

Racist Hate Speech at School and its Association with Bystanders' Active Defending – The Protective Role of Immigrant Background

Melisa Castellanos and Alexander Wettstein

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

Ludwig Bilz

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

Sebastian Wachs

Institute of Education, University of Münster, Münster, Germany


Abstract


There are scarce studies on bystanders' responses to racist hate speech. The present study was rooted in the general aggression model to examine the direct and indirect associations between witnessing racist hate speech and bystanders' active defending behaviors. Moreover, we examined if these associations varied as a function of the immigration background of bystanders. The sample included 3,225 7th, 8th, and 9th graders (51.7% girls, 37.2% with an immigrant background) from 40 schools in Switzerland and Germany. A multilevel mediation path model revealed that witnessing racist hate speech negatively predicted bystanders' active defending. It also had indirect associations with bystanders' active defending via low levels of empathy. Finally, associations were not observed in students with an immigrant background. The findings regarding the potential contribution of these results to preventing hate speech in school are discussed.


Keywords


Hate speech, general aggression model, bystanders' responses, desensitization, migration background

Author Note

Melisa Castellanos  <https://orcid.org/0000-0002-6980-0845>

Alexander Wettstein  <https://orcid.org/0000-0002-5612-3144>

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

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

This research was funded by the German Research Foundation (DFG), grant number WA 4275/2-1; BI 1046/9-1; SCHU1370/4-1 and the University of Teacher Education Bern, grant number 19s 0008 01.

Introduction

According to the general aggression model, exposure to violence might produce an effect of *desensiti-*

Correspondence concerning this article should be addressed to Prof. Dr. Alexander Wettstein, Institute for Research, Development, and Evaluation, Bern University of Teacher Education, Fabrikstrasse 8. CH-3012, room D135, Bern, Switzerland. E-mail: alexander.wettstein@phbern.ch

zation through a reduction in the “emotion-related physiological reactivity to real violence” (Carnagey et al., 2007, p. 490). Consequently, affective and cognitive outcomes change, such that violence is perceived as less harmful, empathy toward victims is reduced, negative attitudes towards violence also decrease, and individuals perceive violence as normative (Miles-Novelo & Anderson, 2020). Subsequently, bystanders of violence are more likely to perpetrate aggression and decrease or delay their prosocial intervention (Carnagey et al., 2007; Miles-Novelo & Anderson, 2020). A phenomenon adolescents are exposed to is hate speech, defined as direct or vicarious derogatory expressions about others based on assigned group features (i.e., nationality, gender, sexual orientation, etc.) (Kansok-Dusche et al., 2022). It includes multiple expressions (e.g., words, online posts, graffiti, etc.) and is intended to cause harm (Kansok-Dusche et al., 2022). Although hate speech and bullying share commonalities (e.g., in both, there is a power imbalance), they differ regarding targets, frequency, and perpetrators’ goals (Kansok-Dusche et al., 2022). Mostly, hate speech has a racist component, as it is often based on skin color, origin, ethnicity, or immigrant background (Castellanos et al., 2023; Caravita et al., 2020; Reichelmann et al., 2021; United States Government Accountability Office, 2021). To contribute to the literature, the present study is rooted in the general aggression model (Bushman & Anderson, 2020) to examine the direct and indirect associations between exposure to racist hate speech and active defending bystander responses in adolescents from two European countries. Moreover, we explored if empathy helped explain these associations. Finally, we examined if these associations varied depending on bystanders’ immigrant background.

The General Aggression Model and Desensitization

Research about the effects of exposure to violence and desensitization is centered on explaining perpetration outcomes (see Prescott et al., 2018, for a meta-analysis). However, the general aggression model poses that exposure to violence can also impact prosociality (Miles-Novelo & Anderson, 2020). For instance, witnessing violent media is associated with decreased bystander prosocial intervention (Bushman & Anderson, 2009; Prot et al., 2015). Although recent evidence has shown that exposure to online hate speech increased adolescents’ likeability to per-

petrate hate speech (Wachs et al., 2022a), its effects on prosocial bystander behaviors are, to date, unknown. Moreover, studies about the consequences of witnessing *offline* hate speech (i.e., perpetrated without digital media; Kansok-Dusche et al., 2022) are scarce. To narrow these gaps in the literature, the first objective of the present study was to explore the direct and indirect associations between witnessing racist hate speech in the classroom and active defending bystander responses. Active defending bystander responses refer to prosocial behavior to stop aggression among peers, including taking sides with the victims, assertively confronting the perpetrators, and seeking support from peers or adults (Salmivalli, 2010). In line with the general aggression model, we hypothesized that witnessing racist hate speech in the classroom would negatively predict three active defending bystander responses: comforting the victim, seeking help, and countering hate speech (hypothesis 1).

