

Associations Between Classroom Climate, Empathy, Self-Efficacy, and Countering Hate Speech Among Adolescents: A Multilevel Mediation Analysis

Journal of Interpersonal Violence

1–25

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/08862605221120905

journals.sagepub.com/home/jiv



Sebastian Wachs^{1,2} , Melisa Castellanos³,
Alexander Wettstein³, Ludwig Bilz⁴ 
and Manuel Gámez-Guadix⁵ 

Abstract

Although hate speech is widely recognized as an online phenomenon, very few studies have investigated hate speech among adolescents in offline settings (e.g., schools). At the same time, not much is known about countering hate speech (counterspeech) among adolescents and which factors are associated with it. To this end, the present study used the socio-ecological framework to investigate the direct and indirect links among one contextual factor (i.e., classroom climate) and two intrapersonal factors (i.e.,

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 Biological and Health Psychology, Autonomous University of Madrid, Madrid, Spain

Corresponding Author:

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

Email: wachs@uni-potsdam.de

empathy for victims of hate speech, self-efficacy regarding intervention in hate speech) to understand counterspeech among adolescents. The sample is based on self-reports of 3,225 students in Grades 7 to 9 (51.7% self-identified as female) from 36 schools in Germany and Switzerland. Self-report questionnaires were administered to measure classroom climate, empathy, self-efficacy, and counterspeech. After controlling for adolescents' grade, gender, immigrant background, and socioeconomic status (SES), the 2-(1-1)-1 multilevel mediation analysis showed that classroom climate (L2), empathy for victims of hate speech (L1), and self-efficacy toward intervention in hate speech (L1) had a positive effect on countering hate speech (L1). Classroom climate (L2) was also positively linked to empathy for victims of hate speech (L1), and self-efficacy toward intervention in hate speech (L1). Furthermore, classroom climate (L2) was indirectly associated with countering hate speech (L1) via greater empathy (L1) and self-efficacy (L1). The findings highlight the need to focus on contextual and intrapersonal factors when trying to facilitate adolescents' willingness to face hate speech with civic courage and proactively engage against it.

Keywords

hate speech, counter speech, empathy, self-efficacy, adolescents

Hate speech can be defined as any harmful communicative form of expression that deliberately promotes, justifies, or disseminates hatred or prejudice toward particular social groups and minorities (e.g., LGBTQI+ people, people of color, people with Muslim or Jewish backgrounds, or refugees; Kansok-Dusche et al., 2022). Although hate speech is often discussed as an online phenomenon, it can also be carried out face-to-face (e.g., in schools) and without the use of information and communication technologies (Krause et al., 2021; Lehman, 2019). While current hate speech research involving adolescents has mainly focused on hate speech witnessing, perpetration, and victimization, other forms of involvement, such as countering hate speech (*counterspeech*), remain blind spots (Kansok-Dusche et al., 2022).

Counterspeech refers to direct responses to hate speech with the aim of refuting or undermining it. Counterspeech can take various forms, such as providing facts in opposition to hate speech and misinformation, highlighting logical errors in hostile claims, supporting the victims, and encouraging the audience to speak up (Garland et al., 2022). Understanding the correlates of adolescents' potential or actual engagement in counterspeech is important. Feeling helpless in dealing with hate speech or becoming the target of hate speech (online or offline) can have a negative impact on adolescents'

well-being (Gámez-Guadix et al., 2020; Krause et al., 2021; Wachs, Gámez-Guadix et al., 2022), whereas the countering of hate speech not only reduces negative consequences for members of the target group but also discourages others' engagement in hate speech (Garland et al., 2022; He et al., 2021; Obermaier et al., 2021; Ștefăniță & Buf, 2021). Despite the relevance of understanding adolescents' engagement in counterspeech, the correlates of counterspeech among adolescents are largely unknown.

Addressing these gaps in the literature, the present study applied the socio-ecological model (Bronfenbrenner, 1979) as a theoretical framework to investigate the direct and indirect associations among classroom climate, adolescents' empathy for victims of hate speech, self-efficacy toward intervening in hate speech, and counterspeech. The findings of the present study might inform the development of anti-hate speech prevention programs that aim to encourage adolescents to stand up against hate speech and overcome their tendency to be passive bystanders, thus contributing to a better classroom climate and a more positive and inclusive school environment.

A Socio-Ecological Framework to Understand Counterspeech among Adolescents

The social-ecological model posits that humans develop within a complex and reciprocal interplay between individuals, their interpersonal relationships, and the wider social contexts in which they live (Bronfenbrenner, 1979). Since we are interested in understanding adolescents' engagement in counterspeech by considering contextual and intrapersonal factors, namely classroom climate, empathy, and self-efficacy, and the interplay among those variables in explaining counterspeech, the socio-ecological framework has been applied. The socio-ecological model has been used to understand traditional bullying (Espelage et al., 2013), cyberbullying (Schultze-Krumbholz et al., 2020), bias-based cyberaggression (Strohmeier et al., 2021), and lately hate speech involvement in schools (Wachs et al., 2020). We add to this line of research by investigating counterspeech among adolescents through a socio-ecological lens.

Classroom Climate as Contextual Correlate of Counterspeech

As part of the broader school climate, classroom climate is a multifaceted construct that includes a physical dimension (e.g., safety and comfort), a social dimension (e.g., quality of interpersonal relationships and degree of competition), and an academic dimension (e.g., quality of instruction; Loukas, 2007). Regarding hate speech, the social dimensions of the classroom climate might

be of particular importance. In classrooms with high social cohesion and more positive relationships among students, students might be more likely to counter hate speech than students in classrooms with a negative climate. This assumption is supported by related research documenting that a positive classroom climate is associated with higher prosocial behavior (Luengo Kanacri et al., 2017), lower rates of aggressive behavior (Thomas et al., 2011), less group-focused enmity (e.g., toward immigrants; Eckstein et al., 2021), and higher defending behavior against bullying (Bistrong et al., 2019; Thornberg et al., 2017, 2018). However, research specifically on the associations between classroom climate and adolescents' engagement in hate speech is scarce. One study found that fair and transparent school rules are negatively linked to hate speech witnessing and victimization in schools (Lehman, 2019). Other research documented that social norms within classrooms favoring hate speech are linked positively to adolescents' hate speech perpetration (Ballaschk et al., 2021; Wachs, Wettstein et al., 2021, 2022). Adding to this line of research, the present study investigates whether classroom climate is positively related to counterspeech. We were particularly interested in classroom climate because students in German and Swiss high schools usually remain together with the same classmates throughout the school day and over several school years, indicating that the classroom provides a relatively stable social group and can be considered a relevant socialization context.

Empathy and Self-Efficacy as Intrapersonal Correlates of Counterspeech

Empathy is defined as the ability to care for and comprehend the emotional condition of another person, as well as to show compassion in response to that person's feelings (Batson, 2009). Low levels of empathy have been shown to be related to higher antisocial behavior, intolerance, and prejudice (Boag & Carnelley, 2016; Pettigrew & Tropp, 2008). In contrast, meta-analyses and systematic reviews revealed that high levels of empathy are related to defending victims of bullying against students who perpetrate bullying (e.g., Deng et al., 2021; Van Noorden et al., 2015). Other empirical research on hate speech found that adolescents are less likely to be perpetrators of such behavior when they think that online hate speech leads to negative consequences (Wachs et al., 2022a, 2022b). What is unknown so far is whether empathy for victims of hate speech is positively associated with counterspeech among adolescents.

