ACCEPTED MANUSCRIPT

This is an author-produced PDF of a paper published in Policy Futures in Education*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination. Citation for the published paper:*

Jääskeläinen T (2019) Countering hate speech through arts and arts education – Addressing intersections and policy implications. In: Anttila E and Martin R (eds) *Special Issue "Arts and Culture in Education: Questioning and Reimagining Current Policies and Practices". Policy Futures in Education*.

URL: https://doi.org/10.1177/1478210319848953

Publisher: SAGE

Countering hate speech through arts and arts education
– Addressing intersections and policy implications

Tuula Jääskeläinen
University of the Arts Helsinki, Finland

Abstract

Hate speech has become a growing topic of discussion and debate on a global scale, especially as advances in the internet transform communication on many levels. Among scholars, hate speech has been defined as any form of expression – for example by means of speech, images, videos, or online activity – that has the capacity to increase hatred against a person or people because of a characteristic they share, or a group to which they belong. In order to maintain the integrity of a functioning democracy, it is important to identify the best balance between allowing freedom of expression and protecting other human rights by countering hate speech. In addition to strengthening the legal framework to address the cases when hate speech can be considered criminal, and developing automated monitoring of online systems to prevent the spreading of cyberhate, counter narratives can be utilised by the targets of hate speech and their communities to create campaigns against hate speech. The employment of artists' expression and arts education have great potential for creating different counter narratives to challenge one-sided narratives and hate speakers' simplified generalisations. Because hate speech is not an easy issue to address in schools, clear research evidence, concrete guidelines, and practical examples can help teachers to contribute, along with their students, in combating it. A great body of evidence supporting the beneficial social impacts of the arts and culture fields is already available, but much more research, backed by sufficient resources, is needed to evaluate the impact and effectiveness of intervention strategies in countering hate speech through arts education.

Keywords

arts education, counter narrative, dignity, freedom of expression, hate speech

Introduction

Hate speech has become a growing topic of discussion and debate on a global scale, especially as advances in the internet transform communication on many levels, including user-generated and anonymous online platforms where hate speech can be easily shared (Chetty and Alathur, 2018; Gomes, 2016, 2017; Saleem et al., 2017). Very often the aim of hate speech is to harm the reputation of vulnerable people from minority groups characterized by disability, ethnicity, gender, gender identity, race, religion, sexual

Tuula Jääskeläinen (M.Ed.) has twenty years working experience in the higher education administration in Finland. She has a general teacher qualification in arts university pedagogy, and is specialized in developing human rights education through arts. Tuula is a doctoral candidate in music education in the Sibelius Academy, University of the Arts Helsinki, Finland, and in the ArtsEqual Research Initiative associated with the Center for Educational Research and Academic Development in the Arts (CERADA). Her research interests are in the field of higher education, especially experiences of learning in the arts education. Correspondence to: Tuula Jääskeläinen, University of the Arts Helsinki, P.O. Box 1, 00097 Uniarts, Finland. Email: tuula.jaaskelainen@uniarts.fi

orientation, or other equivalent characteristics, by making them seem worthless in the social sphere (Gomes, 2016; Waldron, 2012). The most pressing issues arising from the debate on hate speech are: Is hate speech harmful? Can words hurt as much as physical attacks? If so, what can be done about it? (Heinze, 2016).

In their reviews of online hate speech, both Chetty and Alathur (2018) and Blaya (2018) found that it is necessary to produce research, policies, and methods to identify, prevent, and control increased hate speech in online activities. To counter hate speech, they suggest intervention programs such as strengthening the legal framework, developing automated monitoring of online systems, utilizing education for public awareness, and empowering young people to produce counter speech.

Strengthening the legal framework for combating hate speech requires the social and political context of a specific country to be considered, as there are different legislations already in place in different countries (Bonotti, 2017). For example, in Germany and Canada the law considers hate speech to be a crime, whereas in the United States of America hate speech is permitted if the hate speaker does not threaten or use violence or incite others to it (McConnell, 2012). As an example of a solution for automated identification of the high volume of online hateful speech, Saleem et al. (2017) propose an approach that uses content produced by self-identifying hateful communities instead of keyword-based methods, which have been found insufficient for reliable detection. Molnar (2012) suggests that art, education, and other cultural activities, which have minimal risk of unintended side effects, can help prevent hate speech in the cases where hate speech does not present an imminent threat of violence. A report by Silverman et al. (2016) indicates that content creators collaborating with social media companies and private sector partners can create cost-effective counter narrative campaigns which increase awareness of, engagement in, and impact on combating hate speech.

Countering hate speech can also be connected with supporting human rights, for example through Article 1 of Universal Declaration of Human Rights: “All human beings are born free and equal in dignity and rights” (UN General Assembly, 1948). According to Feldman (2013), hate speech annihilates dignity, and in that way detracts from an individual’s assurance of political and legal equality and inclusiveness in society. Intuitively, it seems obvious that human beings should aim to treat each other equally, thus protecting the public dignity of our societal order. However, because of the multiplicity of perspectives related to hate speech, it is not easy for policy makers and practitioners – for example teachers with their students – to contribute to combating hate speech.

In their review exploring teachers' perceptions and responses towards cyberbullying, Macaulay et al. (2018) found that teachers see the education of their pupils about cyberbullying awareness to be effective, but they need additional training to increase their knowledge of how to reduce involvement in and long-term exposure to bullying. Therefore, in this article I examine the potential of arts education to counter hate speech in the light of recent research and discussion, in order to provide factual

knowledge for policy makers and practical tools for teachers, with their students, to increase their confidence and ability to identify, prevent, and combat hate speech.

