computers at home, and the number of holidays taken in the past 12 months. A composite FAS score is calculated for each participant based on his or her responses to these six items. As recommended, FAS scores were collapsed into a low, medium, and high family affluence indicator.

Cyberhate and cyberbullying victimisation was measured in the same way as perpetration asking participants "How often did it happen in the past 12 months that you have personally been the target of hateful or degrading writings or speech online because of your sex, religious affiliation, race, or sexual orientation?" for cyberhate victimisation and "How many times have you cyberbullied others in the last 12 months?" for cyberbullying victimisation.

2.4 | Data analysis

Moderation analyses with standardised variables was conducted using the Process Macro for SPSS with Model 1 and included 5000 bias-corrected bootstrap samples (Hayes, 2013), where cyberhate perpetration was the dependent measure, cyberbullying perpetration the independent measure, and toxic online disinhibition the moderator. Research on cyberhate involvement among young people is very limited. Previous research has, however, shown that variables such as age, sex, migration background, SES, and victimisation are often associated with perpetrating (cyber-)aggression (Barlett & Coyne, 2014; Benner et al., 2018; Costello & Hawdon, 2018; Strohmeier & Spiel, 2003; Tippett & Wolke, 2014; Vitoroulis & Georgiades, 2017). Therefore, these variables were included as control variables in the moderation analyses. All analyses were conducted in SPSS 25.0.

3 | RESULTS

3.1 | Sample

There were 1,480 participants between 12 and 17 years in age (Mean [*M*] = 14.21, *SD* = 1.22) in Grades 7 through 10. About half of the sample were girls (*n* = 744, 50.3%). Around 10% (*n* = 144) reported that German was not the

main language spoken at home. About one third (n = 483, 33.6%) stated that they lived in families of low affluence, a third (n = 474) in families of middle affluence, and a third (n = 479) in families of high affluence. Table 1 shows the demographic characteristics of the participants.

3.2 | Descriptive analyses

The majority (1,256, 88.7%) of participants reported that they had *never* expressed cyberhate; 7.6% (n = 104) of participants reported this behaviour *very rarely*, 1.8% (n = 26) *occasionally*, 0.9% (n = 13) *frequently*, and 1.2% (n = 17) *very frequently*. Similarly, most participants (n = 1,166, 81.8%) reported that they had *never* engaged in such behaviour, 12.9% (n = 184) reported engaging in this behaviour *once or twice*, 2.7% (n = 38) *once or twice in a month*, 1.1% (n = 15) *about once a week*, and 1.6% (n = 23) *several times a week*.

Bivariate correlations between the main study variables (cyberhate perpetration, cyberbullying perpetration, and toxic online disinhibition) and the control variables (grade, sex, SES, migration background, cyberhate victimisation, and cyberbullying victimisation) are shown in Table 2. Cyberhate was positively associated with cyberbullying, as was toxic online disinhibition. Furthermore, cyberbullying perpetration was positively associated with toxic online disinhibition. There were also some significant associations between the control and the main study variables. Specifically, grade was positively associated with cyberhate perpetration and toxic online disinhibition. Boys in comparison with girls reported greater cyberhate perpetration and toxic online disinhibition. Cyberbullying and cyberhate victimisation were positively related to cyberhate perpetration, cyberbullying perpetration, and toxic online disinhibition.

3.3 | Moderation analyses

The overall model was significant, $F(9, 1310) = 43.53$, $p < .001$, $R^2 = .23$, indicating a large effect (Cohen's $f^2 = 0.60$). As Table 3 shows, a number of hypothesized predictors significantly explained variance in cyberhate perpetration. This included cyberbullying perpetration ($\beta = .19$, $p < .001$), toxic online disinhibition ($\beta = .13$, $p < .001$), but also grade level ($\beta = .05$, $p = .026$), and being male ($\beta = .15$, $p < .001$). Migration and/or socio-economic background were, however, unrelated to online cyberhate perpetration. Finally, both cyberhate victimisation ($\beta = .22$, $p < .001$) and cyberbullying victimisation ($\beta = .06$, $p = .013$) were positively associated with cyberhate perpetration.

