





ARTICLE




<https://doi.org/10.1057/s41599-022-01222-4>

OPEN

Future teachers confronting extremism and hate speech

Jordi Castellví ¹, Mariona Massip Sabater², Gustavo A. González-Valencia² & Antoni Santisteban²

Hate speech has become a social problem that needs to be addressed urgently. In many cases, these discourses and ideologies arrive through the media and the internet, and they are transferred to educational contexts. Debates of this type should be addressed at school and should be channelled into a democratic debate, and into the definition of shared objectives through the development of counterspeeches and alternative narratives based on Human Rights. In this research, we investigate the capacity of future teachers ($n = 114$) to identify online hate speech and how they develop counterspeeches. The results show that the majority are able to identify hate speech. However, future teachers present more difficulties developing counterspeeches or complex alternative narratives, which can be transferred to educational practices. We conclude that teacher training needs to be redesigned if we want them to be able to face these problems in their future educational practice.

¹Universidad Internacional de la Rioja, Logroño, Spain. ²Universitat Autònoma de Barcelona, Barcelona, Spain. email: Jordi.castellvimata@unir.net

Introduction

In recent years, extremism and hate speech has become a controversial topic and a problem that needs to be urgently addressed in democratic societies (Alsagheer et al., 2022; European Council, 2014, 2017; Tryggvason, 2018; Wodak, 2015, 2019). According to a study by the Pew Research Center, in recent years there has been a rise in online misogyny, and also an increasing hostility towards Muslims and Jews in Europe, not only from citizens but also from institutions (Kishi, 2017). The globalisation of technology and access to the internet has increased the possibility and the capacity for interaction, and its consequent exposure to online hate narratives and groups (Keipi et al., 2018).

Beyond political, legal and technological strategies, counter hate speech is considered to be an encouraging solution to fight hate speech, as it promotes human rights and democratic values and debates without limiting the principles of freedom of speech (Alsagheer et al., 2022; Mathew et al., 2018). As Alsagheer et al. (2022) argue, counter hate speech takes place in civil society rather than in state-initiated legal methods, and that makes it a promising tool to educate people to throw peer interaction. Counter hate speech, then, interpellates educational responsibility to fight hate speech: if internet users are expected to intervene throw reasoned narratives as committed citizens, they need specific training to do so.

Teachers' role seems to be especially relevant for this educational and social challenge. Before they can train young students to detect and counter hate speech, we wonder about their capacity to do so. In this research, we analyse the capacity of Spanish future primary teachers ($n = 114$) to identify online hate speech and how they develop counter speeches. The results may give us some clues not only about how young people deal with hate speech and counter hate speech, but also about how can we improve training teachers programmes to emphasise Social Justice commitment from all educational spheres.

Social concern about raising hate speech

The increase of far-right policies from democratic institutions, and from the elite (Van Dijk et al., 2002), is an alarming reality that was reflected in Donald Trump's term of office in the presidency of the United States, in the populist pro-Brexit policies in the United Kingdom, in Bolsonaro's government in Brazil and in that of Orbán in Hungary, among many more (Wodak, 2019). However, the fact that these people have come to power means that they are supported by a large proportion of citizens who saw in them a solution to the economic recession, the migratory and political crisis, and the crisis of values that we are experiencing in the 21st century. The current extremism is the result of a systemic crisis of liberal democracy and capitalism (Petrie et al., 2019), which has led to major transformations that question the current system. According to (Eatwell and Goodwin, 2018), some of these transformations can be summarised as (1) mistrust towards politicians and institutions; (2) fear of the destruction of culture and national identities; and (3) the growing inequalities stemming from global neoliberal economies. Other authors have suggested that socioeconomic factors are not sufficient to explain this phenomenon and that we need to consider the use of emotions such as fear and anger, so that these type of policies are successful (Kemmer et al., 2019; Kinnvall, 2018; Salmela and von Scheve, 2017; Zembylas, 2019).

Thus, the rise in national populism (Eatwell and Goodwin, 2018) or authoritarian populism (Norris and Inglehart, 2019) is characterised by certain policies and by rhetoric that is built around a series of common pillars: (1) the 'us' and 'them' rhetoric helps to create an own identity and a feeling of belonging to a group of people in a difficult situation, which through authoritarian policies will recover the splendour of the past; (2) placing the nation above

everything else, praising national sentiments in the face of an external enemy, from whom we must defend ourselves and who is viewed as a virus or disease that can topple national values and culture; and (3) using demagogic narratives to lay the foundations for populist discourses; narratives based on personal experiences often distributed through mass media and social networks.

Our societies also face another type of non-institutionalised extremist discourses, represented by small radicalised groups organised into cells, or isolated individuals with a great ability to attack societies through the fear and terror they instil in the population. These groups have very different origins, for example, white supremacists or Islamic fundamentalists. However, extremism is also spread among ordinary people developing different intensities of *micro-fascism* (Zembylas, 2020). According to Saltman (2011), and Banaji and Buckingham (2013) young people can develop extremist ideologies, especially online, when exploring alternative forms of communication and relation. Ranieri (2016, p. 2) shows concern about the impact of hate speech on the youth because "as intense users of Internet, arguably young people are exposed to discriminatory content while their cognitive and affective development is still unstable". Exposure to hate material online affects youth negatively, from a psychological perspective (Keipi et al., 2018). Studies show that social networks such as Facebook, YouTube, Twitter and others, have played a key role in spreading extremist ideologies and hate speech (Sauer and Pingaud, 2016). Recent studies show how the use of social networks such as Twitter is linked to increasing hate speech targeting migrant people (Arcila-Calderón et al., 2022). Additionally, extremist groups are usually more active online than groups of moderate ideologies (Banaji and Buckingham, 2013). The internet allows anonymity, which provides unprecedented freedom for extremism (Gagliardone et al., 2015). Finally, digital fora provide young people with a space to show their dissent and non-conformity and an opportunity to meet other people who think like them.

