







# Adolescents' motivations to perpetrate hate speech and links with social norms

## Motivos del discurso de odio en la adolescencia y su relación con las normas sociales

-  Dr. Sebastian Wachs. Research Associate, Department of Educational Studies, University of Potsdam (Germany) (wachs@uni-potsdam.de) (<https://orcid.org/0000-0003-2787-6646>)
-  Dr. Alexander Wettstein. Professor, Institute for Research, Development and Evaluation, Bern University of Teacher Education (Switzerland) (alexander.wettstein@phbern.ch) (<https://orcid.org/0000-0002-5612-3144>)
-  Dr. Ludwig Bilz. Professor, Department of Health Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Senftenberg (Germany) (ludwig.bilz@b-tu.de) (<https://orcid.org/0000-0003-3017-0492>)
-  Dr. Manuel Gámez-Guadix. Professor, Department of Biological and Health Psychology, Autonomous University of Madrid (Spain) (manuel.gomez@uam.es) (<https://orcid.org/0000-0002-1575-1662>)

### ABSTRACT

Hate speech has become a widespread phenomenon, however, it remains largely unclear why adolescents engage in it and which factors are associated with their motivations for perpetrating hate speech. To this end, we developed the multidimensional "Motivations for Hate Speech Perpetration Scale" (MHATE) and evaluated the psychometric properties. We also explored the associations between social norms and adolescents' motivations for hate speech perpetration. The sample consisted of 346 adolescents from Switzerland (54.6% boys;  $M_{age}=14$ ;  $SD=0.96$ ) who reported engagement in hate speech as perpetrators. The analyses revealed good psychometric properties for the MHATE, including good internal consistency. The most frequently endorsed subscale was revenge, followed by ideology, group conformity, status enhancement, exhilaration, and power. The results also showed that descriptive norms and peer pressure were related to a wide range of different motivations for perpetrating hate speech. Injunctive norms, however, were only associated with power. In conclusion, findings indicate that hate speech fulfills various functions. We argue that knowing the specific motivations that underlie hate speech could help us derive individually tailored prevention strategies (e.g., anger management, promoting an inclusive classroom climate). Furthermore, we suggest that practitioners working in the field of hate speech prevention give special attention to social norms surrounding adolescents.

### RESUMEN

El discurso de odio se ha convertido en un fenómeno generalizado. Sin embargo, todavía no está claro por qué los adolescentes se involucran en el discurso de odio y qué factores están asociados con las motivaciones para perpetrarlo. Con esta finalidad, desarrollamos una medida multidimensional, la «Escala de Motivaciones para Perpetrar Discurso de Odio» (MHATE), y evaluamos sus propiedades psicométricas. Asimismo, investigamos las asociaciones entre las normas sociales y las motivaciones para participar en el discurso de odio. La muestra estuvo compuesta por 346 adolescentes suizos (54,6% chicos;  $M_{edad}=14$ ;  $DT=0,96$ ) que informaron haber perpetrado discurso de odio. Los análisis revelaron buenas propiedades psicométricas de MHATE, incluyendo adecuada consistencia interna. La subescala con mayor frecuencia fue venganza, seguida de las de ideología, conformidad con el grupo, mejora del estatus, regocijo y poder. Las normas descriptivas y la presión de iguales estuvieron relacionadas con varias motivaciones para perpetrar discurso de odio. Las normas prescriptivas, sin embargo, solo se asociaron con el poder. En conclusión, los hallazgos indican que el discurso de odio cumple varias funciones. Conocer las motivaciones específicas para el discurso de odio ayuda a derivar estrategias de intervención individualmente adaptadas (ej., manejo de la ira, promover un clima escolar inclusivo). Además, sugerimos que los profesionales que trabajan en la prevención del discurso de odio presten especial atención a las normas sociales que rodean a los adolescentes.

### KEYWORDS | PALABRAS CLAVE

Hate speech, cyberhate, motives, social norms, injunctive norms, peer pressure.  
Discurso de odio, ciberodio, motivos, normas sociales, normas cautelares, presión de los pares.

## 1. Introduction

Hate speech can be defined as any communicative form of expression that deliberately promotes, justifies, or disseminates the exclusion, contempt, or devaluation of particular social groups (e.g., due to ethnicity, nationality, religion, sexual orientation, disability, gender etc. Lehman, 2020; Wachs et al., 2020). We know that adolescents are not only the witnesses and victims of hate speech but also that they actively engage in hate speech as perpetrators (Wachs et al., 2021). But there is an important gap in the literature regarding the motivations why adolescents perpetrate hate speech and which factors are associated with the self-reported motivations for devaluing and harassing particular social groups (Ballaschk et al., 2021). Furthermore, not much is known about factors that influence adolescents' motivations for perpetrating hate speech. However, the relevance of social norms for adolescents' deviant behavior and prejudices has been well-established (Allport, 1954; Van-de-Bongardt et al., 2015; Váradi et al., 2021). Whether or not social norms also drive adolescents' motivation to perpetrate hate speech, however, has not yet been investigated. Thus, the present study was conducted to fill these gaps in the literature.

Pinker (2011) developed a taxonomy of motivations for people's violent behavior that includes revenge, power, sadism, ideology, and practical violence. As hate speech can be considered a specific form of violence, it is likely that hate speech has comparable motivations. Ballaschk et al. (2021) found—based on qualitative interviews with students, teachers, and school social workers—evidence for similar causes of hate speech as those proposed by Pinker (2011). Additionally, their analyses highlighted the relevance of peer group conformity and status enhancement among peers as additional motivations that should be considered. We will therefore combine the findings of Pinker (2011) and Ballaschk et al. (2021) as the basis of our investigation into the motivations for hate speech perpetration among adolescents.

