

Playing by the Rules? An Investigation of the Relationship Between Social Norms and Adolescents' Hate Speech Perpetration in Schools

Journal of Interpersonal Violence
2022, Vol. 37(21-22) NP21143–NP21164
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/08862605211056032
journals.sagepub.com/home/jiv



Sebastian Wachs^{1,2} , Alexander Wettstein³,
Ludwig Bilz⁴ , Norman Krause¹, Cindy Ballaschk¹,
Julia Kansok-Dusche⁴, and Michelle F. Wright^{5,6}

Abstract

Hate speech is often discussed and investigated as an online phenomenon, while hate speech among adolescents in the real world (e.g., in schools) has rarely been researched. Consequently, not much is known about potential interpersonal correlates and theoretical frameworks that might help us better understand why adolescents engage in hate speech in offline environments. To add to the literature, this study investigates hate speech perpetration among

¹Department of Educational Sciences, University of Potsdam, Potsdam, Germany

²National Anti-Bullying Research and Resource Centre, Institute of Education, Dublin City University, Dublin, Ireland

³Institute for Research, Development and Evaluation, Bern University of Teacher Education, Bern, Switzerland

⁴Department of Health Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany

⁵Department of Psychology, Penn State University, Pennsylvania, PA, USA

⁶Department of Psychology, DePaul University, Chicago, IL, USA

Corresponding Author:

Sebastian Wachs, Department of Educational Sciences, University of Potsdam, Karl-Liebknecht Str. 24-25, 14476 Potsdam, Germany.

Email: wachs@uni-potsdam.de

young people by utilizing the Social Cognitive Theory; more specifically, the association between witnessing and perpetrating hate speech in schools, and whether this relation is weakened or strengthened by anti-hate speech injunctive norms and deviant peer pressure. The sample consists of 1719 young people (45.1% boys, 53.6% girls, 1.3% gender diverse) between 11 and 18 years old ($M_{\text{age}} = 13.96$; $SD = .98$) from 22 schools in Switzerland. Self-report questionnaires were administered to measure hate speech involvement and adolescents' perception of social norms (i.e., anti-hate speech injunctive norms and deviant peer pressure). Multilevel regression analyses revealed that witnessing and perpetrating hate speech were positively associated. Moreover, injunctive anti-hate speech norms were negatively related and deviant peer pressure positively related to hate speech perpetration. Finally, increasing levels of injunctive anti-hate speech norms weakened the relation between witnessing and perpetrating hate speech, whereas higher levels of deviant peer pressure strengthened this association. The findings demonstrate that the Social Cognitive Theory is a useful framework for understanding hate speech perpetration in schools. The results also highlight the importance of taking into account social norms and interpersonal relationships (e.g., within the class) when designing anti-hate speech prevention programs and not focusing solely on intrapersonal factors.

Keywords

hate crimes, hate speech, social norms, school violence, discrimination, adolescents

The school setting is a central socialization context in which young people develop their personal identity, academic, social, and emotional skills. Schools also address various societal needs through, for example, social placement and supporting the transmission of cultural values and norms that young people need for social integration and functioning in society (Fend, 1974; Schwarzenthal et al., 2020; Steinberg & Morris, 2001). Schools are, however, also a place where students experience interpersonal conflicts and victimization, such as hate speech and discrimination. Such experiences pose risks to the personal development of those students involved and the overall school community, thereby reducing the effectiveness of schools' societal function (Krause et al., 2021; Lehman, 2019, 2020; Thapa et al., 2013).

Hate speech can be defined as a communicative form of expression that deliberately promotes, justifies, or disseminates exclusion, contempt, and devaluation of particular social groups (e.g., LGBTQI+ people, people of

color, people with Muslim or Jewish backgrounds, refugees, etc.; [Wachs et al., 2020](#)). Despite the increasing scientific attention to online hate speech (e.g., [Döring & Mohseni, 2020](#); [Harriman et al., 2020](#); [Wachs et al., 2021](#)), the current literature on offline hate speech among young people in schools is rather limited. Consequently, not much is known about determinants of adolescents' hate speech perpetration in schools and theoretical frameworks that help to explain why adolescents engage in it.

The Social Cognitive Theory ([Bandura, 1989](#)) has gained much popularity in understanding adolescents' deviant behavior through observational learning (e.g., bullying, drug abuse, and sexual risk behavior) and the influence of social (peer) norms ([Farrell et al., 2017](#); [Henneberger et al., 2020](#); [Salmivalli, 2010](#); [Sijtsema & Lindenberg, 2018](#); [Van de Bongardt et al., 2015](#)). Hence, this study uses the Social Cognitive Theory ([Bandura, 1989](#)) to investigate the associations between adolescents' exposure to hate speech (i.e., witnessing hate speech), social norms, and hate speech perpetration, as well as the interaction of witnessing hate speech and social norms to predict hate speech perpetration by adolescents. The findings can be used to develop theory-driven and evidence-based anti-hate speech prevention programs and to support schools in creating an inclusive and safe school environment where all students can thrive.

Understanding Hate Speech Through the Lens of the Social Cognitive Theory

In the 1960s, Bandura and his colleagues conducted a series of social psychological experiments to challenge the common view that learning could only be accounted for through behaviorism (i.e., respondent and operant responses; [Bandura et al., 1963](#); [Bandura & McDonald, 1963](#)). According to this research on the acquisition and maintenance of aggressive behavior and moral values, people do not learn only through direct instruction but also, in part, by observing models, which can be other people's actual behavior but also those of fictitious characters in books or movies ([Bandura et al., 1963](#); [Bandura & McDonald, 1963](#)). As a result of this seminal work, the Social Cognitive Theory was developed (SCT; [Bandura, 1989](#)) combining the person-environment interaction (social component) and the ability to reflect and self-motivation (cognitive component) to understand behavior change through observation.

Following the SCT, it can be assumed that if adolescents (observers) see that their peers (models) engage in hate speech, they might be more likely to imitate this behavior and become hate speech perpetrators themselves (observational learning). Indeed, a vast body of literature has shown that witnessing aggressive behavior in peers is positively associated with aggressive behavior in adolescents ([Evans et al., 2019](#); [Jung et al., 2019](#); [Rivers et al.,](#)

2009). More concretely, research on online hate speech has revealed that witnessing and perpetrating online hate speech are positively related (Bernatzky et al., 2021; Wachs et al., 2021; Wachs & Wright, 2018). What is not clear is whether exposure to hate speech in schools is related to hate speech perpetration in schools and it thus warrants further examination. Bandura, 1977 placed specific emphasis of the SCT on certain cognitive conditions that are prerequisites for observational learning. The most crucial cognitive condition of observational learning is that the observer must be motivated to carry out the modeled behavior (Bandura, 1977). The motivation to imitate hate speech perpetration is influenced by a variety of motives, such as fun, political-ideological convictions, compensation for feelings of frustration and inferiority, need for power, and fear of diminishing status (Ballaschk et al., 2021). Another motive for imitating hate speech might be adolescents' perceptions of social norms.