Hate speech addresses those people who are considered to be different from their group (exogroup), creating the image of the enemy (Spillmann and Spillmann, 1991). Hate speech originated beyond institutions, organised groups and youth, it is also built on informal dialogues, everyday experiences and particular opinions, and it is reproduced—subtly or explicitly—(Parekh, 2006) in the media, on social networks, in any formal or informal debate forum, and even in textbooks (Van Dijk et al., 2002) and schools (Zembylas, 2020), constituting a narrative that may be linked to culture and hegemonic discourses. Therefore, hate speech emerges from extremism and shares its characteristics, such as the use of biased and simplistic language and the perpetuation of dichotomic concepts about society. In this regard, authors such as Ranieri (2016) understand that hate speech is built based on alteration strategies used by people and organisations with far-right ideologies to situate 'others' outside of 'people' and turn them into a subject of discrimination and exclusion. In this vein, Djuric et al. (2015) describe hate speech as abusive discourses that attack the characteristics of specific social groups such as ethnicity, religion and gender. Hate speech is always projected onto 'the other'. Ideas or values are not discussed; but rather the physical or social characteristics of people (Sponholz, 2016). If we analyse hate speech, it has well-defined characteristics. According to Parekh (2006), it is characterised by (1) demarcating an individual or group of individuals, object of stigmatisation; (2) assigning this group a series of qualities considered undesirable; (3) generalising stereotypes for all the members of the social group; and (4) placing the hated group outside the space of normal relationships, where their presence is deemed to be hostile and unacceptable. Therefore, hate speech is much more than words (Van Dijk et al., 2002); it has the power to

attack the human dignity of the hated group, but also of the entire community (Waldron, 2012). This attack can lead, ultimately, to the subjects being stripped of the human condition, and to dehumanisation (Massip Sabater et al., 2021).

Teachers' role in the interruption of extremism and hate speech

The complexity of the problem of extremism and hate speech has triggered different opinions about the measures required to tackle them. Authors such as Parekh (2006) suggest that the most appropriate measure is to establish limits based on the law, which serve to restrict this type of ideology and discourse. This author maintains that the law should prohibit all types of hate speech, as a guarantee to preserve freedom of expression since "hate speech strikes at the root of the shared communal life and represents a gross misuse of the right to free speech" (p. 222), and thus

"when hate speech is banned in order to create and maintain these conditions, we restrict free speech not only in the interest of other values but also its own. Indeed, while restricting it at one level, we consolidate and deepen it at another" (p. 223).

However, Parekh (2006) recognises that the use of the law to stop hate speech is more effective when it is part of a strategy in which formal education is included. Other authors have a very different perspective regarding how to combat extremism and hate speech. Authors such as Mouffe (2000, 2005), Davies (2004, 2008, 2014) and Zembylas (2019) suggest that controlled spaces of expression are needed for these opinions to be able to dismantle in a context of democratic debate. This perspective has been popularised by Mouffe under the term 'agonistic pluralism'. For Mouffe (2000), conflict or passions should not be avoided, but must be turned into a force to be channelled towards political and democratic commitments. From a similar perspective, Davies (2004, 2008) supports what she calls 'interruptive democracy', which consists of fostering spaces for dialogue and dissidence based on horizontal networking. She maintains that "young people need to be able to speak openly with teachers and youth workers about the issues they feel strongly about". Silencing uncomfortable discourses in favour of "respect" can conceal errors and actions that go beyond the discourse (Davies, 2014, p. 454). Mouffe (2005) states that removing conflict and emotions from the debate can potentially lead to a more destructive antagonistic conflict. However, she also highlights that divergent stances established in identities must not be constructed in existentialist terms (such as ethnicity or gender), but based on the definition of shared objectives. Nor should these stances be argued in moral terms, but in political ones. Racist claims, for example, are not excluded because they are irrational or malevolent, but because they clash with democratic principles, which can also be reviewed (Zembylas, 2019). These perspectives share the goal of educating to place hate speech in a democratic debate, in which opposing positions debate fairly about proposals, developing resilience to offence and moving the object of the criticism from the person (enemy) to the idea (adversary) (Davies, 2014; Mouffe, 2005).

Other authors seek to establish connections between the deliberative stances (which have their origin in the thinking of Dewey and Habermas) and agonist stances, applied to discussions about controversial topics in the classroom (Tryggvason, 2018). The deliberative approaches in the political debate at school question the role of emotions in ideological discussions, as they can lead to a personal clash between identities, pushing aside the political problem in question (Englund, 2016) and moving the parties further away from consensus. Tryggvason (2018) argues that both stances agree that emotions must be maintained in the political sphere, and not in the moral sphere, linked to social issues and not to personal

topics. However, he recognises that while agonist positions situate the democratic conflict on the horizon, deliberative stances use it as a starting point for the construction of consensus.

Although extremism and hate speech are not only educational problems (Estellés and Castellví, 2020), all stances consider education as one of the pillars to combat them. Thus, there is a need for well-trained teachers, who do not try to impose political correction, but who have developed the necessary skills to navigate the emotions and debates that arise in the classroom (Davies, 2014), to redirect the existentialist positions toward legitimate positions in the democratic debate (Tryggvason, 2018). For this reason, we need teachers who are prepared to detect hate speech and extremism in society and their classroom in particular, and who have the necessary skills to build counter-narratives and to teach how to build them.

Training teachers to spot hate speech and build counterspeech

There is no consensus surrounding the concept of hate speech (Gagliardone et al., 2015). Several authors highlight that its definition has narrow and broad interpretations (Gagliardone et al., 2015). These authors indicate that the narrow interpretations of hate speech, at times called 'dangerous speech' and 'fear speech', refer to the discourse that incites hatred against a certain person or a vulnerable social group and that constitutes a crime. The broad interpretations could consider hatred any discriminatory or offensive discourse against minorities or vulnerable individuals, which may not be a crime, but it may be morally condemned. Identifying hate speech is not always easy (Gagliardone et al., 2015). However, for practical reasons, we refer to the European Council definition (2017) of hate speech and to the characteristics that Parekh (2006) attributes to it and that we have mentioned earlier. The European Council (2017, p. 31) states that hate speech includes:

all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.