The drive to revenge oneself is a natural human impulse (Pinker, 2011). Some adolescents may use hate speech as a means of revenge when they feel threatened, unfairly treated, frustrated, or angry. Indeed, feelings of frustration and inferiority have been found to be positively related to hate speech perpetration among adolescents (Ballaschk et al., 2021). In addition to this, research has shown that hate speech victimization and perpetration are correlated, which might also indicate that some victims intend to take revenge on their perpetrators or members of other social groups (Wachs & Wright, 2021). The notion that revenge is a cause of hate speech is also supported by research that has highlighted that revenge is particularly relevant in traditional bullying and cyberbullying perpetration (Fluck, 2017; Runions et al., 2018; Tanrikulu & Erdur-Baker, 2021). Power can be understood as an individual's ability to influence others, to experience themselves as strong, and desire for dominance and control over others (McClelland, 1975; Pinker, 2011).

The desire for power is centered on experiencing strength and self-confidence and is countered by the fear of losing this power (Sokolowski & Heckhausen, 2010). It is well known that violence can be traced back to a need to exercise power (Gradinger et al., 2012; Tanrikulu & Erdur-Baker, 2021; Wettstein, 2008). And yet, there is limited evidence about the role of power in hate speech perpetration. Only one study indicates that a desire for power is related to hate speech perpetration among adolescents (Ballaschk et al., 2021). Pinker (2011) cited sadism as a motivation for perpetrating violence, describing it as a feeling of pleasure from watching another person's suffering, discomfort, or pain. To avoid any confusion with the psychological disorder of sadism, we use the term exhilaration in the present study. There is some empirical evidence that exhilaration can explain why people engage in hate speech (Ballaschk et al., 2021; Erjavec & Kovač, 2012). Exhilaration has also been shown to be a relevant cause of violence in research on aggression and bullying (Gradinger et al., 2012; Wettstein, 2008).

Hate speech is rooted in in- and out-group processes (Ballaschk et al., 2021). The social groups that people feel part of and identify with are the in-groups, whereas social groups people do not feel a part of or cannot identify with are the out-groups. The positive evaluation of the in-group often goes hand-in-hand with a tendency to devalue the out-group (Tajfel & Turner, 2004). Members of out-groups are generally assigned negative characteristics and members of the in-group feel contempt towards those not included in it. This rejection is often justified by inhumane and discriminatory positions and threatening behaviors. This in-group favoritism can lead to violence, discriminatory behavior, and hate speech between members of in- and out-groups (Tajfel & Turner, 2004). Inhumane worldviews are also prevalent among adolescents

(Möller et al., 2016). Hate speech perpetrators normally target people who are assigned to the out-group and label these groups as enemies while trying to defend their political and ideological values against others through hatred (Erjavec & Kovač, 2012). More recent research on online hate speech has identified ideological beliefs as the driving force behind such behavior (Ballaschk et al., 2021). Whereas Fluck (2017) found that ideology was not a relevant cause for perpetrating bullying, ideology might play a major role for hate speech perpetration due to the reasons outlined here. Adolescents seek a sense of belonging (Baumeister & Leary, 1995) and acceptance from their peer groups, leading to conforming behavior with existing group norms, even if those norms are anti-social and discriminatory against certain groups (Allport, 1954).

Hence, it can be assumed that group conformity might also drive adolescents to perpetrate hate speech. This assumption is supported by initial research on hate speech (Ballaschk et al., 2021) and cyberbullying (Gradinger et al., 2012), but clearly warrants more empirical evidence. Pinker (2011) noted that people use violence to achieve a goal they cannot reach through non-violent means and labelled this practical or predatory violence. Based on Ballaschk et al.'s (2021) findings, we will operationalize this motivation as a form of status enhancement among peers, as this study revealed that hate speech is used instrumentally to enhance one's own status within the peer group. Peer groups are hierarchically organized systems in which adolescents strive for high social status (Salmivalli & Peets, 2009). Studies on bullying also show that the pursuit of high-status positions within the peer group is an essential motivation for bullying (Olthoff et al., 2011; Salmivalli & Peets, 2009).

In sum, it can be assumed that there are various motivations for hate speech perpetration. Only one qualitative study, however, has investigated the motivations for perpetrating hate speech among adolescents. Consequently, no validated scale exists that measures motivations for hate speech perpetration as a multidimensional construct. And in the only existing study in which the participants reported why they believe adolescents perpetrate hate speech, no hate speech perpetrators were interviewed. We are thus missing research that investigates self-reported motivations for engaging in hate speech. Our first research aim was consequently to test the psychometric properties of a newly developed multidimensional scale for measuring motivations for hate speech perpetration.

Three types of social norms can be distinguished. Firstly, injunctive norms, which describe those that inform us about people's perception of attitudes, i.e., what behaviors are typically approved or disapproved by others. Then descriptive norms, which reflect people's perceptions of behaviors typically performed by others. And finally, peer pressure, which can be defined as peers' active encouragement to exhibit a certain behavior (Cialdini & Trost, 1998; Van-de-Bongardt et al., 2015). In his seminal work on the nature of prejudice, Allport (1954) argued that the majority of negative attitudes and feelings towards social groups arise from social norms and attempts to conform to them. What has not yet been the subject of empirical research is the question of whether social norms might also be relevant variables in explaining the motivations for hate speech perpetration. However, initial research revealed a positive relation between social norms and hate speech perpetration (Ballaschk et al., 2021; Wachs et al., in press). In addition, related research showed a positive association between social norms, prejudices, and aggression (Bastiaensens et al., 2016; Cook et al., 2010; Lee & Wong, 2009; Váradi et al., 2021). To add to the current knowledge on hate speech perpetration, the second aim of the present study was to investigate associations between social norms and adolescents' motivations for perpetrating hate speech.