A prerequisite of social behavior is the idea that one can act in social situations efficaciously (Connolly, 1989). Students who have poor self-efficacy, on the other hand, often avoid difficult tasks because they see them as threats

rather than challenges that can be mastered (Bandura et al., 1999). Thus, it can be assumed that self-efficacy might be important to take action against hate speech, as people are more likely to intervene if they have confidence in their capacity to do so successfully. Some empirical evidence supports the idea that more self-efficacious adolescents are likelier to intervene in social conflicts, such as bullying (Sjögren et al., 2020; Thornberg et al., 2020; Wachs et al., 2018). The present study aims to extend the current knowledge on this topic by investigating the relation between self-efficacy and counterspeech.

The Interplay Between Classroom Climate, Empathy, and Self-Efficacy in Understanding Counterspeech

Beyond the proposed direct associations among classroom climate, empathy, self-efficacy, and counterspeech, previous research supports the assumption that there is an indirect relation among these variables. More specifically, empathy develops as a result of positive relationships, which are characterized by warm and supportive behavior, being available in adolescents' social environments (Eisenberg et al., 2003). Although relationships with adults are an important resource for adolescents' development of empathy, a recent meta-analysis revealed that peers are of especially vital importance (Boele et al., 2019). In addition, research showed that various aspects of classroom climate (e.g., affirming diversity and peer connectedness) were positively related to empathy among young people (Montero-Carretero et al., 2021; Safaria & Suyono, 2020). Hence, it can be assumed that a positive classroom climate is also linked to adolescents' empathy for victims of hate speech.

Like empathy, the development of self-efficacy is influenced by significant others, such as parents, siblings, teachers, and peers. During adolescence, peer acceptance and relationships are especially crucial for the development of self-efficacy because peers have a significant role in adolescents' socialization and self-perception (Schunk & Meece, 2006; Schunk & Miller, 2002). Adolescent peers provide each other with opportunities to learn from one another and function as role models (Ruggeri et al., 2018). Classrooms with a positive climate are characterized by classmates who respect, support, and care about each other and share common pro-social values, beliefs, and attitudes. Such classrooms might constitute an important context in which to develop, exercise, and maintain the skills necessary to engage in counterspeech. As a result, it is reasonable to assume that students acquire self-efficacy in classrooms in which there is a high level of cohesion and mutual sympathy among their peers. Indeed, initial research found a positive link between a supportive classroom climate and students' self-efficacy

toward handling interpersonal conflicts (Sjögren et al., 2021; Wachs et al., 2018). Therefore, a positive relation between classroom climate and adolescents' self-efficacy toward intervening in hate speech can be expected.

In sum, research has indicated that adolescents who are socialized in classrooms with a positive peer culture are likelier to become sensitive to the emotions and well-being of others and self-efficacious toward handling social conflicts (e.g., bullying). Also, both empathy and self-efficacy have been shown to be hallmark characteristics of adolescents' willingness to intervene in interpersonal conflicts (e.g., bullying). Hence, it can be assumed that classroom climate is indirectly associated with counterspeech via empathy and self-efficacy. Considering not only direct but also indirect associations might provide a deeper understanding of the underlying mechanisms that explain adolescents' civic engagement in peer interactions.

The Current Study

Hitherto, little attention has been given to factors that explain counterspeech among adolescents in schools. Adolescents' ability to deal with hate speech is not only related to their own well-being but can also have an impact on the frequency of hate speech in classrooms. Therefore, research on adolescents' engagement in counterspeech is needed to inform prevention programs with the aim of empowering adolescents to deal with hate speech among their peers. Hence, the following hypotheses guided this study:

Hypothesis 1: Classroom climate will be positively associated with counterspeech.

Hypothesis 2: Adolescents' empathy for victims of hate speech and their self-efficacy toward intervening in hate speech will be positively related to their counterspeech.

Hypothesis 3: Classroom climate will be positively and indirectly linked to counterspeech via empathy for the victims of hate speech and self-efficacy toward intervening in hate speech.

Methods

Participants

The present sample is based on 3,225 adolescents from Germany ($n = 1,841$; 57.1%) and Switzerland ($n = 1,384$; 42.9%). Participants were in Grades 7 to 9 (7th grade: 33.2%, $n = 1,070$; 8th grade: 35.6%, $n = 1,147$; 9th grade: 31.3%, $n = 1,008$). In terms of gender, 46.1% ($n = 1,487$) self-identified as male,

51.7% ($n=1,668$) self-identified female, 2% ($n=64$) self-identified as gender diverse, and 0.2 ($n=6$) did not indicate their gender. Overall, 62.8% ($n=2,025$) had no immigrant background (neither themselves nor their parents were born in a foreign country), 14.2% ($n=458$) reported that one parent was born in a country other than Germany or Switzerland, 15.1% ($n=488$) reported that both parents were born in a foreign country, and 7.9% ($n=254$) reported that themselves and their parents were born in a foreign country. In total, 30.8% ($n=994$) of students reported living in families of low affluence, 35.8% ($n=1,155$) in families of medium affluence, and 32.4% ($n=1,046$) in families of high affluence. For 0.9% ($n=30$) of all participants, socioeconomic status (SES) could not be established due to missing values.

Measures

Counterspeech. The instrument to measure counterspeech was developed after conducting qualitative research (Ballaschk et al., 2021; Krause et al., 2021) and pre-testing the instrument ($N=75$). Participants were presented with a vignette that described a hate speech incident. It read as follows: “Please imagine the following situation: At your school, a student makes public, insulting statements about people of a certain skin color or origin.” The participants were then asked: “What would you do in the situation described or what have you done if you have experienced such a situation before?” After reading this question, the participants were asked to respond to several items. For counterspeech, participants rated the following four items: “I tell the person that such statements are hurtful”, “I ask the person to stop”, “I try to get the person to think by asking specific questions”, and “I say that the person is spreading false information (fake news).” All items could be answered on a five-point response scale from “strongly disagree” (1) to “strongly agree” (5). Cronbach’s α was .81 $CI_{95\%}$ [0.80, 0.82] and McDonald’s ω was .81 $CI_{95\%}$ [0.80, 0.82].

Classroom climate. The quality of students’ relationships with their classmates was measured using one scale consisting of three items (e.g., “Most of the students in my class are friendly and supportive”; Currie et al., 2014). Responses could be rated on a five-point scale from “absolutely disagree” (1) to “absolutely agree” (5). Cronbach’s α was .81 $CI_{95\%}$ [0.79, 0.83] and McDonald’s ω was .82 $CI_{95\%}$ [0.80, 0.83]. As recommended by previous research (e.g., Bardach et al., 2020), classroom climate was included at the classroom level because this variable measures shared perceptions of classroom climate using items that refer to the classroom, not to the individual.