Social norms refer to rules, beliefs, attitudes, and behaviors that are perceived as (un)accepted, (un)supported, or (un)desired in the closer or broader social environment (Cialdini & Trost, 1998). Three types of social norms are often distinguished, namely, injunctive norms, descriptive norms, and peer pressure (Van de Bongardt et al., 2015). Cialdini and Trost (1998) introduced the distinction between norms that tell us something about people's perception of attitudes regarding what behaviors are typically approved of or disapproved of by others (injunctive norms) and those that reflect people's perceptions of behaviors that are typically performed by others (descriptive norms). The third type of norm is peer pressure, which refers to the active encouragement of peers to display a specific behavior (Van de Bongardt et al., 2015). Both injunctive and descriptive norms indirectly influence adolescents' behavior by their observation and interpretation of those around them, whereas peer pressure influences adolescents' behavior more directly through the active encouragement of peers (Van de Bongardt et al., 2015). In the present study, we will only focus on injunctive norms and peer pressure. We did not include an additional measure for descriptive norms because of the conceptual overlap between the measurement of witnessing hate speech among peers and descriptive norms regarding hate speech among peers.

Research that has investigated the relevance of social norms for hate speech perpetration by adolescents is scarce. Initial qualitative research by Ballaschk et al. (2021) highlighted the crucial role of social norms for hate speech perpetration (e.g., peer group pressure). These first findings on hate speech are in line with a large number of studies on the positive link between anti-social norms and aggressive behavior, risk-taking behavior, and attitudes and beliefs about the appropriateness and acceptability of deviant behavior among young people (Cohen & Prinstein, 2006; Fluck, 2017; Henneberger et al., 2020; Jung et al., 2019; Pozzoli & Gini, 2010;

Ruggeri et al., 2018; Van de Bongardt et al., 2015; Van Hoorn et al., 2017; Wilton & Campbell, 2011). For example, in a sample of Flemish adolescents, injunctive norms approving cyberbullying of significant others (e.g., friends, class members, and parents) and social pressure to join in on cyberbullying were all positively correlated with cyberbullying perpetration (Bastiaensens et al., 2016). Similarly, in a study with primary and secondary school students from Hong Kong, peer pressure was found to be positively associated with bullying perpetration (Lee & Wong, 2009). Furthermore, based on a meta-analysis including 153 studies, Cook et al. (2010) found that peer influences (e.g., negative influence through deviant peer groups) were the most important predictor for bullying perpetration.

Social norms might not only have a direct effect on adolescents' engagement in hate speech but also moderate the association between witnessing and perpetrating hate speech in schools by increasing witnesses' motivation to perpetrate hate speech. If adolescents witness hate speech and believe that significant others in their personal environment (i.e., peers, parents, and teachers) condone or support such behavior, witnesses might be more likely to perpetrate hate speech themselves in order not to be left out of their peer group. In some cases, perpetrating hate speech may not even be a reflection of personal attitudes but represent an effort to conform to the social norms and to express solidarity to in-group members. In such cases, the expectation of social acceptance or an increase in social status within the peer group might be considered a social reward that might directly influence adolescents' engagement in hate speech but also moderate the association between witnessing and perpetrating hate speech. Again, empirical evidence on this potentially moderating effect does not exist. However, the SCT gives theoretical support to such effects. In addition, previous research on online hate speech among adolescents found that intrapersonal factors (e.g., lack of coping strategies, feeling less inhibited online) moderates the association between witnessing and perpetrating online hate speech (Wachs & Wright, 2018; Wachs et al., 2019). Hence, it is proposed that environmental factors (i.e., social norms) might also play a crucial role in understanding the witnesses-perpetrators relation.

The Present Study

In sum, according to the SCT (Bandura, 1977), humans learn in part through observation. We thus might assume that adolescents who witness hate speech might be more likely to perpetrate hate speech. Whereas scholars have increasingly acknowledged that adolescents' aggressive behavior is influenced by the beliefs, attitudes, and behaviors that are perceived as accepted, supported, or desired by the people they are surrounded by, a lack of literature

exists on these associations when it comes to hate speech. Depending on whether social norms are perceived as condemning hate speech or approving hate speech, adolescents might be more or less likely to perpetrate hate speech themselves. Hence, injunctive anti-hate speech norms might be negatively associated with hate speech perpetration because adolescents might expect punishment if they engage in hate speech. In contrast, if adolescents experience deviant peer pressure, they might be more likely to perpetrate hate speech as they might want to avoid rejection or punishment by their peers. Another open question about the role of social norms concerns the extent to which social norms are not only directly related to hate speech perpetration but also weaken or strengthen the association between witnessing and perpetrating hate speech. Therefore, the present study aims to add to the literature by investigating the following hypotheses:

H1. Witnessing hate speech and peer pressure will be positively linked to hate speech perpetration and injunctive anti-hate speech norms will be negatively related to hate speech perpetration.

H2. Adolescents who witness hate speech in schools will be less likely to perpetrate hate speech if they report higher levels of injunctive anti-hate speech norms compared to those who report lower levels of injunctive anti-hate speech norms.

H3. Adolescents who witness hate speech in schools will be more likely to perpetrate hate speech if they report higher levels of deviant peer pressure compared to those who report lower levels of deviant peer pressure.

Method

Participants

The participants were 1719 adolescents (45.1% male, 53.6% female 1.3% gender diverse) from 138 classes in 22 schools, located across two German-speaking cantons in Switzerland. Participants were between 11 and 18 years old ($M_{\text{age}} = 13.96$; $SD = .98$). Participants were in grades 7–9 (seventh grade: 27.7%, $n = 477$; eighth grade: 26.6%, $n = 457$; ninth grade: 26.2%, $n = 450$). In addition, 335 participants (19.5%) were in mixed grades in which students between 14 and 16 years' old were taught together. Slightly less than the half of the participants (48.3%; $n = 831$) had a migration background (the measure of ethnic minority status in German-speaking countries; a description of how it is measured is given below in "Demographics"). Concerning socio-economic status (SES), 33.4% ($n = 574$) of participants were living in families of low affluence, 33.9% ($n = 583$) in families of medium affluence, and 32.7% ($n = 562$) in families of high affluence.