The final aim of the training of teachers in the treatment of hate speech must be to identify it in the classroom and address its complexity, promoting the debate on responsibility, protagonists or arguments exposed. It is about dismantling the hate speech based on the construction of counterspeeches, based on adding "more speech to the conversation and try to change the mindset of the hate speaker" (Mathew et al., 2018). In this regard, it is very important to identify hate speech on the internet, digital media and social media. Because of its characteristics, the internet has become a favourable place for ideological debate, emotions, satire, rumours and morally condemnable messages. The internet has become a place where hate speech is spread and shared (Ranieri, 2016). Some studies conclude that the main problems with hate speech occur in the most connected countries (Gagliardone et al., 2015). The increasingly common proliferation of hate speech online entails new challenges and difficulties for a democratic society, which is obliged to find suitable responses to this phenomenon. Some of these difficulties are the itinerancy of contents on the internet, the anonymous identity of many of the users and the fact that the internet is a transnational network (Gagliardone et al., 2015). Beyond the legal responses to such a phenomenon, the educational responses to hate discourse and extremism must be one of the tools against its propagation. The approaches of critical media and information literacy have been useful in recent decades to achieve it. These approaches are aimed at training citizens to develop capacities that enable them to perform a critical analysis of mass media,

advertising and political messages. However, in the present digital context, approaches toward the development of critical thinking are beginning to be demanded, such as critical digital literacy (Castellví et al., 2020) and digital citizenship (Gagliardone et al., 2015). It is considered that the media are not the only producers of information, but rather, all people are. In this new paradigm, anyone can produce and consume information.

In the school setting, one of the responses to hate speech can be to conceal or block it. However, the 'No Hate Speech' movement promoted by the European Council (2017) transmits that it is more effective to respond to an oppressive narrative with another narrative, which helps to construct an alternative framework to interpret reality. As Ortega-Sánchez et al. (2021, p. 14) propose, we need "to work with freedom of expression through democratic principals and values within the classroom". Wachs et al. (2022) defend that demeritocratic education is needed to prevent hate speech, along with the promotion of an inclusive atmosphere in the classroom. According to Mathew et al. (2018, p. 9), "counterspeech is considered to be a promising solution as it can help controlling the hate speech problem and, at the same time, it supports free speech". Counterspeeches or alternative narratives combat hate speech by discrediting and deconstructing its content and by basing its arguments on human rights (European Council, 2017). The desired goals are not achieved by just providing information opposing hate speech. The European Council (2017) suggests developing counter-narratives that connect with students and with their lives, through emotions, personal contact and humour. They can use several resources to reach their students. Tuck and Silverman (2016, p. 4) propose deconstructing and delegitimising hate speech through the following strategies:

- Highlighting how extremist activities negatively impact the people they claim to represent.
- Demonstrating the hypocrisy of extremist groups and how their actions are often inconsistent with their own stated beliefs.
- Emphasising factual inaccuracies used in extremist propaganda and setting the record straight.
- Mocking or satirising extremist propaganda to undermine its credibility.

These strategies do not seek to replace hegemonic accounts with others, but to construct strong alternatives, which facilitate dialogue between the different social groups, in order to tackle extremism and hate speech and develop critical thinking (European Council, 2017). Counterspeech has been studied on different social media sites, finding out that some arguments between strangers such as Twitter users do make changes in discourses and also in attitudes (Mathew et al., 2019). Recent studies show how support for solidarity citizenship norms might moderate the effect of exposure to hate comments on the willingness to engage with committed online intervention and counterspeech (Kunst et al., 2021), something relevant when thinking about Demeritocratic Citizenship Education. Furthermore, it seems that hate speeches are frequently countered with massive response, and that strategies to do so often vary depending on the targeted social group (Mathew et al., 2018; Mathew et al., 2019).

We agree with Izquierdo (2019) and Massip Sabater et al. (2021) that future teachers must be able to develop counterspeeches for hatred and be able to transfer this process to the classroom. The present study aims to determine the capacities of future teachers to identify online hate speech, and to develop counterspeech.

Method

This study is part of a research project funded by the Ministry of Science and Innovation in Spain¹ about teaching and learning

contemporary issues. Three main topics structure this project: critical literacy, global citizenship education, and hate speeches. The present study is about the hate speeches topic.

Sample and research instrument

The study used a non-probabilistic convenience sampling (Sabariego, 2019). The sample studied were future teachers ($n = 114$) from the 2nd and 3rd years of the Primary Education Degree at the University of Malaga and the Autonomous University of Barcelona, in Spain. Both institutions were active members of the research project. According to the research ethical criteria of the Autonomous University of Barcelona, trainee teachers were informed about the implication of their participation.

All participants responded a writing dossier about a real Twitter conversation, generated after the terrorist attacks in Barcelona in August 2017. The future teachers had 45 min to complete individual activities structured around this initial conversation, and especially around a concrete publication which constitutes an example of online hate speech:

May no one forget that these Islamists are not left-wing or right-wing, they're after all of us, our way of life, our freedom and our democracy. Here we're not judging Muslims, but it's their religion that's killing us. There's no room for the good ones. It's our society or theirs². (GREDICS, 2018)

The participants were asked

- (1) to identify and explain their feeling while reading the conversation
- (2) to analyse the use of *us* and *them* terms
- (3) to explain whether they identified the publication as a hate speech or not and
- (4) to write their hypothetical contribution to the debate

To do so they could use any digital technology available to them to verify the truthfulness and reliability of the information provided.

Data treatment and analysis

The names of the participants have been modified to ensure anonymity, in accordance with the code of research ethics of the Autonomous University of Barcelona (2020). Their answers have been translated from Catalan and Spanish original writings. As no significant differences were found between the results of the two universities, the data are analysed together.