Empathy for victims of hate speech. The instrument for measuring empathy was adapted from Knauf et al. (2018). Hate speech is referred to in the introduction of the instrument: “When I see classmates being insulted or attacked by other classmates because of their skin color, origin, religion, sexual orientation, or gender. . .” For empathy, six items were included to reflect different aspects of empathy, including perspective-taking (e.g., “I realize how badly they are doing”), affective empathy (e.g., “It hurts me, too”), and empathic concern (e.g., “It makes me want to comfort them”). Items were rated on a five-point scale from “absolutely disagree” (1) to “absolutely agree” (5). Cronbach’s α was .91 $CI_{95\%}$ [0.90, 0.92] and McDonald’s ω was .91 $CI_{95\%}$ [0.79, 0.92].

Self-efficacy toward intervening in hate speech. One item was developed to measure self-efficacy toward intervening in hate speech: “I know what to do to get the classmate to stop spreading hate speech.” Responses could be rated on a five-point scale from “absolutely disagree” (1) to “absolutely agree” (5).

Control variables. Participants were asked to report their grade and gender (self-identifying as male, self-identifying as female, or gender diverse). This variable was dichotomized into “self-identifying as male” (1) and “self-identifying as female” (2). Immigrant background was assessed by asking whether the participant, one parent, or both of their parents were born in a country other than Germany or Switzerland. This variable was also dichotomized “no immigrant background” (1; when neither the participant nor at least one parent was born in a country other than Germany or Switzerland) and “immigrant background” (2; when the participant or at least one parent was born in a foreign country). This approach followed the official German definition of having an immigrant background (Statistisches Bundesamt (Destatis), 2022). SES was measured using the Family Affluence Scale (FAS; Hartley et al., 2016). Participants answered six items regarding, for example, their family’s car, whether children have their own bedrooms, the number of bathrooms they have in their home, the number of computers at home, and the number of holidays taken in the past 12 months. Based on a composite FAS score, an individual FAS category was calculated for each participant (i.e., low, medium, and high SES).

Procedure and Sampling Technique

Approval for this study was obtained from the data protection officer, the educational authority of the Federal State of Berlin and Brandenburg, Germany, and the University of Potsdam Ethics Committee (UP65/2018). In

Germany, data were collected between October 2020 and June 2021. In Switzerland, data were collected between December 2020 and April 2021. In Germany, a tablet-based questionnaire was administered by trained research assistants during a school lesson. In Switzerland, the participants received an access code to the survey via e-mail and subsequently completed the questionnaire on their own digital devices during a school lesson. Overall, the average completion time was 36 minutes.

In Germany, the acquisition pool of sample schools was composed using a stratified and randomized probability-proportional-to-size scheme (Yates & Grundy, 1953). All schools were initially stratified by federal state (Berlin and Brandenburg) and type of school (e.g., grammar secondary school or non-academic-track secondary school). Then, for each school type, schools were randomly selected proportional to their size. This sampling procedure ensured that all students had the same likelihood of being included in the sample, regardless of the size of their schools. In Switzerland, the acquisition pool of sample schools was designed via a contrastive sampling scheme based on high/low immigrant backgrounds and on rural/urban geography.

From the resulting acquisition pools, 100 schools (Germany: $n=76$; Switzerland: $n=24$) were informed via phone calls and e-mail that they had been randomly selected to participate in the study. Acquisition stopped as soon as the sampling plans were fulfilled. In total, 40 schools (Germany: $n=18$; Switzerland: $n=22$) agreed to participate. The participation rate at the school level was 40% across the whole sample (Germany: 24%; Switzerland: 92%). Due to the COVID-19 pandemic, a significant number of German schools declined participation, citing a lack of resources or high regional infection rates at the time of contact.

Students from Grades 7 to 9 were invited to take part in the survey. At each participating school in Germany, two randomly selected classes per grade were invited to participate. In Switzerland, all available classes across Grades 7 to 9 were invited. Students in mixed grades were also invited to take part. In total, 264 classes were invited to participate in the study (Germany: 106; Switzerland: 158), and 236 of them chose to participate (Germany: $n=98$; Switzerland: $n=138$). The response rate at the classroom level was 89% for the whole sample (Germany: 92%; Switzerland: 87%). From a total of 5,836 students (Germany: $n=2,495$; Switzerland: $n=3,341$), 3,560 students (Germany: $n=1,841$; Switzerland: $n=1,719$) participated in the study. The response rate at the student level was 61% for the whole sample (Germany: 74%; Switzerland: 51%). For the present study, the 335 Swiss students from mixed classrooms in four schools were excluded from the overall sample because being in mixed grade was confounded with being Swiss.

Data Analyses

Power analysis. A power analysis conducted a priori with G*Power (Faul et al., 2007) revealed that to detect small to medium correlational effect sizes, the present study would require a sample consisting of at least 782 participants ($\alpha = .05$, power = 0.80). Considering the hierarchical structure of the sample and non-response rates, the resulting minimal sample size was $N = 1,944$ students in 108 classes at 18 schools (Teerenstra et al., 2010). Thus, the present sample size was sufficient to investigate the hypotheses.

Missing data analysis. Overall, missing data were between 1.1% ($n = 36$; counterspeech) and 2% ($n = 66$; self-efficacy toward intervening in hate speech). Little's MCAR test revealed that the data were missing completely at random ($\chi^2 = 11.49$, $df = 9$, $p = .243$). Hence, missing data were handled using the full information maximum likelihood approach in Mplus (Muthén & Muthén, 1998–2017).

Statistical analyses. Before conducting the main analyses, descriptive statistics and bivariate correlations were investigated. A multilevel mediation analysis was conducted using Mplus 8.7 (Muthén & Muthén, 1998–2017). More specifically, the analysis tested a 2-(1-1)-1 model whereby the independent variable of classroom climate was entered at Level 2 (L2), and the mediating variables (i.e., empathy for victims of hate speech and self-efficacy toward intervening in hate speech) and the outcome variable (i.e., counterspeech) were entered at Level 1 (L1).

The following steps were used to test the multilevel mediation analysis. First, a random intercept model (Model 0) was estimated to determine the intraclass correlation coefficients (ICC). Then, in Model 1, the control variables—namely, grade, gender, immigrant background, and SES—were entered as predictors of empathy, self-efficacy, and counterspeech. In Model 2, direct associations among classroom climate (aggregated at the classroom level), empathy, self-efficacy, and counterspeech were entered and tested simultaneously. Finally, in Model 3, two indirect effects were added and tested simultaneously: (1) the indirect effect of classroom climate on counterspeech via empathy, and (2) the indirect effect of classroom climate on counterspeech via self-efficacy.

Given the multilevel nature of the data, the improvement of model fit was evaluated by relative decreases in the Akaike information criterion (AIC) according to which lower AIC values indicate a better model fit (Akaike, 1974). When the change in AIC is larger 10, the model has essentially no support

Table 1. Pearson's Bivariate Correlations and Descriptive Statistics.

Variable	2.	3.	4.	<i>M</i> (<i>SD</i>)
Classroom climate	0.20**	0.07**	0.16**	3.85 (0.43)
Empathy	—	0.22**	0.48**	3.63 (0.96)
Self-efficacy		—	0.31**	3.05 (1.11)
Counterspeech			—	3.28 (1.06)

** $p < .001$.

(Burnham & Anderson, 2004). The tested direct and indirect effects were assessed using a statistical significance test and the syntax for Mplus by Preacher et al. (2011). Because the effect of the mediators (i.e., empathy and self-efficacy) on the outcome (i.e., counterspeech) was expected to be the same within and between classrooms, the variables were not centered. This is a common practice in 2-1-1 multilevel mediation models (Preacher et al., 2011).

