



Teachers' intervention strategies for handling hate-speech incidents in schools

Ludwig Bilz¹ · Saskia M. Fischer^{1,2} · Julia Kansok-Dusche¹ · Sebastian Wachs³ · Alexander Wettstein⁴

Received: 17 August 2023 / Accepted: 28 May 2024 / Published online: 12 June 2024
© The Author(s) 2024

Abstract

Despite ongoing global efforts to combat discrimination, hate speech is a growing problem in many countries. Hate speech is also widespread among young people and thus also affects schools. Because of the negative consequences of hate speech for the targeted individuals and for society in general, schools and teachers are generally encouraged to moderate hate speech. This binational study examines intervention strategies for hate speech in schools using the newly developed Hate-Speech Interventions Scale for Teachers. Data from German and Swiss teachers ($N=486$) were analyzed. Self-reported strategies were best described using a three-factor structure. Teachers primarily direct their interventions at those directly involved in hate speech and frequently employ teaching-oriented strategies. In comparison, external partners and colleagues are relatively rarely involved. Measurement invariance testing supported the scales' construct validity. A link was found between the use of these strategies, teachers' task-specific self-efficacy, gender, and professional experience. These results indicate that schools should play an important role in moderating hate speech and that increasing teachers' confidence in their own ability to address this phenomenon could be an important element of teacher training.

Keywords Hate speech · Intervention · Teacher · School · Self-efficacy · Professional experience

✉ Ludwig Bilz
ludwig.bilz@b-tu.de

¹ Department of Health Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Platz der Deutschen Einheit 1, 03046 Cottbus, Germany

² Faculty of Applied Social Sciences, RheinMain University of Applied Sciences, Wiesbaden, Germany

³ Institute of Education, University of Münster, Münster, Germany

⁴ Institute for Research, Development and Evaluation, Bern University of Teacher Education, Bern, Switzerland

1 Introduction

Hate speech, understood here as an intentionally harmful and derogatory expression about people (directly or vicariously) based on assigned group characteristics (e.g., ethnicity, sexual orientation, religious affiliation), is widespread among young people (Castellanos et al., 2023; Kansok-Dusche et al., 2023) and thus also affects schools. In contrast to bullying, which refers to an individual person by definition, the term hate speech implies a devaluation related to social group categories which, unlike bullying, does not necessarily require a personalized victim (Kansok-Dusche et al., 2023; Smith et al., 2013). There is growing empirical evidence that hate speech has negative consequences for the targeted individuals (Näsi et al., 2015; Stahel & Baier, 2023; Wachs et al., 2022) and for society in general (Bilewicz & Soral, 2020). Therefore, schools and teachers are generally encouraged to moderate hate speech. We know, from research on related phenomena such as bullying, that teacher behavior in these situations plays an important role and that teachers' interventions can help minimize problem behaviors (Fischer et al., 2022; Wachs, Bilz et al., 2019; Yoon & Bauman, 2014). To date, there is very limited evidence on how teachers deal with incidents of hate speech in schools (Krause et al., 2023). Against this background, the aim of this binational study was to present a newly developed instrument to assess teacher interventions in hate speech in schools, describing which strategies teachers use, and analyzing which individual and contextual variables are associated with the use of these strategies.

1.1 Hate speech in schools

The phenomenon of hate speech has been the subject of increased empirical research in recent years. Nevertheless, there have only been a few studies, mainly from Western countries, on hate speech in school-aged children and young people (e.g., Lehman, 2020; Wachs et al., 2021). A recent systematic review identified 18 publications that addressed the prevalence of hate speech (online and offline) in children and adolescents aged 5 to 21 years. In these studies, frequency rates for the last 12 months ranged from 26 to 68% for witnessing hate speech, 7–18% for being a victim of hate speech, and 5–32% for perpetrating hate speech (Kansok-Dusche et al., 2023). This wide range is due to differences in definitions, survey instruments, and samples. Nevertheless, these results showed that many children and adolescents have experienced hate speech, potentially at a developmental stage in which they are particularly vulnerable to negative consequences. Exposure to hate speech can be associated with lower well-being (Stahel & Baier, 2023; UK Safer Internet Centre, 2016), reduced levels of social trust (Näsi et al., 2015), and political radicalization (Bilewicz & Soral, 2020).

Because of these psychological consequences and the high prevalence of hate speech among school-aged children and adolescents, it is also a challenge for schools. Schools represent a suitable location to address hate speech incidents and to prevent future hate speech. There is also preliminary evidence from studies that school contextual characteristics are associated with the occurrence of hate speech (Krause et al., 2023; Wachs, et al., 2023). Hence, just as with the related phenomenon

of bullying (Hong & Espelage, 2012), a socio-ecological perspective on hate speech seems appropriate; and teachers represent a critical element of a school's social ecology. Through teaching and interacting with students, teachers shape the school and classroom climate and the attitudes of students (Yoon & Barton, 2008). It is therefore reasonable to assume that teachers' handling of hate-speech incidents in schools may have an impact on student behavior and could possibly help to moderate hate speech.

1.2 Teachers' interventions in hate-speech incidents in schools

The few publications available to date that looked at teacher interventions for hate speech in schools represent either theoretical reflections on the appropriateness of arts education and moral imagination (Arneback, 2014; Jääskeläinen, 2020), studies on how to address online hate speech or hate postings (Blaya, 2019; Strohmeier & Gradinger, 2021), or qualitative studies with small samples (Krause et al., 2023).

Further publications are available regarding the broader question of how teachers can address prejudice and racism in schools (Lynch et al., 2017). One example is affirmative action pedagogy (Boler, 2004). This approach is about how marginalized groups can have a voice in the classroom, "even at the small cost of limiting dominant voices" (p. 4). However, these approaches have a prescriptive rather than a descriptive character (an exception to this is the study by Arneback & Jämte, 2022) and are of limited help in answering the question of what teachers actually do to address the specific phenomenon of hate speech in schools.

A review of intervention programs to combat cyberhate primarily identified legislative and technological intervention strategies; the educational programs cited are preventive in nature, are almost exclusively aimed at promoting media literacy, and have rarely been evaluated (Blaya, 2019). A review of German-language anti-hate-speech programs also flagged that these programs focus exclusively on online hate speech and suggested that more evaluation is needed (Seemann-Herz et al., 2022). In their study, Strohmeier and Gradinger (2021) examined how teachers deal with the specific phenomenon of online hate postings. A total of 130 Austrian teachers surveyed reported their reactions to a hypothetical "hate posting" incident using a newly developed questionnaire. Six dimensions were identified using exploratory factor analysis: alerting other colleagues, victim-oriented rehabilitation, alerting the perpetrators' parents, authority-based sanctions, seeking help from external professionals, and ignoring. In their qualitative study, Krause et al. (2023) interviewed 46 German teachers and students on teachers' interventions in hate speech in schools. They identified eight intervention strategies, some of which overlap with those of Strohmeier and Gradinger (2021): punishment, involving the police, involving the parents, involving colleagues, mediation, and working with the whole class. However, others were found only in this study: teaching-related strategies and out-of-school projects and trainings. These differences could be due to the different target variables (hate speech vs. hate postings) and different methodological approaches.

Research on how teachers react to bullying is much more advanced. Here, there are widely used survey instruments for teachers' interventions (e.g., the Handling Bullying Questionnaire; Bauman et al., 2008), there is evidence about what teachers do in bullying situations (e.g., Burger et al., 2015), which strategies are more effec-

tive (e.g., Wachs, Bilz et al., 2019), and which competencies teachers should have to intervene successfully. For example, a consistent finding of several studies is that higher task-specific teacher self-efficacy is associated with teachers ignoring bullying less frequently and a more intensive use of intervention strategies (e.g., Fischer & Bilz, 2019; Fischer et al., 2021). Self-efficacy is the belief that one possesses the required capabilities to accomplish a given task successfully (Bandura, 1994). Teacher self-efficacy has been linked to many student outcomes and is considered an important component of teacher competence (Zee & Koomen, 2016).

Empirical evidence shows that hate speech can occur outside and within sequential bullying processes (Kansok-Dusche et al., 2023). Regarding participation roles, moderate correlations have been found between victimization through traditional bullying and online hate speech (Blaya et al., 2022) as well as between the perpetration of cyberbullying and the perpetration of online hate speech (Wachs, Wright et al., 2019). Despite this moderate empirical overlap between hate speech and bullying, they are distinct phenomena and findings from one field cannot be applied unconditionally to the other. Specific research findings on teachers' responses in the context of hate speech could be an important prerequisite for the development of prevention and intervention measures and for teacher training.

