GSDRC Applied Knowledge Services
Governance · Social Development · Humanitarian · Conflict

www.gsdrc.org
helpdesk@gsdrc.org

Helpdesk Research Report

Interventions to counter hate speech

Anna Louise Strachan

23.06.2014

Question

What evidence is there that programming interventions on countering hate speech have been effective, and what examples are there of proven successful interventions?

Contents

1. Overview

There is limited evidence on the effectiveness of interventions to counter hate speech. There is a lack of rigorous impact evaluations in this area and those that do exist tend to focus on individual case studies. Reasons for the lack of evidence on the effectiveness of interventions to counter hate speech include difficulties associated with attributing any changes in the level of hate speech to a particular project. Moreover, changes as a result of such interventions can take a long time to manifest themselves. The impact of a project aiming to achieve significant behavioural change is unlikely to be obvious during, or in the immediate aftermath, of an intervention.

It is not always easy to identify hate speech. This is because context matters when identifying inflammatory language and the level of danger it presents (Taylor and Dolan, 2013, p. 3). Subtle forms of hate speech can be particularly hard to identify (Vollhardt et al, 2006, p. 26). Two of the key characteristics of hate speech are:

- **Dehumanisation:** This can be blatant, such as referring to other groups or individuals with animal names or it can be more subtle (Vollhardt et al, 2006, p. 26). One subtle form of dehumanisation is 'animalistic dehumanisation.' This involves denying a group the characteristics of human uniqueness such as civility, moral sensibility, rationality and maturity (Vollhardt et al, 2006, p. 27). The other subtle form of dehumanisation is 'mechanistic dehumanisation.' This involves denying a group characteristics that constitute human nature, such as emotional responsiveness, cognitive openness, and agency or individuality (Vollhardt et al, 2006, p. 27).

- **Simplistic and unbalanced communication:** Hate speech violates standards of argumentative integrity. For example, the arguments put forward by target groups are often misrepresented and other groups are often blamed for political events or societal problems for which they are not responsible. Perpetrators of hate speech often present subjective arguments as objective truth and they often refuse rational discussion of strong ideological statements (Vollhardt et al, 2006, p. 26).

Interventions to counter hate speech which have had some success include:

- **Television programmes:** In Kenya four episodes of a popular television series were broadcast. All the episodes focused on hate speech and incitement to violence. An independent evaluation of the intervention suggests that the programmes made citizens in areas prone to violence more sceptical of political leaders who use inflammatory language.
- **Radio programmes:** A Dutch NGO called Radio La Benevolencija has used radio dramas, discussions and educational programmes to enable vulnerable citizens in conflict-affected countries to recognise and respond to inflammatory speech. The Search for Common Ground organisation has also used this approach in Côte d'Ivoire.
- **Text messages:** Civil Society Organisations (CSOs) in Indonesia and Kenya have successfully used text messages to counter rumours and inflammatory speech in areas prone to ethnic violence.
- **Monitoring hate speech:** Monitoring hate speech is often used as a foundation for other interventions to counter hate speech. In Kenya, the Umati project created a database of hate speech in the run-up to the country’s 2013 election.
- **Self-regulatory media systems**: In Iraq, the United States Institute of Peace (USIP) supported local media stakeholders in the establishment of a self-regulatory media system, in order to reduce the prevalence of hate speech in the media. However, in Kenya many experts viewed self-regulation or self-censorship at the time of the country’s 2013 general election as negative, as they felt that the media was not fulfilling its role as a watchdog.

2. Are interventions to counter hate speech effective?

There is limited evidence on the effectiveness of projects that target hate speech and inflammatory language. There are a number of reasons for the lack of evidence on this topic. One expert notes that it is rarely possible to attribute a decline in hate speech to a particular project, due to the number of projects being undertaken at the same time (Expert comment). Another expert notes that the success or failure of interventions to counter hate speech is also context dependent (Expert comment). It is therefore difficult to assess the effectiveness of interventions to counter hate speech as a whole. Moreover, interventions to counter hate speech require a long-term perspective. Their impact is unlikely to be immediately obvious and monitoring and evaluation needs to take place over a period of three to five years. Funding cycles can make this difficult (Expert comment).