Results

Preliminary Analyses

The Pearson's bivariate correlations and descriptive statistics for the study's main variables are shown in Table 1. All correlations were in the expected direction (see Table 1). Higher levels of classroom climate were positively correlated with empathy for victims of hate speech ($r = .20, p < .001$), self-efficacy toward intervening in hate speech ($r = .07, p < .001$), and counterspeech ($r = .16, p < .001$). Empathy for victims of hate speech was positively correlated with self-efficacy toward intervening in hate speech ($r = .22, p < .001$) and counterspeech ($r = .48, p < .001$). Self-efficacy toward intervening in hate speech was positively correlated with counterspeech ($r = .31, p < .001$).

Testing the Direct and Indirect Associations among Classroom Climate, Empathy, Self-Efficacy, and Counterspeech

Regarding ICCs, the results of the random intercept model (Model 0) showed that 18% of the variance in classroom climate, 10.9% of the variance in empathy for hate speech victims, 2.1% of the variance in self-efficacy toward intervening in hate speech, and 8.3% of the variance in counterspeech can be

Table 2. Changes in the Indices of Goodness of Fit for the Multilevel Mediation Models.

	Model 0	Model 1	Model 2	Model 3
R_2 (individual)		3.4%	23.2%	24.0%
R_2 (class)		2.3%	76%	71.0%
AIC	35,094.55	26,602.97	25,435.40	25,418.78
Δ AIC		8,591.58	1,198.57	16.62

Note. AIC = Akaike information criterion; Δ AIC = change in AIC.

attributed to the fact that students are nested within classrooms, indicating that multilevel analyses were required. With every step, the model fit improved, as can be shown by decreases in AIC (see Table 2), therefore Model 3 was used for subsequent analyses. Table 3 summarizes the results of testing Model 1 to Model 3.

As shown in Table 3 (Model 3), classroom climate (L2) was positively associated with counterspeech (L1; unstandardized $\gamma = .14$, $p = .002$ CI 95% [0.06, 0.21]), providing support for Hypothesis 1. Additionally, classroom climate (L2) was positively associated to empathy for victims of hate speech (L1; unstandardized $\gamma = .44$, $p < .001$, CI 95% [0.34, 0.53]) and self-efficacy toward intervening in hate speech (L1; unstandardized $\gamma = .21$, $p < .001$, CI 95% [0.18, 0.24]). Empathy for victims of hate speech (L1; unstandardized $\gamma = .45$, $p < .001$, CI 95% [0.42, 0.49]) and self-efficacy toward intervening in hate speech (L1; unstandardized $\gamma = .16$, $p = .004$, CI 95% [0.07, 0.25]) were both positively linked to counterspeech (L1), confirming Hypothesis 2. Finally, the indirect effects of classroom climate (L2) on counterspeech (L1) via their empathy for hate speech victims (L1; unstandardized γ indirect = .20, $p = .005$, CI 95% [0.15, 0.24]) and self-efficacy toward intervening in hate speech (L1; unstandardized γ indirect = .03, $p < .001$, CI 95% [0.01, 0.05]) were significant. Thus, Hypothesis 3 was confirmed (Figure 1).

Several control variables showed a significant effect. Compared to being in Grade 9, being in Grade 7 (unstandardized $\gamma = .11$, $p = .046$, CI 95% [0.02, 0.20]) was positively associated with empathy for victims of hate speech. Self-identifying as female was positively related to counterspeech (L1; unstandardized $\gamma = .09$, $p = .011$, CI 95% [0.03, 0.15]) and empathy for victims of hate speech (L1; unstandardized $\gamma = .61$, $p < .001$, CI 95% [0.56, 0.67]). Having an immigrant background was negatively associated with counterspeech (L1; unstandardized $\gamma = -.08$, $p = .029$, CI 95% [-0.14, -0.02]), and positively associated with self-efficacy toward intervening in hate speech (L1; unstandardized $\gamma = .12$, $p < .001$, CI 95% [0.04, 0.19]). Students' SES

Table 3. Results of the Multilevel Mediation Model.

Predictor	Mediating Variable	Outcome	Model 1			Model 2			Model 3		
			Estr. (SE) [CI 95%]	p	Std. Est.	Estr. (SE) [CI 95%]	p	Std. Est.	Estr. (SE) [CI 95%]	p	Std. Est.
Direct effects											
Classroom climate	Empathy	Counterspeech	0.14 (0.04) [0.07, 0.22]	.001	0.19	0.14 (0.04) [0.06, 0.21]	.002	0.20	0.14 (0.04) [0.06, 0.21]	.002	0.20
Classroom climate	Self-efficacy	Counterspeech	0.59 (0.06) [0.50, 0.68]	<.001	0.63	0.44 (0.06) [0.34, 0.53]	<.001	0.53	0.44 (0.06) [0.34, 0.53]	<.001	0.53
Classroom climate	Empathy	Counterspeech	0.43 (0.02) [0.40, 0.47]	<.001	0.54	0.21 (0.02) [0.18, 0.24]	<.001	0.22	0.21 (0.02) [0.18, 0.24]	<.001	0.22
	Self-efficacy	Counterspeech	0.43 (0.02) [0.40, 0.47]	<.001	0.39	0.45 (0.02) [0.42, 0.49]	<.001	0.40	0.45 (0.02) [0.42, 0.49]	<.001	0.40
	Self-efficacy	Counterspeech	0.21 (0.02) [0.18, 0.24]	<.001	0.23	0.16 (0.06) [0.07, 0.25]	.004	0.25	0.16 (0.06) [0.07, 0.25]	.004	0.25
Indirect effects											
Classroom climate	Empathy	Counterspeech				0.20 (0.03) [0.15, 0.24]	<.001		0.20 (0.03) [0.15, 0.24]	<.001	
Classroom climate	Self-efficacy	Counterspeech				0.03 (0.01) [0.01, 0.05]	.005		0.03 (0.01) [0.01, 0.05]	.005	
Control variables											
7th grade		Counterspeech	0.11 (0.07) [-0.01, 0.23]	.119	0.18	0.02 (0.05) [-0.06, 0.11]	.649	0.03	0.02 (0.05) [-0.07, 0.11]	.678	0.03
8th grade		Counterspeech	0.06 (0.07) [-0.06, 0.17]	.419	0.09	0.03 (0.05) [-0.06, 0.11]	.610	0.03	0.02 (0.05) [-0.06, 0.11]	.625	0.04
7th grade	Empathy	Counterspeech	0.10 (0.07) [-0.01, 0.21]	.116	0.12	0.11 (0.06) [0.02, 0.20]	.050	0.12	0.11 (0.06) [0.02, 0.20]	.046	0.14
8th grade	Empathy	Counterspeech	0.04 (0.06) [-0.05, 0.14]	.469	0.05	0.06 (0.06) [-0.04, 0.15]	.327	0.06	0.05 (0.05) [-0.04, 0.14]	.341	0.06
7th grade	Self-efficacy	Counterspeech	0.07 (0.06) [-0.02, 0.16]	.200	0.12	0.07 (0.06) [-0.02, 0.17]	.210	0.10	0.08 (0.06) [-0.02, 0.17]	.169	0.12
8th grade	Self-efficacy	Counterspeech	0.05 (0.06) [-0.04, 0.14]	.396	0.08	0.06 (0.06) [-0.04, 0.15]	.344	0.07	0.05 (0.06) [-0.04, 0.14]	.362	0.08
Gender: self-identifying as female		Counterspeech	0.37 (0.04) [0.30, 0.44]	<.001	0.18	0.11 (0.04) [0.05, 0.17]	.003	0.05	0.09 (0.04) [0.03, 0.15]	.011	0.05
Gender: self-identifying as female	Empathy	Counterspeech	0.61 (0.04) [0.56, 0.67]	<.001	0.34	0.61 (0.04) [0.56, 0.67]	<.001	0.34	0.61 (0.04) [0.56, 0.67]	<.001	0.34
Gender: self-identifying as female	Self-efficacy	Counterspeech	-0.06 (0.04) [-0.13, 0.01]	.130	-0.03	-0.06 (0.04) [-0.13, 0.01]	.146	-0.03	-0.06 (0.04) [-0.13, 0.01]	.162	-0.03
Immigrant background/yes		Counterspeech	-0.05 (0.04) [-0.12, 0.02]	.235	-0.02	-0.08 (0.04) [-0.14, -0.02]	.033	-0.04	-0.08 (0.04) [-0.14, -0.02]	.029	-0.04
Immigrant background/no	Empathy	Counterspeech	0.02 (0.04) [-0.05, 0.08]	.676	0.01	0.03 (0.04) [-0.04, 0.08]	.498	0.01	0.02 (0.04) [-0.04, 0.08]	.226	0.01
Immigrant background/yes	Self-efficacy	Counterspeech	0.11 (0.05) [0.04, 0.19]	.011	0.05	0.13 (0.05) [0.05, 0.20]	.005	0.06	0.12 (0.05) [0.04, 0.19]	<.001	0.05
SES		Counterspeech	0.01 (0.01) [-0.01, 0.02]	.377	0.02	0.01 (0.01) [0.00, 0.02]	.233	0.02	0.01 (0.01) [0.00, 0.02]	.226	0.02
SES	Empathy	Counterspeech	-0.03 (0.01) [-0.04, 0.01]	.004	-0.06	-0.03 (0.01) [-0.04, -0.02]	.001	-0.07	-0.03 (0.01) [-0.04, -0.02]	.001	-0.07
SES	Self-efficacy	Counterspeech	0.02 (0.01) [0.00, 0.03]	.120	0.03	0.02 (0.01) [0.00, 0.03]	.158	0.03	0.02 (0.01) [0.00, 0.04]	.072	0.04