Measures

As part of a larger study, participants completed measures on hate speech witnessing and perpetration, injunctive anti-hate speech norms, deviant peer pressure, and demographics.

Hate speech witnessing and perpetration. Based on prior qualitative research (Ballaschk et al., 2021; Krause et al., 2021), we developed a new instrument to measure involvement in hate speech as witness and/or perpetrator. The participants were first shown a short video presenting a definition of hate speech. The text-based definition reads as follows:

“Hate speech is directed against groups of people because of, for example, their origin, skin color, religion, sexual orientation, or gender. Hate speech is offensive and purposefully hurtful. Sometimes hate speech is aimed directly at someone who is present because they belong to a certain group. However, hate speech can also be expressed without directly referring to a person who is present. Hate speech takes place in public. That is, it takes place in front of people who can hear it. Hate speech does not always have to be expressed in words. We also talk about hate speech when groups are insulted by graffiti of offensive symbols (e.g., the swastika).”

For witnessing hate speech, we asked: “*In the past 12 months, how often have you witnessed hate speech at your school?*” And for perpetration: “*In the past 12 months, how often have you perpetrated hate speech at your school?*” Answer options were: *never* (1), *one or two times within the last month* (2), *two or three times per month* (3), *about one time a week* (4), and *several times a week* (5).

Injunctive anti-hate speech norms. Perceived injunctive anti-hate speech norms were measured by three items, which the participants were asked to rate after reading a text-based vignette that described a hate speech scenario. The three items were: “*My close friends don’t like it when you say that about other people,*” “*In my family, they don’t like it when you say this about other people,*” and “*The teachers at my school don’t like it when you say that about other people.*” All items were answered on a five-point scale from *absolutely disagree* (1) to *absolutely agree* (5). The McDonald’s ω was acceptable: .74.

Deviant Peer Pressure. To measure susceptibility to deviant peer pressure, the following four items were modified from Santor et al. (2000) by referring to classmates and not to general friends and groups: “*At times, I’ve broken rules because classmates have urged me to,*” “*At times, I’ve done dangerous or foolish things because classmates dared me to,*” “*I’ve skipped classes when classmates have urged me to,*” and “*If a group of classmates at school asked me to do something forbidden, it would be difficult to say no.*” All items were

answered on a five-point scale from *absolutely disagree* (1) to *absolutely agree* (5). The McDonald's ω was acceptable: .78.

Demographics. Participants were asked about their age, grade, and gender (male, female, and diverse). Migration background was assessed by asking whether the participants themselves and/or one parent or both were born in another country other than Switzerland. Socio-economic status (SES) was measured using the Family Affluence Scale (FAS; [Hartley et al., 2016](#)). Participants answered six items regarding whether they had a dishwasher at home, a family car, whether children had their own bedroom, the number of bathrooms at home, the number of computers at home, and the number of holidays taken in the year 2019. Based on a composite FAS score, an individual FAS category was calculated for each participant (low, medium, and high socioeconomic status).

Procedure

Approval for the study was obtained from the University of Potsdam Ethics Committee (UP65/2018). In total, 24 schools from Switzerland from two German-speaking cantons were selected based on their migration background (high vs. low) and their geographical location (municipal vs. rural) to achieve a representative sample of the young Swiss population, according to these criteria. These 24 schools were contacted via phone calls and email and 22 schools agreed to participate in the study (participation rate at the school level: 93%). All students in grades 7–9 from these 22 schools were asked to participate in the study. In total, 175 classes were invited to take part and 138 classes took part (participation rate at the class level: 79%). From 3,341 students, 1719 participated in the present study (participation rate at the individual level: 51%). The students and their parent(s)/legal guardians had to sign a written consent letter in order to participate in the study. Data collection took place online via an online survey during school time. Participants were told that partaking in the study was optional and participation in the survey could be stopped at any time, without the need for giving a reason and without fear of negative consequences. Completion time was 37 minutes on average.

A priori conducted power analysis with G*Power ([Faul et al., 2007](#)) revealed that for detecting small to medium correlational effect sizes the present study needed a sample consisting of 1944 students in 108 classes at 18 schools ([Teerenstra et al., 2010](#)). Accordingly, the present sample size was sufficient to investigate the hypotheses.

Data Analysis

Before testing our hypotheses, descriptive statistics and bivariate correlations were computed to investigate the study’s variables. We conducted multilevel regression analyses in Mplus 8 (Muthén & Muthén, 2018) to test the study’s hypotheses. A two-level model with random intercept was explored with students (individual level; L1) nested within classrooms (classroom level; L2). Due to the skewed outcome variable, namely, hate speech perpetration, the maximum likelihood robust (MLR) estimator was used (Muthén & Muthén, 1998-2017). Initially a baseline model that did not contain any predictors at both levels was used to estimate the intra-class correlations (ICC; Model 0). Then we estimated Model 1 and entered only control variables, namely, age, gender, SES, and country of origin on the student level. On the classroom level, we included grade as a dummy-coded variable and used mixed grade as a reference category for seventh, eighth, and ninth grade. In Model 2, we added witnessing hate speech, injunctive anti-hate speech norms, and deviant peer pressure on the student level to the model with control variables. Finally, in Model 3 we added the interaction terms between witnessing hate speech and injunctive anti-hate speech norms as well as the interaction between and witnessing hate speech and deviant peer pressure to the model. We evaluated the model fit of the multilevel models using the Akaike information criterion (AIC) indices, according to which lower AIC values indicate a better model fit (Akaike, 1974).

Results

Preliminary Analyses

All correlations were in the expected direction (see Table 1). Higher levels of perpetrating hate speech were positively correlated with witnessing hate

Table 1. Means, Standard Deviations, and Bivariate and Partial Correlations between Witnessing Hate Speech, Perpetrating Hate Speech, Injunctive Anti-Hate Speech Norms, and Deviant Peer Pressure.