The Role of Empathy

To better understand these associations, we explored if empathy could explain the links between witnessing racist hate speech and bystanders’ responses. Empathy is the capacity to comprehend and feel another person’s emotional state (Batson, 2009). A study by Wachs et al. (2023b) found that empathetic adolescents are more likely to counter hate speech (i.e., direct responses to refute and discourage hate speech, such as providing information to oppose perpetrators’ claims; Garland et al., 2022). According to the general aggression model, when exposed to violence, individuals can be desensitized by decreasing their empathetic responses, which might result in a lower likeability to engage in active defending (Miles-Novelo & Anderson, 2020). To our knowledge, no studies investigated this for offline hate speech. To fill this gap in the literature, the second objective of the present study was to estimate if empathy explained the association between witnessing racist hate speech in the classroom and active defending bystander responses. Based on the desensitization effect posed by the general aggression model and the well-documented association between empathy and bystanders’ active defending (Gönültaş et al., 2020; Machackova & Pfetsch, 2016), we hypothesized that witnessing racist hate speech in the classroom would be negatively and indirectly associated with the three active defending bystander responses via low levels of empathy (hypothesis 2).

Differential Impact of Witnessing Racist Hate Speech

Multiple studies show that the most common hate-related words observed and received were associated with race and place of origin (Castellanos et al., 2023; Caravita et al., 2020; Reichelmann et al., 2020; United States Government Accountability Office, 2021). Thus, evidence about the impact of exposure to racist hate speech on adolescents with an immigrant background is needed to inform prevention and intervention programs. According to previous evidence, immigrant background moderates how adolescents react when witnessing racist hate speech. Nonetheless, the findings are not conclusive. Some studies show that when witnessing racist hate speech, adolescents with an immigrant background are less likely to defend victims actively (Wachs et al., 2023b). Other studies from the related field of bullying based on migration status showed that adolescents with an immigrant background had a higher disposition to actively intervene (e.g., get help from someone) when the victim also had an immigrant background (Gönültaş & Mulvey, 2021; 2022).

These differential reactions can be explained by the social identity or the component of self-concept related to the knowledge and perception of belonging to one or many social groups (Tajfel, 1982). This identification with a collective identity has an emotional value for individuals, resulting in in-group affiliation and out-group differentiation (Tajfel, 1982). Accordingly, when witnessing racist hate speech, adolescents with an immigrant background might feel their social identity more affected than those without an immigrant background. In line with this proposition, a study found that Muslim adults perceived a higher threat to their identity and reported a higher disposition to intervene when witnessing Islamophobic than non-Islamophobic hate speech (Obermaier et al., 2021).

To better understand these differential reactions, our third objective was to explore if the direct and indirect associations between witnessing racist hate speech on active defending bystander responses varied as a function of adolescents' immigrant background. In line with the social identity theory and the empirical findings from related fields of research, we hypothesized that the negative indirect association between witnessing racist hate speech and the three active defending bystander responses via low levels of empathy would be *weaker* for the group of students with an immigrant background (Hypothesis 3).

Method

Participants

A total of 3,225 adolescents ($n_{Germany} = 1,841$, 57.1%; $n_{Switzerland} = 1,384$, 42.9%) from 40 schools in Germany and Switzerland ($n_{Germany} = 18$, $n_{Switzerland} = 22$) equally distributed by grade ($n_{7th} = 1,070$, 33.2%; $n_{8th} = 1,147$, 35.6%; $n_{9th} = 1,008$, 31.3%), and gender ($n_{female} = 1,668$, 51.7%; $n_{male} = 1,487$, 46.1%; $n_{diverse} = 64$, 2%; $n_{no_response} = 6$, 0.2%) participated in the study. Moreover, 37.2% ($n = 1,200$) of them had an immigrant background (e.g., they or one of their parents being born outside Germany/Switzerland). On average, the classroom proportion of students with an immigrant background was 0.35 in Germany and 0.53 in Switzerland. From the participants, 30.8% ($n = 994$) reported living in families of low affluence, 35.8% ($n = 1,155$) in families of medium affluence, 32.4% ($n = 1,046$) in families of high affluence, and 0.9% ($n = 30$) had missing values on the socioeconomic status items (Family Affluence Scale, Currie et al., 2008).

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). The acquisition pool of sample schools was composed using a stratified and randomized probability-proportional-to-size scheme (Yates & Grundy, 1953) in Germany and by a contrastive sampling scheme including immigrant background and location (rural/urban) in Switzerland. An a priori power analysis with G*Power (Faul et al., 2007) indicated that a sample size of at least 782 participants would be needed to detect small to medium correlational effect sizes for an $\alpha = 0.05$ and power = 0.80. Anticipating non-response and accounting for the nested structure of the sample, the required minimum sample size was $N = 1,944$ students in 108 classes at 18 schools (Teerenstra et al., 2010). Then, 100 schools from two federal states in Germany and six German-speaking cantons in Switzerland were invited to participate in the study. Among them, 40 agreed to participate ($n_{Germany} = 18$, $n_{Switzerland} = 22$).

A total of 5,836 students from 7th to 9th grade and mixed classrooms (i.e., Swiss classes where 14 to 16 years old are taught together) were informed

about the study and invited to participate. In total, 3,560 participated in the study ($n_{Germany} = 1,841$; $n_{Switzerland} = 1,719$). Parents or legal guardians provided consent for all of them except those aged 14 and older in Berlin, following the legal regulation in the state, which poses that at this age, students have the legal autonomy to decide about participating in research studies. For the present study, 335 participants from the mixed classrooms were excluded from the analyses, as attending a school in Switzerland could be confounded with being part of these classes. Between 1% and 1.9% of data were missing. Little's MCAR test revealed that the data were not missing completely at random ($\chi^2(26) = 41.94, p = .025$). This data was handled using the full information maximum likelihood (FIML) approach (Muthén & Muthén, 2012–2021).