## 2. Material and methods

The total sample consisted of 1,381 participants (44.5% boys, 54% girls, 1.4% gender diverse) between 11 and 18 years old ( $M_{age}=13.92$ ;  $SD=0.98$ ) from 7<sup>th</sup> to 9<sup>th</sup> grade (7<sup>th</sup> grade: 29.9%,  $n=413$ ; 8<sup>th</sup> grade: 27.1%,  $n=374$ ; 9<sup>th</sup> grade: 27.8%,  $n=384$ ; age-mixed grade: 15.2%,  $n=210$ ) from 119 classes of 22 German-speaking schools across two cantons in Switzerland. Based on prior qualitative research (Ballaschk et al., 2021; Krause et al., 2021), an item was created to measure the frequency of hate speech perpetration in schools. The participants were first shown a short video presenting a definition of hate speech. Afterwards they were asked to rate the following item, "In the past 12 months how often have you perpetrated hate speech at your school?", on a five-point scale ("never", "1 or 2 times within the last

month”, “2 or 3 times per month”, “about once a week”, “several times a week”). If students answered that they had perpetrated hate speech at least once, they were asked for their motivations. The obtained subsample included 346 adolescents (54.6% boys, 43.6% girls, 1.7% gender diverse) between the ages of 12 and 18 ( $M_{age}=14$ ;  $SD=0.96$ ) from 106 classes in 19 schools. Regarding migration background, 52.3% ( $n=165$ ) reported that they had a migration background. Concerning socio-economic status (SES), 31.2% ( $n=108$ ) of participants reported living in families of low affluence, 34.4% ( $n=119$ ) in families of medium affluence, and 34.4% ( $n=119$ ) in families of high affluence.

As no existing scale to investigate the motivations for hate speech perpetration was available, we developed the Motivations for Hate Speech Perpetration Scale (MHATE). To develop the MHATE, qualitative interviews with students and school staff were conducted focused on the motivations why students engaged in hate speech perpetration at school (Ballaschk et al., 2021). We then reviewed the literature on motives for perpetrating aggression, bullying, and cyberbullying. Based on these findings, a set of 12 items was developed, which were assumed to reflect six different motivations, namely revenge, power, exhilaration, ideology, group conformity, and status enhancement. Table 1 gives a full overview of all items. Each item could be rated on a five-point scale from “absolutely disagree” (1) to “absolutely agree” (5).

Descriptive hate speech norms were measured by asking participants if they had witnessed hate speech. The scale was developed based on qualitative research (Ballaschk et al., 2021; Krause et al., 2021) and in reference to existing instruments (Reichelmann et al., 2020). Items included, for example: “Someone distributed discriminatory symbols, stickers, pictures, memes, or videos against a specific group of people,” and: “Someone made insulting jokes about a certain group of people.” Possible answers ranged from “never” (1) to “often” (4). The McDonald’s  $\omega$  was good: .83. Perceived injunctive anti-hate speech norms were measured using three items, which the participants were asked to rate after reading a text-based vignette describing an online or offline hate speech scenario. The three items were: “My close friends don’t like it when you say that about other people,” “My family doesn’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 could be answered on a five-point scale from “absolutely disagree” (1) to “absolutely agree” (5). The McDonald’s  $\omega$  was acceptable: .74. To measure deviant peer pressure, the following four items were modified from Santor et al. (2000): “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 could be answered on a five-point scale from “absolutely disagree” (1) to “absolutely agree” (5). The McDonald’s  $\omega$  was acceptable: .78.

Participants were asked for their age and gender (male, female, diverse). Grade was added afterwards according to class codes that were used to fill out the questionnaires. Migration background (the standard measure of ethnic diversity in German-speaking countries) was assessed by asking whether the participants themselves, one, or both parents were born in a country other than Switzerland. The socio-economic status (SES) was measured using the Family Affluence Scale (FAS; Hartley et al., 2016). Based on a composite FAS score, an individual FAS category was calculated for each participant (low, medium, and high socioeconomic status). After obtaining ethical clearance from the University of Potsdam Ethics Committee (UP65/2018), 22 Swiss schools were invited to take part in this study. From these 22 schools, 20 agreed to participate in the study (participation rate at the school level: 91%). From 2,593 invited students, 1,381 participated (participation rate at the individual level: 53%). The students and their parent or legal guardian had to sign a written consent form to participate in the study. Data collection took place via an online survey during a school lesson between December 2020 and April 2021. Participants received an access code to the survey via e-mail and subsequently completed the online questionnaire. Participants were told that partaking in the study was optional and participation in the survey could be stopped at any time, without needing to give a reason and without fear of negative consequences. Completion time was 37 minutes on average.