Hate speech and freedom of expression

What actually constitutes hate speech is not a simple matter to define. The problem in defining hate speech is not in the hateful thoughts, but in the concrete harm that follows from the publication and dissemination of hate speech. Many international courts do not define the term hate speech, which makes it difficult to state where and when emotions and incitements become hatred (Mendel, 2012). Among scholars (e.g. Gomes, 2016; Mendel, 2012), hate speech has been defined as any form of expression – for example by means of speech, images, videos, or online activity – that has the capacity to increase hatred against a person or people because of a characteristic they share, or a group to which they belong. In defining hate speech, it is necessary to clarify the difference between insults (or offense), which are connected to an individual's emotions, and punishable hate speech (or defamation) (Feldman, 2013). According to Waldron (2012), in order to be punishable, speech must attack social aspects of an individual in a society, such as the status or reputation or dignity of the group, rather than effect how things feel to them.

In order to maintain the integrity of a functioning democracy, the government must protect both the equal human dignity of every person and free speech with open debate engaging all viewpoints, as a precondition for democratic citizenship (Koltay, 2016; Tsesis,1999). Therefore, in combating hate speech it is important, and at the same time challenging, to identify the best balance between allowing freedom of expression[1] and protecting other human rights (Gomes, 2016; Heinze, 2016). Some scholars (e.g. Bonotti, 2017; Heinze, 2016) argue that hate speech laws are not a solution to combating hate speech in the cases where hate speech does not cause an imminent threat of violence. They agree that hate speech bans may, under some circumstances, promote security in order to preserve democracy for all citizens. However, in their opinion, hate speech bans do not promote democracy, because within a democracy public discourse is the constitutional foundation which allows citizens to express their opinions for and against any policies, without being censored or penalized, even in cases in which their viewpoint is considered hateful. It is a serious concern that hate speech bans can be abused by politically powerful factions to censor speech that criticizes them (McConnell, 2012). Hate speech bans are also often ineffective, because hate speakers can reformulate their hateful speech in euphemistic and indirect forms, which can be as harmful as direct hate speech (Bonotti, 2017; Heinze, 2016). Also, punishing hate speakers does not necessary directly support the ability of their targets to speak in response (Gelber, 2012).

On the contrary, some scholars (e.g. Feldman, 2013; Koltay, 2016; Tsesis, 1999; Waldron, 2012) argue that although free speech is an important value as an individual right, and essential to democratic citizenship, freedom of expression cannot negatively impact human dignity, equality, and reputation. Thus, just as protection against actual physical attacks should be guaranteed in a democracy, so also should there be a formal, symbolic recognition of human dignity, even when hate speech does not cause any

obvious harm to the members of the community attacked (Koltay, 2016). The right to free speech cannot safeguard hate speech, because supporting hate speakers' verbal freedom can weaken a pluralist democracy, when outgroup members do not feel safe enough to equally exercise their political and constitutional rights in a society (Tsesis, 1999; Waldron, 2012). Therefore, the emphasis on equal human dignity and citizenship for all individuals, and thus all groups and communities, should be an essentially pluralist democratic concern, and hate speech laws can protect a minority individual's ability to participate fully as a democratic citizen (Feldman, 2013; Koltay, 2016; Waldron, 2012).

Harm from hate speech

What kind of harm can hate speech cause to the individuals, groups, and communities? Some scholars (e.g. Gelber, 2012; Gomes, 2016) have reported that hate speech is damaging in itself, and creates conditions for further and more serious harm, such as human rights violations, discrimination, mental and emotional damage, disempowerment, marginalization, silencing and suppression, and violence. According to Gelber (2012), sometimes people, especially children being influenced by their peer groups, can use hate speech without intending to harm, when they do not realise that they are using hate speech, or when they do not understand the message. Often the response of the targets of hate speech is to become angry and to defend themselves, but the response can also be to become an activist in a society instead of becoming victimized.

There is also a debate among scholars as to whether hate speech can cause long-term harm. Tsesis (1999) argues that the Holocaust, the Native American dislocation, and Black slavery were made possible by repeated hate propaganda, which formulated over long periods of time a foundation for a conceptual framework to promote systematic intolerance, oppression, discrimination, destruction, and racist policies. As an example of the long-term harm of hate speech, Hancock (1991) has outlined a chronology of Gypsy history in which he shows the origins of the Holocaust against the Romani beginning in the 15th century and leading stage by stage to the genocide of the European Romani during World War II.

To the contrary, however, some scholars (e.g. Desai, 2003; Heinze, 2016) see the claims of a causal relationship between hate speech and long-term direct effects or indirect harm to individuals, groups, and communities as too simple and straightforward. They argue that there is not enough legal or scientific evidence to indicate that those incidents in history were caused by hate speech. In this view, hate speech might be part of the process, but other factors, such as government actions and policy, have had a stronger impact on those consequences.