1.3 The present study

Since there are no quantitative research findings on teachers' responses to hate speech in schools, the first aim of this binational study was to develop a reliable and valid survey instrument. Such an instrument can be a valuable tool for further research or in evaluating teacher training interventions. For this purpose, the instrument's factor structure and measurement invariance were examined. The second aim was to describe teacher interventions to hate speech by considering associations with teachers' gender, professional experience, and country of residence. Finally, the specific hate-speech-related self-efficacy level and its relationship with teacher interventions were examined. Due to the lack of existing knowledge, the study was mainly exploratory in nature. Nevertheless, since dealing with hate speech does not play a major role in teacher training and school-based prevention work in German-speaking countries, we assumed that teachers may not be adequately prepared to deal with this phenomenon. Therefore, we expected teachers to have low self-efficacy expectations when dealing with hate speech. Given the results of bullying research on the subject, we expected a positive association between teachers' task-specific self-efficacy and the frequency with which they intervene in incidents of hate speech.

2 Methods

2.1 Participants

The sample consists of 486 German and Swiss teachers. Of those, 230 teachers (47%) worked at schools in the German speaking part of Switzerland and 256 teachers (53%) worked at schools in Germany (from the federal state of Berlin and the federal

state of Brandenburg). The teachers were on average $M=43.0$ years old ($SD=11.4$ years). The majority reported a female gender ($n=282$, 58%), $n=199$ reported a male gender (41%), only one indicated a diverse gender (0.2%), and four did not report their gender (0.8%). Less than a quarter ($n=102$, 21%) reported having an immigrant background—the standard measure of ethnic minority status in German-speaking countries is defined as either oneself or at least one parent being born outside of the country in question. In 2021, 27% of the population in Germany (Statistisches Bundesamt, 2023) and 39% of the population in Switzerland (Federal Statistical Office Switzerland, 2023) had an immigrant background. The teachers had worked on average for 15.7 years in their profession ($SD=12$ years).

2.2 Procedure

2.2.1 Scale development procedure

As a starting point, we took the German version of the Handling Bullying Questionnaire (HBQ, Burger et al., 2015) with its five dimensions (work with bully, work with victim, enlist adults, ignore, authority-based interventions). We removed items that captured bullying-specific interventions (e.g., *"I would help the bully achieve greater self-esteem so that he or she would no longer want to bully anyone."*), and changed the wording of the items by replacing "bullying" or "bully" with "hate speech" or "person who carried out hate speech," respectively. We then conducted 46 qualitative interviews with students and teachers, asking them: *"In what ways do you know to respond to hate speech?"* Prior to asking this question, a definition of hate speech was presented to the participants to increase the validity of their responses (Krause et al., 2023): *"We define hate speech as the public, deliberate and derogatory insulting of a social group. Sometimes an individual person is also insulted because they belong to a certain group. Hate speech can happen at school, but it can also take place online."* Some of the reported new hate-speech-specific intervention strategies could easily be assigned to the existing dimensions of the HBQ (e.g., *"I would work together with external partners [for instance, the anti-discrimination department, organizations etc.]"*). The following three strategies could not be assigned and were used to form the new dimension *"Teaching-oriented strategies"*: *"During class, I would educate students, challenge prejudices, and refute misinformation," "During class, I would teach students about the consequences of the discrimination and exclusion of minorities,"* and *"During class, I would lead a discussion on the fine line between stirring up hatred and free speech."* Eight HBQ bullying-related items were deleted, two items were merged into one item, and seven new items were added. This led to the first version of the Hate-Speech Interventions Scale for Teachers (HIST) with 20 items and six subscales (see Table 1), used for the survey with the present sample. As in the HBQ, a hypothetical scenario (here: a hate speech scenario) was used, to which the questions on the intervention strategies refer.

Table 1 Results of the item analysis and the EFA of the Hate-Speech Interventions Scale for Teachers (HIST)

No.		(full sample)			EFA factor loadings above 0.20 (subsample 1, $N=231$)		
		<i>M</i>	<i>SD</i>	Agreement ^a (%)	Factor 1 ^b	Factor 2 ^c	Factor 3 ^d
1	I would discuss with the perpetrator of hate speech how they could contribute to improving the situation.	3.28	1.00	81.2	0.67		
2	I would talk to the perpetrator of hate speech and explain why these kinds of expressions are hurtful and derogatory.	3.57	0.67	92.9	0.69		
3	I would discuss with the perpetrator of hate speech how they could behave in a more responsible and respectful way.	3.17	0.91	78.9	0.67		
4	I would support those affected by hate speech and offer them help.	3.22	1.02	80.0	0.38		
5	I would talk to those affected by hate speech about how they could react.	3.14	0.97	76.2	0.68		
6	I would advise those affected by hate speech to tell the perpetrators to leave them alone.	2.36	1.21	47.3	not included in EFA		
7	I would report the incident to someone more senior (for instance the principal).	2.43	1.06	45.0			0.68
8	I would contact the parents or guardians of the students involved.	2.32	1.10	44.2			0.63
9	I would plan how to deal with the incident with my colleagues in the school.	2.55	1.05	51.7			0.63
10	I would work together with external partners (for instance, the anti-discrimination department, organizations etc.).	1.85	1.02	25.9		0.25	0.43
11	I would ignore it.	0.24	0.57	1.0	not included in EFA		
12	I would not take the incident very seriously.	0.44	0.84	4.2	not included in EFA		
13	I would let the students sort out the problem themselves.	0.58	0.75	1.9	not included in EFA		
14	I would not get involved in that kind of incident, because my influence as a teacher is very limited.	0.58	0.87	4.1	not included in EFA		
15	I would insist that the perpetrators of hate speech stop doing what they are doing.	3.72	0.71	95.3	not included in EFA		
16	I would make it clear to the perpetrators that their behavior will not be tolerated.	3.80	0.48	97.7	not included in EFA		
17	I would ensure that the perpetrators are appropriately punished.	2.10	1.09	35.1	not included in EFA		
18	During class, I would educate students, challenge prejudices, and refute misinformation.	3.24	0.88	78.9		0.81	

Table 1 (continued)

No.		(full sample)			EFA factor loadings above 0.20 (subsample 1, $N=231$)		
		<i>M</i>	<i>SD</i>	Agreement ^a (%)	Factor 1 ^b	Factor 2 ^c	Factor 3 ^d
19	During class, I would teach students about the consequences of the discrimination and exclusion of minorities.	2.98	1.01	68.2		0.83	
20	During class, I would lead a discussion on the fine line between stirring up hatred and free speech.	2.89	1.03	67.6		0.74	

Note^a Categories “I would probably do that (75% of cases)” and “I would definitely do that (100% of cases).” ^b Working with those directly involved. ^c Teaching-oriented strategies. ^d Collaborating with others

2.2.2 Data collection and sampling procedure

The research was approved by the University of Potsdam Ethics Committee (UP65/2018). Teachers participated in this research on an anonymous and voluntary basis. In Germany, an acquisition pool of sample schools was created using a stratified and randomized probability-proportional-to-size scheme (Yates & Grundy, 1953). The stratification characteristics were federal state (Berlin and Brandenburg) and type of school (grammar secondary school [*Gymnasium*], mixed secondary schools with elements of both academic and non-academic schools [*Gesamtschule* or *Integrierte Sekundarschule*], non-academic-track secondary school [*Oberschule*] or school for special education [*Förderschule*]). In Switzerland, the acquisition pool of sample schools was designed using a contrastive sampling scheme that was based on high/low migrant background and on rural/urban geography. From the acquisition pool, a total of 100 schools (Germany: $n=76$; Switzerland: $n=24$) received phone calls and emails to inform them that their schools had been randomly selected to participate in this research. Due to the coronavirus pandemic, a significant proportion of German schools declined to participate (e.g., because of high regional infection rates or resource constraints). In total, 43 schools (Germany: $n=21$; Switzerland: $n=22$) agreed to participate (participation rate at the school level: 40% in total; 24% for Germany; 90% for Switzerland). In total, 1,621 teachers from these schools (Germany: $n=1,149$; Switzerland: $n=472$) were invited to participate. In Germany, teachers completed a paper-pencil survey between October 2020 and June 2021. In Switzerland, data gathering took place between December 2020 and April 2021. The teachers received an access code to an online questionnaire via email, which they subsequently completed.