To investigate the construct validity of the MHATE, we conducted confirmatory factor analyses (CFA) and analyzed composite reliability (CR). We then conducted a series of multi-group confirmatory

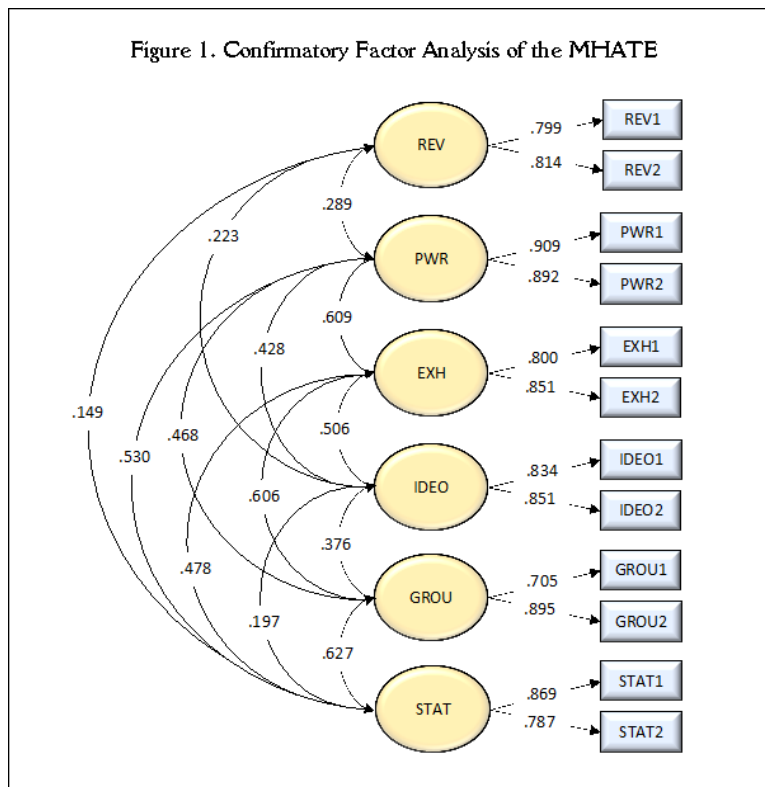
factor analyses (MGCFA) maximum likelihood parameter estimates with standard errors (MLR) in Mplus version 8.3 (Muthén & Muthén, 2017) to analyze the measurement invariance of the MHATE among different demographic groups. To evaluate the measurement invariance assumptions, Chen's (2007) recommendations were followed, according to which decreases in  $\Delta CFI > 0.010$  and increases in  $\Delta RMSEA > 0.015$  indicate that the assumption of measurement invariance is not met. If scalar measurement invariance is established, the meaning of the items can be considered as equal across the groups and latent mean differences can be compared (Van-de-Schoot et al., 2012). Cohen's *d* was used as a measure of effect sizes for latent factor means. Cohen's *d* was calculated by dividing the difference between factor means by the pooled factor standard deviation.

Finally, structural equation modelling was used to test the associations between social norms and the MHATE. The model fit was examined by considering the following fit indices: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). The quality of each model was evaluated using typical cut-off scores reflecting good and adequate fit of the data, respectively: CFI and TLI  $> .95$  and  $.90$ ; RMSEA  $< .06$  and  $.08$ , and SRMR  $< .10$  and  $.05$  (Hu & Bentler, 1999). To account for the multilevel structure of the data (i.e., adolescents nested within schools), standard errors were corrected by using the complex sampling option () in Mplus (Muthén & Muthén, 2017).

### 3. Results

#### 3.1. Construct validity of the motivations for perpetrating hate speech scale

Regarding hate speech perpetration, 74.9% ( $n=1,035$ ) of participants answered that they had never engaged in this behavior; 16.4% ( $n=227$ ) reported this behavior one or two times within the last month; 4.6% ( $n=63$ ) two or three times per month, 1.7% ( $n=23$ ); about once a week; and 2.4% ( $n=33$ ) several times a week.



Note. Only significant correlations are shown. REV=revenge, PWR=power, EXH=exhilaration, IDEO=ideology, GROU=group conformity, STAT=status enhancement.  $N=346$ .

For the following analyses, only the subsample of hate speech perpetrators ( $n=346$ ) was used. First, we tested a model in which all 12 items of the MHATE loaded on one single factor. Results showed that the data did not fit well,  $\chi^2=853.64$ ,  $df=54$ ,  $\chi^2/df=15.52$ ,  $p<.001$ ,  $CFI=.564$ ,  $TLI=0.467$ ,  $RMSEA=0.207$ ,  $SRMR=0.115$ , with standardized factor loadings between 0.18 and 0.70. Then we tested a second model with the proposed six-factor structure of the MHATE. The second model showed a good model fit,  $\chi^2=90.49$ ,  $df=40$ ,  $\chi^2/df=2.26$ ,  $p<.001$ ,  $CFI=.972$ ,  $TLI=0.955$ ,  $RMSEA=0.060$ ,  $SRMR=0.034$ . All standardized factor loadings were significant ( $p<.001$ ) and ranged between 0.78 and 0.90 (Figure 1). The CR was .79 for revenge, .90 for power, .81 for exhilaration, .76 for ideology, .79 for group conformity, and .82 for status enhancement, indicating evidence for factor validity of the MHATE.

### 3.2. Frequency rates of motivations for hate speech perpetration

Table 1 gives a full summary of frequencies, M, and SD of all items and the subscales of the MHATE. The three most frequently reported motivations were: “Because I was made angry by others” (38.4%; REV1), “Because I was hurt or annoyed by others” (33.8%; REV2), and “Because that is the way we talk to each other in my class” (25.4%; GROU1). The three least frequently reported motivations were: “To show who is the boss” (5.7%; PWR2), “To be respected and belong to the cool kids” (7%; STAT1), and “Because it feels good” (7.8%; EXH2). The most frequently endorsed subscale was revenge ( $M=2.78$ ;  $SD=1.34$ ), followed by ideology ( $M=2.17$ ;  $SD=1.15$ ), group conformity ( $M=2.02$ ;  $SD=1.12$ ), status enhancement ( $M=1.73$ ;  $SD=1.03$ ), exhilaration ( $M=1.73$ ;  $SD=1.05$ ), and power ( $M=1.59$ ;  $SD=0.94$ ).