TABLE 1 Frequencies of demographic variables by grade, sex, and socio-economic status

FAS	Sex	Grade									
		7th		8th		9th		10th		Total	
		n	%	*n*	%	*N*	%	*n*	%	*n*	%
High FAS	Girls	77	28.9	85	32	62	23.3	42	15.8	266	100
	Boys	62	28.6	81	37.3	49	22.6	25	11.5	217	100
Average FAS	Girls	77	33.6	55	24	53	23.1	44	19.2	229	100
	Boys	63	25.7	74	30.2	63	25.7	45	18.4	245	100
Low FAS	Girls	63	27.6	53	23.2	70	30.7	42	18.4	228	100
	Boys	82	32.7	63	25.1	67	26.7	39	15.5	251	100
Total		424	29.5	411	28.6	364	25.3	237	16.5	1436	100

Abbreviation: FAS, Family Affluence Scale.

TABLE 2 Means, standard deviations, and correlations between online hate perpetration, online hate victimisation, toxic online disinhibition, and control variables

n = 44	1	2	3	4	5	6	7	8	9
1. Cyberhate perpetration	1								
2. Cyberbull. perpetration	.32**	1							
3. TOD	.24**	.18**	1						
4. Grade	.06*	−.01	.06*	1					
5. Sex male	.12**	.15**	.09**	−.01	1				
6. Migration background yes	.50	.01	.02	−.07**	.07**	1			
7. SES	.01	.05	.01	.03	.07**	−.13**	1		
8. Cyberhate victimisation	.31**	.15**	.14**	.03	.01	.06*	−.09**	1	
9. Cyberbull. victimisation	.19**	.28**	.07*	−.01	−.03	.01	.01	.28**	1
M	0.19	0.28	0.29	—	—	—	—	0.33	0.61
SD	0.62	0.71	0.74	—	—	—	—	0.74	0.73

Note. The correlations among continuous variables, namely, cyberhate perpetration, victimisation, cyberbullying perpetration, victimisation, and toxic online disinhibition, were computed as Pearson's bivariate correlations; correlations between binary variables, such as sex and migration background, and continuous variables, were computed as point-biserial correlations.

Abbreviations: SES, socio-economic status; TOD, toxic online disinhibition.

*$p < .01$.

**$p < .001$.

TABLE 3 Standardised coefficients of the model predicting cyberhate perpetration

	β [a]	*SE*	*t*	*p*
Constant	−0.82 [−1.28 – −0.353]	0.237	−3.44	<.001
CYBP	.19 [0.129–0.241]	0.028	6.51	<.001
TOD	.13 [0.076–0.174]	0.025	5.01	<.001
CYBP × TOD	.11 [0.073–0.144]	0.018	5.94	<.001
Control Variables				
Grade	.05 [0.006–0.096]	0.023	2.22	.026
Sex male	.15 [0.053–0.247]	0.049	3.03	.002
Migration background yes	.10 [−0.054–0.272]	0.083	1.31	.189
SES	.02 [−0.042–0.076]	0.030	0.563	.563
Cyberhate victimisation	.22 [0.164–0.267]	0.026	8.22	<.001
Cyberbullying victimisation	.06 [0.014–0.118]	0.066	2.48	.013

Abbreviations: CYBP, cyberbullying perpetration; SES, socio-economic status; TOD, toxic online disinhibition.

[a]95% BCa bootstrap confidence intervals based on 5,000 samples.

A significant moderation effect was found by toxic online disinhibition on the link between cyberbullying perpetration and cyberhate perpetration (β = .11, $p < .001$). As Figure 1 shows, probing the interaction further revealed that cyberbullying perpetrators reported more cyberhate perpetration at greater levels of toxic online disinhibition (β = .29, $p < .001$ at +1 *SD*) and less frequent cyberhate perpetration at lower levels (β = .11, $p < .001$ at −1 *SD*). Given that the study was cross sectional, and there were theoretical assumptions that cyberhate might lead to