The final sample for the data analysis consisted of 486 teachers (participation rate at the teacher level: 30% in total; 22% for Germany, 49% for Switzerland). In Germany, the teachers ($n=256$) were distributed across different school types (grammar secondary school [*Gymnasium*]: $n=88$, 34.4%; mixed secondary schools with elements of both academic and non-academic schools [*Gesamtschule* or *Integrierte Sekundarschule*]: $n=114$, 44.5%; non-academic secondary school [*Oberschule*]:

$n=40$; 15.6%; school for special education [*Förderschule*]: $n=14$, 5.5%). Swiss teachers ($n=230$) were distributed among the country-specific school types (separated model: $n=101$, 43.9%; cooperative model: $n=58$, 25.2%; integrated model: $n=69$, 30%; missing: $n=2$; 0.9%). The participation on the school level ranged from a minimum of three teachers per school to a maximum of 29 teachers per school.

2.3 Measures

2.3.1 Teachers' interventions in hate speech

The following introduction was given to the participants: *"Please imagine the following situation: In your classroom, students make publicly offensive statements about the skin color, origins, sexual orientation, religious affiliation, or gender of a group of people (e.g., people born in other countries, Muslims, homosexuals, women, trans people)."* The teachers were then asked the question: *"Have you ever experienced a situation like this?"* (response options: "yes" / "no"). Then the teachers were asked: *"What would you do or what have you done in this situation?"* They were then invited to respond to 20 items (see Table 1) using a five-point response scale: 0 = *"I would definitely not do that (0% of cases)"*, 1 = *"I probably would not do that (25% of cases)"*, 2 = *"I would do that every now and then (50% of cases)"*, 3 = *"I probably would do that (75% of cases)"*, 4 = *"I would definitely do that (100% of cases)"*.

2.3.2 Hate-speech-related self-efficacy

We assessed a task-specific form of self-efficacy related to the intervention in hate speech, understanding it as the teacher's individual expectation that they were capable of handling these incidents successfully. For this, we adapted the bullying-specific self-efficacy scale (Fischer & Bilz, 2019) by replacing "bullying" with "hate speech," removing one bullying-specific item, and adding four new items (*"I am confident that I can recognize the boundary between free speech and hate speech," "I know how to support students who have been the target of hate speech," "If there is hate speech at my school, I know who I need to turn to for help," "When it comes to hate speech from students, I feel confident that I can effectively counter it and take a stand."*) These additions were based on the idea that hate speech - more so than bullying - requires teachers to deal more intensively with the content of the derogatory comments. Answers were given on a four-point Likert scale: 0 = *"not at all true"*, 1 = *"hardly true"*, 2 = *"moderately true"*, 3 = *"completely true"*. CFA points to a one-factor-solution as most indices suggest an adequate fit ($\chi^2 [20]=77.88$, $p<.001$; $RMSEA=0.078$ [90% confidence interval CI : 0.060; 0.097; p ($RMSEA\leq 0.05$)=0.006]; $CFI=0.937$; $SRMR=0.042$). The Chi-square test, as well as the CFI, speak against an adequate model fit of the one-factor solution. A model with two factors (factor 1: recognizing hate speech, 4 items from the bullying-specific self-efficacy scale; factor 2: intervening in hate speech, 4 newly developed items) was tested but did not reach a better fit with an increased $RMSEA$ ($\chi^2 [19]=76.97$, $p<.001$; $RMSEA=0.080$ [90% confidence interval CI : 0.062; 0.099; p ($RMSEA\leq 0.05$)=0.004]; $CFI=0.937$; $SRMR=0.042$). As two common indices suggest an adequate fit, the one-factor solution is chosen.

Internal consistency (composite reliability) of the 8-item scale in the present sample was $CR=0.808$.

2.3.3 Control variables

Teachers were asked for their gender (male, female, gender diverse) and their professional experience in schools in number of years (two groups based on median split: 0 to 11 years vs. 12 to 45 years). The country of residence was assigned based on the country in which the data was collected.

2.4 Data analyses

2.4.1 Power analysis

A priori conducted power analysis with G*Power (Faul et al., 2007) revealed that, to detect small to medium correlational effect sizes ($\rho=0.20$), the present study needed a sample consisting of at least 193 participants ($\alpha=0.05$, Power=0.80). Taking the hierarchical structure of the sample (teachers in schools) and the non-response rate into consideration, the resulting minimum sample size is $N=296$ teachers (Teerenstra et al., 2010). Accordingly, the present sample size was sufficient to investigate the hypotheses.

2.4.2 Exploratory and confirmatory factor analyses

After an initial descriptive analysis of the items, the sample was randomly divided into two equally sized subsamples. The first subsample was used to examine the factor structure with an exploratory factor analysis (EFA; maximum likelihood estimation, Promax rotation) using the statistics program SPSS 28. To prevent the occurrence of random factors, a parallel analysis (O'Connor, 2000) was conducted for a randomly generated data matrix of the same size. The result of the EFA then underwent confirmatory testing using the data from Subsample 2. The Chi-square test and RMSEA, CFI, and SRMR were used as fit indices (interpretation according to Schermelleh-Engel et al., 2003: χ^2/df between 0 and 2 for a good fit, between 2 and 3 for an acceptable fit; $RMSEA \leq 0.05$ for a good fit, 0.05 to 0.8 for an adequate fit, 0.08 to 0.10 for a mediocre fit, >0.10 not acceptable; left boundary of the $CI < 0.05$ for a close fit; $CFI: >0.97$ for a good fit, >0.95 for an acceptable fit; $SRMR: <0.05$ for a good fit, <0.10 for an acceptable fit). Confirmatory factor analysis (CFA) was performed using Mplus 8.7 with robust maximum likelihood estimation (MLR) and type=complex to consider clustering.

2.4.3 Measurement invariance testing

Measurement invariance of the scale was tested for gender, professional experience, and country in which the data was collected. To do so, a series of multi-group confirmatory factor analyses was calculated in the total sample in Mplus version 8.7 (configural [free factor loadings and item intercepts, factor variances and means

fixed for model specification], metric [equal factor loading, free factor variances in second group], scalar [equal item intercepts, free factor means], strict [equal item residual variances, free factor means]). Cases with missing data were excluded from the analyses (Listwise Delete). The estimator used was a maximum likelihood estimator with robust standard errors, which is suitable for violations of the normal distribution assumption at the item level (MLMV was chosen; Maydeu-Olivares, 2017). Chi-square tests between the models (difference test) as well as RMSEA, CFI and SRMR were used to check the model fit. The indices were assessed according to Schermelleh-Engel et al. (2003) and Chen (2007; $\Delta CFI < -0.010$, $\Delta RMSEA > 0.015$, $\Delta SRMR > 0.015$). Unlike the calculation of CFA, it was not possible to consider the clustering of school-level data in the measurement invariance test (unreliability of the determined results due to too small sample sizes because of some schools being relatively small).

3 Results

3.1 Item analysis, EFA, and CFA

The mean values and standard deviations of the 20 items can be found in Table 1. Item analysis revealed extremely low item difficulties for items 11, 12, 13, and 14, and extremely high item difficulties for items 15 and 16. These item difficulties fall below the critical threshold of 0.8 or above the critical threshold of 3.2, respectively (representing values of 0.2 and 0.8, respectively, when the possible range of values is standardized to 0 to 1; Lienert & Raatz, 1994). These items are derived from the two HBQ scales *ignore* and *authority-based interventions*, which, in the original version, already exhibited problems in terms of item difficulty and internal consistency (Burger et al., 2015). For this reason, the seven items of these two dimensions were not included in the following analyses.