**Table 1. Percentage, means and standard deviation of each item of the MHATE**

Items	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Item	Subscale
	%	%	%	%	%	M (SD)	M (SD)
Because I was made angry by others. (REV1)	32	9.8	19.7	19.9	18.5	2.82 (1.51)	2.78 (1.34)
Because I was hurt or annoyed by others. (REV2)	30.3	15.3	20.5	17.9	15.9	2.73 (1.45)	
To show who is in charge. (PWR1)	64.2	16.5	11.3	5.2	2.9	1.66 (1.05)	1.59 (0.94)
To show who is the boss. (PWR2)	68.2	17.9	8.1	4	1.7	1.53 (0.92)	
Because I think it was funny. (EXH1)	59.8	15.9	11	5.8	7.5	1.85 (1.26)	1.73 (1.05)
Because it feels good. (EXH2)	65.9	15.5	9.8	4.9	2.9	1.62 (1.03)	
Because it corresponds to my beliefs. (IDEO1)	58.4	18.2	11.6	8.7	3.2	1.80 (1.13)	2.17 (1.15)
Because it is my opinion. (IDEO2)	35.8	15	18.2	19.4	11.6	2.55 (1.43)	
Because that is the way we talk to each other in my class. (GROU1)	43.9	13	17.6	15.9	9.5	2.34 (1.41)	2.02 (1.12)
Because this behavior is expected in my class. (GROU2)	60.4	20.5	10.4	5.8	2.9	1.70 (1.05)	
To be respected and belong to the cool kids. (STAT1)	66.8	15.6	10.7	3.5	3.5	1.61 (1.03)	1.73 (1.03)
To belong and not become a victim myself. (STAT2)	56.4	18.5	13.6	5.8	5.8	1.86 (1.19)	

### 3.3. Measurement invariance testing and factor mean differences

Overall, for all group comparisons (i.e., by gender, grade, migration background, and SES) no substantial reduction in model fit (i.e.,  $\Delta CFI>0.010$  and  $\Delta RMSEA>0.015$ , Chen 2007) between the configural and metric invariance model and between the metric and scalar invariance model was found (Table 2). These findings indicate scalar measurement invariance of the MHATE scale across these groups and imply that latent means can be compared across groups. The scalar measurement invariance model (Model 1) was used to compare latent means between boys and girls. Results showed that boys showed lower means in ideology ( $p=.003$ ,  $d=0.40$ ) compared with girls. The scalar measurement invariance model (Model 2) was used to compare latent means between 7<sup>th</sup>/8<sup>th</sup> graders vs. 9<sup>th</sup> graders.

Results revealed that 9<sup>th</sup> graders showed lower means in group conformity ( $p=.013$ ,  $d=0.31$ ) compared with 7/8<sup>th</sup> graders. Next, the scalar measurement invariance model (Model 3) was used to

compare latent means between adolescents with and without migration background. Results showed that adolescents without migration background showed higher means in exhilaration ( $p < .001$ ,  $d = 0.55$ ) and group conformity ( $p < .001$ ,  $d = 0.46$ ) compared with participants with migration background. Finally, the scalar measurement invariance model was used to compare latent means between adolescents of high vs. low SES. Results showed that adolescents with high SES showed higher means in revenge ( $p = .006$ ,  $d = 0.41$ ) and power ( $p = .019$ ,  $d = 0.28$ ) compared to adolescents with low SES.

Table 2. Measurement invariance testing across gender, grade, migration background, and SES							
Models	$\chi^2$ (df)	p	RMSEA	$\Delta$ RMSEA	CFI	$\Delta$ CFI	Invariance rule accepted
<b>Model 1: Gender (boys vs. girls)</b>							
Configural	153.95 (80)	<.001	.073		.961		
Metric	155.96 (85)	<.001	.069	-0.004	.963	0.002	Yes
Scalar	168.51 (91)	<.001	.070	0.001	.959	-0.004	Yes
<b>Model 2: Grade (7<sup>th</sup>/8<sup>th</sup> graders vs. 9<sup>th</sup> graders)</b>							
Configural	152.82 (80)	<.001	.058		.976		
Metric	132.19 (85)	<.001	.057	-0.001	.975	-0.001	Yes
Scalar	144.34 (91)	<.001	.058	0.001	.972	-0.003	Yes
<b>Model 3: Migration background (with vs. without migration background)</b>							
Configural	123.98 (78)	<.001	.058		.975		
Metric	125.73 (84)	.002	.054	-0.004	.977	0.002	Yes
Scalar	129.33 (90)	.004	.050	-0.004	.978	0.001	Yes
<b>Model 4: SES (high vs. low SES)</b>							
Configural	133.79 (78)	<.001	.079		.957		
Metric	145.78 (84)	<.001	.080	0.001	.952	-0.005	Yes
Scalar	155.24 (90)	<.001	.080	0.000	.950	-0.002	Yes

Note.  $\chi^2$ =chi square test of model fit; CFI=comparative fit index;  $\Delta$ CFI=change in CFI compared to the weaker measurement invariance model above; RMSEA=root mean square error of approximation;  $\Delta$ RMSEA=change in RMSEA compared to the weaker measurement invariance model above.