Data is analysed using mixed methods (Tashakkori and Teddlie, 2003) in order to obtain a global quantitative picture and detailed qualitative understanding. Qualitative data is analysed through thematic analysis of the content (Flick, 2004).

Based on the data obtained, we have identified correlations between different pre-established categories (Table 1), we have compared the results with the theoretical framework considered (Table 2) and, finally, we have studied the position of the future teachers when dealing with hate speech.

In the first analysed activity, the published comment constitutes a paradigmatic example of hate speech against Muslim society, according to Parekh criteria (2006). We asked the future teachers to critically explain whether or not it could be classified as hate speech: "Write a critical analysis of his comment. Do you consider it could be hate speech?" (GREDICS, 2018)

The analysis of the results for this first activity has taken into account the following categories:

Table 1 Categories of analysis for future teachers' responses in activity 1.

It is hate speech	The future teacher identifies the account as hate speech.
It is not hate speech or they are not sure	The future teacher expresses doubts about the nature of the account or believes that it cannot be classified as hate speech.
They reproduce the hate speech	The future teacher adheres to the criticism of the author of the publication and reproduces the hate speech.

Table 2 Categories of analysis for future teachers' responses in activity 2 based on the works of Mouffe (2000, 2005) and Parekh (2006).

Triggers passions	The future teachers consider that it is hate speech because it others the hated group through the generalisation of stereotypes.
Othering	The future teachers believe it is hate speech because it others the hated group, demarcating them and stigmatising them.
Demarcates and stigmatises	
Generalises stereotypes	The future teachers consider that it is hate speech because it others the hated group through the generalisation of stereotypes.
Displaces the hated group	The future teachers consider that it is hate speech because it others the hated group by displacing them from society.

Based on the answers identifying the account as hate speech, we have developed a second analysis category which enables us to identify the attributions that future teachers give to hate speech. Using the ideas of Mouffe (2000, 2005) and Parekh (2006) we classified the responses into the following categories:

In the second activity, we asked the future teachers to write a contribution to the online conversation: "Let's suppose you have decided to participate in this online debate. What comment would you write?" (GREDICS, 2018). Based on a thematic analysis of the content (Flick, 2004) we classified their responses into four categories, using the strategies to develop counter-narratives proposed by Tuck and Silverman (2016), adding a fifth category for alternative narratives (European Council, 2017):

Results

Identifying hate speech and extremist ideologies. The results for the first activity (Fig. 1) show that 89.47% (102 future teachers) identified the proposed narrative as hate speech. On the contrary, 6.96% (8 future teachers) expressed doubts or denied that it is hate speech, while 2.61% (3 future teachers) not only denied that it was hate speech, but reproduced the author's arguments. Finally, 0.87% (1 future teacher) did not respond.

Considering the results that identify the narrative as hate speech, 24.35% (28) of the answers were classified in the category of 'triggers passions' (see Fig. 1). The future teachers who replied in this category consider that the narrative presented is hate speech, because it demonstrates the author's emotions and triggers the passions of other people. This is the case of Diana's response:

It is clear that this comment is hate speech. This person is expressing their feelings of anger and disdain towards this group of people, inciting hatred among those who read the comment.

All the responses in this category identify hate speech as a discourse that triggers passion. However, these passions are not channelled into defending political ideas but are used as existentialist arguments against Muslims. Diana expresses the fact that these passions can be used to convince other people about the truthfulness and legitimacy of these arguments, in such a way that hate speech against Muslims continues. Sara's response is also included in this category. She made the following statement:

Yes [it is hate speech]. And more so if it is read by someone from that religion, as they could feel offended.

Here, Sara highlights the emotions that could be triggered among those who are the object of the hate speech. The reaction that this narrative can trigger is an offence. Furthermore, 51.3% (60) of the responses identify the narrative presented as hate discourse because it others Muslim people. We analysed this type of narrative in greater detail following the characteristics of hate speech defined by Parekh (2006) (see Table 2). As a result of this analysis, we found that 12.17% (14 future teachers) argued that it demarcates a group of individuals and stigmatises them, assigning them a series of qualities that are considered to be undesirable, in this case those of terrorists, fundamentalists and/or radicals (see Fig. 1).

This is the case of Marcos' response:

Yes, it is hate speech. This person is talking out of ignorance and out of the impotence they feel about what has happened. If they were to inform themselves, they would see that the Muslim religion is based on values such as respect and solidarity, not on killing.

Marcos considers that the hate speech presented originates from the impotence and ignorance of the author, who attributes qualities to the Muslim religion and culture, which are not correct. 14.78% (17 future teachers) argue that the author of the narrative others Muslims by generalising stereotypes for all members of the social group (see Fig. 1). When asked if it is hate speech, Ana says the following:

Yes, because they are discriminating against an entire group for something a small percentage of this group did.

Ana considers that among Muslims there are terrorists, but these comprise a very small percentage of the total. Here, she does not focus on the attribution of undesirable characteristics to a social group, but on the generalisation of these characteristics for the whole group. The generalisation of stereotypes is one of the main attributions made by future teachers to hate speech such as that presented in this activity. Although we only consider that 17 of the narratives clearly include this element, it is true that the majority of the responses include some reference that suggests this generalisation. Likewise, 24.35% (28 future teachers) maintain that the narrative displaces the hated group out of normal