	1	2	3	4
1. Perpetrating hate speech	–	.42***	–.13***	.16***
2. Witnessing hate speech	.40***	–	–.05*	.12***
3. Injunctive anti-hate speech norms	–.18***	–.05*	–	–.10***
4. Deviant peer pressure	.20***	.12***	–.15***	–
M	1.39	2.60	4.24	1.79
SD	.85	1.30	.90	.74

Note. *** $p < .001$ * $p < .01$ * $p < .05$. $N = 1719$. Zero-order correlation (below the diagonal) and partial correlation controlling for gender, age, migration background, and socioeconomic status (above the diagonal).

speech ($r_{(1719)} = .40, p < .001$) and deviant peer pressure ($r_{(1719)} = .20, p < .001$), and negatively correlated with injunctive anti-hate speech norms ($r_{(1719)} = -.18, p < .001$). Higher levels of witnessing hate speech were negatively correlated with injunctive anti-hate speech norms ($r_{(1719)} = -.05, p = .039$) and positively correlated with peer pressure ($r_{(1719)} = .12, p < .001$). Finally, higher levels of injunctive anti-hate speech norms were negatively correlated with peer pressure ($r_{(1719)} = -.15, p < .001$). In addition, partial correlation controlling for age, gender, migration background, and socio-economic status were investigated (see Table 1). The findings were relatively similar to the zero-order correlations.

Multilevel Analyses

The analysis of the baseline model (Model 0) revealed that for hate speech perpetration the ICC was .136, indicating that 13.6% of the variance could be explained due to class differences and 86.4% due to individual differences. In Model 1, we included the control variables, namely, age, gender, migration background, and SES, to predict hate speech perpetration and the between-class variation decreased to 10% ($ICC = .101$). In Model 2, we entered witnessing hate speech, injunctive anti-hate speech norms, and deviant peer pressure to the model and the between-class variation decreased to 6% ($ICC = .055$). Finally, in Model 3, we included the both interaction effects (witnessing hate speech X injunctive anti-hate speech norms and witnessing hate speech X deviant peer pressure). The between-class variation was 5% ($ICC = .050$). With every step, the model fit improved as can be shown by the decreasing AIC indices (Model 1: 4156.43; Model 2: 3821.47; Model 3: 3772.60). In addition, the R^2_{within} increased (Model 1: .037; Model 2: .230; Model 3: .252), clearly indicating that Model 3 was compared to Model 1 and Model 2 the best fitting model (see Table 2). Hence, Model 3 was used for the subsequent analysis.

On the individual level, witnessing hate speech ($\hat{\beta} = .39, p < .001$) was positively associated with hate speech perpetration. In addition, injunctive anti-hate speech norms ($\hat{\beta} = -.11, p < .001$) were negatively related to hate speech perpetration and deviant peer pressure ($\hat{\beta} = .09, p < .001$) was positively linked to hate speech perpetration. Therefore, the findings support our first hypothesis that peer pressure and witnessing hate speech were positively associated and injunctive anti-hate speech norms negatively linked to hate speech perpetration in schools (see Table 2). A significant interaction was found between witnessing hate speech and injunctive anti-hate speech norms, when predicting online hate speech perpetration ($\hat{\beta} = -.07, p = .040$). Probing the significant interaction effect further revealed that the unstandardized simple slope coefficients were $b_{\text{simple}} = .31$ ($SE = .03, p < .001$, at $-1 SD$) for low, $b_{\text{simple}} = .26$ ($SE = .02, p < .001$, at $0 SD$) for moderate, and $b_{\text{simple}} = .21$

Table 2. Multilevel Model for Hate Speech Perpetration in Schools.

Estimator	Model 1			Model 2			Model 3		
	$\hat{\beta}$	SE	<i>p</i>	$\hat{\beta}$	SE	<i>p</i>	$\hat{\beta}$	SE	<i>p</i>
Individual level (L1)									
Age	.05	.04	.192	.05	.03	.118	-.04	.03	.183
Gender ^{boys}	-.15	.03	<.001	-.12	.03	<.001	-.12	.03	<.001
Migration background ^{yes}	.09	.03	<.001	.07	.02	.008	.07	.02	.004
SES	.07	.02	.005	.04	.02	.087	.04	.02	.104
Hate speech witnessing (HSW)				.40	.02	<.001	.39	.03	<.001
Injunctive anti-hate speech norms (IN)				-.11	.03	<.001	-.11	.03	<.001
Deviant peer pressure (DPP)				.11	.03	<.001	.09	.03	<.001
HSW × IN							-.07	.04	.040
HSW × DPP							.14	.03	<.001
Classroom level (L2)									
7 th grade ^{yes}	-.02	.18	.898	-.15	.27	.583	.05	.20	.811
8 th grade ^{yes}	-.01	.19	.960	-.08	.27	.997	-.04	.20	.838
9 th grade ^{yes}	.01	.23	.979	-.11	.29	.728	-.05	.25	.851
R ² (L1)	.037	.01	<.001	.230	.02	<.001	.252	.03	<.001
R ² (L2)	.001	.01	.924	.017	.04	.697	.008	.03	.773
AIC	4156.43			3821.47			3772.60		
ICC	.101			.055			.050		

Note. Continuous variables were z-standardized at the individual level. Reference category gender: girls; reference category migration background: no; reference category grade: mixed classroom.

(*SE* = .03, *p* < .001, at +1 *SD*) for high levels of injunctive anti-hate speech norms, indicating that the positive association between witnessing and perpetrating hate speech was weakened as the perception of injunctive anti-hate speech norms increased (see Figure 1). Hence, the findings support our second hypothesis that witnesses of hate speech were less likely to perpetrate hate speech if they reported higher levels of injunctive anti-hate speech norms compared to those who reported lower levels of injunctive anti-hate speech norms. Furthermore, a significant moderation effect was found between witnessing hate speech and deviant peer pressure when predicting hate speech perpetration ($\hat{\beta}$ = .14, *p* < .001). Probing the significant interaction effect further revealed that the unstandardized simple slope coefficients were *b*_{simple} = .25 (*SE* = .04, *p* < .001, at −1 *SD*) for low, *b*_{simple} = .38 (*SE* = .03, *p* < .001, at 0 *SD*) for moderate, and *b*_{simple} = .51 (*SE* = .04, *p* < .001, at +1 *SD*) for high levels of