There are a number of factors that make the effectiveness of interventions to counter hate speech more likely. The majority of the interventions included in this report have been locally-led. This is described by one expert as the only way to ensure effectiveness (Expert comment). In their systematic review of literature on communications-related development interventions in fragile states, Skuse et al (2013) identify developing a comprehensive understanding of conflict as critical to interventions to counter hate speech. This is because hate speech builds on stereotypes, societal beliefs and cultural preconceptions, which should be understood before interventions to counter hate speech can be effective (Skuse et al, 2013, p. 45).

While not focusing exclusively on hate speech, a related review on messages to reduce violent behaviour by changing attitudes, behaviour and norms, found that there is no hard evidence to suggest that this type of intervention is successful (Rao, 2014, p. 1). Rather, it found that this type of intervention is most likely to be effective as part of a wider strategy involving a range of activities such as dialogue and training (Rao, 2014, p. 1). A multi-level strategy is also considered advisable when dealing with hate speech specifically. Some note that structural and political interventions to deal with hate speech need to be combined with interventions at the individual level, such as psychological campaigns that create resistance to hate speech, in order to be effective (Vollhardt et al, 2006, p. 18). One expert notes that interventions to counter hate speech are unlikely to be effective if a lexicon of hate speech used in the specific context is not compiled as part of the initial stage of the intervention (Expert comment).

3. Examples of successful interventions

A number of international organisations and NGOs suggest strategies for countering hate speech, but there is limited evidence of successful interventions. There is a general lack of impact evaluations of interventions to counter hate speech. When impact is measured it tends to be done in the short-term, for example by looking at feedback from individuals involved in conferences, rather than measuring behavioural change or changes in attitudes. Moreover, claims about the impact of specific interventions are often made by the organisations running the interventions and are unsubstantiated.

Television programmes

In Kenya, four episodes of the popular television programme Vioja Mahakamani were created with the aim of making Kenyan audiences less susceptible to inflammatory speech (Benesch, 2013, p. 16). Each episode was filmed in a town or village that had experienced severe inter-communal violence in the aftermath of the country's 2007 elections. The episodes were shown on television in October and November 2012 and again in the run-up to the 2013 elections (Benesch, 2013, p. 17). An independent evaluation of the project was undertaken by the Center for Global Communication at the Annenberg School for Communication at the University of Pennsylvania. The four episodes dealing with inflammatory speech were shown to one group of Kenyans and another set of Vioja Mahakamani episodes were shown to a control group (Benesch, 2013, p. 18). The evaluation found that those who watched the four episodes were more sceptical of inflammatory speech and had a better understanding of the idea that leaders often use such language to their own advantage (Benesch, 2013, p. 18).[1]

An assessment of media support to the Balkans, based on interviews and existing evaluations, found that donor support for new media significantly reduced overt ethno-nationalist propaganda. An example of this

[11] For the full evaluation see: Kogen, L. (2013). *Testing a Media Intervention in Kenya: Vioja Mahakamani, Dangerous Speech, and the Benesch Guidelines*. Center for Global Communication Studies, Annenberg School for Communication, University of Pennsylvania. Available at http://voicesthatpoison.org/vioja-evaluation/

type of project is a Norwegian Aid financed television programme in Croatia, which addressed the contentious issue of the return of Serb refugees who had been driven out or fled the country during ‘Operation Storm’ in 1995 (Rhodes, 2007, p. 25).

Radio programmes

A Dutch NGO called Radio La Benevolencija broadcasts radio soaps, discussions and educational programmes in conflict and post-conflict settings to help citizens to recognise and resist hate speech and manipulation to violence.[2] This approach has been implemented with some success in Burundi, DR Congo, Rwanda and South Sudan.