### 3.4. Associations between social norms and motivations for perpetrating hate speech

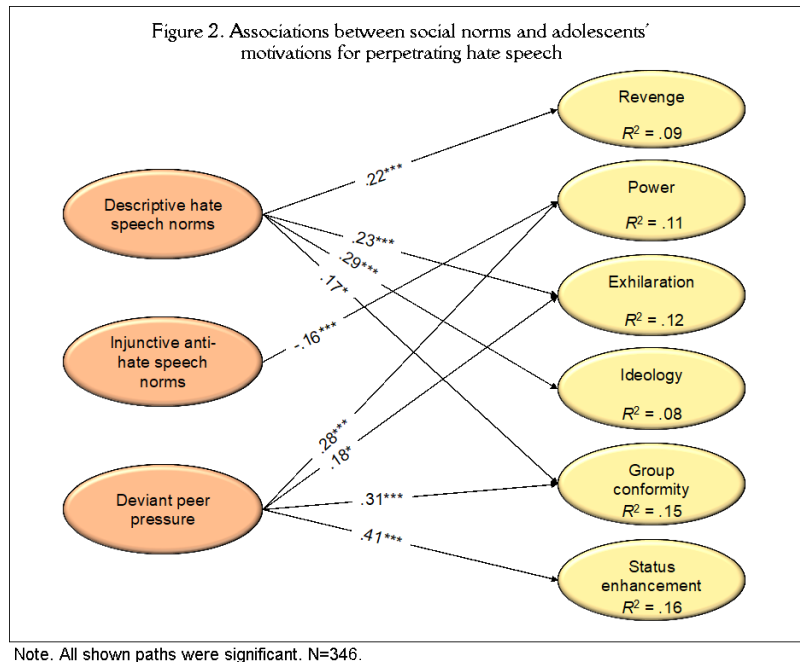
Bivariate correlations among the main variables are shown in Table 3.

Table 3. Bivariate correlations among main study variables										
	1.	2.	3.	4.	5.	6.	7.	8.	9.	
1. Descriptive hate speech norms	–									
2. Injunctive anti-hate speech norms	.03	–								
3. Deviant peer pressure	.12*	-.13*	–							
4. Revenge	.14**	-.12**	.11*	–						
5. Power	.02	-.13**	.27**	.22**	–					
6. Exhilaration	.13**	-.13**	.20**	.04	.50**	–				
7. Ideology	.18**	-.02	-.01	.13*	.29**	.35**	–			
8. Group conformity	.28**	-.06	.29**	.09	.40**	.47**	.23**	–		
9. Status enhancement	.11*	-.06	.30**	.13*	.42**	.37**	.11**	.43**	–	

Note.  $N=346$  \* $p < .05$  \*\* $p < .01$ .

The SEM to investigate the relationships between social norms and motivations for perpetrating hate speech is shown in Figure 2. The model fit was acceptable,  $\chi^2 = 420.45$ ,  $df = 263$ ,  $p < .001$ , CFI = .946, TLI = 0.934, RMSEA = 0.042, SRMR = 0.047, standardized factor loadings ranged between 0.51 to 0.91. The analyses showed that descriptive hate speech norms had a positive effect on revenge (.22 SE = .06,  $p < .001$ ), exhilaration (.23, SE = .08,  $p < .001$ ), ideology (.29, SE = .09,  $p < .001$ ), and group conformity (.17, SE = .09,  $p = .041$ ).

Injunctive anti-hate speech norms had only a negative effect on power (-.16, SE = .03,  $p < .001$ ). Finally, deviant peer pressure had a positive effect on power (.28, SE = .06,  $p < .001$ ), exhilaration (.18, SE = .08,  $p = .030$ ), group conformity (.31, SE = .10,  $p < .001$ ), and status enhancement (.41, SE = .06,  $p < .001$ ). The proposed model explained 9% of variance in the subscale revenge ( $R^2 = .088$ ), 11% of variance in the subscale power ( $R^2 = .111$ ), 12% of variance in the subscale exhilaration ( $R^2 = .124$ ), 8% of variance in the subscale ideology ( $R^2 = .084$ ), 15% of variance in the subscale group conformity ( $R^2 = .153$ ), and 16% of variance in the subscale status enhancement ( $R^2 = .162$ ).



#### 4. Discussion and conclusion

The aims of the present study were twofold. First, to investigate the psychometric properties of the multidimensional MHATE. Second, to investigate associations between social norms and adolescents' motivations for perpetrating hate speech. Regarding our first aim, our analyses supported a six-factor structure of the MHATE. Testing multi-group measurement invariance of the MHATE revealed strong evidence for scalar measurement invariance across different groups (i.e., boys vs. girls, younger vs. older adolescents, adolescents with vs. without migration background, and higher vs. lower SES), indicating an agreement on how motivations of hate speech perpetration manifest between these groups. Our findings support the idea that hate speech fulfills various functions, serving as a defense against or reaction to a perceived threat, conveying a feeling of power, or instrumentally improving one's position in the social group. The motivations to perpetrate hate speech found in this study align with qualitative research on adolescents' motivation (Ballaschk et al., 2021) and adults' motivations (Erjavec & Kovačič, 2012) for hate speech perpetration and with research on aggressive behavior (Fluck, 2017; Gradinger et al., 2012; Pinker 2011; Tanrikulu & Erdur-Baker, 2021; Wettstein, 2008).