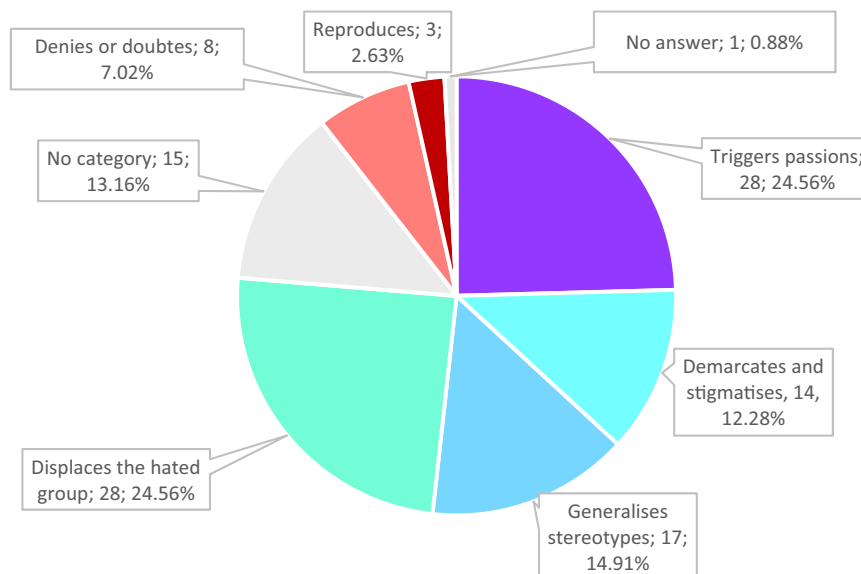


Fig. 1 Future teachers identifying hate speech. Results of activity 1 classified by category.

relationships, where their presence is considered to be hostile and unacceptable (see Fig. 1). The responses with these characteristics refer to racism as a key element in the narrative presented. Thus, we considered that racism includes among its main characteristics, the displacement of the hated group, towards social exclusion and, therefore, we included it in this section. Pedro's response considers the mentioned characteristics:

I think it is a comment that can be classified as racist, because it labels Muslims as terrorists, which means that when this person meets a Muslim on the street, they probably won't be able to walk in the same place, or they will insult them.

Pedro's answer highlights this displacement from normal relations, for example, in public spaces like the street. Finally, 13.91% (15 future teachers) identified the narrative as hate speech (see Fig. 1). However, they do so without providing arguments, or the arguments they use cannot be classified in any of the pre-established categories. On the contrary, 8 future teachers (6.96%) denied or doubted whether it is hate speech, as Marta states:

I don't think it's hate. It is their belief and they are warning the readers and trying to convince them of something; that Islam is a religion that promotes murder.

This type of response shows that some future teachers were not able to identify the narrative presented as hate speech, even when consciously participating in hate speech research. Their responses reproduce the argument and they question whether it is hate speech, and in some cases justify the author. Finally, a minority of future teachers (3; 2.61%) not only deny that it is hate speech, but they support the author's point of view and reproduce the same discourse (see Fig. 1). Juan said the following:

[The narrative] is written by someone who is angry about the situation, but I think it is the most correct comment in the debate.

Carmen's response was also classified in this category. Carmen, like the author of the narrative presented, attributes the blame for the terrorist attacks to the Muslim religion:

Yes, it's the same thing again. To what extent is this a respectable religion?

We must not forget that the above comments, although they are the minority, were made by future teachers. However, 89.75% were able to identify the narrative as hate speech, but this does not mean that they are capable of developing counterspeeches, or that they do this in a coherent and well-argued way. We will analyse the results of the activity of the hate counterspeeches below.

Counter speeches of hate. The results in the second analysed activity (see Fig. 2) show us that 86 future teachers (80.34%) were able to develop counterspeeches. Through a thematic analysis of the content, we classified the responses into different categories (see Table 3). Below, we analyse some of the responses obtained starting with those we classified in the majority categories and finishing with the minority responses. The majority of the responses (50; 43.86%) were classified in the category 'inexact arguments' (see Fig. 2). These counterspeeches maintain that the arguments used by the author of the online publication are based on biased or false facts with a view to promoting hatred towards the Muslim community. This is the case of Cecilia's response. As a future teacher, she would try:

To make students understand that we cannot generalise. The terrorists are jihadists, not Muslims.

Like the majority of counterspeeches in this category, Cecilia dismantles the hate speech, indicating that the arguments used are inexact because it extends an isolated behaviour to the entire Muslim community. Cecilia specifies that the concept 'Muslim' and 'jihadists' are used interchangeably when they are not the same, meaning the hatred is supported by biased arguments.

Furthermore, 20 future teachers (17.54%) developed alternative narratives based on human rights (see Fig. 2). In this case, they did not develop counterspeeches based on hate speech, but they promoted a series of values such as tolerance, equality and peace in order to dismantle extremist ideologies, like that represented by the author of the online publication. A clear example is that of Patricia, who states in her comment that 'we need to promote tolerance'. Paula's response was also classified in this category. Paula argues that

if we want a world in which peace and love prevail, we must put these thoughts and emotions into practice, and not hate and be resentful.

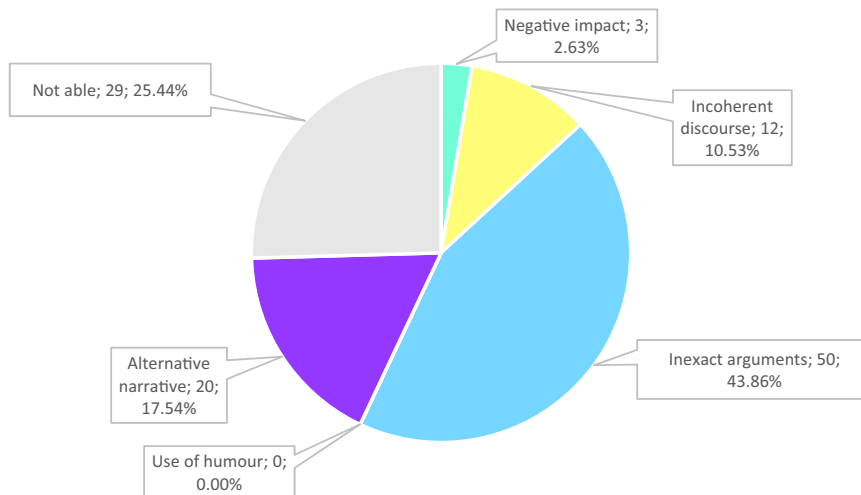


Fig. 2 Future teachers developing counterspeeches. Results of activity 2 classified by category.