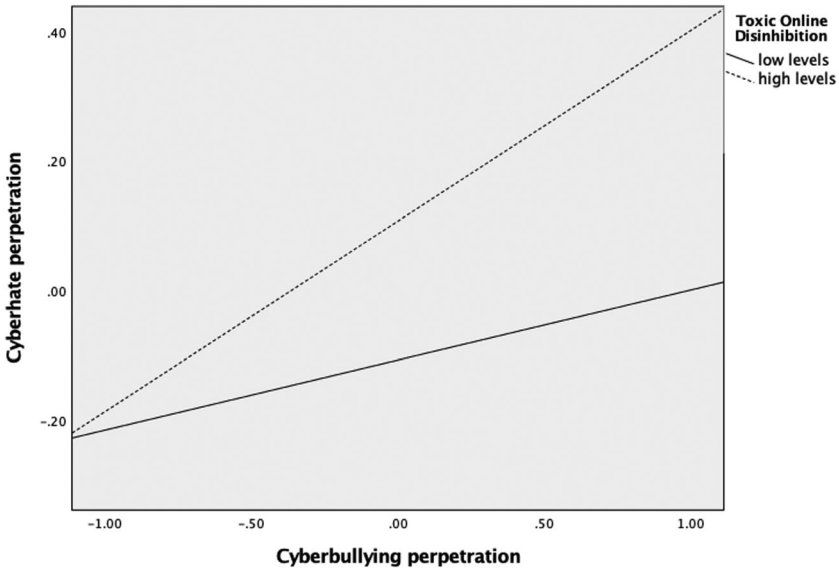


FIGURE 1 Graphical representation of the moderation of toxic online disinhibition on the association between cyberbullying and cyberhate perpetration

cyberbullying, the analysis was repeated with cyberbullying as dependent variable and cyberhate as the independent variable, revealing similar findings, although with slightly weaker effects (details available from the first author).

4 | DISCUSSION

The current study sought to make two main contributions to the literature on cyberhate among young people, first whether cyberbullying and cyberhate perpetration were associated, and second, whether this relationship was mediated by toxic online disinhibition. Both predictions were supported. Cyberbullying and cyberhate perpetration were positively associated (H1). This is both consistent with problem behaviour theory (Jessor & Jessor, 1977), and with previous research, which found that when young people are involved in one form of cyberaggression as perpetrators, then they are more likely to also be involved in others (Olweus, 2012; Wachs et al., 2015; Yahner et al., 2015; Zezulka & Seigfried-Spellar, 2016). This finding can be explained by common set of risk factors or antecedents, as theoretically hypothesised by the theory, including low self-control, anger regulation, or lower levels of empathy, all of which need to be further investigated in future research.

The finding that cyberbullying and cyberhate perpetration are correlated also has some relevance for the development of prevention and intervention programmes in the future. First, findings suggest that prevention programmes focused on cyberbullying should include elements that teach young people about prejudice, democracy education, and intergroup violence. Second, findings support the need for perhaps more comprehensive prevention programmes that aim to promote positive young people development and prevent engagement in aggressive behaviour in a broad sense. An example of such a broad approach is developing life skills, such as ethical media competence, democracy beliefs, responsibility, self-esteem, self-control, and empathy. Life skills might play an important role to mitigate risks by focusing on resiliency against a wide variety of aggressive behaviours. Finally, because the reduction of toxic online disinhibition might prevent young people from becoming cyberhate perpetrators, it seems important to increase awareness among young people about how the online environment influences their own thoughts and actions.

As hypothesised, young people who endorsed cyberbullying perpetration were also more likely to report cyberhate perpetration when they reported higher levels of toxic online disinhibition and less likely to report cyberhate perpetration when they reported lower levels of toxic online disinhibition (H2). This finding can be explained by previous research that found that both toxic online disinhibition and cyberbullying perpetration are associated with feeling less inhibited, limited ability to empathise with others, and to recognise social cues (Voggeser et al., 2018; Zych et al., 2019). Study findings are consistent with previous research that showed that toxic online disinhibition moderated the association between cyberaggression and other types of antisocial behaviours and conduct problems (Wright et al., 2019). It is also consistent with work which showed that toxic online disinhibition moderated the associations between observing, being victimized, and perpetrating cyberhate (Wachs & Wright, 2018, 2019), but also with research that showed associations between toxic online disinhibition and the perpetration of cyberaggression (Udris, 2014; Voggeser et al., 2018). Study findings also show that being a cyberbully alone does not explain engagement in cyberhate as a perpetrator. It appears to be more likely that additional variables need to be taken into consideration to explain this relationship. Thus, an important next step in developing intervention and prevention programmes and understanding the relationship between cyberbullying and cyberhate perpetration is a further examination of additional potential moderators.