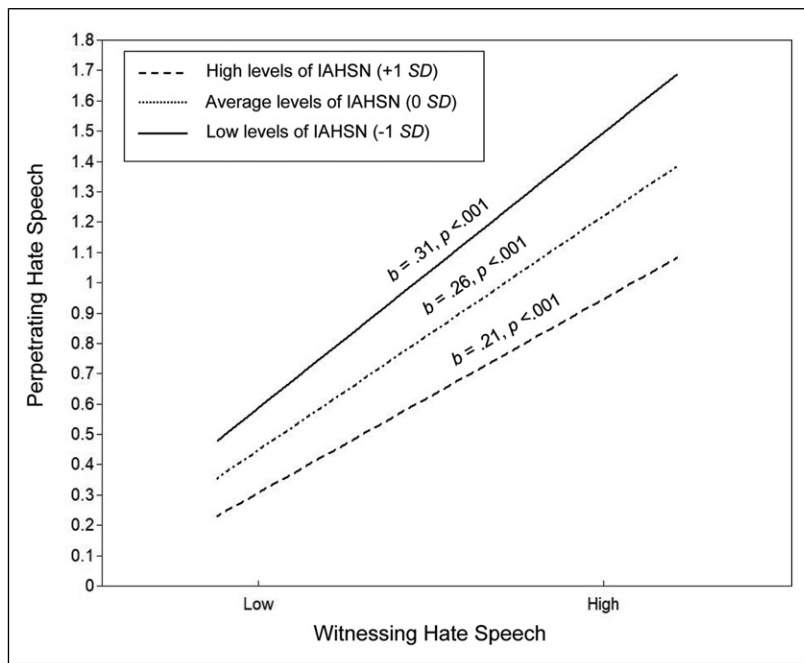


Figure 1. Graphical representation of the moderation of injunctive anti-hate speech norms (IAHSN) on the association between witnessing and perpetrating hate speech.

perceived peer pressure, suggesting that the positive relation between hate speech witnessing and perpetrating was strengthened as the perception of deviant peer pressure increased (see Figure 2). In conclusion, the findings support our third hypothesis that witnesses of hate speech in schools were more likely to perpetrate hate speech if they reported higher levels of deviant peer pressure compared to those who reported lower levels of deviant peer pressure. Two control variables were significant. More specifically, being a boy ($\hat{\beta} = -.12, p < .001$) predicted lower hate speech perpetration, while having a migration background ($\hat{\beta} = .07, p < .001$) predicted higher hate speech perpetration. No significant association was found between age and hate speech perpetration and SES and hate speech perpetration. On the classroom level, grade had no influence on hate speech perpetration. The multilevel model explained approximately 25% of the variance on the individual level and 1% of the variance on classroom level.

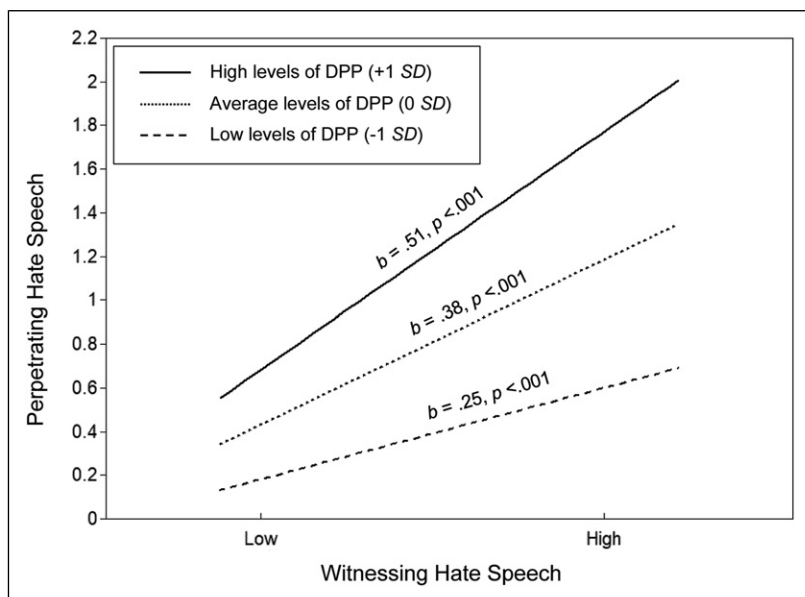


Figure 2. Graphical representation of the moderation of deviant peer pressure (DPP) on the association between witnessing and perpetrating hate speech.

Discussion

Based on a large sample of 1719 Swiss adolescents, the present study sought to understand hate speech perpetration among students in schools. More specifically, the current investigation explored the association between witnessing hate speech, social norms, and perpetrating hate speech. Furthermore, we investigated the moderation effects of social norms (i.e., injunctive anti-hate speech norms, deviant peer pressure) in the witnesses–perpetrators relationship.

After controlling for adolescents' age, gender, migration background, and SES, we found support for our first hypothesis that witnessing and perpetrating hate speech were positively related. Following the SCT, this finding can be explained through observational learning, which posits that the development of deviant behavior among adolescents can be partially explained by exposure to socially deviant role models. This finding is also in line with previous research on online hate speech and research on different forms of aggressive behavior (Bernatzky et al., 2021; Evans et al., 2019; Jung et al., 2019; Rivers et al., 2009; Wachs et al., 2021; Wachs & Wright, 2018).

The findings of the present study also confirmed that social norms were significantly correlated with adolescents' engagement in hate speech perpetration. More concretely, injunctive anti-hate speech norms were negatively

associated and deviant peer pressure positively associated with adolescents' hate speech perpetration, lending further support to our first hypothesis. These findings support initial qualitative research on the relevance of social norms for hate speech perpetration (Ballaschk et al., 2021) and research on the link between social norms and deviant behavior among adolescents (Bastiaensens et al., 2016; Cook et al., 2010; Cohen & Prinstein, 2006; Fluck, 2017; Henneberger et al., 2020; Jung et al., 2019; Pozzoli & Gini, 2010; Ruggeri et al., 2018; Van de Bongardt et al., 2015; Van Hoorn et al., 2017; Wilton & Campbell, 2011). A possible explanation for this finding might be that—as put forward by the SCT—adolescents' perception of (un)accepted, (un)supported or (un)desired rules, beliefs, attitudes, and behaviors influence their motivation to show a certain behavior to avoid penalties or receive rewards by breaking the rules or following them, respectively. That is to say, adolescents' behavior in relation to hate speech is partially guided by social norms and their willingness to adapt to these norms. Follow-up research is needed to understand whether social norms also influence other hate speech roles' (e.g., reinforcers, defenders), reactions to hate speech, and how these roles are transferred among students in schools.