In their independent evaluation of Radio La Benevolencija’s work in post-genocide Rwanda, Paluck and Green (2009) found that the radio soap opera Musekeweya (Kinyarwanda for ‘New Dawn’)[3] had a significant impact, increasing listener’s willingness to express dissent, and improving the ways in which they resolved communal problems. However, despite the programme’s aim of making Rwandans immune to hate speech and other incitement to violence, it had not succeeded in changing attitudes and behaviours to other social groups one year after it began (Paluck and Green, 2009). Changes in individual attitudes, perceived community norms, and deliberative behaviours were assessed using closed-ended interviews, focus group discussions, role-play exercises, and measures of collective decision making (Rao, 2014). The authors conclude that personal convictions about social group boundaries tend to be more difficult to change. However, they argue that personal convictions have a lesser impact on behaviour than social and political norms (Paluck and Green, 2009).

In DRC, Radio La Benevolencija undertook a large-scale media campaign to counter hate speech during the country’s 2006 presidential elections. Part of the campaign consisted of weekly radio broadcasts on Radio Okapi (a UN funded radio station), which aimed to counter the effects of hate speech before the second round of elections (Vollhardt et al, 2006).

An evaluation of Search for Common Ground’s project *Supporting a Conversation on Youth Leadership in Côte d'Ivoire* found that the project raised awareness about political violence and manipulation (Gouley and Kanyatsi, 2010, p. 5). The radio programme component of the project was found to have played an important role in promoting and depoliticising dialogue among diverse youth groups (Gouley and Kanyatsi, 2010, p. 5). The evaluation was based on a document review, a workshop, focus group discussions and surveys (Gouley and Kanyatsi, 2010).

Text messages

In Ambon, Indonesia an inter-faith group called the Peace Provocateurs used text messages and social media to counter messages inciting violence.[4] Whenever rumours of violent incidents began circulating, volunteers were immediately sent to verify the facts, and text messages and messages on social media were circulated to set the record straight, with the aim of preventing revenge attacks. The group is believed to have played a key role in containing violence in Ambon in September 2011.[5]

[2] http://www.labenevolencija.org/la-benevolencija/mission-and-vision/

[3] See also: http://www.labenevolencija.org/rwanda/la-benevolencija-in-rwanda/

[4] http://www.independent.co.uk/news/world/asia/how-peaceprovocateurs-are-defusing-religious-tensions-in-indonesia-7562725.html

[5] http://www.independent.co.uk/news/world/asia/how-peaceprovocateurs-are-defusing-religious-tensions-in-indonesia-7562725.html

A similar approach was used in Kenya in the run-up to the 2013 elections. The *Nipe Ukweli* (Give me truth) campaign educated citizens about the meaning of hate speech and dangerous speech and how to deal with them. *Nipe Ukweli's* message was publicised via Twitter and Facebook, as well as via community radio and word of mouth. Community forums were also held in some of the areas most affected by post-election violence in 2007. Feedback from those targeted suggests that the campaign was well-received (Benesch, 2013, p. 15). However, it is not clear to what extent the intervention reduced hate speech.

Advocating for the removal of hate speech from the internet

In some countries NGOs have succeeded in persuading online authors and internet service providers to remove websites dedicated to hate speech. This approach has been particularly successful in countries with hate speech laws, such as the Netherlands (OSCE, 2009, p. 57). For example, the Dutch Magenta Foundation's Complaints Bureau for Discrimination on the Internet has succeeded in removing thousands of examples of hate speech from the Internet since 1997, by pointing out the illegality of hate speech to authors and owners of sites containing such material. This approach was successful in 95 per cent of cases (OSCE, 2009, p. 57).

Monitoring hate speech

In Kenya in 2012-2013, the Umati project[6] produced a database of hate speech and dangerous speech by monitoring Kenya's online spaces. Findings were classified according to their 'dangerousness' in accordance with the Dangerous Speech Guidelines produced by Susan Benesch.[7] This type of intervention often serves as the foundation for other types of intervention to counter hate speech. For example, Umati forwarded any examples of extremely dangerous speech, as well as any calls for action to Uchaguzi,[8] a multi-stakeholder initiative that enables citizens to report and monitor election-related events on the ground (Umati, 2013, p. 10). An evaluation of the Umati and Uchaguzi projects was undertaken. It was based on interviews with 35 people, and found that reports to Uchaguzi tended to result in an increase in security personnel in areas mentioned in the report (Oddsdóttir, 2014, p. 10). However, information about the impact of the projects is limited and remains largely anecdotal (Oddsdóttir, 2014, p. 10).