The current study has several limitations that should be addressed in future research. First, even though our sample is large, it was not representative. To increase the generalisability of study findings, follow-up research with representative samples including students from across Germany are needed. In addition, the current study only included students living in Germany. More cross-cultural research could help to better understand cultural differences in cyberhate perpetration or whether the associations found in the current study are idiosyncratic. Second, the cross-sectional nature of the study means that no conclusions on the temporal ordering of the study's main variables can be reached. Longitudinal studies are needed to better understand the causal ordering between cyberbullying perpetration and cyberhate perpetration. Third, single-item measures were used for cyberbullying and cyberhate perpetration. To overcome reliability and validity assessment issues, follow-up research should use validated scales. Finally, the measurement of online hate in this study was based on what young people perceive as online hate. Thus, our measurement of cyberhate the based on a subjective rather than objective definition. It is possible that young people differ in how they classify online content as cyberhate and what they judge as mean or threatening. More research is needed to better understand the definitional criteria for cyberhate among young people.

5 | CONCLUSION

The current study tested the potential moderating effect of online disinhibition on the association between cyberbullying and cyberhate perpetration among young people. Study findings supported this link, particularly the moderating effect of toxic online disinhibition. Identifying cyberhate correlates, such as cyberbullying perpetration and online disinhibition, supports the development of intervention programmes aimed at reducing online risks for young people. Study findings identify the need for more comprehensive prevention programmes that perhaps encompass a reduction of various forms of cyberaggression in the same intervention.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Sebastian Wachs https://orcid.org/0000-0003-2787-6646

REFERENCES

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

Benner, A. D., Wang, Y., Shen, Y., Boyle, A. E., Polk, R., & Cheng, Y.-P. (2018). Racial/ethnic discrimination and well-being during adolescence: A meta-analytic review. *American Psychologist*, *73*(7), 855–883. https://doi.org/10.1037/amp0000204

Bergmann, M., & Baier, D. (2018). Prevalence and correlates of cyberbullying perpetration. Findings from a German representative student survey. *International Journal of Environmental Research and Public Health*, *15*(2), 274. https://doi.org/10.3390/ijerph15020274

Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The family affluence scale as a measure of national wealth: Validation of an adolescent self-report measure. *Social Indicators Research*, *78*(3), 473–487. https://doi.org/10.1007/s11205-005-1607-6

Costello, M., & Hawdon, J. (2018). Who are the online extremists among us? Sociodemographic characteristics, social networking, and online experiences of those who produce online hate materials. *Violence and Gender*, *5*(1), 55–60. https://doi.org/10.1089/vio.2017.0048

Feierabend, S., Rathgeb, T., & Reutter, T. (2018). JIM Study 2018. Youth, Information, Media. Basic Investigation on Media Handling of 12- to 19-year-olds. Retrieved from Medienpädagogischer Forschungsverbund Südwest website: https://www.mpfs.de/fileadmin/files/Studien/JIM/2018/Studie/JIM_2018_Gesamt.pdf

Hartley, J. E., Levin, K., & Currie, C. (2016). A new version of the HBSC Family Affluence Scale-FAS III: Scottish qualitative findings from the international FAS development study. *Child Indicators Research*, *9*(1), 233–245. https://doi.org/10.1007/s12187-015-9325-3

Hawdon, J., Oksanen, A., & Räsänen, P. (2015). Online extremism and online hate: Exposure among adolescents and young adults in four nations. *Nordicom Information: Medie-Och Kommunikationsforskning I Norden*, *37*(3–4), 29–37.

Hawdon, J., Oksanen, A., & Räsänen, P. (2017). Exposure to online hate in four nations: A cross-national consideration. *Deviant Behavior*, *38*(3), 254–266. https://doi.org/10.1080/01639625.2016.1196985

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, London: The Guilford Press.

Jessor, R., & Jessor, S. L. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth*. San Diego, CA: Academic Press.