Table 3 Categories of analysis for future teachers' responses in activity 2 based on the works of Tuck and Silverman (2016) and European Council (2017).

Negative impact	Highlighting how extremist activities negatively impact the people they claim to represent.
Incoherent discourse	Demonstrating the hypocrisy of extremist groups and how their actions are often inconsistent with their own stated beliefs.
Inexact arguments	Emphasising factual inaccuracies used in extremist propaganda and setting the record straight.
Use of humour	Mocking or satirizing extremist propaganda to undermine its credibility.
Alternative narrative	Creating an alternative narrative based on the values represented by human rights.

These responses are characterised by being assertive and by promoting an alternative narrative to hate speech. They do not tend to explicitly mention the hate speech or try to respond to its arguments. Instead, they propose an alternative view of society, based on democratic values and human rights. Thirdly, 12 future teachers (10.53%) developed counterspeeches that revealed the incoherence of the hate speech (see Fig. 2). The responses in this category agree that other cultures and religions also have fanatics among their members, and that the entire community is not blamed. Specifically, within the Christian community, which the author of the hate speech claims to represent, there are also people who have committed crimes in the name of religion. Carlos argues that...

I find it surprising that those of you who differentiate between us and them, and accuse the entire Muslim population of being terrorists and murderers, don't realise that you are promoting hatred, inequality, injustice...Are there no Catholic criminals? Has the church not done harm? Have you forgotten about what happened in 1492? So, are all Spaniards murderers because a group ravaged the Inca and Maya cultures? It is important to try and understand that despite it being a very serious event, people are not all the same and, therefore they don't deserve to be treated like that.

We classified a minority of the responses in the category 'negative impact'. In general, the future teachers do not use the argument that these discourses negatively impact the Christian community or western society. Only 3 future teachers (2.63%) use this argument (see Fig. 2). For example, Lorena says the following:

We believe that countries such as the United States or Spain are examples to follow and we look down upon others

where radicals of this type are generated in mass, but the first world countries are providing these radicals with weapons, the same weapons that are then used to kill innocent people in these countries and in those of the first world. Maybe the problem is elsewhere and the solution is much more complex than simply exterminating these people.

Lorena maintains that western countries indirectly supply terrorist groups with weapons. She states that it is a complex topic. The global economic inequalities linked to neo-colonialism are one of the factors leading to the emergence of radicalised people, who attack the western world. 29 of the future teachers (25.44%) who participated in the research did not develop a counterspeech (see Fig. 2). Some did not respond to the second activity while others admitted that they were not able to build a counter-narrative, that they did not participate in discussions of this type, or that it was best to simply interrupt this type of discourse.

Finally, no future teacher used humour or satire as a strategy to develop counterspeeches. This may be due to the seriousness of the social problem considered. We must consider that this hate speech was generated after the terrorist attacks in Barcelona.

Discussion and conclusions

The results obtained in the first activity show that the majority of future teachers are able to identify hate speech like the one we presented to them. The majority attributes the characteristics described by Mouffe (2000, 2005) and Parekh (2006) to hate speech. It is noteworthy that 24.56% relate hate speech to the fact that it displaces the targeted group, and other 24.56% to the fact that triggers passions and emotions. This is an important point as we cannot relate emotions of an ideological nature with something that is intrinsically negative. While in this case it is clearly an example of hate speech, not all discourses that transmit

emotions must be dismantled, provided they are not constructed in existentialist terms (Zembylas, 2019). A considerable number of the responses refer to racism in the discourse, arguing that discourses of this type displace the hated group from normal social relations. Although the future teachers who identified the hate narratives are a majority, it is concerning that 9.65% did not manage to detect them. Recent studies show how internet users' intervention against hate speech is more likely when hate comments contain strongly abusive and hateful language such as insults (Kunst et al., 2021). May it be possible for this nearly 10% of future teachers to not have detected the hate message underlining subtle and disparaging language. Nevertheless, 2 of them reproduced the hate speech. Although it is a very small number, it is shocking that future teachers reproduce hate speech against the Muslim population, in this case. We believe that, to educate critical citizens, we need to have critical teachers. A teacher who elaborates hate speech cannot be a critical teacher, nor a teacher.

In the second activity, the results show that 74.56% were able to develop counter speech, although in the majority of cases (43.86%) they highlight the inexactness of the arguments in the hate speech. Only a minority use other strategies. None of the narratives combine more than one strategy or consider the counterspeech for an explicit translation in their educational practice. It is also important to note that 17.54% developed an alternative narrative revealing proactive strategies, and alternatives to the simple critique of the hate speech. Strategies such as the use of humour or satire were not used by the future teachers, which shows that not all the strategies proposed by Tuck and Silverman (2016) are suitable in any context, or that the simulated situation in an academic context does not favour them. On the contrary, 25.44% did not develop a counterspeech or an alternative narrative, either because they did not think they were able to or because they did not think it was relevant. In recent research, Kunst et al. (2021) highlight how the willingness to engage in intervention against hate comments depends on the attacked social group. Thus it would be interesting to carry other similar studies out, modifying the targeted social group.

These results highlight the lack of training in working with hate speech and the construction of counterspeeches in the universities studied. It is likely that these results will not be very different in other Spanish universities. This statement is consistent with the results obtained in other similar research. Previous studies (Arroyo et al., 2018; García-Ruiz and Zorrilla Luque, 2019; Ortega-Sánchez et al., 2021), show that the number of future teachers who develop critical narratives in relation to hate speech is very low. According to these studies, the majority of future teachers develop simplistic narratives and do not identify the complexity of the social problem being addressed. Moreover, they show difficulties when developing a counterspeech. They also point to the difficulties that arise from interpreting the curriculum, to develop a didactic proposal that is committed to problems in society, as it generates ideological conflicts, linked to social beliefs and representations. Nevertheless, the present study shows more optimistic results in some aspects. On Ortega-Sánchez et al. (2021), there's a minority of trainee teachers' narratives capable of rationalising the incoherences of hate speech; in the present one, nearly 70% of the future teachers not only identify hate speeches but also dismantle their arguments, incoherence and narrative strategies.