A recent example of the harm of hate speech is the case of three persons who made deliberately offensive and provocative online posts – called "trolls" in internet slang – and who were convicted in the District Court in Finland of systematic defamation against a journalist.[2] The court rejected their arguments of exercising the right to freedom of speech, because the trolls' attacks, made as false accusations posted online, continued systematically for more than three years, and the primary motive was to undermine and

destroy the journalist's professional credibility and reputation (Higgins, 2018). The journalist received death threats, was mocked online as a subject of insulting memes, and had her face photoshopped onto pornographic images, and her address, medical records, and contact details were published online (BBC News, 2018b). Another victim in the same case described how, after internet trolls' systematic continuing defamation, she had serious fears; for example, she was afraid to go shopping, she became afraid of arsonists, she had everyday difficulties in sleeping and eating, she was not able to work, and she suffered from anxiety and vomiting (Salminen, 2018).

Waldron (2012) emphasizes that although it is a serious concern that hate speech can create imminent dangers of harmful or illegal conduct, the constituent concern is that hate speech deflates the requisite conditions for a pluralist democratic process. On the other hand, public incidents caused by the hate speakers can increase the empowerment of opposition, and in that way strengthen instead of weaken the assurance of security for the targets of the hate speech (McConnell, 2012). Both Gelber (2012) and Winter and Fürst (2017) suggest that, in cases where hate speech does not cause imminent danger, the appropriate concrete response to hate speech is counter speech, which enables counteractions against the silencing and disempowering effects caused by hate speech on its targets. According to Reagle (2015), counter speech can expose hate, deceit, abuse, and stereotypes by providing clarification, promoting counter narratives, and advancing counter-values, such as sharing experiences and uniting communities.

Counter narratives

How can we overcome the seemingly polarized choices between hate speech bans and free speech, and at the same time support the targets of hate speech and their communities, so that they become capable of responding to hate speech? Gelber (2012) suggests that we should utilize an expanded conception of counter speech, in which freedom is not merely an opportunity but an exercise. This requires a reconceptualization of freedom of expression in participatory terms, such as self-development, and understanding that speech is capable of doing both good and bad things for people. Gelber has adapted this idea from Nussbaum's (e.g. 2003) theory of ethics, which entails human functional capabilities as being necessary to foster human flourishing. When we understand that speech has a constitutive role in the formation of individual capabilities, a supported policy response, including adequate institutional, material, and educational support, is focused on the targets of hate speech and their supporters instead of on the hate speakers.

Counter narratives can be utilised in counter speech to support and enable a response to hate speech, by giving a voice to people who would otherwise not have one. These kinds of narratives aim to dispute and contradict a commonly held belief or truth relating to cultures, people, and institutions by sharing a different point of view, based on human rights and democratic values such as openness, respect for difference, freedom, and equality (Gomes, 2017; Tuck and Silverman, 2016). Counter narratives do not necessarily discredit the beliefs that have been previously established, but rather deconstruct the narratives on which they are based by offering a different way of thinking

about the issues. For example, counter narratives can provide alternative and accurate information against hate speech propaganda, and aim to deconstruct or delegitimise hate speech narratives by using humour, appealing to emotions on the topics involved, and offering different perspectives focusing on what we are for rather than against (Gomes, 2017; Tuck and Silverman, 2016).

In their Counter-Narrative Handbook, Tuck and Silverman (2016) advise how to create counter narratives by planning a campaign, creating and testing the content, running a campaign, advertising, engaging audiences, and evaluating campaigns. An effective campaign is age appropriate, the language should be easily understood and it is pitched at the right level for the audience to reach the right people, not necessarily the most people. The most effective messages do not lecture the audience; instead, they offer something to think about, feel, remember, and reflect on. In some cases, counter narratives can also be misunderstood – in particular, comedy is not necessarily easy to use, because not everyone will find the same things funny. A project by Silverman et al. (2016) shows that the process of creating counter narrative content can be slow, and require an enormous amount of work. Therefore, a good option is also to expand and redirect pre-existing counter narrative content.

A counter narrative campaign can be a counteractive community newsletter, an awareness program, a discussion workshop about the effects of hate speech, a workshop on writing replies and opinions to newspapers, producing radio or television advertisements or an online video, or creating community art projects (Gelber, 2012). As an example, Tuck and Silverman (2016) have illustrated instructions for a counter narrative campaign against extremism.[3] That kind of campaign can highlight how extremist activities negatively impact on the people which they argue to represent. Also, it is possible to demonstrate the hypocrisy of extremist groups, and how their actions are often inconsistent with their own stated beliefs. The factual inaccuracies can be emphasised by showing that something which has been regarded as true is in fact not true, and by satirising extremist propaganda to undermine its credibility.

Tuck and Silverman (2016) also caution that there are security considerations in running counter narrative campaigns. Negative responses and abusive, threatening, or racist comments, even from the extremist groups, can be a consequence of the campaign. Therefore, it is crucial to estimate beforehand whether securing the campaigner’s personal details and social media accounts is needed. In addition, it is worthwhile to consider possible risks before running the campaign, and whether the campaign can be linked to the campaigner’s organisation or not.

Intersections in countering hate speech through arts and arts education

In order to be effective and enticing, counter narratives can combine real and fictional elements. Artistic expression often enjoys a wider degree of freedom of expression than formal speech, and therefore artistic freedom can offer a creative way to navigate between freedom of expression and combating hate speech (McGonagle et al., 2012). Although artistic freedom is often allowed to be provocative, artists are also responsible

for being mindful that their artistic expression does not use hate speech. There is a recent case from 2018 in Spain demonstrating how artistic expression was considered criminal by the government. The Spanish court condemned rapper “Valtonyc” (Josep Miquel Arenas) to three-and-a-half years in jail for incitement to terrorism, insulting the crown, and making threats, based on one of his songs where he criticized the King of Spain (Telesur, 2017). In a pluralist democracy it is problematic if any single narrative is considered to be the only “normal” one, and even more serious if the narrative includes hate speech (Gomes, 2017). The employment of artists’ practices has great potential for creating different counter narratives to challenge one-sided narratives. Also, teachers can utilize arts education in schools to make activities with students to counter hate speech, which may offer a constructive way to handle hate speech.