Instruments

Witnessing Racist Hate Speech

An instrument based on qualitative studies with adolescents and school teachers was used (Ballaschk et al., 2021). Initially, participants were presented with a definition of hate speech as a short video clip, followed by a brief text-based introduction emphasizing that hate speech is intentionally hurtful, takes place in public, is directed towards social groups, and can also take non-verbal forms. Then students were asked to report how often they observed hate speech in school without using digital media in the previous 12 months by indicating on a 5-point scale ranging from “never” to “several times a week”. A follow-up question was formulated for participants who indicated to have witnessed hate speech at least once. With a *yes/no* format, the question inquired if the witnessed hate speech was directed against people with a particular skin color or origin (e.g., dark-skinned people, foreigners). A classroom-level racist hate speech score was calculated by averaging the responses of each classroom's students to capture a more objective operationalization of exposure to racist hate speech.

Empathy for Victims of Hate Speech (adapted from Knauf et al., 2018). Students were introduced: “When I see classmates being insulted or attacked by other classmates because of their skin color, origin, religion, sexual orientation, or gender...”. Then, they were asked to indicate their agreement with six statements that addressed empathy (e.g., “I realize how badly they are doing”). The five-point response

scale ranged from “strongly disagree” to “strongly agree”. Cronbach's alpha of the scale was 0.91.

Multidimensional Bystander Responses to Hate Speech Scale (Wachs et al., 2023a). At first, participants were asked to imagine a situation in which someone from their school made “publicly insulting statements about people of a certain skin color or origin”. After, participants were presented with the question, “What would you do in the situation described or what have you done if you have experienced such a situation before?” and asked to rate their agreement with multiple statements using a five-point scale that ranged from “strongly disagree” to “strongly agree”. The three active defending subscales were used: comforting the victim (e.g., “I comfort the person against whom hate speech was targeted at”), seeking help at school (e.g., “I ask classmates if they can help me to do something about it”), and countering hate speech (e.g., “I say that the person is spreading false information ‘fake news’”). Cronbach's alpha was 0.83 for comforting the victim, 0.83 for seeking help at school, and 0.81 for countering hate speech.

Data Analysis

Because of the hierarchical nature of our data, with students nested within classrooms, a multilevel mediation path model was estimated in Mplus 8.7 (Muthén & Muthén, 2012–2021). The 2-(1)-1-1-model, including predictors at level 2, a mediator variable at level 1, and dependent variables at level 1, was estimated in four steps. First, a random-intercept model (*Null Model*) with no predictors was estimated to compute the intraclass correlation coefficients of the dependent variables comforting the victim, seeking help at school, and countering hate speech. Second, control variables at level 1 (gender and socioeconomic status) and level 2 (classroom proportion of students with an immigrant background and grade) were added as predictors. Third, we tested the direct associations between a) empathy and each active defending bystander response, b) classroom racist hate speech and each active defending bystander response, and c) classroom racist hate speech and empathy. In the last step, we tested the indirect associations between classroom racist hate speech on comforting the victim (indirect effect 1), seeking help at school (indirect effect 2), and countering hate speech (indirect effect 3) via empathy. Finally, to test our third hypothesis, we estimated the model separately for students with and without an immigrant background.

The goodness of fit of the models was assessed by changes in the Akaike information criterion (AIC; Akaike, 1974), suggesting a better fit for reductions of 10 points in each model concerning the previous (Burnham & Anderson, 2004). Direct and indirect associations were assessed using statistical significance tests at an alpha level of 0.05. Predictors were not centered as the effects of the mediator variables on the dependent variable were expected to be the same within and between classrooms (Preacher et al., 2010).

Results

Descriptive Statistics

Most participants ($n = 2,123$, 65.8%) reported that they had observed hate speech in their schools. Among them, 50.3% ($n = 1,068$) reported witnessing racist hate speech. Table 1 of the supplemental material displays some statistically significant and small in magnitude differences observed between students with and without an immigrant background. Table 2 of the supplementary material presents bivariate correlations among the study variables.

Direct and Indirect Associations between Classroom Racist Hate Speech and Bystanders' Active Defending

The intraclass correlation coefficients were 0.14 for witnessing racist hate speech, 0.10 for comforting the victim, 0.10 for seeking help at the school, 0.08 for countering hate speech, and 0.11 for empathy. Therefore, a multilevel approach was justified. The results from each step followed in estimating the models are presented in Tables 3, 4, and 5 of the supplemental material. Table 1 displays the final step results (i.e., the model that tested direct and indirect associations) for the complete sample and separately for students with and without an immigrant background. As shown, the results for the complete sample supported our first hypothesis, as the classroom level of racist hate speech negatively predicted comforting the victim ($B = -0.24$, 95% CI [-0.40, -0.07], $p = 0.019$), seeking help at school ($B = -0.42$, 95% CI [-0.59, -0.25], $p < 0.001$), and countering hate speech ($B = -0.31$, 95% CI [-0.48, -0.14], $p = 0.002$). Moreover, empathy positively predicted comforting the victim ($B = 0.64$, 95% CI [0.61, 0.68], $p < 0.001$), seeking help at school ($B = 0.47$, 95% CI [0.43, 0.50],