Lewis, T. T., Cogburn, C. D., & Williams, D. R. (2015). Self-reported experiences of discrimination and health: Scientific advances, ongoing controversies, and emerging issues. *Annual Review of Clinical Psychology*, *11*, 407–440. https://doi.org/10.1146/annurev-clinpsy-032814-112728

Olweus, D. (2012). Cyberbullying: An overrated phenomenon? *European Journal of Developmental Psychology*, *9*(5), 520–538. https://doi.org/10.1080/17405629.2012.682358

Räsänen, P., Hawdon, J., Holkeri, E., Keipi, T., Näsi, M., & Oksanen, A. (2016). Targets of online hate: Examining determinants of victimization among young Finnish Facebook users. *Violence and Victims*, *31*(4), 708–726. https://doi.org/10.1891/0886-6708.VV-D-14-00079

Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry*, *49*(4), 376–385. https://doi.org/10.1111/j.1469-7610.2007.01846.x

Sponholz, L. (2018). Hate speech in the mass media. In *Theoretical Foundations and Empirical Implementation*. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-15077-8

Strohmeier, D., & Spiel, C. (2003). Immigrant children in Austria: Aggressive behavior and friendship patterns in multicultural school classes. *Journal of Applied School Psychology*, *19*(2), 99–116. https://doi.org/10.1300/J008v19n02_07

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

Tippett, N., & Wolke, D. (2014). Socioeconomic status and bullying: A meta-analysis. *American Journal of Public Health*, *104*(6), 48–59.

Udris, R. (2014). Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale. *Computers in Human Behavior*, *41*, 253–261. https://doi.org/10.1016/j.chb.2014.09.036

Vitoroulis, I., & Georgiades, K. (2017). Bullying among immigrant and non-immigrant early adolescents: School-and student-level effects. *Journal of Adolescence*, *61*, 141–151. https://doi.org/10.1016/j.adolescence.2017.10.008

Voggeser, B. J., Singh, R. K., & Göritz, A. S. (2018). Self-control in online discussions: Disinhibited online behavior as a failure to recognize social cues. *Frontiers in Psychology*, *8*, 2372. https://doi.org/10.3389/fpsyg.2017.02372

Wachs, S., Junger, M., & Sittichai, R. (2015). Traditional, cyber and combined bullying roles: Differences in risky online and offline activities. *Societies*, *5*(1), 109–135. https://doi.org/10.3390/soc5010109

Wachs, S., & Wright, M. F. (2018). Associations between bystanders and perpetrators of online hate: The moderating role of toxic online disinhibition. *International Journal of Environmental Research and Public Health*, *15*(9), 2030. https://doi.org/10.3390/ijerph15092030

Wachs, S., & Wright, M. F. (2019). The moderation of online disinhibition and sex on the relationship between online hate victimization and perpetration. *Cyberpsychology, Behavior and Social Networking*, *22*, 300–306. https://doi.org/10.1089/cyber.2018.0551

Wright, M. F., Harper, B. D., & Wachs, S. (2019). The associations between cyberbullying and callous-unemotional traits among adolescents: The moderating effect of online disinhibition. *Personality and Individual Differences*, *140*, 41–45. https://doi.org/10.1016/j.paid.2018.04.001

Yahner, J., Dank, M., Zweig, J. M., & Lachman, P. (2015). The co-occurrence of physical and cyber dating violence and bullying among teens. *Journal of Interpersonal Violence*, *30*(7), 1079–1089. https://doi.org/10.1177/0886260514540324

Zezulka, L. A., & Seigfried-Spellar, K. C. (2016). Differentiating cyberbullies and Internet trolls by personality characteristics and self-esteem. *Journal of Digital Forensics, Security, and Law*, *11*(3), 5.

Zych, I., Baldry, A. C., Farrington, D. P., & Llorent, V. J. (2019). Are children involved in cyberbullying low on empathy? A systematic review and meta-analysis of research on empathy versus different cyberbullying roles. *Aggression and Violent Behavior*, *45*, 83–97. https://doi.org/10.1016/j.avb.2018.03.004

How to cite this article: Wachs S, Wright MF, Vazsonyi AT. Understanding the overlap between cyberbullying and cyberhate perpetration: Moderating effects of toxic online disinhibition. *Crim Behav Ment Health*. 2019;29:179–188. https://doi.org/10.1002/cbm.2116

APPENDIX

Toxic Online Disinhibition (Udris, 2014)

I do not mind writing insulting things about others online, because it's anonymous.

It is easy to write insulting things online because there are no repercussions.

There are no rules online therefore you can do whatever you want.

Writing insulting things online is not bullying.

Copyright of Criminal Behaviour & Mental Health is the property of John Wiley & Sons, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.