Various artists have already utilised counter narratives in their art works, which can serve as examples of methods for teachers to adapt in arts education in schools. One example of how visual art can be used to create a counter narrative, even without speech, is the artist Ana Teresa Fernández’s artwork “Borrando la Frontera” (Erasing the Border). This project took place in 2016 in three places along the border of the United States of America and Mexico, where members of the cultural organization Border/Arte “removed” parts of the border fence by painting large sections sky blue, allowing the fence to visually blend into the sky and to symbolically erase a long-standing physical barrier separating families and causing harm and sorrow to them (Taylor, 2016). Although counter narratives usually aim to construct something new, in the artistic activism by the “Hate Destroyer”, Irmela Mensah-Schramm, graffiti-erasing is used as a way to counter hate speech. She is 72 years old, and since 1985 she has been going out every morning in Berlin looking for racist, homophobic, or anti-Semitic graffiti or stickers, to permanently erase them, scratch them off, or cover them with paint (Caruso, 2017).

An art installation can also provide a space where people can participate in cooperatively building a counter narrative through dynamic conversation, instead of being isolated with their stressful emotions. Artist Matthew “Levee” Chavez noticed how people suffered post-election anxiety and uncertainty during the day after Republican Donald Trump’s presidential election win over Democratic nominee Hillary Clinton in the United States of America in 2016. He set up a therapy desk in a bypass subway tunnel in New York and offered hundreds of travellers post-it notes, encouraging them to put their thoughts and drawings on the wall, such as “9/11 Never Forget, 11/9 Always Regret” (Leigh, 2016). As another example, a counter narrative against hate speech in a massive art work at a public festival can reach many people. “Wall of Hope” was a 15 meters wide and 2 meters high art work made by artists EGS and Jani Leinonen at the World Village Festival 2017 in Finland, as a part of Amnesty International Finland's (2017) campaign against hate speech. The wall consisted of pieces of hate speech sent to Amnesty International, which were covered over by the artists’ works illustrating hope. A human figure in the art work framed the hate speech, showing how every individual is responsible for expressing their emotions in a constructive way, instead of through discriminatory hate speech.

The Council of Europe (2018) introduces creative ways in which young people can counter hate speech in different contexts, by combining various art forms and methods such as participatory theatre, storytelling, pictures, and videos, in order to address different types of hate speech. For example, the Living Library is a participatory work meant to challenge prejudice, stereotypes, and discrimination, by offering the possibility to borrow *people*, who can be, for example, victims of hate speech or activists in combating hate speech, instead of books. With the message "Don't judge a book by its cover!" it shows that despite our differences, people share a common humanity with similar concerns and hopes. In Finland, the ByHelpers (2017) community fights against the bystander effect by encouraging people to help strangers in everyday life, with the slogan "Act when you encounter hate speech instead of giving your silent approval to the situation. Don't be a bystander, be a #ByHelper!". They also utilize art to gather people together, for example by organising a community art painting event, the "Wall of Art", in a park in Helsinki in 2017, where people could imprint the figure of their hand with different colours.

Visual artist Eetu Kevarinmäki (2017) started to investigate aggressive chatting on Facebook, and made art works based on the comments, including hate speech, which were shown in his exhibition "Vihapuheen Estetiikka" (Aesthetics of Hate Speech) in Helsinki in 2017. The exhibition included a sound art work and 400 photos, in which he had opened the code behind the Facebook profile photo of hate speakers as a text file, and added the hate speech text in between the code. As a result, there was an abstract and broken profile picture, which illustrated how humanity is fragmented by hate speech. A photo can also use counter narratives to raise awareness and hope. In her "Precious Baby Project", photographer Angela Forker has photographed medically fragile babies, or babies with disabilities, in her home studio in Indiana, as a way to show strength, potential, and love. She uses fabric and other ordinary items in her work, and places them to create a unique environment meant to show the potential of each baby, for example by giving the appearance of the baby flying, steering a boat, or running (Stumbo, 2018).

Music can also be used to enhance counter narrative activities against hate speech. "Love Music Hate Racism" (2018) started in 2002 in the United Kingdom as a response to rising levels of racism, and over concerns about the success of the British National Party. The movement uses music to promote diversity and a multicultural society, and to involve people in anti-racist activities at their music events, from local gigs to large outdoor festivals.

Addressing implications for practice and policy

Arts education can offer a creative and effective way for policy makers and practitioners to combat hate speech, as they try to balance between respect for human dignity on the one hand, and freedom of expression on the other, as human rights and preconditions for democratic citizenship. Recent research by Van de Vyver and Abrams (2018) provides evidence that people's greater engagement with the arts predicates greater pro-sociality through volunteering and charitable giving; therefore, art can act as an important social psychological catalyst towards a cohesive and socially prosperous society. A literature

review by Menzer (2015) suggests that music, drama, and visual arts activities are positively related to both social and emotional competencies in early childhood. Catterall (2009) and Catterall et al. (2012) arrived at similar conclusions in their studies, which indicate that young people who have arts-rich experiences in school become more active and engaged citizens than their less artistically involved peers in voting, volunteering, and generally participating in society.