The same study with secondary school students (Massip et al., 2021) shows for the first activity that 60% of the students are able to identify hate speech. However, only 53.5% of secondary school students are able to develop counter-narratives. Secondary students showed a noteworthy predisposition to generate counterspeeches, although most of them based their arguments on the impossibility of generalising, and only a few (6%) articulated

concrete arguments (Massip et al., 2021). For both activities, the results show that future teachers are better prepared to identify hate speech and develop counterspeeches than secondary students are. These data were as expected since future teachers have better training and are more mature than secondary school students. However, we still believe their capacities are to improve.

There is great concern among different educational institutions, in academia and in schools about the proliferation of hate speech (European Council, 2014, 2017; Tryggvason, 2018; Wodak, 2015, 2019), but the response must not be exclusively educational. Schools have a large responsibility in the dissemination of human rights, but society must be very conscious of the growing problem that exists. Within educational solutions, one of the main tools is teacher training. We need critical teachers who are capable of exposing hate speech and extremist ideologies, and who are able to address these narratives, which are so present on the internet and in their classes. The results of this study show that, while future teachers are capable of identifying hate speech, they present real difficulties when developing counterspeeches. Their responses are very simplistic and reveal a lack of strategies to address this type of discourse and ideology. Better teacher training is needed, which explicitly tackles the problem of hate speech and extremism in society. We need to train teachers who dare to open spaces for political debate in their classes, where emotions and ideological positions are expressed. We need teachers who are able to promote inclusive education for social justice, coexistence and peace.

Data availability

The datasets generated during and/or analysed during the current study are not publicly available due to preserve the privacy of participants but are available from the corresponding author on reasonable request.

Received: 9 June 2021; Accepted: 7 June 2022;

Published online: 15 June 2022

Notes

- 1 Research Project funded by the Spanish Ministry of Science and Innovation (R&D PID2019-107383RB-I00, PI: Dr. Antoni Santisteban).
- 2 Translated from the original publication, in Spanish.

References

- Alsagheer D, Hadi M, Weidong S (2022) Counter hate speech in social media: a survey. arXiv. <https://doi.org/10.48550/arxiv.2203.03584>
- Arcila-Calderón C, Sánchez-Holgado P, Quintana-Moreno C, Amores J, Blanco-Herrero D (2022) Hate speech and social acceptance of migrants in Europe: analysis of tweets with geolocation. *Comunicar* 71:21–35. <https://doi.org/10.3916/C71-2022-02>
- Arroyo A, Ballbé M, Canals R, Llusà J, López M, Oller M, García CR, Santisteban A (2018) El discurso del odio: una investigación en la formación inicial. In: López E, García CR, Sánchez M (eds) *Buscando formas de enseñar: investigar para innovar en Didáctica de las Ciencias Sociales*. Universidad de Valladolid/AUPDCS, Valladolid, pp. 413–424
- Banaji S, Buckingham D (2013) *The civic web: young people, the Internet and civic participation*. MIT Press, Cambridge
- Castellví J, Díez-Bedmar MC, Santisteban A (2020) Pre-service teachers' critical digital literacy skills and attitudes to address social problems. *Soc Sci* 9(8):134. <https://doi.org/10.3390/socsci9080134>
- Council of Europe (2014) *Bookmarks. A manual for combating hate speech online through human rights education*. No Hate Speech Movement. European Youth Centre, Strasbourg
- Council of Europe (2017) *We can! Taking Action against Hate speech through Counter Alternative Narratives*. No Hate Speech Movement. European Youth Centre, Strasbourg
- Davies L (2004) *Education and conflict: complexity and chaos*. Routledge, London