$p < 0.001$), and countering hate speech ($B = 0.50$, 95% CI [0.47, 0.54], $p < 0.001$). Furthermore, our second hypothesis was confirmed, as classroom racist hate speech had an indirect negative association with comforting the victim ($B = -0.17$, 95% CI [-0.29, -0.05], $p = 0.017$), seeking help at school ($B = -0.12$, 95% CI [-0.21, -0.04], $p = 0.017$), and countering hate speech ($B = -0.13$, 95% CI [-0.23, -0.04], $p = 0.016$) via low levels of empathy.

Differential Associations by Immigrant Background of the Students

In line with our third hypothesis, the associations observed for the complete sample remained the same for the students without an immigrant background (see Table 1). Furthermore, in this group, the model's indirect associations were statistically significant, indicating that classroom racist hate speech had a negative indirect association with comforting the victim ($B = -0.29$, 95% CI [-0.43, -0.16], $p < 0.001$), seeking help at school ($B = -0.20$, 95% CI [-0.29, -0.11], $p < 0.001$), and countering hate speech ($B = -0.23$, 95% CI [-0.33, -0.12], $p < 0.001$), via low levels of empathy. In contrast, classroom racist hate speech did not predict the levels of empathy of students with an immigrant background ($B = -0.03$, 95% CI [-0.30, 0.21], $p = 0.854$) (See Figure 1). The association between the predictor and the mediator is an assumption of the mediation analyses (Baron & Kenny, 1986), so the indirect associations were not estimated for this group.

Discussion

Our study provided three main findings that contribute to our current understanding of hate speech in adolescence and added evidence about the effects of exposure to violence. First, in classrooms where racist hate speech was frequent, bystanders were less likely to comfort the victims, seek help at school, and counter hate speech compared to bystanders from classrooms where this problem was less common. This result aligns with findings about bullying (Lucas-Molina et al., 2018) and extends findings from studies about the effects of exposure to online hate speech on online hate speech perpetration (Wachs et al., 2022a) and increases in prejudice (Soral et al., 2018). In line with the general aggression model (Miles-Novelo & Anderson, 2020), students who constantly witness racist hate speech in their class-

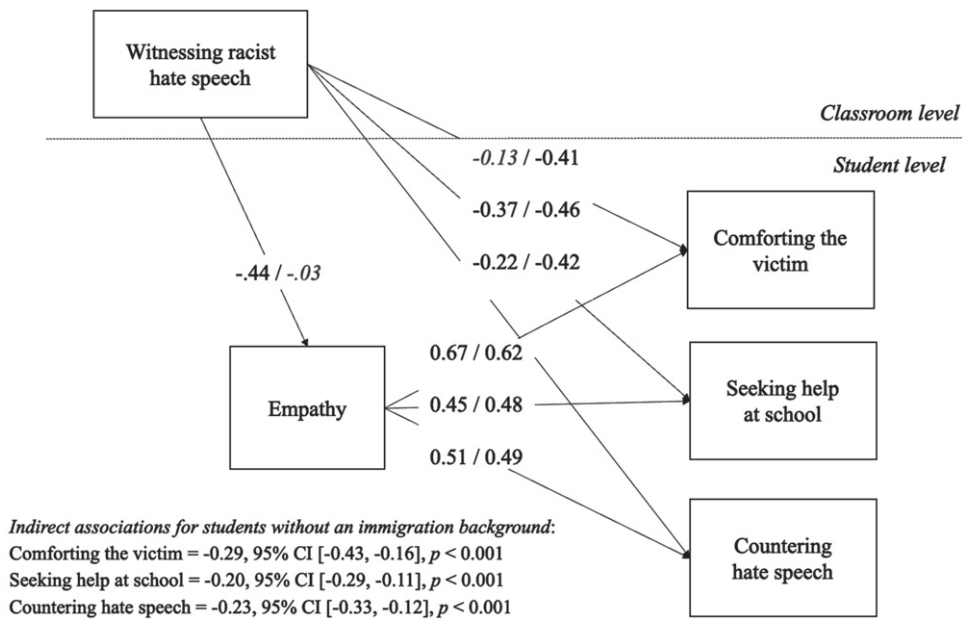
Table 1
 Multilevel Mediation Model to Predict Active Defending Bystander Responses