To explore the factor structure of the remaining 13 items by means of EFA, subsample 1 ($n=243$) was used. The parallel analysis indicated the adequacy of a three-factor solution, because the eigenvalue of the fourth factor was smaller than the corresponding criterion value of the comparison random matrix. The scree plot and the Kaiser criterion (eigenvalue > 1) also suggested a three-factor solution. Table 1 lists the factor loadings of the rotated pattern matrix of this three-factor-solution (Maximum likelihood, Promax rotation). The first factor ("*Working with those directly involved*") explained 30.7% of the variance and received high loadings from items 1 to 5, which originally stem from the two HBQ scales *work with the bully* and *work with the victim*. The new items 18 to 20 capturing "Teaching-oriented strategies" had high loadings on the second factor (explained variance: 7.6%), and the adapted items 7 to 10 from the HBQ scale *enlist adults* showed high loadings on the third factor ("Collaboration with others," explained variance: 8.0%). Because Item 6 had only a small loading on the first factor and, in addition, had a very low communality of 0.15, it was not considered for the following analyses.

The result of the EFA was tested in the second subsample ($n=243$) using CFA. The 3-factor solution obtained in the EFA showed a good global model fit

in the CFA (χ^2 [51]=77.16, $p=.011$; $RMSEA=0.048$ [90% CI: 0.024; 0.068; p ($RMSEA \leq 0.05$)=0.549]; $CFI=0.970$; $SRMR=0.047$). Figure 1 shows the standardized factor loadings, intercepts, residual variances, and the correlations of the latent factors. The matrix of the standardized covariance residuals (z-scores) is presented in the Online Resource. The matrix points to single problems concerning item 20 (which loads on factor 2, Teaching-oriented strategies) and item 7 and item 10 (which load on factor 3, Collaboration with others).

Composite Reliabilities of the three factors were above 0.7, indicating a reliable measure (factor 1, Working with directly involved students: $CR=0.815$; factor 2, Teaching-oriented strategies: $CR=0.829$; factor 3, Collaboration with others: $CR=0.767$). The Average Variance Extracted (AVE) is acceptable for factor 2 (factor 2, Teaching-oriented strategies: $AVE=0.619$), but is slightly below the level of 0.5 that is usually considered necessary for convergent valid measures (factor 1, Working with directly involved students: $AVE=0.477$; factor 3, Collaboration with others: $AVE=0.454$). However, as the AVE for factor 1 and factor 3 are close to the value of 0.5 and the CR for both factor 1 and factor 3 are above the threshold of 0.7, we consider the measure as generally reliable and valid. This is especially the case as the AVE is considered a more conservative measure than CR (e.g., Fornell & Larcker, 1981) and as the measure was thoroughly tested in an EFA before.

3.2 Measurement invariance testing

The results of the measurement invariance test are presented in Table 2. The results showed partial metric, partial scalar, and partial strict measurement invariance with respect to gender. Each of these could be achieved when the equality restrictions were removed for one of the four items of the factor “Collaboration with others” (item 9). The model comparison between the configural and the partial metric model was with

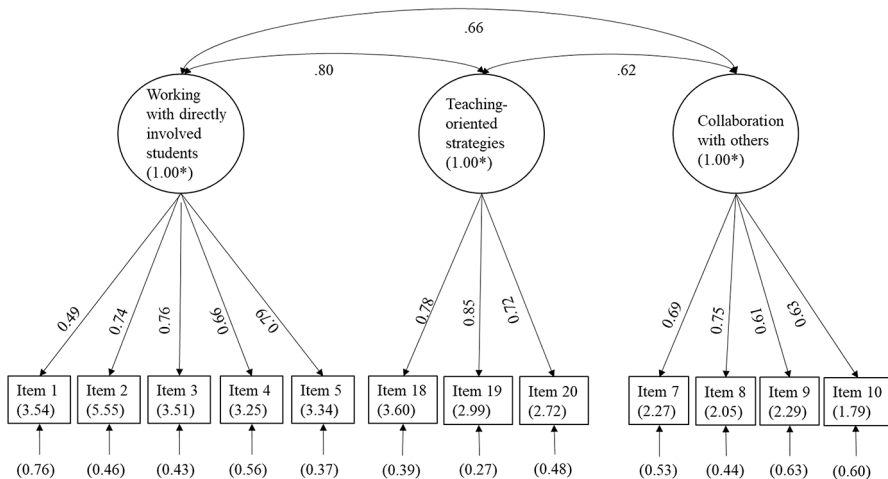


Fig. 1 Results of the Confirmatory Factor Analysis of the Hate Speech Interventions Scale for Teachers in Subsample 2 ($N=226$) (standardized values)

Table 2 Results of the measurement invariance testing of the Hate-Speech Interventions Scale for Teachers (HIST) across gender, professional experience, and country of residence

	χ^2 model fit	χ^2 difference test (model compared to lower level of mea- surement invariance)	RMSEA (90% CI)	CFI	SRMR
<i>Gender (N=454, male: n=184, female: n=270)</i>					
Configural invariance	χ^2 (102)=160.66; $p<.001$	-	0.050 (0.035; 0.065)	0.949	0.055
Metric invariance	χ^2 (111)=180.39; $p<.001$	$\Delta\chi^2$ (9)=24.71; $p=.003$	0.052 (0.038; 0.066)	0.940	0.067
Metric: Partial with removed equality restric- tions for item 9	χ^2 (110)=173.22; $p<.001$	$\Delta\chi^2$ (8)=15.28; $p=.054$	0.050 (0.035; 0.064)	0.945	0.062
Scalar: Partial with removed equality restric- tions for item 9	χ^2 (118)=181.44; $p<.001$	$\Delta\chi^2$ (8)=8.07; $p=.426$	0.049 (0.034; 0.062)	0.945	0.062
Strict: Partial with removed equality restric- tions for item 9	χ^2 (129)=190.05; $p<.001$	$\Delta\chi^2$ (11)=10.49; $p=.487$	0.046 (0.031; 0.059)	0.947	0.065
<i>Professional experience (N=457, 0 to 11 years: n=227, 12 to 45 years: n=230)</i>					
Configural invariance	χ^2 (102)=185.13; $p<.001$	-	0.060 (0.046; 0.073)	0.930	0.059
Metric invariance	χ^2 (111)=192.57; $p<.001$	$\Delta\chi^2$ (9)=7.67; $p=.567$	0.057 (0.043; 0.070)	0.931	0.063
Scalar invariance	χ^2 (120)=209.16; $p<.001$	$\Delta\chi^2$ (9)=20.63; $p=.014$	0.057 (0.044; 0.070)	0.925	0.065
Scalar: Partial with removed equality restric- tions for item 3 and item 5	χ^2 (118)=203.29; $p<.001$	$\Delta\chi^2$ (7)=12.41; $p=.088$	0.056 (0.043; 0.069)	0.928	0.064
Strict: Partial with removed equality restric- tions for item 3 and item 5	χ^2 (128)=209.62; $p<.001$	$\Delta\chi^2$ (10)=8.12; $p=.618$	0.053 (0.040; 0.065)	0.931	0.066
<i>Country of residence (N=459, Germany: n=230, Switzerland: n=229)</i>					
Configural invariance	χ^2 (102)=161.32; $p<.001$	-	0.050 (0.035; 0.065)	0.949	0.053
Metric invariance	χ^2 (111)=166.69; $p=.001$	$\Delta\chi^2$ (9)=5.13; $p=.823$	0.047 (0.031; 0.061)	0.952	0.056
Scalar invariance ^a	χ^2 (120)=211.47; $p<.001$	$\Delta\chi^2$ (9)=66.04; $p<.001$	0.058 (0.045; 0.070)	0.921	0.065
<i>Model only for factor 1 and factor 2: Country of residence (N=463, Germany: n=233, Swit- zerland: n=230)^b</i>					
Configural invariance (without factor 3)	χ^2 (38)=52.72; $p=.057$	-	0.041 (0.000; 0.066)	0.982	0.037
Metric invariance (with- out factor 3)	χ^2 (44)=55.93; $p=.107$	$\Delta\chi^2$ (6)=2.84; $p=.828$	0.034 (0.000; 0.059)	0.986	0.043
Scalar invariance (with- out factor 3)	χ^2 (50)=85.72; $p=.001$	$\Delta\chi^2$ (6)=43.57; $p<.001$	0.056 (0.035; 0.075)	0.957	0.057

Table 2 (continued)

	χ^2 model fit	χ^2 difference test (model compared to lower level of mea- surement invariance)	RMSEA (90% CI)	CFI	SRMR
Scalar: Partial with removed equality restric- tions for item 1 and item 4	χ^2 (48)=63.37; $p=.068$	$\Delta\chi^2$ (4)=9.33; $p=.053$	0.037 (0.000; 0.060)	0.981	0.045
Strict: Partial with removed equality restric- tions for item 1 and item 4	χ^2 (54)=78.52; $p=.016$	$\Delta\chi^2$ (6)=16.93; $p=.010$	0.044 (0.020; 0.065)	0.970	0.051

^a Stepwise removal of equality restrictions on the intercepts of items 4, 7, 8, 9, 20; in addition allowance of residual correlations between items 7 and 8 and between items 8 and 10 and between items 7 and 9; χ^2 difference test in all analyses $p \leq .011$

^b Testing only for partial scalar measurement invariance without factor 3 (Collaboration with others), as most problems occurred in this factor

$p=.054$ only slightly above the established significance threshold of $p=.05$. However, for all other indices (RMSEA, CFI, SRMR, χ^2/df) there was also no significant deterioration in the partial metric compared to the configural model ($\Delta \leq 0.007$) so that in combination with the reported chi-square difference test it was assumed that the corresponding higher level of measurement invariance was reached.