Research by Rose et al. (2017) shows that artists and cultural organizations can have a remarkable role in equity change work through diverse and avant-garde forms, such as bringing creative visions, forming political resistance against poverty and human rights abuses, unifying and healing communities, and advocating for equitable economies. Arts and cultural activities can bring many benefits and high value to both individuals and society by creating the conditions for change, such as creating spaces for experimentation and risk-taking and developing the ability to reflect in a safer and less direct way on personal, community, and societal challenges (Crossik and Kaszynska, 2016). In addition, a literature review of interdisciplinary studies exploring the social impacts of arts and culture by the Department of Canadian Heritage (2016) found arts and culture to have multiple and positive impacts on and benefits for society; however, the measurement of these characteristics is very difficult and there is no current consensus around the conclusions.

Although there is a great body of evidence available on the social impact of arts and culture, research by Silverman et al. (2016) shows that with regard to increasing the understanding of the impact of interventions in countering hate speech, much more research is needed. They suggest the use of offline market research techniques to better understand web users' online content, offline opinions, and behaviour changes. This should also include in-depth interviews with intervention providers who work with young people in order to deepen our understanding of youth attitudes and behaviour. Testing and comparing the impact of counter narratives is one way to increase the available scientific evidence on countering hate speech. Because hate speech is not at all a simple and easy issue to address at schools, research evidence, concrete guidelines, and practical examples can help teachers in their efforts to combat it.

There is a recent example (BBC News, 2018a) of the complexity of this issue in Finland, from a secondary school which was drawn into an argument with a Nationalist member of parliament, who accused it of encouraging hatred. Three 15-year-olds designed a poster as part of a city-wide event to highlight social issues. They chose immigration as their theme, presenting migrants in a cramped boat, facing a choice of who to turn to. To the left of the boat, under the name "Suomeen" (to Finland), the students set photos of the President of Finland and a Greens member of parliament, while the Nationalist member of parliament and her party leader were put to the right of the boat under the caption "kuoleen" (to death). The poster caused a heated debate over whether it was appropriate for a social studies project at school.[4]

Rather than focusing on the public accusation of inciting hatred, the episode around the poster can be seen as a call for training teachers to better handle issues around hate speech in schools. For example, in place of the use of more extreme language, the

teacher could have steered the students towards a more sensitive message, including diverse perspectives that did not detract from the overall meaning. It is understandable that overreactions occur when handling burning political topics with young people. Because of the ethical ambiguity that exists in hate speech discourse, any communication is, in reality, not always so simple. It is expected that teachers encounter challenges and resistance from some of the students, their parents, and other teachers from diverse backgrounds and beliefs, and with varying positions of power and perspective, when addressing issues around hate speech. Even tiny differences of opinion within the conversation on hate speech can lead to intense disputation, and this is often the reason why it is safer and more comfortable not to interfere with the topic in a school environment. However, fear of missteps and public blaming should not discourage teachers from activating their students to address the issues around hate speech.

Emcke (2016) highlights that those who do not interfere and attempt to tackle hate speech, actually allow the space for hate to grow by tolerating it with their silent acceptance. That is why practical work with democratic values, such as openness, inclusion, equality, and justice, is one of the most important ways to counter hate speech. Hate as an emotion is not an efficient response to ideological hate speech. Instead, using tools which hate speakers cannot use may undermine hate speakers' credibility. Those kinds of tools can be many things, from deciding not to join the call of hate, to taking the time, again and again, to carefully elaborate ourselves and our differences, backgrounds, and frameworks related to hate, even before the hate is expressed. Education is an important factor in deconstructing intolerance, prejudice, and discriminatory attitudes and behaviours, and thus teachers at schools may have a crucial role to play in encouraging young people to combat hate speech. Because hate speech bans cannot reach the roots of hatred, arts education can offer ways to disclose what is hidden, and to examine the ignorance, misunderstandings, and false beliefs within the historical and cultural contexts of hate speech (Molnar, 2012).

In countering hate speech, counter narratives can be considered as a method which works best when combined with other policy approaches (Gelber, 2012). Waldron (2012) points out that counter speech alone is not a sufficient response to hate speech, because it legitimates the issue by suggesting that we should be engaged in conversation with hate speakers, trying to convince them and others that minorities should be treated as full and equal citizens. In a functioning pluralist democratic process all citizens, including minorities, are worthy of equal citizenship without such conversations. Also, Coustick-Deal (2017) reminds us that counter speech is often defined by those who already have the privilege and freedom to exercise it without fear or harm. For example, research by Munger (2017) showed that counter speech as a reply to racist Tweets reduced racist hate speech, but only if people thought that the reply was written by a white male avatar. The counter speech was produced by automated Twitter bots, and included one sentence: "Hey man, just remember that there are real people who are hurt when you harass them with that kind of language." If the avatar was thought to be a person of colour, the counter speech showed no measurable impact, and in fact the avatar was more likely to receive a negative response. Therefore, work for structural changes is needed to create spaces where everyone can feel equally safe to counteract these influences, including in arts education (Jääskeläinen, 2016).

Conclusions

In addition to strengthening the legal framework for addressing cases when hate speech can be considered criminal, and developing automated monitoring of online systems to prevent cyberhate, utilizing arts education to create culturally sensitive and effective counter narratives can provide a practical and creative way for policy makers to increase awareness of these issues, and for teachers to empower students to counter hate speech.