- Davies L (2008) Educating against extremism. Stoke on Trent, Trentham
- Davies L (2014) Interrupting extremism by creating educative turbulence. *Curric Inq* 44:450–468
- Djuric N, Zhou J, Morris R, Grbovic M, Radosavljevic V, Bhamidipati N (2015) Hate speech detection with comment embeddings. In: Gangemi A, Leonardi S, Panconesi A (eds) Proceedings of the WWW '15: 24. Association for Computing Machinery, International conference World Wide Web, Florence, Italy, May 2015
- Eatwell R, Goodwin M (2018) National populism. The revolt against liberal democracy. Pelican, London
- Englund T (2016) On moral education through deliberative communication. *J Curric Stud* 48(1):58–76
- Estellés M, Castellví J (2020) The educational implications of populism, emotions and digital hate speech: a dialogue with scholars from Canada, Chile, Spain, the UK, and the US. *Sustainability* 12(15):6034. <https://doi.org/10.3390/su12156034>
- Flick U (2004) *Introducción a la investigación cualitativa*. Morata, Madrid
- Gagliardone I, Gal D, Alves T, Martinez G (2015) Countering online hate speech. UNESCO, Paris
- García-Ruiz CR, Zorrilla Luque JL (2019) Educar para la ciudadanía frente a discursos de odio desde la prensa digital. Una experiencia en formación inicial del profesorado de educación secundaria. In: Hortas MJ, Dias A, De Alba N (eds) Enseñar y aprender didáctica de las ciencias sociales: la formación del profesorado desde una perspectiva sociocrítica. AUPDCS – ESE/PL, Lisbon, pp. 659–666
- GREDDICS (2018) (Contra)relatos del odio. Dossier de actividades. <https://bit.ly/3NP8juP>
- Izquierdo A (2019) Literacidad crítica y discurso del odio: una investigación en Educación Secundaria. *REIDICS* 5:42–55. <https://doi.org/10.17398/2531-0968.05.42>
- Keipi T, Räsänen P, Oksanen A, Hawdon J, Näsi M (2018) Exposure to online hate material and subjective well-being: a comparative study of American and Finnish youth. *Online Inf Rev* 42(1):2–15. <https://doi.org/10.1108/OIR-05-2016-0133>
- Kemmer L, Peters CH, Weber V, Anderson B, Mühlhoff R (2019) On right-wing movements, spheres, and resonances: an interview with Ben Anderson and Rainer Mühlhoff Distinktion. *J Soc Theory* 20:25–41
- Kinnvall C (2018) Ontological insecurities and postcolonial imaginaries: the emotional appeal of populism. *Humanit Soc* 42:523–543
- Kishi K (2017) Muslims, Jews faced social hostilities in seven-in-ten European countries in 2015. Pew Research Center. <https://pewrsr.ch/3aVA18D>
- Kunst M, Porten-Cheé P, Emmer M, Eilders C (2021) Do “Good Citizens” fight hate speech online? Effects of solidarity citizenship norms on user responses to hate comments. *J Inf Technol Politics* 18(3):258–273
- Massip Sabater M, García-Ruiz CR, González-Monfort N (2021) Contrariar el odio: Los relatos del odio en los medios digitales y la construcción de discursos alternativos en alumnado de Educación Secundaria. *Bellaterra J Teach Learn Language Lit* 14(2):e909. <https://doi.org/10.5565/rev/jtl3.909>
- Mathew B, Kumar, N, Goyal P, Mukherjee A (2018) Analyzing the hate and counterspeech accounts on Twitter. Indian Institute of Technology.
- Mathew B, Saha P, Tharad H, Rajgaria S, Singharia P, Maity S, Goyal P, Makherjee A (2019) Thou shalt not hate. Countering Online Hate Speech. Indian Institute of Engineering Science and Technology. <https://doi.org/10.48550/arXiv.1808.04409>
- Mouffe C (2000) *The democratic paradox*. Verso, London
- Mouffe C (2005) *On the political*. Routledge, New York, NY
- Norris P, Inglehart R (2019) *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge University Press, Cambridge
- Ortega-Sánchez D, Pagès Blanch J, Ibáñez Quintana J, Sanz de la Cal E, de la Fuente -Anunciabay R (2021) Hate speech, emotions, and gender identities: a study of social narratives on Twitter with Trainee Teachers. *Int J Environ Res Public Health* 18:4055. <https://doi.org/10.3390/ijerph18084055>
- Parekh B (2006) Hate speech. is there a case for banning? *Investig Políticas Públ* 12(4):213–223. <https://doi.org/10.1111/j.1070-3535.2005.00405.x>
- Petrie M, McGregor C, Crowther J (2019) Populism, democracy and a pedagogy of renewal. *Int J Lifel Educ* 38:488–502. <https://doi.org/10.1080/02601370.2019.1617798>
- Ranieri M (ed) (2016) *Populims, media and education. Challenging discrimination in contemporary digital societies*. Routledge, New York, NY
- Sabariego M (2019). *El proceso de investigación II*. R. Bisquerra, R. (Coord.) Metodología de la Investigación Educativa. La Muralla, pp. 123–160.
- Salmela M, von Scheve C (2017) Emotional roots of right-wing political populism. *Soc Sci Inf* 56:567–595
- Saltman E (2011) Radical right culture and the youth: the development of contemporary Hungarian political culture. *Slovo* 23:114–131
- Sauer B, Pingaud E (2016) Framing differences. Theorizing new populist communicative strategies on the Internet. In: Ranieri M (ed) *Populism, media and education. Challenging discrimination in contemporary digital societies*. Routledge, New York, NY, pp. 26–43
- Spillmann K, Spillmann K (1991) La imagen del enemigo y la escalada de los conflictos. *Rev Int Cienc Soc* 127:59–80
- Sponholz L (2016) Islamophobic hate speech: What is the point of counter-speech? The case of Oriana Fallaci and *The Rage and the Pride*. *J Muslim Minor Aff* 36(4):502–522. <https://doi.org/10.1080/13602004.2016.1259054>
- Tashakkori A, Teddlie C (eds.) (2003) *Handbook of mixed methods in social and behavioral research*. Sage, Thousand Oaks, CA
- Tryggvason Á (2018) Democratic education and agonism. exploring the critique from deliberative theory. *Democr Educ* 26(1):1–9
- Tuck H, Silverman T (2016) *The counter-narrative handbook*. Institute for strategic dialogue, London
- Van Dijk T (2002) Discourse and racism. In: Goldberg DT, Solomos J (eds) *A companion to racial and ethnic studies*. Blackwell Publishing, Malden & Oxford, pp. 145–159
- Wachs S, Wettstein A, Bilz L, Gámez-Guadix M (2022) Adolescents’ motivations to perpetrate hate speech and links with social norms. *Comunicar* 71:9–20. <https://doi.org/10.3916/C71-2022-01>
- Waldron J (2012) *The harm in hate speech*. Harvard University Press, Cambridge
- Wodak R (2015) *The politics of fear: what right-wing populist discourses mean*. Sage, Los Angeles
- Wodak R (2019) Entering the ‘post-shame era’: the rise of illiberal democracy, populism and neo-authoritarianism in Europe. *Global Discourse* 9(1):195–213
- Zembylas M (2019) The affective dimension of far right rhetoric in the classroom: the promise of agonistic emotions and affects in countering extremism. *Discourse* 42(2):267–281. <https://doi.org/10.1080/01596306.2019.1613959>
- Zembylas M (2020) Affect, biopower, and ‘the fascist inside you’: the (Un-)making of microfascism in schools and classrooms. *J Curric Stud* 53(1):1–15. <https://doi.org/10.1080/00220272.2020.1810780>

Acknowledgements

Research Project funded by the Spanish Ministry of Science and Innovation (R&D PID2019-107383RB-I00, PI: Antoni Santisteban Fernández).

Author contributions

All the authors contributed equally to this work.

Competing interests

The authors declare no competing interests.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards, and also with the standards of research ethical criteria of the Autonomous University of Barcelona.

Informed consent

Informed participation is guaranteed, as well as publication consent. All participants were adults and their personal data has been anonymized.

Additional information

Correspondence and requests for materials should be addressed to Jordi Castellví.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022