Predictor	All the sample			Students without an immigrant background			Students with an immigration background ⁴		
	B [95% CI]	p	β	B [95% CI]	p	β	B [95% CI]	p	β
<i>Outcome: Comforting the Victim</i>									
Gender ¹	0.25 [0.20, 0.31]	<.001	0.11	0.29 [0.23, 0.36]	<0.001	0.14	0.19 [0.09, 0.29]	0.002	0.08
SES ¹	0.02 [<.01, 0.03]	0.033	0.03	0.01 [-0.01, 0.03]	0.498	0.01	0.02 [<.01, 0.04]	0.101	0.04
Empathy ¹	0.64 [0.61, 0.68]	<.001	0.52	0.67 [0.63, 0.72]	<0.001	0.55	0.62 [0.56, 0.67]	<0.001	0.51
Grade 7 ²	0.10 [0.01, 0.19]	0.077	0.1	0.09 [-0.03, 0.21]	0.207	0.09	0.13 [-0.02, 0.29]	0.155	0.17
Grade 8 ²	0.10 [0.00, 0.19]	0.087	0.1	0.11 [<.01, 0.23]	0.098	0.11	0.07 [-0.10, 0.23]	0.494	0.08
Proportion students IB ^{2,3}	-0.23 [-0.38, -0.08]	0.013	-0.14	0.05 [-0.17, 0.27]	0.696	0.03	-0.53 [-0.77, -0.29]	<0.001	-0.37
Witnessing racist hate speech ²	-0.24 [-0.40, -0.07]	0.019	-0.14	-0.13 [-0.32, 0.07]	0.276	-0.07	-0.41 [-0.70, -0.11]	0.024	-0.28
Indirect association	-0.17 [-0.29, -0.05]	0.017		-0.29 [-0.43, -0.16]	<0.001				
<i>Outcome: Seeking Help at the School</i>									
Gender ¹	0.05 [-0.02, 0.13]	0.22	0.02	0.06 [-0.03, 0.15]	0.266	0.03	0.03 [-0.10, 0.15]	0.733	0.01
SES ¹	-0.01 [-0.03, 0.00]	0.127	-0.03	-0.01 [-0.03, 0.01]	0.504	-0.01	-0.03 [-0.06, 0.00]	0.058	-0.06
Empathy ¹	0.47 [0.43, 0.50]	<0.001	0.38	0.45 [0.41, 0.50]	<0.001	0.37	0.48 [0.42, 0.54]	<0.001	0.4
Grade 7 ²	0.47 [0.36, 0.57]	<0.001	0.36	0.43 [0.29, 0.56]	<0.001	0.34	0.55 [0.38, 0.72]	<0.001	0.48
Grade 8 ²	0.22 [0.12, 0.32]	<0.001	0.16	0.27 [0.14, 0.40]	0.001	0.21	0.13 [-0.02, 0.29]	0.146	0.11
Proportion students IB ²	-0.04 [-0.20, 0.12]	0.688	-0.02	0.13 [-0.13, 0.39]	0.405	0.06	-0.16 [-0.41, 0.10]	0.305	-0.08
Witnessing racist hate speech ²	-0.42 [-0.59, -0.25]	<.001	-0.18	-0.37 [-0.58, -0.16]	0.003	-0.16	-0.46 [-0.74, -0.18]	0.006	-0.22
Indirect association	-0.12 [-0.21, -0.04]	0.017		-0.20 [-0.29, -0.11]	<0.001				
<i>Outcome: Countering Hate Speech</i>									
Gender ¹	0.07 [0.01, 0.13]	0.066	0.03	0.01 [-0.06, 0.09]	0.765	0.01	0.13 [.03, 0.23]	0.025	0.06
SES ¹	0.02 [0.01, 0.04]	0.004	0.05	0.03 [0.01, 0.05]	0.017	0.06	0.02 [.00, 0.04]	0.156	0.04
Empathy ¹	0.50 [0.47, 0.54]	<0.001	0.46	0.51 [0.47, 0.56]	<0.001	0.47	0.49 [.44, 0.55]	<0.001	0.46
Grade 7 ²	0.03 [-0.07, 0.13]	0.649	0.02	0.01 [-0.11, 0.13]	0.924	0.01	0.07 [-0.09, 0.23]	0.468	0.06
Grade 8 ²	0.02 [-0.08, 0.11]	0.797	0.01	0.06 [-0.05, 0.18]	0.35	0.05	-0.06 [-0.22, 0.09]	0.510	-0.05
Proportion students IB ²	-0.07 [-0.23, 0.09]	0.451	-0.03	-0.07 [-0.29, 0.15]	0.605	-0.03	-0.22 [-0.47, 0.04]	0.162	-0.11
Witnessing racist hate speech ²	-0.31 [-0.48, -0.14]	0.002	-0.14	-0.22 [-0.41, -0.04]	0.046	-0.1	-0.42 [-0.71, -0.14]	0.015	-0.2
Indirect association	-0.13 [-0.23, -0.04]	0.016		-0.23 [-0.33, -0.13]	<.001				

Note. ¹ student level variable, ² classroom level variable. ³ IB = immigrant background. ⁴ The displayed results correspond to the direct associations-only model.

Figure 1

Direct and Indirect Associations between Classroom Racist Hate Speech and Active Defending Bystander Responses, via Empathy – Differences by Immigration Background



Note. Unstandardized coefficients are shown. Control variables are not displayed for clarity. For each path, the number on the left displays the coefficient for students without an immigration background ($n = 1,873$), and the number on the right displays the coefficient for students with a migration background ($n = 1,300$). All coefficients shown are statistically significant except for those displayed in italics.

room are likely to reduce their empathy, pay less attention to it, and judge it as normative and not severe. In addition, adolescents do not defend their victimized peers because they become afraid of being victimized (Kollerová et al., 2018; Strindberg et al., 2020). Thus, when observing that racist hate speech is common, they might conclude that it is accepted by their peers (or, at least, not negatively sanctioned). In sum, exposure to racist hate speech is associated with changes in emotional and cognitive outcomes that favor defending victims, which might also shape adolescents' perceptions about their classroom social dynamics. The latter assertion deserves further empirical exploration.