Note. L1 variables: Empathy, self-efficacy, counterspeech, gender, migration background, SES. L2 variables: Classroom climate, grade. Reference category: Grade: 9th grade; gender: self-identifying as male; immigrating background: no. Est. = unstandardized estimate; SE = standard error; Std. Est. = standardized estimate; CI = confidence intervals; SES = socioeconomic status.

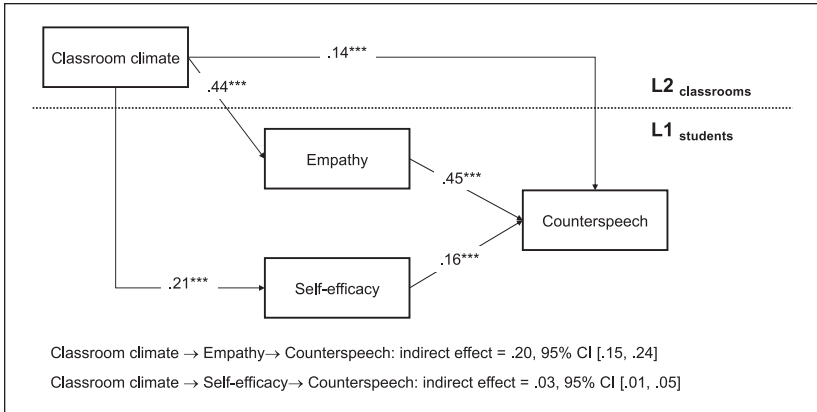


Figure 1. Direct and indirect associations among classroom climate (L2) and counterspeech (L1) via empathy and self-efficacy (L1).

Note. Control variables are omitted due to clarity.

*** $p < .001$.

was negatively associated with empathy for victims of hate speech (L1; unstandardized $\gamma = -.03$, $p = .001$, CI 95% [-0.04 , -0.02]). Finally, the analyses were repeated by country of origin. Generally, the analyses for each country revealed the same results, indicating that the three hypotheses could be confirmed in the German and Swiss subsample. The findings can be requested from the first author.

Discussion

Using the socio-ecological Model (Bronfenbrenner, 1979), the present study sought to understand correlates of counterspeech in a large sample of adolescents from Germany and Switzerland. More specifically, a multilevel mediation analysis was conducted to investigate the direct and indirect associations among classroom climate, empathy for victims of hate speech, self-efficacy toward intervening in hate speech, and counterspeech. An important consideration when investigating these associations is that a lack in adolescents' capacity to deal productively with hate speech is negatively related to their well-being (Gómez-Guadix et al., 2020; Krause et al., 2021; Wachs, Gómez-Guadix et al., 2022), whereas counterspeech can reduce negative outcomes for victims of hate speech and reduce people's engagement in hate speech as perpetrators (Garland et al., 2022; He et al., 2021; Obermaier et al., 2021; Ștefăniță & Buf, 2021).

Hypothesis 1: Direct Associations between Classroom Climate and Counterspeech

Concerning Hypothesis 1, the findings confirmed the assumption that classroom climate is positively linked to counterspeech. It is proposed that adolescents in classrooms characterized by positive peer relationships, connectedness, and warm and supportive behavior are more willing to show prosocial behavior, which also includes the engagement in countering hate speech. This finding is in line with research showing that antisocial behavior and group-focused enmity are less prevalent in classrooms with a positive atmosphere (Eckstein et al., 2021; Thomas et al., 2011) and that victims of bullying are defended more frequently in classrooms with a positive climate (Bistrong et al., 2019; Thornberg et al., 2017, 2018). This result might also partially explain the initial research that revealed a negative association between a better school climate and less hate speech witnessing and victimization in schools (Lehman, 2019). Furthermore, the present study extends previous research that has found that social norms within classrooms are related to adolescents' engagement in hate speech as perpetrators (Ballaschk et al., 2021; Wachs, Wettstein et al., 2021, 2022).

Hypothesis 2: Direct Associations between Empathy, Self-Efficacy, and Counterspeech

Supporting Hypothesis 2, this study revealed that empathy for victims of hate speech and self-efficacy toward intervening in hate speech are positively related to their counterspeech. These results suggest that adolescents who show empathy and believe that hate speech is hurtful to the targeted person might perceive hate speech as more objectionable and, thus, are more willing to intervene. This finding is aligned with research showing that empathy is negatively related to antisocial behavior, intolerance, and prejudice (Boag & Carnelley, 2016; Pettigrew & Tropp, 2008) and positively associated with defending victims of bullying (Deng et al., 2021; Van Noorden et al., 2015). Moreover, the present study extends previous research showing that empathy is negatively related to online hate speech perpetration among adolescents (Wachs et al., 2022a, 2022b).