Regarding professional experience, complete metric, and partial scalar as well as partial strict measurement invariance was shown to be achieved in each case when the equality restrictions were removed for two of the five items of the factor “*Working with those directly involved*” (items 3 and 5).

With respect to the country of residence, complete metric but no scalar measurement invariance could be achieved. The problems were constantly related to the factor “*Collaboration with others*.” For this reason, the measurement invariance regarding the country of residence was additionally only checked for the factors “*Working with those directly involved*” and “*Teaching-oriented strategies*.” For these two factors, complete metric and partial scalar measurement invariance was found with respect to the country, which could be achieved in each case when the equality restrictions were removed for two items (one item of each factor; items 1 and 4). However, strict measurement invariance was not achieved. Due to the lower number of items, no further restrictions were removed to potentially achieve partial strict measurement invariance.

For all models, the fit indices indicated at least an acceptable fit. The only exception to this was the CFI value, which was often below 0.95, indicating an unacceptable fit. However, since CFI values are often smaller for maximum likelihood estimation methods (Shi & Maydeu-Olivares, 2020) and all other indices indicated at least an acceptable fit, each with different underlying methods for estimating goodness of fit (Lai & Green, 2016; Schermelleh-Engel et al., 2003), a sufficient model fit was assumed.

Overall, the results of the measurement invariance test showed that mean comparisons with respect to gender and professional experience are acceptable for all three

subscales. Regarding country of residence, mean comparisons are only possible for two of the three subscales and not for “*Collaboration with others.*”

3.3 Group comparisons and correlations

The analysis of the means of the three subscales showed that teachers would most likely work with the students who were directly involved ($M=3.28$, $SD=0.68$), followed by using teaching-oriented strategies ($M=3.04$, $SD=0.83$), and collaborating with others ($M=2.28$, $SD=0.79$). When asked whether they had ever experienced the situation described in the hypothetical scenario, 60.5% of teachers answered yes, and 39.5% answered no. Teachers who had never experienced such a hate speech incident were significantly more likely to collaborate with others ($M=2.41$, $SD=0.77$) than teachers with this experience ($M=2.20$, $SD=0.79$; $t[470]=2.86$, $p<.01$). There were no significant differences between the two groups for “*Working with those directly involved*” and “*Teaching-oriented strategies*”.

The subgroup comparisons regarding gender, professional experience, and country of residence are presented in Table 3. The analyses revealed that, when confronted with hate speech incidents, female teachers were significantly more likely to collaborate with others ($M=2.40$, $SD=0.74$) than male teachers ($M=2.12$, $SD=0.83$), and professionally experienced teachers were more likely to work with the students directly involved ($M=3.36$, $SD=0.66$) than less professionally experienced teachers ($M=3.20$, $SD=0.70$). The effect sizes of these group differences were small. All other comparisons were not significant.

The overall mean for hate-speech-related self-efficacy was $M=2.17$ ($SD=0.43$), indicating a mean response between the categories “*quite true*” and “*completely true.*” Significantly higher hate-speech-related self-efficacy was reported by male ($M=2.22$, $SD=0.38$) compared to female teachers ($M=2.13$, $SD=0.46$), more professionally experienced ($M=2.22$, $SD=0.43$) compared to less professionally experienced teachers ($M=2.12$, $SD=0.42$), and Swiss ($M=2.24$, $SD=0.41$) compared to German teachers ($M=2.10$, $SD=0.44$). The effect sizes of these group differences were small (Table 3).

The intercorrelations of the HIST scales were in the medium range for the correlation between “*Working with those directly involved*” and “*Collaboration*” ($r=.43$, $p<.001$) and for “*Teaching-oriented strategies*” and “*Collaboration*” ($r=.44$, $p<.001$); only the correlation between “*Working with those directly involved*” and “*Teaching-oriented strategies*” was in the high range with $r=.56$ ($p<.001$).

Multiple linear regression analyses were performed to examine the associations between hate-speech-related self-efficacy and the HIST subscales. Gender, professional experience, and country of residence were included as control variables (Table 4). The results revealed that all three subscales had a positive significant correlation with hate-speech-related self-efficacy. The associations with the subscales “*Working with those directly involved*” ($\beta = 0.36$, $p<.001$) and “*Teaching-oriented strategies*” ($\beta = 0.34$, $p<.001$) were in the medium range, and in the low range with the subscale “*Collaboration*” ($\beta = 0.13$, $p<.01$).

Table 3 Univariate comparisons of the three HIST scales and hate-speech-related self-efficacy regarding gender, professional experience, and country of residence

	Gender				Professional experience						Country of residence			
	Females		Males		0–11 years		12–45 years		t (Cohen's d)	r (Cohen's d)	Germany		Switzerland	
	M	SD	M	SD	M	SD	M	SD			M	SD	M	SD
Working with those directly involved	3.33	0.62	3.21	0.74	1.88 (0.18)	3.20	0.70	3.36	0.66	2.52* (0.23)	3.23	0.71	3.33	0.65
Teaching-oriented strategies	3.08	0.80	2.99	0.86	1.10 (0.10)	3.00	0.86	3.07	0.81	0.95 (0.09)	3.05	0.85	3.02	0.82
Collaboration	2.40	0.74	2.12	0.83	3.81*** (0.36)	2.21	0.79	2.35	0.79	1.87 (0.17)	<i>no scalar measurement invariance</i>			
Hate-speech-related self-efficacy	2.13	0.46	2.22	0.38	2.12* (0.20)	2.12	0.42	2.22	0.43	2.72** (0.25)	2.10	0.44	2.24	0.41

Note *** $p < .001$. ** $p < .01$. * $p < .05$

4 Discussion

The starting point for this study was, on the one hand, the growing significance of hate speech in schools and, on the other hand, the lack of empirical findings on teachers' intervention strategies. Against this background, we developed a new survey instrument, which we tested psychometrically. In addition, we analyzed the characteristics of three different intervention strategies, taking into account differences in gender, professional experience, and country of residence. Finally, we examined the correlations between the three intervention strategies and hate-speech-related self-efficacy.

4.1 Factor structure of the Hate-Speech Interventions Scale for Teachers (HIST)

This study demonstrates that teachers' intervention strategies in response to hate speech incidents can be validly and reliably captured by HIST with three dimensions: "*Working with those directly involved*," "*Teaching-oriented strategies*," and "*Collaboration with others*." The first dimension combines the two adapted scales "*Work with bully/victim*" from the HBQ, which is used in bullying research. Since, by definition, hate speech—unlike bullying—does not always necessarily involve identifiable victims who are present, it is not possible to separate, empirically, intervention strategies based on whether they target perpetrators or victims. The identification of victim-related strategies in Strohmeier and Grading's (2021) study of teacher interventions for hate postings could be related to the hypothetical scenario used in the study, which describes an incident in which an individual student is targeted. This is thus closer to the phenomenon of bullying than to hate speech, in which, as mentioned, there is not always an obvious binary of perpetrator and victim.

The second dimension captures teaching-oriented strategies in which teachers challenge prejudice or address the boundary between free speech and hate speech. These strategies appear to be specific to dealing with hate speech and have not previously appeared in research on bullying (Bauman et al., 2008; Burger et al., 2015), but have been mentioned in approaches for dealing with racism and prejudice in schools (Lynch et al., 2017). In contrast, the third dimension, which addresses collaboration with others (parents, external partners, colleagues), is used in bullying research and is also described by Strohmeier and Grading (2021) in their study on hate postings.