The most frequently endorsed subscale was revenge, followed by ideology, group conformity, status enhancement and exhilaration (with both having the same mean), and then power. The high prevalence of the revenge and ideology subscales may be attributed to possible justification strategies of the perpetrators. Related research showed that revenge was also one of the most frequently reported motivation for traditional bullying and cyberbullying (Gradinger et al., 2012; Fluck, 2017; Tanrikulu & Erdur-Baker, 2021). Whereas Fluck's (2017) study showed that ideology was not a relevant motivation for cyberbullying, we found that ideology was an important motivation for hate speech perpetration. An explanation might be that ideological motives may be more relevant for hate speech than for cyberbullying. In addition, we found that power was the least frequently reported motivation for hate speech perpetration. This finding corresponds to Gradinger et al.'s (2012) study on cyberbullying but contradicts findings from Tanrikulu and Erdur-Baker (2021), who found that power was the second most often reported motivation for cyberbullying perpetration.

We also found that the various subscales of the MHATE showed differential intercorrelation patterns. Exhilaration and power, as well as exhilaration and group conformity, were most strongly correlated. These findings indicate that hate speech is often ostensibly expressed in the form of jokes to gain power



and to conform to groups within adolescents' peer networks. In contrast, the revenge subscale showed rather weak correlations, or even non-significant correlations with the other subscales. An explanation might be that revenge is a relatively independent motivation and an expression of perceived threat, while group-related hate speech represents controlled attempts to instrumentally attain goals or increase one's status within a given group.

Based on these findings, it appears to be imperative to educate adolescents on strategies of conflict management, the ability to deal with negative emotions and handle frustration to prevent adolescents from using hate speech as a means of revenge. The high endorsement of the ideology subscale highlights the need for democratic education in schools, thereby preventing hate speech perpetration. The approval ratings for group-motivated motivations (i.e., group conformity, status enhancement) indicate that promoting a positive and inclusive classroom climate and encouraging students to reflect on rules and norms are also important starting points for preventing hate speech among adolescents.

Our second aim was to investigate associations between social norms and adolescents' motivations for perpetrating hate speech. Generally, our results confirmed that social norms are associated with adolescents' motivations for hate speech perpetration. This finding is in line with previous research on prejudices and aggressive behavior (Bastiaensens et al., 2016; Cook et al., 2010; Lee & Wong, 2009; Váradi et al., 2021) and initial research on hate speech (Ballaschk et al., 2021). The relationship between descriptive hate speech norms and adolescents' motivation to engage in hate speech might be explained by observational learning (Wachs, in press). An explanation for the association between adolescents' susceptibility to deviant peer pressure and their motivations to perpetrate hate speech might be that these adolescents want to avoid rejection or punishment by their peers. More specifically, the results indicated that motivations for hate speech perpetration are not associated with injunctive norms as much as it is with descriptive norms, as well as perceived peer pressure. It seems plausible that adolescents' motivations are more strongly influenced by their perception of behaviors that are normally performed in their social environment (descriptive norms) or to which they are actively encouraged to (peer pressure) compared to their perception of behaviors that are perceived as typically approved or disapproved (injunctive norms). However, these findings should be interpreted with caution as the three scales for measuring social norms differed from each other, which might limit the comparability. According to this study, the motivation to perpetrate hate speech arises at least in part from the social dynamics of a given social environment, where reciprocal reactive processes, social recognition, social pressure, and imitation, may play an important role. These social dynamics should therefore also be the starting point for prevention measures.

This study has a few limitations which should be addressed by future research. Firstly, the present study tested important psychometric properties of the MHATE, namely construct validity and reliability. Future research should aim to test additional psychometric properties (e.g., predictive validity). In addition to this, the six factors included only two items each, which is often considered problematic for reliability. However, if the two indicators are highly correlated, two items are often sufficient (Hayduk & Littvay, 2012). Secondly, although the sample included a relatively large number of hate speech perpetrators, it cannot be considered representative for Swiss adolescents. Follow-up research should include a representative sample and try to replicate the findings of the present study. Thirdly, due to the cross-sectional study design, we were not able to determine temporal relationships between social norms and motivations for hate speech perpetration. Longitudinal studies are needed to overcome these methodological limitations. Fourthly, we included only interpersonal correlates of adolescents' motivations for hate speech perpetration. Future research should also investigate intrapersonal factors (e.g., moral disengagement). Finally, we measured hate speech perpetration by using self-reports which might have affected the identification of perpetrators, as some participants might have been reluctant to report a socially undesirable behavior such as perpetration. Follow-up research should also use peer reports to overcome such methodological problems.

In conclusion, the present study contributes to the literature by presenting a new instrument for measuring motivations for hate speech perpetration among adolescents and investigating the relationships with social norms. The findings highlight the need to understand hate speech perpetration by taking several motivations and relations to social norms into consideration. The most frequently reported motivations

were revenge, ideology, and group conformity. Prevention programs that aim to tackle hate speech need to consider various strategies, including educating adolescents on social skills and democratic values and creating social environments in which adolescents are not exposed to hate speech or actively encouraged to engage in hate speech.

### Authors' Contribution

Idea, S.W.; Literature review (state of the art), S.W., A.W.; Methodology, S.W. M.G.-G.; Data analysis, S.W. M.G.-G.; Results, S.W.; Discussion and conclusions, S.W., A.W.; Writing (original draft), S.W.; Final revisions, S.W.; Project design and funding agency, S.W., L.B., A.W.

### Funding Agency

This research was funded by the German Research Foundation (DFG), grant number WA 4275/2-1, BI 1046/9-1 and the University of Teacher Education Bern, grant number 19s 0008 01.