While other research has highlighted the crucial role of self-efficacy in defending victims of bullying (Sjögren et al., 2020; Thornberg et al., 2020; Wachs et al., 2018), this study extends those findings to counterspeech by showing that adolescents' beliefs in their ability to successfully intervene in hate speech are relevant to countering hate speech.

Hypothesis 3: Indirect Associations between Classroom Climate and Counterspeech via Empathy and Self-Efficacy

Finally, in accordance with Hypothesis 3, the present study provides evidence for the assumption that classroom climate is positively and indirectly linked to counterspeech via empathy for victims of hate speech and self-efficacy toward intervening in hate speech. These findings highlight the crucial role of the classroom climate in developing students' empathy and self-efficacy, which is in line with research that has stressed the importance of peers in the development of young people's empathy and self-efficacy (Boele et al., 2019; Montero-Carretero et al., 2021; Safaria & Suyono, 2020; Sjögren et al., 2021; Wachs et al., 2018). Furthermore, the results concerning our third hypothesis shed light on mechanisms that might explain the relation between classroom climate and engagement in counterspeech. Hence, the results suggest that the interplay between contextual and intrapersonal factors is important to consider in understanding what motivates adolescents to counter hate speech.

Limitations and Future Research

Several limitations should be mentioned. First, the present study was based on a cross-sectional design. Thus, the temporal ordering of classroom climate, empathy, self-efficacy, and counterspeech cannot be determined. Longitudinal and experimental research is required to completely comprehend the temporal ordering of these variables. Second, empathy for victims of hate speech was measured as a unidimensional construct, self-efficacy toward intervening in hate speech was measured with only one global item, and school climate was measured by only one of its many facets. Follow-up research should consider the cognitive and affective subdimensions of empathy to gain a deeper understanding of the complex interplay between empathy and engaging in counterspeech. Along the same lines, to overcome the measurement problems typical of single-item measures (e.g., low content validity, sensitivity, and lack of a measure of internal consistency reliability), such research should include scales for measuring self-efficacy. Follow-up research should also include more facets of classroom climate (e.g., student-teacher relationships) to gain a deeper understanding of the relation between classroom climate and counterspeech among adolescents. Third, we measured how adolescents engaged in counterspeech or would engage in counterspeech without being able to distinguish between actual or hypothetical behavior. Follow-up studies should measure both separately and compare the findings. To avoid self-reporting biases, one possibility might be to include peer nominations regarding engagement in countering hate speech and then

investigate the association between the nominations and adolescents' self-reported intervention willingness, or actual behavior. Finally, more research is needed to understand the nature of counterspeech among adolescents. This includes, for example, an investigation on what they normally say when they practice counterspeech and what they perceive as an effective strategy to deal with hate speech among peers. At the same time, such research could investigate what typical barriers are when adolescents decide to intervene or not after witnessing hate speech among peers. A theoretical framework to understand the barriers might be the bystander intervention model (Latané & Darley, 1970) which has been shown to be useful in understanding bystander behavior in bullying and sexual harassment (Nickerson et al. 2014).

Practical Implications

Despite these limitations, the current study has important practical implications. First, prevention programs with the aim of increasing counterspeech among adolescents should address the classroom climate. Building positive peer-to-peer relationships within the classroom seems to be a promising avenue both for encouraging counterspeech and fostering adolescents' empathy and self-efficacy. Consequently, teachers, social workers, and school psychologists should be aware of how the classroom climate might influence adolescents' empathy, self-efficacy, and counterspeech. Furthermore, professionals working in schools with young people need to be supported to develop effective whole-class activities that encourage students to establish and sustain warm, caring, and supportive relationships with their classmates. Such whole-class activities might include discussions about respect and tolerance, the differences between hate speech and free speech, and how to respond to hate speech appropriately (Krause et al., 2022).

Second, anti-hate speech prevention programs should include empathy training that increases adolescents' emotional sharing and perspective-taking abilities. This might be accomplished by addressing hypothetical or actual hate speech cases and talking with adolescents about how hate speech targets might feel, thus assisting them in comprehending the potential harms of hate speech victimization. Adolescents may also be urged to consider how they would feel if they were the victims of hate speech, thus learning how to show sympathy for the victims' pain. Similar strategies have been shown to be effective in anti-bullying prevention programs, such as "Media Heroes" (Schultze-Krumbholz et al., 2016) and "KiVa" (Kärnä et al., 2011) and are currently applied in the anti-hate speech prevention program "HateLess. Together against hatred" (Krause et al., 2022).

Third, self-efficacy toward intervening in hate speech needs to be taken into consideration when designing and implementing anti-hate speech programs. Such programs should promote adolescents' ability to resolve interpersonal conflicts using negotiation rather than hate speech. One way to do so might be the implementation of cooperative learning during school lessons, which has been shown to increase adolescents' self-efficacy (In'am & Sutrisno, 2021), prosocial behavior (Manzano-Sánchez & Valero-Valenzuela, 2019), and democratic attitudes (Erbil & Kocabaş, 2018). More specifically, adolescents should be educated about how they can intervene directly and indirectly in hate speech and should be given opportunities to practice different forms of intervention (e.g., through role play). This approach is used in the anti-hate speech prevention program "HateLess. Together against hatred" (Krause et al., 2022).

Finally, a general practical implication of this study is the need to address both contextual factors (e.g., classroom climate) and individual factors (e.g., empathy and self-efficacy) to tackle hate speech among adolescents in schools. That is, by combining universal and indicated actions, prevention programs need to focus on the whole classroom and not solely on certain students who have shown (dis)engagement in hate speech or counterspeech.

Conclusion

The present study investigated the direct and indirect associations among classroom climate (L2), empathy for victims of hate speech (L1), self-efficacy toward intervening in hate speech (L1), and counterspeech among adolescents (L1). The findings revealed that classroom climate, empathy for victims, self-efficacy toward intervening were positively associated with counterspeech. Furthermore, the results confirmed an indirect relation between classroom climate and counterspeech via empathy and self-efficacy. That is, as suggested by the socio-ecological model it appears to be important to address both contextual and individual factors to understand counterspeech. Generally, this study suggests that prevention programs need to include all students in one classroom and combine universal and indicated actions to increase counterspeech. Although the present study offers much potential for understanding counterspeech and developing prevention programs, to understand the temporal ordering of the study's main variables, there is a need for more longitudinal research that considers both potential and actual engagement in counterspeech separately. Finally, more research is needed to understand the nature of counterspeech among adolescents, effective strategies, and barriers to intervening in hate speech.