The second finding was that witnessing racist hate speech had an indirect association with the active defending bystanders' responses via low levels of empathy. Though a few studies have documented that exposure to peer aggression reduces prosocial behaviors (Bushman & Anderson, 2009; Prot et al., 2015), to our knowledge, this is the first made specifically for bystanders' active defending behaviors in racist hate speech. In line with the general aggression model, this finding can be explained by reduced automatic responses to witnessed hate speech, namely

desensitization (Miles-Novelo & Anderson, 2020). Accordingly, exposure to racist hate speech might affect the affective component of empathy (e.g., feeling concerned) more than other components (e.g., understanding how victims might feel). Additionally, the association between bystanders' active defending and empathy is stronger for its affective than for its cognitive component (Deng et al., 2021). In sum, bystanders who constantly witness racist hate speech are less likely to defend victims actively due to a diminished emotional reaction. Studies differentiating among multiple dimensions of empathy might enlighten our understanding of the impact of constant exposure to racist hate speech on adolescents.

The third finding was that, in line with our prediction, the levels of empathy of students with an immigrant background did not vary as a function of how much racist hate speech they witnessed in their classrooms. Importantly, these observed differential associations cannot be attributed to between-group differences in the analyzed outcomes. Contrary to previous studies about intergroup-based bullying (Bayram Özdemir et al., 2022), students with an immigrant background did not have higher levels of empathy nor intervened more often than their

classmates without an immigrant background. This finding corroborated that although students have the same disposition to intervene and experience empathy to the same extent, immigrant background is a protective factor against the detrimental consequences of witnessing racist hate speech. In line with the social identity theory, this might be associated with increases in in-group belonging and preference due to out-group interactions (Tajfel, 1982). Since hate speech is based on a collective identity (e.g., skin color, nationality), adolescents with an immigrant background's resistance to desensitization might result from a perceived threat to their social identity, either directly (i.e., being victims) or indirectly (i.e., being bystanders). In contrast to the dominant approach of studying having an immigrant background under the lens of vulnerability, we approached it as a resource in the present study.

With this conclusion, we do not pretend to opaque the vulnerability many students might experience due to their social identity as immigrants. However, approaches that view immigration background exclusively as a risk factor fall short. Our study shows that having an immigration background may protect students from the nocive effects of hate speech exposure on empathy. It is, therefore, important not to view immigration background exclusively as a risk factor but also to consider its potential protective functions. Likewise, an important point that must be considered when interpreting our findings is that desensitization occurred in the group of students without an immigrant background. According to the results, they were affected by the constant exposure to racist hate speech, such that this exposure negatively predicted their empathetic response. In line with our explanations above, these adolescents might be more susceptible to desensitization because their social identity is not compromised by the racist hate speech they witness.

Practical Implications

Our findings provide two lines of action. One, promoting empathy as a protective factor against the desensitization that can result from constantly witnessing hate speech in the school. The empirically tested school prevention program *HateLess* includes intergroup contact, provision of knowledge, and the development of social skills, as tools to develop empathy toward victims of hate speech (Wachs et al., 2023c). Similarly, fostering empathy is a central component of successful programs to tackle bullying

(e.g., *KiVa*; Garandeanu et al., 2022; and *Medienhelden* [Media Heroes]; Schultze-Krumbholz et al., 2018). Second, we found that immigrant background was a relevant factor in understanding desensitization. Therefore, fostering mutual positive social relationships characterized by respect, tolerance, and cooperation might increase students' awareness of multiple social identities (e.g., based on their belonging to a classroom or their preferences and hobbies). Accordingly, it would be easier for students to empathize with hate speech victims due to a perception of a shared social identity that goes beyond salient characteristics such as nationality or skin color.

Limitations and Future Studies

Despite the valuable contributions of the present study, some limitations must be acknowledged. First, within the general aggression model literature, empathy is, perhaps, the most studied affective outcome impacted by exposure to violence. Nonetheless, other emotional and cognitive factors can also be affected, such as increases in attitudes in favor of aggression and reductions in the ability to recognize violence as such (Miles-Novelo & Anderson, 2020). Models including these variables would allow answering important questions for prevention efforts, such as which factors are more susceptible to desensitization and how this impacts bystanders' active defending responses.

Second, we only addressed reductions in empathy. However, exposure to and perpetration of violence can also incite positive affect. Among other reasons, adolescents perpetrate online hate speech for exhilaration (Wachs et al., 2022b). Therefore, if perpetrators display these positive emotional states constantly, bystanders might interpret racist hate speech as amusing, minimizing its consequences for victims. Third, our conclusions are limited to witnessed face-to-face hate speech. Nonetheless, a systematic review found that the percentage of adolescents who have witnessed online hate speech varies between 26% and 39.2% (Kansok-Dusche et al., 2022). Future studies could deepen in this direction, studying if exposure to online racist hate speech has a differential impact on bystanders depending on their immigrant background. Finally, the interpretations of indirect associations are partially limited by the cross-sectional nature of our data and the potential desirability bias involved in questions regarding hypothetical situations. Therefore, our asseverations

regarding explanatory mechanisms must be interpreted with caution. Longitudinal and observational studies could offer a more comprehensive view of our findings.