This study also showed that considering social norms contributes to our understanding of the association between witnessing and perpetrating hate speech in school. We found support for our second hypothesis that witnesses of hate speech in schools were less likely to perpetrate hate speech if they reported higher levels of injunctive anti-hate speech norms compared to those who reported lower levels of injunctive anti-hate speech norms. This finding suggests that, if witnesses of hate speech believe that hate speech is violating social norms, then they are less likely to engage in it. In accordance with our third hypothesis, adolescents who witness hate speech in schools were more likely to perpetrate hate speech if they reported higher levels of deviant peer pressure, compared to those who reported lower levels of deviant peer pressure. Overall, these results are supported by the SCT, which highlights the crucial role of the social environment in people's behavior.

The results illustrate, that witnesses consider social norms and adapt themselves to social norms prior to taking action. These findings help us to understand the complex link between witnessing and perpetrating hate speech by highlighting the relevance of witnesses' perception of their social environment in hate speech perpetration. Our findings extend current knowledge on intrapersonal factors (e.g., coping strategies, online disinhibition effect), which have been shown to moderate the witnesses–perpetrator relationship in online hate speech (Wachs & Wright, 2018; Wachs et al., 2019). In conclusion, both individual and contextual factors need to be taken into consideration to understand the witness–perpetrator hate speech relation among adolescents.

Practical Implications

To begin with, this study indicates that hate speech should be considered as a group phenomenon. That is to say, prevention programs that seek to understand hate speech exclusively as an individual phenomenon, and neglect the social ecology of this problem behavior, are falling short. To add to this, hate speech perpetration is influenced by adolescents' perceptions of social norms around them. Hence, to prevent hate speech perpetration, actively counteracting the formation of pro-hate speech peer norms, promoting resistance to negative peer influences, and providing positive role models might be fruitful directions for future prevention endeavors. In addition to this, the creation of a classroom climate in which all students, regardless of their religion, sexual orientation, ethnicity, or ability, feel valued, encouraged, and included might help to promote anti-hate speech peer norms and help students resist negative peer influences that encourage hate speech, ultimately reducing hate speech. Reflecting on normative notions and collaborative learning may allow intergroup interactions and alleviate biased in- and out-group dynamics within the classroom, both of which might help to achieve a more positive school climate (Hong & Garbarino, 2012).

The risk of hate speech witnessing should not be underestimated, as this study showed a strong association between witnessing and perpetrating hate speech. Furthermore, research showed that, with increasing exposure to hate speech, people can become desensitized to derogatory communication (Soral et al., 2018) and that hate speech in school classes initiated by individual students can become part of the class culture and can gain momentum (Ballaschk et al., 2021). Therefore, schools need to develop a comprehensive approach to combating hate speech in schools by working together with parents, teachers and staff, students, and community groups. Any anti-hate speech policies should be examined routinely to determine accuracy, clarity, and legal compliance, and such policies should be communicated to school personnel, students, and parents. While having effective and clear policies is important, so too is having clear intervention plans to minimize the disruption of the educational process by incidents of hate speech.

Limitations and Future Research

Although the present study is a step toward a better comprehension of the correlates and theoretical frameworks of hate speech perpetration among adolescents in schools, a few limitations of the current study must be mentioned and warrant attention in future research. Firstly, due to the cross-sectional research design, it is impossible to understand the temporal ordering of the main study variables. Longitudinal studies would be needed to determine the temporal ordering of the links investigated in the current study.

Secondly, the present study relies exclusively on self-reports, and as a result, social desirability bias likely influenced the answers that students gave. Follow-up research should try to replicate the present findings by using a combination of peer-, parents-, teacher-, and self-reports, which might significantly increase the validity of the findings. Thirdly, we were not able to investigate all assumptions of the SCT but focused on a few aspects. Indeed, one limitation of the SCT is that it is impossible to address all aspects of this theory in one study. Hence, more research is needed that operationalizes other aspects of the SCT (e.g., cognitive processes such as attention, retention, and reproduction). In addition, follow-up research should also test the usefulness of other theories to explain hate speech perpetration (e.g., Self Determination Theory; Deci & Ryan, 2012). Fourthly, we used only single items for measuring hate speech involvement. Follow-up studies need to develop and validate scales to overcome typical methodological issues that are aligned to using single-item measures (e.g., validity, accuracy, and reliability). In addition, our single-item measures did not allow to draw any conclusions on whether witnessing hate speech was related to hate speech directed at oneself or classmates. In the same line, the way how hate speech perpetration was measured gives no information on whether the students engaged in hate speech because they observed it or maybe as a mean of retaliation. To address this limitation partially, analyses were repeated with hate speech victimization as further control variable. The findings remained the same (results can be requested by the first author).

Fifthly, we used different instruments for measuring injunctive anti-hate speech norms and deviant peer pressure. In the same vein, injunctive anti-hate speech norms considered several groups of significant others (i.e., peers, family members, and teachers), whereas deviant peer pressure was only related to the peers. These methodological differences do not allow us to compare the varying effects. Follow-up research should use similar scales with similar items for measuring the different forms of social norms to understand and compare potentially varying effects. Such research could help to design prevention programs, as it might give us information on which social norms need to be focused on in anti-hate speech prevention programs. And, finally, we measured adolescents' susceptibility to deviant peer influence rather than the levels of peer pressure that exists.

Conclusion

This study investigates the association between witnessing and perpetrating hate speech in schools and the moderating effects of social norms in this relationship. The results show that witnessing and perpetrating hate speech are positively related. Injunctive anti-hate speech norms were negatively associated and deviant peer pressure positively associated with hate speech perpetration.

Moreover, the findings confirm that injunctive anti-hate speech norms weaken and deviant peer pressure strengthens this relation. The findings support the need to focus on social dynamics and social norms in anti-hate speech prevention programs. Follow-up studies that allow the comparison of the varying effects of social norms on adolescents' hate speech engagement are needed. Moreover, more research is needed that clarifies the most effective strategies to implement anti-hate speech norms and more information on how these norms are effectively transmitted in the school context.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the German Research Foundation (DFG), grant number WA 4275/2-1; BI 1046/9-1; SCHU 1370/4-1 and the University of Teacher Education Bern, grant number 19s 0008 01.