Authority-based strategies and ignoring the problem, which are strategies known from bullying research, cannot be confirmed as distinguishable dimensions for interventions in incidents of hate speech. The HBQ has already shown problems with internal consistency for these two scales (Burger et al., 2015). In this study, it is not possible to reliably capture the strategy of ignoring the problem. Although agreement rates on these items are encouragingly very low in our study, it can be of great practical value in capturing tendencies of individual teachers to ignore hate speech in schools. For this purpose, we suggest additionally using item no. 12 ("I would not take the incident very seriously") separately from the three dimensions of HIST, as the agreement rate was still somewhat higher for this item.

In contrast, almost all teachers agreed with the statements that described authority-based interventions. The issue here for items 15 and 16, in particular, was probably that the items did not succeed in capturing the authoritarian nature of these interven-

Table 4 Regression coefficients of hate-speech-related self-efficacy and control variables on HIST scales

	Working with those directly involved			Teaching-oriented strategies			Collaboration		
	<i>B</i>	β	<i>SE</i>	<i>B</i>	β	<i>SE</i>	<i>B</i>	β	<i>SE</i>
Constant	1.58***		0.21	1.59***		0.26	1.49***		0.25
Hate-speech-related self-efficacy	0.59***	0.36	0.07	0.66***	0.34	0.09	0.25**	0.13	0.09
Gender ^a	0.17**	0.13	0.06	0.12	0.08	0.07	0.29***	0.18	0.07
Professional experience ^b	0.11	0.08	0.06	0.01	0.00	0.07	0.12	0.08	0.07
Country of residence ^c	0.00	0.00	0.06	-0.13	-0.08	0.07	-0.25***	-0.16	0.07
<i>R</i> ²	0.15			0.12			0.08		
<i>N</i>	453			465			461		

Note^a Male = 1, female = 2. ^b 0–11 years = 1, 12–45 years = 2. ^c Germany = 1, Switzerland = 2. *** $p < .001$, ** $p < .01$

tions (e.g., punishing, not explaining sanctions) that we had intended them to incorporate. These items focus more on stopping hate speech quickly, meaning that many teachers may, presumably, have agreed with these statements, without it necessarily revealing an authoritarian tendency in their behaviors. For further development of the instrument with regard to this aspect, it could be helpful to take established instruments of parenting-style research as a starting point.

The testing of the measurement invariance shows that the three HIST subscales can be used to compare the mean values of female and male teachers, teachers with little vs. a lot of professional experience, and teachers from Germany and Switzerland (though, in the case of the latter, not for the subscale “*Collaborating with others*”).

4.2 Teachers' self-reported intervention strategies in hate-speech incidents

If they imagine a case of hate speech in their classroom, most teachers would direct their interventions toward the students directly involved. Teaching-oriented strategies ranked second, and collaboration with parents, colleagues, or external partners ranked last. These findings are not consistent with the few previous findings on teacher responses to hate postings, where teachers were most likely to involve other colleagues, followed by victim-oriented strategies, involving parents, authority-based sanctions, and lastly seeking help from external professionals (Strohmeier & Gradinger, 2021). Only two dimensions (“*Working with those directly involved*” and “*Collaboration with others*”) can be compared to research on teacher interventions for bullying; when compared to teachers' self-reports in these areas, we find poor agreement (Bauman et al., 2008; Burger et al., 2015) but greater agreement when students report on their teachers' interventions (Wachs, Bilz et al., 2019). The lower preference for involving others and the preference for teaching-oriented strategies found in this study may indicate that teachers attribute different causal factors to hate speech compared to bullying. While bullying may be understood more as a social phenomenon that can only be countered through joint efforts, the causes of hate speech are more likely to be seen in personal aspects like knowledge deficits and prejudices. Also, because hate speech is more closely linked to social inequalities, teachers may be more inclined to address these incidents in their teaching. However,

because teachers' causal attributions for hate speech and bullying were not explored in this study, these considerations remain speculative.

To qualify the findings on group differences by gender, professional experience, and country of residence, it is necessary to refer to findings on related phenomena, since no findings are available on hate speech. The finding that female teachers are more likely to involve others in their interventions aligns with Bauman et al.'s (2008) findings on U.S. teachers' interventions for bullying. Burger et al. (2015), on the other hand, could not confirm the same with Austrian teachers. And there were also no gender differences in how teachers involve others when dealing with hate postings (Strohmeier & Gradinger, 2021). The role of professional experience has only been examined in the context of bullying interventions. As was the case in this study, Burger et al. (2015) reported a stronger focus on interventions with those directly involved among more professionally experienced Austrian teachers, whereas these differences did not exist among U.S. teachers (Bauman et al., 2008). However, it should be noted that the differences found in the present study are relatively small in magnitude.

The correlation analyses show a close relationship between the two HIST strategies "*Working with those directly involved*" and "*Teaching-oriented strategies*." In contrast, "*Collaborating with others*" seems to be less strongly correlated with these strategies. One possible explanation for this could be that collaborating with colleagues, parents, and external partners on hate-speech incidents requires more specific competencies that not all teachers possess. This is consistent with the finding that these strategies are used comparatively seldom.

4.3 Associations with hate-speech-related self-efficacy

The hypothesis that teachers would have comparatively low self-efficacy when it comes to dealing with hate speech cannot be confirmed. Teachers feel confident in recognizing and countering hate speech and supporting those affected. This is particularly true for male teachers with more professional experience and more so for Swiss teachers than for German teachers. One reason could be that, although dealing with hate speech is not yet a prominent part of teacher training, teachers may experience it in practice (Kansok-Dusche et al., 2023) and thus develop confidence in their competencies in dealing with it. Another reason for the high level of self-efficacy could be that teachers can identify links between the content of their professional training and the competencies needed to deal with hate speech. For example, the standards for teacher training that are binding for Germany list some relevant competencies for prospective teachers (e.g., "Teachers teach values and norms, an attitude of appreciation and recognition of diversity, and support self-determined, reflected judgment and action by students," KMK, 2019, p. 10).

As expected, teachers with higher self-efficacy expectations do use all intervention strategies for hate speech more often than teachers with lower self-efficacy. This association is stronger for the two strategies "*Working with those directly involved*" and "*Teaching-oriented strategies*," and only very weak with the strategy "*Collaboration with others*." The first two strategies may be more likely to be composed of practices that teachers use in other contexts, while more specific skills are needed to

collaborate with others on hate speech incidents (especially when external partners are involved). As such, confidence in one's own competencies in this area may not be sufficient. In the area of bullying, it is particularly the collaboration between colleagues at all levels that is an important part of prevention, for example in the whole-school approach of Dan Olweus (Olweus & Limber, 2010). Also, the link between teacher interventions and self-efficacy has already been demonstrated in many studies on bullying (Bradshaw et al., 2007; Fischer & Bilz, 2019; Williford & Depaolis, 2016). Wachs, et al. (2023) also confirm the importance of self-efficacy for students' use of counter-speech in hate-speech incidents in school, stressing the importance of further investigation into hate-speech-related self-efficacy, both in students and teachers.