Countering hate speech requires us to use other ways than expressing hate ourselves, and according to recent studies on various types of beneficial social impacts, the arts have the potential to provide a more positive means of communication. Arts education can be utilised to create efficient counter narratives, which can provide space to support diverse viewpoints that can question hate speakers' simplified generalisations. However, much more research, supported by sufficient resources, is needed to evaluate the impact and effectiveness of intervention strategies in countering hate speech through arts education. In order to address policy recommendations around arts education-based intervention strategies, more focused exploration needs to be undertaken into the specifics of what counter narratives could look like in arts education in diverse cultural and educational contexts, how they are facilitated in practice, and how these actions are connected to policy implications.

Although there is not yet enough evidence on the impact of using arts education in countering hate speech, the brave art works of artists creating influential counter narratives can encourage people to join together and act. Just as was written by an anonymous by-passer on one of the post-it notes in the art installation in the New York subway station (Leigh, 2016): "LET'S USE THIS ANGER. LET'S ORGANIZE!".

Funding

This work was undertaken at the Center for Educational Research and Academic Development in the Arts (CERADA), University of the Arts Helsinki, Finland, as part of ArtsEqual Research Initiative, supported by the Academy of Finland's Strategic Research Council under Grant 293199/2015.

Acknowledgments

I thank Dr. Christopher Ten Wolde for valuable comments on the initial manuscript.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Notes

1. The Article 19 of Universal Declaration of Human Rights: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers" (UN General Assembly, 1948).
2. "The concept of punishable hate speech or hate speech crime is not contained in legislation [in Finland]. Cases investigated by the Hate Speech Investigation Team are categorised as ethnic agitation, aggravated ethnic agitation or infringing the right to practice a religion in peace. The Helsinki Hate Speech Investigation Team also investigates cases of online defamation, aggravated defamation, illegal threats and other crimes, if the act is committed against someone on the basis of their race, skin colour, descent, national or ethnic origin, religion or beliefs, sexual orientation or other equivalent grounds." (Karuselli uutiset, 2017)
3. "Extremism is the vocal or active opposition to our fundamental values, including democracy, the rule of law, individual liberty and the mutual respect and tolerance of different faiths and beliefs. We also regard calls for the death of members of our armed forces as extremist." (HM Government, 2015: 9)
4. According to the National Core Curriculum for Basic Education in Finland (FNCC, 2014: 15–16), "Discussions of values with the pupils guide the pupils to recognize values and attitudes they encounter and to also think about them critically" and "Basic education promotes well-being, democracy and active agency in civil society".

References

Amnesty International Finland (2017). *Vihaa vastaan [Against hate].* In: Amnesty International Finland. [online] Available at: www.amnesty.fi/vihaa-vastaan/ (accessed 31 October 2018).

BBC News (2018a) Finland school poster sparks migration row. *BBC News*, 3 October. [online] Available at: www.bbc.co.uk/news/world-europe-45731741 (accessed 31 October 2018).

BBC News (2018b) Jessikka Aro: Finn jailed over pro-Russia hate campaign against journalist. *BBC News*, 18 October. [online] Available at: www.bbc.co.uk/news/world-europe-45902496 (accessed 31 October 2018).

Blaya C (2018) Cyberhate: A review and content analysis of intervention strategies. *Aggression and Violent Behavior*. Epub ahead of print 5 May 2018. DOI: 10.1016/j.avb.2018.05.006.

Bonotti M (2017) Book review: Hate speech and democratic citizenship. *Social & Legal Studies* 26(2), 276–280. DOI: 10.1177/0964663917704734a.

ByHelpers (2017) In: *ByHelpers*. [online] Available at: www.facebook.com/ByHelpers-398505493857271/ (accessed 31 October 2018).

Caruso V (2017) *The Hate Destroyer*. In: Film Commission Torino Piemonte. [online] Available at: www.fctp.it/movie_item.php?id=2167&lang=_en (accessed 31 October 2018).

Catterall JS (2009) *Doing Well and Doing Good by Doing Art: The Effects of Education in the Visual and Performing Arts on the Achievements and Values of Young Adults*. Los Angeles/London: I-Group Books.

Catterall JS, Dumais SA and Hampden-Thompson G (2012) *The Arts and Achievement in At-Risk Youth: Findings from Four Longitudinal Studies*. Research Report# 55, National Endowment for the Arts, USA, March. [online] Available at: www.arts.gov/sites/default/files/Arts-At-Risk-Youth.pdf (accessed 31 October 2018).

Chetty N and Alathur S (2018) Hate speech review in the context of online social networks. *Aggression and Violent Behavior* 40, 108–118. DOI: 10.1016/j.avb.2018.05.003.

Council of Europe (2018) *Campaign examples.* In: Council of Europe. [online] Available at: www.coe.int/en/web/no-hate-campaign/campaign-examples1#{%2226873630%22:[17]} (accessed 31 October 2018).

Coustick-Deal R (2017) *What's wrong with counter speech?* In: Ruth Coustick-Deal. [online] Available at: https://medium.com/@ruthcoustickdeal/https-medium-com-whats-wrong-with-counter-speech-f5e972b13e5e (accessed 31 October 2018).