Supplementary Material

The supplementary material is available in the electronic version of this article: <https://dx.doi.org/10.3233/DEV-230341>.

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., Bilz, L., & Schubarth, W. (2021). "Dann machen halt alle mit." Eine qualitative Studie zu Beweggründen und Motiven für Hatespeech unter Schüler*innen. *Diskurs Kindheits- und Jugendforschung / Discourse. Journal of Childhood and Adolescence Research*, *16*(4-2021), 463-480. <https://doi.org/10.3224/diskurs.v16i4.01>
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- 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. 3-16). The MIT Press. <https://doi.org/10.7551/mitpress/9780262012973.003.0002>
- Bayram Özdemir, S., Yanagida, T., & Özdemir, M. (2022). Bystanders of ethnic victimization: Do classroom context and teachers' approach matter for how adolescents intend to act? *Child Development*, *93*, 1540- 1558. <https://doi.org/10.1111/cdev.13822>
- 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>
- Bushman, B. J., & Anderson, C. A. (2020). General aggression model. In J. Bulck (Ed.), *The international encyclopedia of media psychology*. Wiley. <https://doi.org/10.1002/9781119011071.iemp0154>
- Bushman, B. J., & Anderson, C. A. (2009). Comfortably numb: Desensitizing effects of violent media on helping others. *Psychological Science*, *20*(3), 273-277. <https://doi.org/10.1111/j.1467-9280.2009.02287.x>
- Caravita, S. C. S., Stefanelli, S., Mazzone, A., Cadei, L., Thornberg, R., & Ambrosini, B. (2020). When the bullied peer is native-born vs. immigrant: A mixed-method study with a sample of native-born and immigrant adolescents. *Scandinavian Journal of Psychology*, *61*(1), 97-107. <https://doi.org/10.1111/sjop.12565>
- Carnagey, N. L., Anderson, C. A., & Bushman, B. J. (2007). The effect of video game violence on physiological desensitization to real-life violence. *Journal of Experimental Social Psychology*, *43*(3), 489-496. <https://doi.org/10.1016/j.jesp.2006.05.003>
- Castellanos, M., Wettstein, A., Wachs, S., Kansok-Dusche, J., Ballaschk, C., Krause, N. & Bilz, L. (2023) Hate speech in adolescents: A binational study on prevalence and demographic differences. *Frontiers in Education* (8),1076249. <https://doi.org/10.3389/feduc.2023.1076249>
- Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) family affluence scale. *Social Science & Medicine*, *66*(6), 1429-1436. <https://doi.org/10.1016/j.socscimed.2007.11.024>
- 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>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. <https://doi.org/10.3758/BF03193146>
- Garandea, C. F., Laninga-Wijnen, L., & Salmivalli, C. (2022). Effects of the KiVa anti-bullying program on affective and cognitive empathy in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *51*(4), 515-529.
- 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>
- Gönültaş, S., & Mulvey, K. L. (2021). The role of immigration background, intergroup processes, and social-cognitive skills in bystanders' responses to bias-based bullying toward immigrants during adolescence. *Child Development*, *92*(3). <https://doi.org/10.1111/cdev.13476>
- Gönültaş, S., & Mulvey, K. L. (2022). Do adolescents intervene in intergroup bias-based bullying? Bystander judgments and responses to intergroup bias-based bullying of refugees. *Journal of Research on Adolescence*, *33*(1), 4-23. <https://doi.org/10.1111/jora.12752>
- Gönültaş, S., Mulvey, K. L., Irdam, G., Goff, E., Irvin, M. J., Carlson, R., & DiStefano, C. (2020). The role of social-emotional factors in bystanders' judgments and responses to peer aggression and following retaliation in adolescence. *Journal of Emotional and Behavioral Disorders*, *28*(4), 195-208. <https://doi.org/10.1177/1063426619870492>
- 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 phenomenon. *Trauma, Violence & Abuse*. Advanced online publication. <https://doi.org/10.1177/15248380221108070>
- 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). <https://doi.org/10.5817/CP2018-4-3>
- Kollerová, L., Yanagida, T., Mazzone, A., Soukup, P., & Strohmeier, D. (2018). "They think that I should defend": Effects of peer and teacher injunctive norms