Support for the creation of self-regulatory media systems

In Iraq, USIP (United States Institute of Peace) and its partners helped create a locally driven self-regulatory media system to prevent media incitement to violence. Taylor and Dolan (2013, p. 11) describe the five step process involved in establishing this system:

1) Iraqi journalists identified and defined terms that had the potential to incite violence in the run-up to the country's 2010 elections. The list of terms was distributed to journalists and editors before the elections to help avoid inflammatory reporting.
2) USIP undertook a content analysis of the 2010 election coverage to identify the prevalence, intensity, and location of the terms identified in stage one.
3) USIP shared the content analysis with Iraqi media stakeholders. A small group of them then learned how to undertake their own conflict analyses. During this stage Iraqi media stakeholders

[6] See: http://www.ihub.co.ke/umati

[7] See: http://voicesthatpoison.org/

[8] See: https://uchaguzi.co.ke/main

also produced a style guide for conflict reporting, which serves as a practical framework to minimise the use of inflammatory language in reporting.

4) The number of organisations implementing the self-regulatory tools was increased via a range of capacity building activities, such as locally driven workshops to introduce the content analysis and style guide to additional media organisations and CSOs (Taylor and Dolan, 2013, p. 11).
5) The emergent Adaa' Media Monitoring network was to build on the previous stages and to increase civil society oversight of the media (Taylor and Dolan, 2013, p. 11). This final stage was not realised to its full potential due to funding issues (Expert comment).

An impact evaluation of this project has not been undertaken, so it is difficult to gauge to what extent the project was successful.

Self-regulation or self-censorship can also be viewed as negative. For example, some have criticised the Kenyan media for too much self-censorship prior to, during, and after the country's 2013 general election in its efforts to avoid inciting violence (Muriithi and Page, 2013, p. 18). It was felt that by undertaking self-censorship the media had not fulfilled its responsibility to act as a watchdog and to make the public aware of leaders' wrongdoings and failures (Muriithi and Page, 2013, p. 20). Due to the number of interventions to prevent the incitement of violence at the time of the Kenyan elections, it is difficult to ascertain whether this self-censorship had an impact on the peacefulness of the elections.

Youth conferences

In 2013, USAID held a five day Future Leaders Conference in Galle, Sri Lanka to Stand Up Against Hate Speech (USAID, 2013, p. 1). 518 students from all of Sri Lanka's districts and ethnic and religious groups participated in the conference. The conference appears to have been viewed as a success by participants, but there is no evidence on its wider or long-term impact.

4. References

Bajraktari, Y. and Hsu, E. (2007). *Developing Media in Stabilisation and Reconstruction Operations* (Special Report). Washington D. C.: United States Institute of Peace. http://www.usip.org/sites/default/files/srs7.pdf

Burgess, J. (2013). *Media literacy 2.0: A sampling of programs around the world.* Washington D. C.: Center for International Media Assistance, National Endowment for Democracy. http://cima.ned.org/sites/default/files/CIMA-Media%20Literacy%202_0-%2011-21-2013.pdf

Benesch, S. (2014). *Countering dangerous speech to prevent mass violence during Kenya's 2013 elections*. http://voicesthatpoison.org/kenya-2013/

Gouley, C. and Kanyatsi, Q. (2010). *Final evaluation of the project "Supporting a Conversation on Youth Leadership in Côte d'Ivoire*." http://www.sfcg.org/programmes/ilt/evaluations/CIV_EV_Aug10_Final%20Evaluation%20Report%20-%20Supporting%20a%20Conversation%20with%20Youth%20on%20Leadership%20%282%29.pdf

Muriithi, A. G. and Page, G. (2013). *The Kenyan election 2013: the role of the factual discussion programme Sema Kenya (Kenya speaks)* (Working Paper No. 5). BBC Media Action. http://downloads.bbc.co.uk/mediaaction/pdf/kenya_election_2013_working_paper.pdf