ORCID iDs

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

Ludwig Bilz  <https://orcid.org/0000-0003-3017-0492>

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723. <https://doi.org/10.1109/TAC.1974.1100705>.
- Ballaschk, C., Wachs, S., Krause, N., Schulze-Reichelt, F., Kansok-Dusche, J., Bilz, L., & Schubarth, W. (2021). “Dann machen halt alle mit.” Eine qualitative Studie zu Beweggründen und Motiven für Hatespeech unter Schüler*innen. [“Then everyone just goes along with it.” A qualitative study on reasons and motives of hate speech among students]. *Diskurs Kindheits- und Jugendforschung/Discourse Journal of Childhood and Adolescence Research*, 16(4), 1–18. <https://doi.org/10.3224/diskurs.v16i4.01>.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, 44(9), 1175–1184. <https://doi.org/10.1037/0003-066X.44.9>.
- Bandura, A., & McDonald, F. J. (1963). Influence of social reinforcement and the behavior of models in shaping children's judgment. *The Journal of Abnormal and Social Psychology*, 67(3), 274–281. <https://doi.org/10.1037/h0044714>.

- Bandura, A., Ross, D., & Ross, S. A. (1963). Imitation of film-mediated aggressive models. *The Journal of Abnormal and Social Psychology*, 66(1), 3–11. <https://doi.org/10.1037/h0048687>.
- Bastiaensens, S., Pabian, S., Vandebosch, H., Poels, K., Van Cleemput, K., DeSmet, A., & De Bourdeaudhuij, I. (2016). From normative influence to social pressure: How relevant others affect whether bystanders join in cyberbullying. *Social Development*, 25(1), 193–211. <https://doi.org/10.1111/sode.12134>.
- Bernatzky, C., Costello, M., & Hawdon, J. (2021). Who produces online hate?: An examination of the effects of self-control, social structure, & social learning. *American Journal of Criminal Justice*, 1–20. <https://doi.org/10.1007/s12103-020-09597-3>.
- Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 151–192). McGraw-Hill.
- Cohen, G. L., & Prinstein, M. J. (2006). Peer contagion of aggression and health risk behavior among adolescent males: An experimental investigation of effects on public conduct and private attitudes. *Child Development*, 77(4), 967–983. <https://doi.org/10.1111/j.1467-8624.2006.00913.x>.
- Cook, C. R., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly*, 25(2), 65–83. <https://doi.org/10.1037/a0020149>.
- Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 416–436). Sage Publications Ltd. <https://doi.org/10.4135/9781446249215.n21>.
- Döring, N., & Mohseni, M. R. (2020). Gendered hate speech in YouTube and YouNow comments: Results of two content analyses. *SCM Studies in Communication and Media*, 9(1), 62–88. <https://doi.org/10.5771/2192-4007-2020-1-62>.
- Evans, C. B. R., Smokowski, P. R., Rose, R. A., Mercado, M. C., & Marshall, K. J. (2019). Cumulative bullying experiences, adolescent behavioral and mental health, and academic achievement: An integrative model of perpetration, victimization, and bystander behavior. *Journal of Child and Family Studies*, 28(9), 2415–2428. <https://doi.org/10.1007/s10826-018-1078-4>.
- Farrell, A. D., Thompson, E. L., & Mehari, K. R. (2017). Dimensions of peer influences and their relationship to adolescents' aggression, other problem behaviors and prosocial behavior. *Journal of Youth and Adolescence*, 46(6), 1351–1369. <https://doi.org/10.1007/s10964-016-0601-4>.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>.

- Fend, H. (1974). *Gesellschaftliche Bedingungen schulischer Sozialisation*. Weinheim [u.a.]: Beltz.
- Fluck, J. (2017). Why do students bully? An analysis of motives behind violence in schools. *Youth & Society*, 49(5), 567–587. <https://doi.org/10.1177/0044118X14547876>.
- Harriman, N., Shortland, N., Su, M., Cote, T., Testa, M. A., & Savoia, E. (2020). Youth exposure to hate in the online space: An exploratory analysis. *International Journal of Environmental Research and Public Health*, 17(22), 8531. <https://doi.org/10.3390/ijerph17228531>.
- Hartley, J. E. K., 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>.
- Henneberger, A. K., Mushonga, D. R., & Preston, A. M. (2020). Peer influence and adolescent substance use: A systematic review of dynamic social network research. *Adolescent Research Review* 6(9), 57–73. <https://doi.org/10.1007/s40894-019-00130-0>.
- Hong, J. S., & Garbarino, J. (2012). Risk and protective factors for homophobic bullying in schools: An application of the social–ecological framework. *Educ Psychol Rev*, 24, 271–285. <https://doi.org/10.1007/s10648-012-9194-y>.
- Jung, J., Busching, R., & Krahé, B. (2019). Catching aggression from one's peers: A longitudinal and multilevel analysis. *Social and Personality Psychology Compass*, 13(2), e12433. <https://doi.org/10.1111/spc3.12433>.
- Krause, N., Ballaschk, C., Schulze-Reichelt, F., Kansok-Dusche, J., Wachs, S., Schubarth, W., & Bilz, L. (2021). “Ich lass mich da nicht klein machen!” Eine qualitative Studie zur Bewältigung von Hatespeech durch Schüler/innen. [“I don't let them get me down!”—A qualitative study on students' coping with hate speech] *Zeitschrift für Bildungsforschung*, 11(1), 169–185. <https://doi.org/10.1007/s35834-021-00291-w>.
- Lee, S. S. T., & Wong, D. S. W. (2009). School, parents, and peer factors in relation to Hong Kong students' bullying. *International Journal of Adolescence and Youth*, 15(3), 217–233. <https://doi.org/10.1080/02673843.2009.9748030>.
- Lehman, B. (2019). Stopping the hate: Applying insights on bullying victimization to understand and reduce the emergence of hate in schools. *Sociological Inquiry*, 89(3), 532–555. <https://doi.org/10.1111/soin.12296>.
- Lehman, B. (2020). Hate at school: Victimization and disorder associated with school avoidance. *Sociological Spectrum*, 40(3), 172–190. <https://doi.org/10.1080/02732173.2020.1734890>.
- Muthén, L.K., & Muthén, B.O. (1998-2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, B., & Muthén, L. (2018). *Mplus 8.1. [Computer Software]*. Los Angeles, CA: Muthén and Muthén.