4.4 Limitations, strengths, and further directions

Despite the new findings in the current study, several limitations should be taken into account. Firstly, the data used here are based on teachers' self-reports and thus may be susceptible to self-referential bias. Future studies should also consider student reports of their teachers' intervention behaviors, as is already common in bullying research (Bilz & Fischer, 2020; Wachs, Bilz et al., 2019). Secondly, teachers' intervention behavior is assessed using a hypothetical scenario. Although this scenario encompasses various forms of hate speech, it may not represent the full range of hate speech incidents that teachers may encounter in practice. Given that 60% of teachers stated that they had already experienced such a situation, this suggests that many based their responses on actual experiences with hate speech. However, only 40% expressed behavioral intentions. Even though our analyses point to only small differences between the two groups, this must be considered a methodological limitation. Studies using teacher reports of real hate-speech incidents could potentially come to different conclusions and would also be able to consider more situational and contextual determinants of teacher interventions. Another limitation is the low response rate at the school level and the teacher level, especially in the German subsample. Therefore, further caution is needed when generalizing the study results. In addition, using a four-point response scale when assessing hate-speech-related self-efficacy could be associated with lower psychometric precision compared to a five- or six-point response scale (Simms et al., 2019). The items and item structure of HIST have thoroughly been analyzed in exploratory and confirmatory factor analyses based on two sub-samples. The standardized covariance residuals point to problems with three items (items 7, 10, 20). In addition, the AVE for factors 1 and 3 are too low. As most standardized covariance residuals and the CR for all three factors point to a reliable and valid measure, no further changes have been made to the item structure that had been identified in the EFA. However, the described problems suggest that the measure could be improved by re-formulating particular items and adding further items. Finally, the fact that strategies that are authority-based or involve ignoring what is going on cannot be confirmed as part of teacher responses to hate speech in this study may have been due to the wording of the corresponding items. The challenge for future studies would be to find better ways of describing these issues that do not result in such extreme item difficulties. The examination of contextual variables

could also be a worthwhile objective of future studies in this area. The differences found between German and Swiss teachers point in this direction. It can be assumed that school- and class-specific factors are also related to how teachers deal with hate speech. In addition to variable-oriented data analysis, person-oriented analysis strategies such as latent profile analysis could provide interesting insights into the behavior of intervention within specific groups of teachers in future studies. A strength of this study is the rigorous psychometric testing of a new instrument in two different countries and the demonstration that the HIST can be used in a reliable and valid way to assess teacher interventions for hate speech in schools. Nevertheless, it would be desirable to establish further evidence of validity.

4.5 Practical implications

The finding that most teachers would use a wide range of intervention strategies in response to hate speech and report high confidence in their own abilities to intervene in hate speech is an encouraging result and shows that teachers and schools should play an important role in reducing hate speech. Teachers' real-life experiences in using strategies, such as transferring knowledge, challenging prejudices, and countering misinformation, could be used to develop new evidence-based prevention programs. The relationship found between the use of these strategies and teacher self-efficacy could also motivate efforts to increase teachers' confidence in their own competencies. Bullying research revealed that teacher self-efficacy can be increased through training (Byers et al., 2011; Crooks et al., 2017; Newman-Carlson & Horne, 2004). The comparatively low use of collaborative strategies and the weaker association with self-efficacy may indicate that teachers need more support in this specific area. The setup of support networks with external partners or established forms of staff cooperation within the school could be helpful in this regard.

4.6 Conclusions

This study examines teachers' intervention strategies in response to hate speech in school. We present and test HIST as a newly developed survey instrument that can be used to reliably and accurately measure intervention strategies along three dimensions. Data from teachers in two different countries show that teachers focus their interventions primarily on those directly involved in hate speech and that they frequently use teaching-oriented strategies. In comparison, external partners or colleagues are rarely involved in interventions. These results, and the association found with self-efficacy, indicate that schools should play an important role in moderating hate speech and that increasing teachers' confidence in their own ability to address this phenomenon could be an important element of teacher training.

Author contributions Ludwig Bilz: Conceptualization, Methodology, Investigation, Writing Original Draft, Writing– Review & Editing, Funding acquisition. Saskia M. Fischer: Methodology, Writing– Review & Editing. Julia Kansok-Dusche: Investigation, Writing– Review & Editing. Sebastian Wachs: Investigation, Writing– Review & Editing, Funding acquisition. Alexander Wettstein: Investigation, Writing– Review & Editing, Funding acquisition.

Funding This work was supported by the German Research Foundation (BI 1046/9–1; SCHU 1370/4–1; WA 4275/2–1) and the University of Teacher Education Bern (Grant number 19s 0008 01). Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arneback, E. (2014). Moral imagination in education: A deweyan proposal for teachers responding to hate speech. *Journal of Moral Education*, 43(3), 269–281. <https://doi.org/10.1080/03057240.2014.918876>
- Arneback, E., & Jämte, J. (2022). How to counteract racism in education– A typology of teachers' anti-racist actions. *Race Ethnicity and Education*, 25(2), 192–211. <https://doi.org/10.1080/13613324.2021.1890566>
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (pp. 71–81). Academic.
- Bauman, S., Rigby, K., & Hoppa, K. (2008). US teachers' and school counsellors' strategies for handling school bullying incidents. *Educational Psychology*, 28(7), 837–856. <https://doi.org/10.1080/01443410802379085>
- Bilewicz, M., & Soral, W. (2020). Hate speech epidemic: The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, 41(1), 3–33. <https://doi.org/10.1111/pops.12670>
- Bilz, L., & Fischer, S. M. (2020). Interventionsstrategien und Interventionserfolg von Lehrkräften bei Cybermobbing und traditionellem Mobbing aus Schülersicht. *Kindheit und Entwicklung*, 29(2), 84–91. <https://doi.org/10.1026/0942-5403/a000306>
- Blaya, C. (2019). Cyberhate: A review and content analysis of intervention strategies. *Aggression and Violent Behavior*, 45, 163–172. <https://doi.org/10.1016/j.avb.2018.05.006>
- Blaya, C., Audrin, C., & Skrzypiec, G. (2022). School bullying, perpetration and cyberhate: Overlapping issues. *Contemporary School Psychology*, 26(3), 341–349. <https://doi.org/10.1007/s40688-020-00318-5>
- Boler, M. (2004). All speech is not free: The ethics of affirmative action pedagogy. In M. Boler (Ed.), *Democratic dialogue in education. Troubling speech, disturbing silence* (pp. 3–14). Peter Lang.
- Bradshaw, C. P., Sawyer, A. L., & O'Brennan, L. M. (2007). Bullying and peer victimization at school: Perceptual differences between students and school staff. *School Psychology Review*, 36(3), 361–382. <https://doi.org/10.1080/02796015.2007.12087929>
- Burger, C., Strohmeier, D., Spröber, N., Bauman, S., & Rigby, K. (2015). How teachers respond to school bullying: An examination of self-reported intervention strategy use, moderator effects, and concurrent use of multiple strategies. *Teaching and Teacher Education*, 51, 191–202. <https://doi.org/10.1016/j.tate.2015.07.004>
- Byers, D. L., Caltabiano, N., & Caltabiano, M. (2011). Teachers' attitudes towards overt and covert bullying, and perceived efficacy to intervene. *Australian Journal of Teacher Education*, 36(11), 105–119. <https://doi.org/10.14221/ajte.2011v36n11.1>

- 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. <https://doi.org/10.3389/educ.2023.1076249>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Crooks, C. V., Jaffe, P. G., & Rodriguez, A. (2017). Increasing knowledge and self-efficacy through a pre-service course on promoting positive school climate: The crucial role of reducing moral disengagement. *Advances in School Mental Health Promotion*, 10(1), 49–64. <https://doi.org/10.1080/1754730X.2016.1249383>
- 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>
- Federal Statistical Office Switzerland (2023). Migration and integration—At a glance. Retrieved July 25, 2023, from <https://www.bfs.admin.ch/bfs/en/home/statistics/population/migration-integration/by-migration-status.html>.
- Fischer, S. M., & Bilz, L. (2019). Teachers' self-efficacy in bullying interventions and their probability of intervention. *Psychology in the Schools*, 56(5), 751–764. <https://doi.org/10.1002/pits.22229>
- Fischer, S. M., John, N., & Bilz, L. (2021). Teachers' self-efficacy in preventing and intervening in school bullying: A systematic review. *International Journal of Bullying Prevention*, 3(3), 196–212. <https://doi.org/10.1007/s42380-020-00079-y>
- Fischer, S. M., Woods, H. A., & Bilz, L. (2022). Class teachers' bullying-related self-efficacy and their students' bullying victimization, bullying perpetration, and combined victimization and perpetration. *Journal of Aggression Maltreatment & Trauma*, 31(2), 184–203. <https://doi.org/10.1080/10926771.2021.1933290>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.1177/002224378101800104>
- Hong, J. S., & Espelage, D. L. (2012). A review of research on bullying and peer victimization in school: An ecological system analysis. *Aggression and Violent Behavior*, 17(4), 311–322. <https://doi.org/10.1016/j.avb.2012.03.003>
- Jääskeläinen, T. (2020). Countering hate speech through arts and arts education: Addressing intersections and policy implications. *Policy Futures in Education*, 18(3), 344–357. <https://doi.org/10.1177/1478210319848953>
- Kansok-Dusche, J., Ballaschk, C., Krause, N., Zeißig, A., Seemann, L., Wachs, S., & Bilz, L. (2023). A systematic review on hate speech among children and adolescents: Definitions, prevalence, and overlap with related phenomena. *Trauma Violence & Abuse*, 24(4), 2598–2615. <https://doi.org/10.1177%2F15248380221108070>
- KMK (2019). *Standards für die Lehrerbildung: Bildungswissenschaften*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Retrieved August 17, 2023, from https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf.
- Krause, N., Wachs, S., Bilz, L., & Schubarth, W. (2023). Dem Hass entgeg(n)en— Eine qualitative Studie zu den Interventionsstrategien, -zielen und -effekten durch pädagogisches Schulpersonal bei Hatespeech in der Schule. *Zeitschrift für Pädagogik*, 69(3), 310–332. <https://doi.org/10.3262/ZP2303310>
- Lai, K., & Green, S. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research*, 51(2–3), 220–239. <https://doi.org/10.1080/00273171.2015.1134306>
- Lehman, B. (2020). Hate at school: Victimization and disorder associated with school avoidance. *Sociological Spectrum*, 40(3), 172–190. <https://doi.org/10.1080/02732173.2020.1734890>
- Lienert, G. A., & Raatz, U. (1994). *Testaufbau und Testanalyse*. Beltz.
- Lynch, I., Swartz, S., & Isaacs, D. (2017). Anti-racist moral education: A review of approaches, impact and theoretical underpinnings from 2000 to 2015. *Journal of Moral Education*, 46(2), 129–144. <https://doi.org/10.1080/03057240.2016.1273825>
- Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling*, 24(3), 383–394. <https://doi.org/10.1080/10705511.2016.1269606>

- Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E., & Oksanen, A. (2015). Exposure to online hate material and social trust among Finnish youth. *Information Technology & People*, 28(3), 607–622. <https://doi.org/10.1108/ITP-09-2014-0198>
- Newman-Carlson, D., & Horne, A. M. (2004). Bully busters: A psychoeducational intervention for reducing bullying behaviour in middle school students. *Journal of Counseling & Development*, 82(3), 259–267. <https://doi.org/10.1002/j.1556-6678.2004.tb00309.x>
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods Instruments & Computers*, 32(3), 396–401. <https://doi.org/10.3758/bf03200807>
- Olweus, D., & Limber, S. P. (2010). The Olweus Bullying Prevention Program. Implementation and evaluation over two decades. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *The handbook of bullying in schools: An international perspective* (pp. 377–401). Routledge.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit for structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 28–74.
- Seemann-Herz, L., Kansok-Dusche, J., Dix, A., Wachs, S., Krause, N., Ballaschk, C., Schulze-Reichelt, F., & Bilz, L. (2022). Schulbezogene Programme zum Umgang mit Hatespeech– Eine kriteriengeleitete Bestandsaufnahme. *Zeitschrift für Bildungsforschung*, 12(3), 597–614. <https://doi.org/10.1007/s35834-022-00348-4>
- Shi, D., & Maydeu-Olivares, A. (2020). The effect of estimation methods on SEM fit indices. *Educational and Psychological Measurement*, 80(3), 421–445. <https://doi.org/10.1177/0013164419885164>
- Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, 31(4), 557–566. <https://doi.org/10.1037/pas0000648>
- Smith, P. K., del Barrio, C., & Tokunaga, R. S. (2013). Definitions of bullying and cyberbullying: How useful are the terms? In S. Bauman, D. Cross, & J. Walker (Eds.), *Principles of cyberbullying research: Definitions, measures, and methodology* (pp. 26–40). Routledge / Taylor & Francis Group.
- Stahel, L., & Baier, D. (2023). Digital hate speech experiences across age groups and their impact on well-being: A nationally representative survey in Switzerland. *Cyberpsychology Behavior and Social Networking*, 26(7), 519–526. <https://doi.org/10.1089/cyber.2022.0185>
- Statistisches Bundesamt (2023). *Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2021* Retrieved July 25, 2023, from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/Downloads-Migration/migrationshintergrundergebnisse-2010220217004.pdf?__blob=publicationFile.
- Strohmeier, D., & Gradinger, P. (2021). Teachers' knowledge and intervention strategies to handle hate-postings. *European Journal of Developmental Psychology*, 18(6), 865–879. <https://doi.org/10.1080/17405629.2021.1877130>
- 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>
- UK Safer Internet Centre (2016). *Creating a better internet for all: Young people's experiences of online empowerment + online hate* UK Safer Internet Centre. Retrieved August 16, 2023, from <https://child-netsic.s3.amazonaws.com/ufiles/SID2016/Creating%20a%20Better%20Internet%20for%20All.pdf>.
- Wachs, S., Bilz, L., Niproschke, S., & Schubarth, W. (2019). Bullying intervention in schools: A multilevel analysis of teachers' success in handling bullying from the students' perspective. *Journal of Early Adolescence*, 39(5), 642–668. <https://doi.org/10.1177/02F0272431618780423>
- Wachs, S., Wright, M. F., & Vazsonyi, A. T. (2019). Understanding the overlap between cyberbullying and cyberhate perpetration: Moderating effects of toxic online disinhibition. *Criminal Behaviour and Mental Health*, 29(3), 179–188. <https://doi.org/10.1002/cbm.2116>
- Wachs, S., Costello, M., Wright, M. F., Flora, K., Daskalou, V., Maziridou, E., Kwon, Y., Na, Y. E., Sittichai, R., Biswal, R., Singh, R., Almendros, C., Gámez-Guadix, M., Gözrig, A., & Hong, S. J. (2021). DNT LET 'EM H8 U! Applying the routine activity framework to understand cyberhate victimization among adolescents across eight countries. *Computers & Education*, 160, 104026. <https://doi.org/10.1016/j.compedu.2020.104026>
- Wachs, S., Gámez-Guadix, M., & Wright, M. F. (2022). Online 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., Castellanos, M., Wettstein, A., Bilz, L., & Gámez-Guadix, M. (2023). Associations between classroom climate, empathy, self-efficacy, and countering hate speech among adolescents: A multilevel mediation analysis. *Journal of Interpersonal Violence*, 38(5–6), 5067–5091. <https://doi.org/10.1177/08862605221120905>
- Williford, A., & Depaolis, K. J. (2016). Predictors of cyberbullying intervention among elementary school staff: The moderating effect of staff status. *Psychology in the Schools*, 53(10), 1032–1044. <https://doi.org/10.1002/pits.21973>
- 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.
- Yoon, J. S., & Barton, E. (2008). The role of teachers in school violence and bullying prevention. In T. W. Miller (Ed.), *School Violence and primary Prevention* (pp. 249–275). Springer.
- Yoon, J., & Bauman, S. (2014). Teachers: A critical but overlooked component of bullying prevention and intervention. *Theory into Practice*, 53(4), 308–314. <https://doi.org/10.1080/00405841.2014.947226>
- Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, 86(4), 981–1015. <https://doi.org/10.3102/0034654315626801>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ludwig Bilz is a Professor of Educational Psychology at the Faculty of Human Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Germany. His research interests encompass aggressive and bullying behavior, school mental health, teacher competence, school health promotion, and school violence prevention.

Saskia M. Fischer is a research assistant at the Faculty of Human Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Germany. Currently, she is also a visiting professor for Developmental and Social Psychology in the Context of Social Work and Health at the RheinMain University of Applied Sciences in Wiesbaden, Germany. Her research interests include bullying at school and the role of teachers in school bullying, parent-teacher collaboration in bullying prevention and intervention, mental health of children and adolescents and influencing factors at the family level, and perceived social support of children and adolescents.

Julia Kansok-Dusche is a research assistant and Ph.D. student at the Faculty of Human Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Germany. Her current research focuses on educational approaches to address hate speech with young people. She is also interested in the overlap between hate speech and other phenomena (e.g., bullying) and in school-related factors that are associated with the health of children and adolescents.

Sebastian Wachs is a Professor of Education with a specialization in Digitalization at the Institute of Education, 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 developing and evaluating prevention programs.

Alexander Wettstein is a Professor of Educational Psychology and head of the research group "Social Interaction" at the University of Teacher Education Bern. He is working on the ontogenesis of aggressive and prosocial behavior, peer group influence, teacher-student relationships, and psychobiological teacher stress, applying ambulatory assessment approaches.