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>

Manuel Gámez-Guadix  <https://orcid.org/0000-0002-1575-1662>

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Ballaschk, C., Wachs, S., Krause, N., Schulze-Reichelt, F., Kansok-Dusche, J., Schubarth, W., & Bilz, L. (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*, 4, 463–480. <https://doi.org/10.3224/diskurs.v16i4.01>
- Bandura, A., Pastorelli, C., Barbaranelli, C., & Caprara, G. V. (1999). Self-efficacy pathways to childhood depression. *Journal of Personality and Social Psychology*, 76(2), 258–269. <https://doi.org/10.1037/0022-3514.76.2.258>
- Bardach, L., Yanagida, T., & Lüftenegger, M. (2020). Studying classroom climate effects in the context of multi-level structural equation modelling: An application-focused theoretical discussion and empirical demonstration. *International Journal of Research & Method in Education*, 43(4), 348–363. <https://doi.org/10.1080/1743727X.2020.1791071>
- Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–15). MIT Press.
- Bistrong, E., Bottiani, J. H., & Bradshaw, C. P. (2019). Youth reactions to bullying: Exploring the factors associated with students’ willingness to intervene. *Journal of School Violence*, 18(4), 522–535. <https://doi.org/10.1080/15388220.2019.1576048>
- Boag, E. M., & Carnelley, K. B. (2016). Attachment and prejudice: The mediating role of empathy. *British Journal of Social Psychology*, 55(2), 337–356. <https://doi.org/10.1111/bjso.12132>

- Boele, S., Van der Graaff, J., de Wied, M., Van der Valk, I. E., Crocetti, E., & Branje, S. (2019). Linking parent-child and peer relationship quality to empathy in adolescence: A multilevel meta-analysis. *Journal of Youth and Adolescence, 48*(6), 1033–1055. <https://doi.org/10.1007/s10964-019-00993-5>
- Bronfenbrenner, U. (1979). *Ecology of human development: Experiments by nature and design*. Harvard University Press.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Connolly, J. (1989). Social self-efficacy in adolescence: Relations with self-concept, social adjustment, and mental health. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 21*(3), 258–269. <https://doi.org/10.1037/h0079809>
- Currie, C., Inchley, J., Molcho, M., Lenzi, M., Veselska, Z., & Wild, F. (Eds.) (2014). Health behaviour in school-aged children (HBSC) study protocol: Background, methodology and mandatory items for the 2013/14 survey. *Child and Adolescent Health Research Unit (CAHRU)*.
- Deng, X., Yang, J., & Wu, Y. (2021). Adolescent empathy influences bystander defending in school bullying: A three-level meta-analysis. *Frontiers in Psychology, 12*, 690898. <https://doi.org/10.3389/fpsyg.2021.690898>
- Eckstein, K., Miklikowska, M., & Noack, P. (2021). School matters: The effects of school experiences on youth's attitudes toward immigrants. *Journal of Youth and Adolescence, 50*, 2208–2223. <https://doi.org/10.1007/s10964-021-01497-x>
- Eisenberg, N., Spinrad, T. L., & Morris, A. (2003). Empathy-related responding in children. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 184–207). Psychology Press.
- Erbil, D. G., & Kocabaş, A. (2018). Cooperative learning as a democratic learning method. *Journal of Research in Childhood Education, 32*(1), 81–93. <https://doi.org/10.1080/02568543.2017.1385548>
- Espelage, D. L., Rao, M. A., & De La Rue, L. (2013). Current research on school-based bullying: A social-ecological perspective. *Journal of Social Distress and the Homelessness, 22*(1), 21–27. <https://doi.org/10.1179/1053078913Z.0000000002>
- 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*, 175–191. <https://doi.org/10.3758/BF03193146>
- Gámez-Guadix, M., Wachs, S., & Wright, M. F. (2020). “Haters back off!” Psychometric properties of the coping with cyberhate questionnaire and the relationship with well-being among Spanish adolescents. *Psicothema, 32*(4), 567–574. <https://doi.org/10.7334/psicothema2020.219>
- Garland, J., Ghazi-Zahedi, K., Young, J.-G., Hébert-Dufresne, L., & Galesic, M. (2022). Impact and dynamics of hate and counter speech online. *EPJ Data Science, 11*(1), 3. <https://doi.org/10.1140/epjds/s13688-021-00314-6>
- 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>
- He, B., Ziems, C., Soni, S., Ramakrishnan, N., Yang, D., & Kumar, S. (2021, November). *Racism is a virus: Anti-Asian hate and counterspeech in social media during the COVID-19 crisis* (pp. 90–94). [Conference session]. Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. <https://doi.org/10.1145/3487351.3488324>
- In'am, A., & Sutrisno, E. S. (2021). Strengthening students' self-efficacy and motivation in learning mathematics through the cooperative learning model. *International Journal of Instruction*, 14(1), 395–410. <https://doi.org/10.29333/iji.2021.14123a>
- Kansok-Dusche, J., Ballaschk, C., Krause, N., Zeißig, A., Seemann-Herz, L., Wachs, S., & Bilz, L. (2022). A systematic review on hate speech among children and adolescents: definitions, prevalence, and overlap with related phenomena. *Trauma, Violence, & Abuse*. Advance online publication. <https://doi.org/10.1177/15248380221108070>
- Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa antibullying program: Grades 4 to 6. *Child Development*, 82(1), 311–330. <https://doi.org/10.1111/j.1467-8624.2010.01557.x>
- Knauf, R. K., Eschenbeck, H., & Hock, M. (2018). Bystanders of bullying: Social-cognitive and affective reactions to school bullying and cyberbullying. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 12(4), 3. <https://doi.org/10.5817/CP2018-4-3>
- 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, 169–185. <https://doi.org/10.1007/s35834-021-00291-w>
- Krause, N., Richter, M., & Wachs, S. (2022). *HateLess. Gemeinsam gegen hass. Ein trainingsmanual für die prävention von hatespeech in der schule*. [HateLess. Together against hate. A training manual for the prevention of hate speech in schools.]. Universität Potsdam.
- Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Prentice Hall.
- 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>
- Loukas, A. (2007). What is school climate. *Leadership Compass*, 5(1), 1–3.
- Luengo Kanacri, B. P., Eisenberg, N., Thartori, E., Pastorelli, C., Uribe Tirado, L. M., Gerbino, M., & Caprara, G. V. (2017). Longitudinal relations among positivity, perceived positive school climate, and prosocial behavior in Colombian adolescents. *Child Development*, 88(4), 1100–1114. <https://doi.org/10.1111/cdev.12863>