Oddsdóttir, F. (2014). *Peaceful or violent? Online hate speech during Kenya's general elections 2013* (Conference Paper). https://www.academia.edu/7396058/Peaceful_or_Violent_Online_hate_speech_during_Kenyas_general_elections_2013

OSCE/ODIHR. (2009). *Preventing and responding to hate crimes: A resource guide for NGOs in the OSCE region*. Warsaw: OSCE/ODIHR. http://www.osce.org/odihr/39821?download=true

Paluck, E. L. and Green, D. P. (2009). Deference, dissent, and dispute resolution: An experimental intervention using mass media to change norms and behaviour in Rwanda. *American Political Science Review,* 103:4. http://isps.yale.edu/sites/default/files/publication/2012/12/ISPS09-024.pdf

Rao, S. (2014). *Sending messages to reduce violent conflict* (GSDRC Helpdesk Research Report 1050). Birmingham, UK: GSDRC, University of Birmingham. http://www.gsdrc.org/go/display&type=Helpdesk&id=1050

Rhodes, A. (2007). *Ten years of media support to the Balkans: An assessment*. Media Task Force of the Stability Pact for South Eastern Europe. http://www.medienhilfe.ch/fileadmin/media/images/dossier/mediasupport_Balkan.pdf

Skuse, A., Rodger, D., Power, G., Mbus, D. F., and Brimacombe, T. (2013). *Communication for development interventions in fragile states: A systematic review*. Adelaide: University of Adelaide. https://www.adelaide.edu.au/accru/projects/c4dfragilestates/SR566final.pdf

Taylor, M. and Dolan, T. (2013). *Mitigating media incitement to violence in Iraq: A locally driven approach* (Special Report). Washington D.C.: United States Institute of Peace. http://www.usip.org/sites/default/files/SR329-Mitigating-Media-Incitement-to-Violence-in-Iraq.pdf

Umati. (2013). *Umati: Monitoring dangerous speech online – October 2012-January 2013.* http://www.ihub.co.ke/uploads/default/files/umati/Umati_Report_Oct-Jan_2013.pdf

USAID Sri Lanka. (2013). *Helping youth leaders counter hate* (Snapshot). http://www.usaid.gov/results-data/success-stories/helping-youth-leaders-counter-hate-sri-lanka

Vollhardt, J., Coutin, M., Staub, E., Weiss, G. and Deflander, J. (2006). Deconstructing Hate Speech in the DRC: A Psychological Media Sensitization Campaign. *Journal of Hate Studies*, 5:1, 15-35. http://journals.gonzaga.edu/index.php/johs/article/view/74

Key websites

- Article 19:
  http://www.article19.org/
- Voices that Poison
  http://voicesthatpoison.org/

Expert contributors

Susan Benesch, American University
Theo Dolan, United States Institute of Peace
Nabila Ghanea-Hercock, Kellogg College, University of Oxford
Mark Lattimer, Minority Rights Group
Freyja Oddsdóttir, GSDRC
Shane Perkinson, USAID

Suggested citation

Strachan, A.L. (2014). *Interventions to counter hate speech* (GSDRC Helpdesk Research Report 1116). Birmingham, UK: GSDRC, University of Birmingham.

About this report

This report is based on three days of desk-based research. It was prepared for the UK Government's Department for International Development, © DFID Crown Copyright 2014. This report is licensed under the Open Government Licence (www.nationalarchives.gov.uk/doc/open-government-licence). The views expressed in this report are those of the author, and do not necessarily reflect the opinions of GSDRC, its partner agencies or DFID.

The GSDRC Research Helpdesk provides rapid syntheses of key literature and of expert thinking in response to specific questions on governance, social development, humanitarian and conflict issues. Its concise reports draw on a selection of the best recent literature available and on input from international experts. Each GSDRC Helpdesk Research Report is peer-reviewed by a member of the GSDRC team. Search over 400 reports at www.gsdrc.org/go/research-helpdesk. Contact: helpdesk@gsdrc.org.