- Pozzoli, T., & Gini, G. (2010). Active defending and passive bystanding behavior in bullying: The role of personal characteristics and perceived peer pressure. *Journal of Abnormal Child Psychology*, 38(6), 815–827. <https://doi.org/10.1007/s10802-010-9399-9>.
- Rivers, I., Poteat, V. P., Noret, N., & Ashurst, N. (2009). Observing bullying at school: The mental health implications of witness status. *School Psychology Quarterly*, 24(4), 211–223. <https://doi.org/10.1037/a0018164>.
- Ruggeri, A., Luan, S., Keller, M., & Gummerum, M. (2018). The influence of adult and peer role models on children' and adolescents' sharing decisions. *Child Development*, 89(5), 1589–1598. <https://doi.org/10.1111/cdev.12916>.
- Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior*, 15(2), 112–120. <https://doi.org/10.1016/j.avb.2009.08.007>.
- Santor, D. A., Messervey, D., & Kusumakar, V. (2000). Measuring peer pressure, popularity, and conformity in adolescent boys and girls: Predicting school performance, sexual attitudes, and substance abuse. *Journal of Youth and Adolescence*, 29(2), 163–182. <https://doi.org/10.1023/A:1005152515264>.
- Schwarzenthal, M., Schachner, M. K., Juang, L. P., & Van De Vijver, F. J. (2020). Reaping the benefits of cultural diversity: Classroom cultural diversity climate and students' intercultural competence. *European Journal of Social Psychology*, 50(2), 323–346. <https://doi.org/10.1002/ejsp.2617>.
- Sijtsema, J. J., & Lindenberg, S. M. (2018). Peer influence in the development of adolescent antisocial behavior: Advances from dynamic social network studies. *Developmental Review*, 50, 140–154. <https://doi.org/10.1016/j.dr.2018.08.002>.
- Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, 44(2), 136–146. <https://doi.org/10.1002/ab.21737>.
- Steinberg, L., & Morris, A. S. (2001). Adolescent development. *Annual Review of Psychology*, 52, 83–110. <https://doi.org/10.1146/annurev.psych.52.1.83>.
- Teerenstra, S., Lu, B., Preisser, J. S., van Achterberg, T., & Borm, G. F. (2010). Sample size considerations for GEE analyses of three-level cluster randomized trials. *Biometrics*, 66(4), 1230–1237. <https://doi.org/10.1111/j.1541-0420.2009.01374.x>.
- Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, 83(3), 357–385. <https://doi.org/10.3102/0034654313483907>.
- Van de Bongardt, D., Reitz, E., Sandfort, T., & Deković, M. (2015). A meta-analysis of the relations between three types of peer norms and adolescent sexual behavior. *Personality and Social Psychology Review*, 19(3), 203–234. <https://doi.org/10.1177/1088868314544223>.
- Van Hoom, J., Crone, E. A., & Van Leijenhorst, L. (2017). Hanging out with the right crowd: Peer influence on risk-taking behavior in adolescence. *Journal of Research on Adolescence*, 27(1), 189–200. <https://doi.org/10.1111/jora.12265>.

- Wachs, S., Mazzone, A., Milosevic, T., Wright, M. F., Blaya, C., Gámez-Guadix, M., & O'Higgins Norman, J. (2021). Online correlates of cyberhate involvement among young people from ten European countries: An application of the routine activity and problem behaviour theory. *Computers in Human Behavior*, 123, 106872. <https://doi.org/10.1016/j.chb.2021.106872>.
- Wachs, S., Schubarth, W., & Bilz, L. (2020). Hate speech als schulproblem? Erziehungswissenschaftliche perspektiven auf ein aktuelles phänomen. In I van Ackeren, H. Bremer, F. Kessler, H. C. Koller, N. Pfaff, C. Rotter, D. Klein, & U. Salaschek (Eds.). *Bewegungen – Beiträge aus dem 26. Kongress der Deutschen Gesellschaft für Erziehungswissenschaft* (pp. 223–236). Verlag Barbara Budrich. <https://doi.org/10.2307/j.ctv10h9fjc.19>.
- 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., Sittichai, R., Singh, R., Biswal, R., Kim, E., Yang, S., Gámez-Guadix, M., Almendros, C., Flora, K., Daskalou, V., & Maziridou, E. (2019). Associations between witnessing and perpetrating online hate in eight countries: The buffering effects of problem-focused coping. *International Journal of Environmental Research and Public Health*, 16(20), 3992. <https://doi.org/10.3390/ijerph16203992>.
- Wilton, C., & Campbell, M. (2011). An exploration of the reasons why adolescents engage in traditional and cyber bullying. *Journal of Educational Sciences and Psychology*, 1(2), 101–109.

Author Biographies

Dr. Sebastian Wachs is a Deputy Professor for Education and Socialization Theory at the Department of Educational Studies at the University of Potsdam and honorary research fellow at the National Anti-Bullying Research and Resource Centre. Sebastian has led several international projects on emotional and social development of adolescents, adjustment, well-being, risky behavior, and aggression (e.g., bullying in schools and cyberspace, hate speech, cybergrooming), with particular attention to development in context (e.g., family, school, and culture).

Prof. Dr. Alexander Wettstein is a Professor of Educational Psychology and Head of the Research Group 'Social Interaction in Educational Settings' at the University of Teacher Education Bern. He is working on the ontogenesis of aggressive and prosocial behavior, peer group influence, teacher-student relationships, psychobiological teacher stress, applying ambulatory assessment approaches.

Prof. Dr. Ludwig Bilz, is a Professor of Educational Psychology at the Brandenburg University of Technology Cottbus-Senftenberg. Main areas of work: school violence, bullying, school mental health, health behavior in school-aged children, teacher competence, teacher health, school health promotion, school violence prevention.

Norman Krause, M.A., is a Research Assistant and PhD student at the University of Potsdam. His research focuses on hate speech and discrimination among adolescents and how adolescents cope with it.

Cindy Ballaschk, M.A., is a Research Assistant at the University of Potsdam and PhD student at the University of Innsbruck. Her research focuses on hate speech and discrimination among adolescents, sex education, intersectionality and queer studies.

Julia Kansok-Dusche, M.A., is a Research Assistant and PhD student at the Brandenburg University of Technology Cottbus-Senftenberg. Her research focuses on discrimination and aggression (e.g., hate speech) among adolescents and on school-based approaches (e.g., teacher interventions) to deal with these issues.

Michelle F. Wright, PhD, is a Research Associate at the Pennsylvania State University and DePaul University. Her research focus is on the contextual factors which influence children's, adolescents', and young adults' involvement in aggressive behaviors, with a special interest in social goals, peer status, and cultural values. She has published on these topics, with her most recent work focused on culture and anonymity, and their role in cyberbullying among adolescents.