- Manzano-Sánchez, D., & Valero-Valenzuela, A. (2019). Implementation of a model-based programme to promote personal and social responsibility and its effects on motivation, prosocial behaviours, violence and classroom climate in primary and secondary education. *International Journal of Environmental Research and Public Health*, *16*(21), 4259. <https://doi.org/10.3390/ijerph16214259>
- Montero-Carretero, C., Pastor, D., Santos-Rosa, F. J., & Cervelló, E. (2021). School climate, moral disengagement and, empathy as predictors of bullying in adolescents. *Frontiers in Psychology*, *12*, 656775. <https://doi.org/10.3389/fpsyg.2021.656775>
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide*. Muthén & Muthén.
- Nickerson, A. B., Aloe, A. M., Livingston, J. A., & Feeley, T. H. (2014). Measurement of the bystander intervention model for bullying and sexual harassment. *Journal of Adolescence*, *37*(4), 391–400. <https://doi.org/10.1016/j.adolescence.2014.03.003>
- Obermaier, M., Schmuck, D., & Saleem, M. (2021). I'll be there for you? Effects of islamophobic online hate speech and counterspeech on Muslim in-group bystanders' intention to intervene. *New Media & Society*. Advance online publication. <https://doi.org/10.1177/14614448211017527>
- Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, *38*(6), 922–934. <https://doi.org/10.1002/ejsp.504>
- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, *18*(2), 161–182. <https://doi.org/10.1080/10705511.2011.557329>
- 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>
- Safaria, T., & Suyono, H. (2020). The role of parent-child relationship, school climate, happiness, and empathy to predict cyberbullying behavior. *International Journal of Evaluation and Research in Education*, *9*(3), 548–557. <https://doi.org/10.11591/ijere.v9i3.20299>
- Schultze-Krumbholz, A., Schultze, M., Zagorscak, P., Wölfer, R., & Scheithauer, H. (2016). Feeling cybervictims' pain—The effect of empathy training on cyberbullying. *Aggressive Behavior*, *42*(2), 147–156. <https://doi.org/10.1002/ab.21613>
- Schultze-Krumbholz, A., Zagorscak, P., Hess, M., & Scheithauer, H. (2020). The influence of school climate and empathy on cyberbystanders' intention to assist or defend in cyberbullying. *International Journal of Bullying Prevention*, *2*(1), 16–28. <https://doi.org/10.1007/s42380-019-00040-8>
- Schunk, D. H., & Meece, J. L. (2006). Self-efficacy development in adolescence. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 71–96). Information Age Publishing.
- Schunk, D. H., & Miller, S. D. (2002). Self-efficacy and adolescents' motivation. In F. Pajares & T. Urdan (Eds.), *Academic motivation of adolescents* (Vol. 2, pp. 29–52). Information Age Publishing.

- Sjögren, B., Thornberg, R., Wänström, L., & Gini, G. (2020). Associations between individual and collective efficacy beliefs and students' bystander behavior. *Psychology in the Schools, 57*(11), 1710–1723. <https://doi.org/10.1002/pits.22412>
- Sjögren, B., Thornberg, R., Wänström, L., & Gini, G. (2021). Bystander behaviour in peer victimisation: Moral disengagement, defender self-efficacy and student-teacher relationship quality. *Research Papers in Education, 36*(5), 588–610. <https://doi.org/10.1080/02671522.2020.1723679>
- Statistisches Bundesamt (Destatis). (2022, June 27). Migration and Integration. *Migrationshintergrund*. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Glossar/migrationshintergrund.html>
- Ștefăniță, O., & Buf, D.-M. (2021). Hate speech in social media and its effects on the LGBT community: A review of the current research. *Romanian Journal of Communication and Public Relations, 23*(1), 47–55. <https://doi.org/10.21018/rjcp.2021.1.322>
- Strohmeier, D., Gradinger, P., & Yanagida, T. (2021). The role of intrapersonal-, interpersonal-, family-, and school-level variables in predicting bias-based cyber-victimization. *The Journal of Early Adolescence*. Advance online publication. <https://doi.org/10.1177/02724316211010335>
- 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>
- Thomas, D. E., Bierman, K. L., Powers, C. J. & The Conduct Problems Prevention Research Group. (2011). The influence of classroom aggression and classroom climate on aggressive–disruptive behavior. *Child Development, 82*(3), 751–757. <https://doi.org/10.1111/j.1467-8624.2011.01586.x>
- Thornberg, R., Wänström, L., Elmelid, R., Johansson, A., & Mellander, E. (2020). Standing up for the victim or supporting the bully? Bystander responses and their associations with moral disengagement, defender self-efficacy, and collective efficacy. *Social Psychology of Education, 23*(3), 563–581. <https://doi.org/10.1007/s11218-020-09549-z>
- Thornberg, R., Wänström, L., Hong, J. S., & Espelage, D. L. (2017). Classroom relationship qualities and social-cognitive correlates of defending and passive bystanding in school bullying in Sweden: A multilevel analysis. *Journal of School Psychology, 63*, 49–62. <https://doi.org/10.1016/j.jsp.2017.03.002>
- Thornberg, R., Wänström, L., & Jungert, T. (2018). Authoritative classroom climate and its relations to bullying victimization and bystander behaviors. *School Psychology International, 39*(6), 663–680. <https://doi.org/10.1177/0143034318809762>
- Van Noorden, T. H. J., Haselager, G. J. T., Cillessen, A. H. N., & Bukowski, W. M. (2015). Empathy and involvement in bullying in children and adolescents: A systematic review. *Journal of Youth and Adolescence, 44*(3), 637–657. <https://doi.org/10.1007/s10964-014-0135-6>
- Wachs, S., Bilz, L., Fischer, S. M., Schubarth, W., & Wright, M. F. (2018). Students' willingness to intervene in bullying: Direct and indirect associations with classroom

- cohesion and self-efficacy. *International Journal of Environmental Research and Public Health*, 15(11), 2577. <https://doi.org/10.3390/ijerph15112577>
- Wachs, S., Bilz, L., Wettstein, A., Ballaschk, C., Krause, N., & Kansok-Dusche, J. (2022a). The online hate speech cycle of violence: Moderating effects of moral disengagement and empathy in the victim-to-perpetrator relationship. *Cyberpsychology, Behavior, and Social Networking*, 25(4), 223–229. <https://doi.org/10.1089/cyber.2021.0159>
- Wachs, S., Bilz, L., Wettstein, A., Krause, N., Ballaschk, C., & Kansok-Dusche, J. (2022b). Associations between witnessing and perpetrating online hate speech among adolescents: Testing moderation effects of moral disengagement and empathy. *Psychology of Violence*. Advance online publication. <https://doi.org/10.1037/vio0000422>
- Wachs, S., Gámez-Guadix, M., & Wright, M. F. (2022). Hate speech victimization and depressive symptoms among adolescents: The protective role of resilience. *Cyberpsychology, Behavior, and Social Networking*, 25(7), 416–423. <https://doi.org/10.1089/cyber.2022.0009>
- Wachs, S., Schubarth, W., & Bilz, L. (2020). Hate speech als schulproblem? Erziehungswissenschaftliche perspektiven auf ein aktuelles phänomen. In I. van Ackeren, H. Bremer, F. Kessel, 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.
- Wachs, S., Wettstein, A., Bilz, L., & Gámez-Guadix, M. (2022). Adolescents' motivations to perpetrate hate speech and links with social norms. *Comunicar*, 71, 9–20. <https://doi.org/10.3916/C71-2022-01>
- Wachs, S., Wettstein, A., Bilz, L., Krause, N., Ballaschk, C., Kansok-Dusche, J., & Wright, M. F. (2021). Playing by the rules? An investigation of the relationship between social norms and adolescents' hate speech perpetration in schools. *Journal of Interpersonal Violence*. Advance online publication. <https://doi.org/10.1177/08862605211056032>
- Yates, F., & Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society*, 15(2), 253–261.

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 (family, school, and culture).

Melisa Castellanos, PhD, is a Research Associate at the Research Group “Social Interaction in Educational Settings” at the University of Teacher Education Bern. Her research interest is social development in the context of peer relationships.

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.

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.

Manuel Gámez-Guadix, PhD, is a Professor of Psychology at the Department of Psychology, Autonomous University of Madrid. His research focus on the study of the nature, correlates, and consequences of problematic Internet use, both in terms of harassment or cyberbullying, as well as abuse or addiction to communication and information technologies. He also conducted research on family violence (e.g., intimate partner violence, physical punishment, child-to-parent violence) and on educational guidelines and parenting practices.