### References

- Allport, G.W. (1954). *The nature of prejudice*. Addison-Wesley. <https://bit.ly/3s5s35M>
- 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, 1-18. <https://doi.org/10.3224/diskurs.v16i4.01>
- Bastiaensens, S., Pabian, S., Vandebosch, H., Poels, K., Van Cleemput, K., Desmet, A., & Bourdeaudhuij, I.D. (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>
- Baumeister, R.F., & Leary, M.R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497-529. <https://doi.org/10.1037/0033-2909.117.3.497>
- Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
- 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.
- 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>
- Erjavec, K., & Kovač, M.P. (2012). You don't understand, this is a new war!" Analysis of hate speech in news web sites' comments. *Mass Communication and Society*, 15, 899-920. <https://doi.org/10.1080/15205436.2011.619679>
- 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>
- Gradinger, P., Strohmeier, D., & Spiel, C. (2012). Motives for bullying others in cyberspace: A study on bullies and bully-victims in Austria. In Q. Li, D. Cross, & P. K. Smith (Eds.), *Cyberbullying in the global playground: Research from international perspectives* (pp. 263-284). Wiley Blackwell. <https://doi.org/10.1002/9781119954484.ch13>
- 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 developmental study. *Child Indicators Research*, 9(1), 233-245. <https://doi.org/10.1007/s12187-015-9325-3>
- Hayduk, L.A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, 12(1), 1-17. <https://doi.org/10.1186/1471-2288-12-159>
- Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- 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. (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>
- McClelland, D.C. (1975). *Power: The inner experience*. Irvington Publishers.
- Möller, K., Grote, J., Nolde, K., & Schuhmacher, N. (2016). *Die kann ich nicht ab! Ablehnung, Diskriminierung und Gewalt bei Jugendlichen in der (Post-) Migrationsgesellschaft. Analysen zu gesellschaftlicher Integration und Desintegration*. Springer. <https://doi.org/10.1007/978-3-658-02302-7>
- Muthén, L.K., & Muthén, B.O. (2017). *Mplus user's guide*. Muthén & Muthén. <https://bit.ly/3lJorwi>

- Olthof, T., Goossens, F.A., Vermande, M.M., Aleva, E.A., & Van-Der-Meulen, M. (2011). Bullying as strategic behavior: Relations with desired and acquired dominance in the peer group. *Journal of School Psychology, 49*(3), 339-359. <https://doi.org/10.1016/j.jsp.2011.03.003>
- Pinker, S. (2011). *The better angels of our nature: Why violence has declined*. Viking.
- Reichelmann, A., Flawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Rasanen, P., & Zych, I. (2020). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior, 42*(9), 1100-1111. <https://doi.org/10.1080/01639625.2020.1722337>
- Runions, K.C., Salmivalli, C., Shaw, T., Burns, S., & Cross, D. (2018). Beyond the reactive-proactive dichotomy: Rage, revenge, reward, and recreational aggression predict early high school bully and bully/victim status. *Aggressive Behavior, 44*(5), 501-511. <https://doi.org/10.1002/ab.21770>
- Salmivalli, C., & Peets, K. (2009). Bullies, victims, and bully-victim relationships in middle childhood and early adolescence. In K. H. Rubin, W. M. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships, and groups* (pp. 322-340). The Guilford Press.
- 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>
- Sokolowski, K., & Heckhausen, H. (2010). Soziale Bindung: Anschlussmotivation und Intimitätmotivation. In J. Heckhausen, & H. Heckhausen (Eds.), *Motivation und Handeln [Motivation and Action]* (pp. 193-210). Springer. [https://doi.org/10.1007/978-3-642-12693-2\\_7](https://doi.org/10.1007/978-3-642-12693-2_7)
- Tajfel, H., & Turner, J.C. (2004). The social identity theory of intergroup behavior. In J. T. Jost, & J. Sidanius (Eds.), *Political psychology: Key readings* (pp. 276-293). Psychology Press. <https://doi.org/10.4324/9780203505984-16>
- Tanrikulu, I., & Erdur-Baker, O. (2021). Motives behind cyberbullying perpetration: A test of uses and gratifications theory. *Journal of Interpersonal Violence, 36*, 6699-6724. <https://doi.org/10.1177/0886260518819882>
- Van-De-Bongardt, D., Reitz, E., Sandfort, T., & Dekovi, 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-De-Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486-492. <https://doi.org/10.1080/17405629.2012.686740>
- Váradi, L., Barna, I., & Németh, R. (2021). Whose norms, whose prejudice? The dynamics of perceived group norms and prejudice in new secondary school classes. *Frontiers in Psychology, 11*, 3621-3621. <https://doi.org/10.3389/fpsyg.2020.524547>
- Wachs, S., Mazzone, A., Milosevic, T., Wright, M.F., Blaya, C., Gámez-Guadix, M., & Norman, J.H. (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. Kessl, H. Koller, N. Pfaff, ... 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., Wetzstein, A., Bilz, L., Krause, N., Ballaschk, C., Kansok-Dusche, J., & Wright, M.F. Playing by the Rules? An investigation of the relationship between social norms and adolescents' hate speech perpetration in schools. *Journal of Interpersonal Violence, 36*, 1056032. <https://doi.org/10.1177/08862605211056032>
- Wachs, S., & Wright, M.F. (2021). Associations between online hate victimization and perpetration: The buffering effects of technical and assertive coping. *Medienpädagogik, 16*, 109-128. <https://doi.org/10.21240/mpaed/jb16/2021.01.14.X>
- Wetzstein, A. (2008). *Beobachtungssystem zur Analyse aggressiven Verhaltens in schulischen Settings (BASYS)*. [Observation system for the analysis of aggressive behavior in school settings]. Huber.