Crossik G and Kaszynska P (2016) *Understanding the Value of Arts & Culture: The AHRC Cultural Value Project*. Swindon, Wiltshire: Arts and Humanities Research Council. [online] Available at: ahrc.ukri.org/documents/publications/cultural-value-project-final-report/ (accessed 31 October 2018).

Department of Canadian Heritage (2016) *Social Impacts and Benefits of Arts and Culture: A Literature Review. Final Version.* Report, Open Government Canada, Canada, February. [online] Available at: publications.gc.ca/collections/collection_2018/pch/CH4-187-2016-eng.pdf (accessed 31 October 2018).

Desai A (2003) Attacking Brandenburg with history: Does the long-term harm of biased speech justify a criminal statute suppressing it? *Federal Communications Law Journal* 55(2), 352–394. [online] Available at: www.repository.law.indiana.edu/fclj/vol55/iss2/8 (accessed 31 October 2018).

Emcke C (2016) *Gegen den Hass [Against Hate]*. Frankfurt am Main: S. Fischer Verlag.

Feldman SM (2013) Review essay: Hate speech and democracy. *Criminal Justice Ethics* 32(1), 78–90. DOI: 10.1080/0731129X.2013.777254.

FNCC (2014) *National Core Curriculum for Basic Education*. Helsinki: National Board of Education.

Gelber K (2012) Reconceptualizing counterspeech in hate-speech policy (with a focus on Australia). In: Herz ME and Molnar P (eds) *The Content and Context of Hate Speech: Rethinking Regulation and Responses.* New York: Cambridge University Press, pp. 198–216.

Gomes R (ed) (2016) *Bookmarks. A Manual for Combating Hate Speech Online Through Human Rights Education*. Revised edition. Budapest: Council of Europe Publishing. [online] Available at: rm.coe.int/168065dac7 (accessed 31 October 2018).

Gomes R (ed) (2017) *WE CAN! Taking Action against Hate Speech through Counter and Alternative Narratives.* Budapest: Council of Europe Publishing. [online] Available at: rm.coe.int/wecan-eng-final-23052017-web/168071ba08 (accessed 31 October 2018).

Hancock I (1991) Gypsy history in Germany and neighboring lands: A chronology to the Holocaust and beyond. *Nationalities Papers. The Journal of Nationalism and Ethnicity* 19(3), 395–412. DOI: 10.1080/00905999108408210.

Heinze E (2016) *Hate Speech and Democratic Citizenship*. New York: Oxford University Press.

Higgins A (2018) Three internet trolls convicted of systematic defamation against journalist in Finland. *The New York Times*, 18 October. [online] Available at: www.nytimes.com/2018/10/19/world/europe/finland-internet-trolls-defamation.html (accessed 31 October 2018).

HM Government (2015) *Counter-Extremism Strategy*. London: HMSO. [online] Available at: assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/470088/51859_Cm9148_Accessible.pdf (accessed 31 October 2018).

Jääskeläinen T (2016) Vahvistanko hiljaisuudellani rasistisia rakenteita? Toiseuden kokemuksen tunnistamisesta kohti toiminnallista tasa-arvon edistämistä [Is structural racism strengthened by my silence? From recognising experiences of Otherness, towards activism for enhancing equality]. *The Finnish Journal of Music Education* 19, 2, 64–72. [online] Available at: sites.uniarts.fi/fi/web/fjme/-/vol-19-02-2016 (accessed 31 October 2018).

Karuselli uutiset (2017) *The police's Hate Speech Investigation Team is taking action against abusive online statements which constitute crimes.* In: Poliisi. Police of

Finland. [online] Available at: www.poliisi.fi/prime101/prime101.aspx/1/0/the_police_s_hate_speech_investigation_team_is_taking_action_against_abusive_online_statements_which_constitute_crimes_58608 (accessed 31 October 2018).

Kevarinmäki E (2017) In: *Eetu Kevarinmäki. Vihapuheen estetiikka [Aesthetics of hate speech].* [online] Available at: www.facebook.com/events/219745185179509/ (accessed 31 October 2018).

Koltay A (2016) Book review: Hate speech and democratic citizenship. *Journal of Media Law* 8(2), 302–306. DOI: 10.1080/17577632.2016.1209318.

Leigh S (2016) Post-its on New York subway provide post-election therapy – in pictures. *The Guardian*, 12 November. [online] Available at: www.theguardian.com/us-news/gallery/2016/nov/12/post-its-new-york-14th-street-subway-donald-trump-protest (accessed 31 October 2018).

Love Music Hate Racism (2018) In: *Love Music Hate Racism.* [online] Available at: www.lovemusichateracism.com/about-us/ (accessed 31 October 2018).

Macaulay PJ, Betts L, Stiller J and Kellezi B (2018) Perceptions and responses towards cyberbullying: a systematic review of teachers in the education system. *Aggression and Violent Behavior* 43, 1–12. DOI: 10.1016/j.avb.2018.08.004.

McConnell M (2012) Sunday book review: You can't say that. *The New York Times*, 22 June. [online] Available at: www.nytimes.com/2012/06/24/books/review/the-harm-in-hate-speech-by-jeremy-waldron.html?ref=review (accessed 31 October 2018).

McGonagle T, Herz M and Molnar P (2012) A survey and critical analysis of Council of Europe strategies for countering "Hate Speech". In: Herz ME and Molnar P (eds) *The Content and Context of Hate Speech: Rethinking Regulation and Responses.* New York: Cambridge University Press, pp 456–498.