- on defending victimized classmates in early adolescents. *Journal of Youth and Adolescence*, 47(11), 2424-2439. <https://doi.org/10.1007/s10964-018-0918-2>
- Lucas-Molina, B., Giménez-Dasí, M., Fonseca-Pedrero, E., & Pérez-Albéniz, A. (2018). What makes a defender? A multilevel study of individual correlates and classroom norms in explaining defending behaviors. *School Psychology Review*, 47(1), 34-44. <https://doi.org/10.17105/SPR-2017-0011.V47-1>
- Machackova, H., & Pfetsch, J. (2016). Bystanders' responses to offline bullying and cyberbullying: The role of empathy and normative beliefs about aggression. *Scandinavian Journal of Psychology*, 57(2), 169-176. <https://doi.org/10.1111/sjop.12277>
- Miles-Novelo, A., & Anderson, C. A. (2020). Desensitization. In J. Bulck (Ed.), *The international encyclopedia of media psychology* (pp. 1-11). Wiley. <https://doi.org/10.1002/9781119011071.iemp0056>
- Muthén, B., & Muthén, L. (2012-2021). *MPlus* (Version 8.7.) [Computer Software].
- Obermaier, M., Schmuck, D., & Saleem, M. (2021). I'll be there for you? Effects of Islamophobic online hate speech and counter-speech on Muslim in-group bystanders' intention to intervene. *New Media & Society*, 14614448211017527. <https://doi.org/10.1177/14614448211017527>
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 209-233. <https://doi.org/10.1037/a0020141>
- Prescott, A. T., Sargent, J. D., & Hull, J. G. (2018). Meta-analysis of the relationship between violent video game play and physical aggression over time. *Proceedings of the National Academy of Sciences*, 115(40), 9882-9888. <https://doi.org/10.1073/pnas.1611617114>
- Prot, S., Anderson, C. A., Gentile, D. A., Warburton, W., Saleem, M., Groves, C. L., & Brown, S. C. (2015). Media as agents of socialization. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization* (2nd ed., pp. 276-300). Guilford Press.
- Reichelmann, A., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Räsänen, P., & Zych, I. (2021). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior*, 42(9), 1100-1111. <https://doi.org/10.1080/01639625.2020.1722337>
- Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior*, 15(2), 112-120. <https://doi.org/10.1016/j.avb.2009.08.007>
- Schultze-Krumbholz, A., Zagorscak, P., & Scheithauer, H. (2018). A school-based cyberbullying preventive intervention approach: The Media Heroes program. In M. Campbell & S. Bauman (Eds.), *Reducing cyberbullying in schools* (pp. 145-158). Elsevier. <https://doi.org/10.1016/B978-0-12-811423-0.00011-0>
- Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, 44(2), 136-146. <https://doi.org/10.1002/ab.21737>
- Strindberg, J., Horton, P., & Thornberg, R. (2020). The fear of being singled out: Pupils' perspectives on victimization and bystanding in bullying situations. *British Journal of Sociology of Education*, 41(7), 942-957. <https://doi.org/10.1080/01425692.2020.1789846>
- Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33, 1-39. <https://doi.org/10.1146/annurev.ps.33.020182.000245>
- 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>
- United States Government Accountability Office (2021). *K-12 EDUCATION Students' experiences with bullying, hate speech, hate crimes, and victimization in schools*. Retrieved October 14, 2022, from <https://www.gao.gov/products/gao-22-104341>
- Wachs, S., Bilz, L., Wettstein, A., Wright, M. F., Kansok-Dusche, J., Krause, N., & Ballaschk, C. (2022a). Associations between witnessing and perpetrating online hate speech among adolescents: Testing moderation effects of moral disengagement and empathy. *Psychology of Violence*, 12(6), 371-381. <https://doi.org/10.1037/vio0000422>
- Wachs, S., Wettstein, A., Bilz, L., & Gámez-Guadix, M. (2022b). Adolescents' motivations to perpetrate hate speech and links with social norms. *Comunicar*, 30(71), 9-20. <https://doi.org/10.3916/C71-2022-01>
- Wachs, S., Bilz, L., Wettstein, A., & Spelage, D. (2023a). Validation of the multidimensional bystander responses to racist hate speech scale and its association with empathy and moral disengagement among adolescents. *Aggressive Behavior*. Advanced online publication. <https://doi.org/10.1002/ab.22105>
- Wachs, S., Castellanos, M., Wettstein, A., Bilz, L., & Gámez-Guadix, M. (2023b). Associations between classroom climate, empathy, self-efficacy, and countering hate speech among adolescents: A multilevel mediation analysis. *Journal of Interpersonal Violence* (38), 5067-5091. <https://doi.org/10.1177/08862605221120905>
- Wachs, S., Krause, N., Wright, M. F., & Gámez-Guadix, M. (2023c). Effects of the prevention program "HateLess. together against hatred" on adolescents' empathy, self-efficacy, and countering hate speech. *Journal of Youth and Adolescence*, 52, 1115-1128. <https://doi.org/10.1007/s10964-023-01753-2>
- Yates, F., & Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society: Series B (Methodological)*, 15(2), 253-261. <https://doi.org/10.1111/j.2517-6161.1953.tb00140.x>

Bio Sketches

Dr. Melisa Castellanos is a Research Associate at the Research Group “Social Interaction in Educational Settings” at the University of Teacher Education Bern. Her research interest includes children and adolescents’ peer relationships, friendship, aggression, prosociality, hate speech, and immigration.

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, and 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, and school violence prevention.

Dr. Sebastian Wachs is a Professor of Educational Science with a Specialization in Digitalization at the University of Münster, Germany. His research interests focus on young people’s emotional and social development, well-being, risky behavior, and aggressive behavior in the online and offline context and the development and evaluation of prevention programs.