Mendel T (2012) Does international law provide for consistent rules on hate speech? In: Herz ME and Molnar P (eds) *The Content and Context of Hate Speech: Rethinking Regulation and Responses.* New York: Cambridge University Press, pp. 417–429.

Menzer M (2015) *The Arts in Early Childhood: Social and Emotional Benefits of Arts Participation: A Literature Review and Gap-analysis (2000–2015).* Washington, DC: National Endowment for the Arts. [online] Available at: www.arts.gov/sites/default/files/arts-in-early-childhood-dec2015-rev.pdf (accessed 31 October 2018).

Molnar P (2012) Responding to “hate speech” with art, education, and the imminent danger test. In: Herz ME and Molnar P (eds) *The Content and Context of Hate*

Speech: Rethinking Regulation and Responses. New York: Cambridge University Press, pp. 183–197.

Munger K (2017) Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior* 39(3), 629–649. DOI: 10.1007/s11109-016-9373-5.

Nussbaum M (2003) Capabilities as fundamental entitlements: Sen and social justice. *Feminist Economics* 9(2–3), 33–59. DOI: 10.1080/1354570022000077926.

Reagle J (2015) *Counter speech*. In: Joseph Reagle. [online] Available at: reagle.org/joseph/2015/07/counter-speech.html (accessed 31 October 2018).

Rose K, Daniel M and Liu J (2017) *Creating Change through Arts, Culture, and Equitable Development: A Policy and Practice Primer.* Oakland, CA: PolicyLink. [online] Available at: www.policylink.org/sites/default/files/report_arts_culture_equitable-development.pdf (accessed 31 October 2018).

Saleem HM, Dillon KP, Benesch S and Ruths D (2017) A web of hate: Tackling hateful speech in online social spaces. In: *Proceedings of the first workshop on text analytics for cybersecurity and online safety (TA-COS 2016)* (eds G De Pauw, B Verhoeven, B Desmet and E Lefever), Portorož, Slovenia, 23 May 2016, pp. 1–9. Portorož: European Language Resources Association (ELRA). [online] Available at: www.ta-cos.org/sites/ta-cos.org/files/master.pdf (accessed 31 October 2018).

Salminen S (2018) ”Ilja, pystyisitkö laittamaan tämän?” Näin MV-lehden viharikokset vaikuttivat uhreihin – Oksenteleva perheenäiti ohjeisti lastaan pahimman varalle [”Ilja, could you add this?” In this way the hate crimes by the MV Magazine influenced the victims – A vomiting mother told her child to be prepared for the worst]. *Iltalehti*, 18 October. [online] Available at: www.iltalehti.fi/kotimaa/d30b3e83-4cf0-407b-bdb6-3c4a3703ba5d_u0.shtml (accessed 31 October 2018).

Silverman T, Stewart CJ, Birdwell J and Amanullah Z (2016) *The Impact of Counter-Narratives. Insights from a Year-Long Cross-Platform Pilot Study of Counter-Narrative Curation, Targeting, Evaluation and Impact*. Report, Institute for Strategic Dialogue, UK, August. [online] Available at: www.isdglobal.org/wp-content/uploads/2016/08/Impact-of-Counter-Narratives_ONLINE_1.pdf (accessed 31 October 2018).

Stumbo E (2018) *Why this photo series of babies with disabilities has the power to be revolutionary.* In: The Mighty. [online] Available at: themighty.com/2018/10/the-precious-baby-project-angela-forker/ (accessed 31 October 2018).

Taylor A (2016) Erasing the Border. *The Atlantic*, 12 April. [online] Available at: www.theatlantic.com/photo/2016/04/erasing-the-border/477891/ (accessed 31 October 2018).

Telesur (2017) Rapper sentenced to 3 years in jail for rap about King of Spain. In: *Telesur*, 23 February. [online] Available at: www.telesurtv.net/english/news/Rapper-Sentenced-to-3-Years-in-Jail-for-Rap-about-King-of-Spain-20170223-0047.html (accessed 31 October 2018).

Tsesis A (1999) The empirical shortcomings of first amendment jurisprudence: A historical perspective on the power of hate speech. *Santa Clara Law Review* 40, 728–786. [online] Available at: digitalcommons.law.scu.edu/lawreview/vol40/iss3/3 (accessed 31 October 2018).

Tuck H and Silverman T (2016) *The Counter-Narrative Handbook*. London: Institute for Strategic Dialogue. [online] Available at: www.isdglobal.org/wp-content/uploads/2016/06/Counter-narrative-Handbook_1.pdf (accessed 31 October 2018).

UN General Assembly (1948) *Universal declaration of human rights (217 [III] A).* Paris. [online] Available at: www.un.org/en/universal-declaration-human-rights/ (accessed 31 October 2018).

Van de Vyver J and Abrams D (2018) The arts as a catalyst for human prosociality and cooperation. *Social Psychological and Personality Science* 9(6), 664–674. DOI: 10.1177/1948550617720275.

Waldron J (2012) *The Harm in Hate Speech*. Cambridge, Massachusetts: Harvard University Press.

Winter C and Fürst J (2017) *Challenging Hate: Counter-Speech Practices in Europe.* Report, International Centre for the Study of Radicalisation (ICSR), UK, March. [online] Available at: icsr.info/wp-content/uploads/2018/03/ICSR-Report-Challenging-Hate-Counter-speech-Practices-in-Europe.pdf (accessed 31 October 2018).