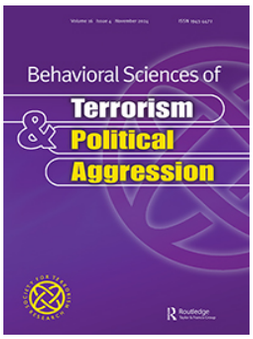

Routledge
Taylor & Francis Group

Behavioral Sciences of Terrorism and Political Aggression

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rirt20

A systematic integrative review of counter-messaging communication campaigns targeting terrorism or violent extremism

Samantha Treacy, Alastair Reed & Andrew Glazzard

To cite this article: Samantha Treacy, Alastair Reed & Andrew Glazzard (07 Nov 2024): A systematic integrative review of counter-messaging communication campaigns targeting terrorism or violent extremism, Behavioral Sciences of Terrorism and Political Aggression, DOI: 10.1080/19434472.2024.2419390

To link to this article: https://doi.org/10.1080/19434472.2024.2419390

BEHAVIORAL SCIENCES OF TERRORISM AND POLITICAL AGGRESSION
https://doi.org/10.1080/19434472.2024.2419390

Routledge
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

A systematic integrative review of counter-messaging communication campaigns targeting terrorism or violent extremism

Samantha Treacy[a], Alastair Reed[b] and Andrew Glazzard[c]

[a]Department of Criminology, Sociology & Social Policy, Swansea University, Swansea, Wales, UK; [b]Hillary Rodham Clinton School of Law, Swansea University, Swansea, Wales, UK; [c]Centre for Peace and Security, Coventry University, Coventry, England, UK

ABSTRACT

In response to the rise of the so-called Islamic State and its effective use of propaganda to recruit and radicalise followers, countering the communication strategies of a range of terrorist and violent extremist actors has become a significant challenge for policymakers and practitioners. Over the past decade, counter-messaging campaigns have increasingly played a central role within Countering Violent Extremism (CVE) programming. However, evidence supporting the effectiveness of these campaigns remains limited. While previous systematic reviews have focused on empirical and theoretical contributions, this paper makes a valuable contribution by conducting the first systematic review of evaluations of 'live' CVE counter-messaging campaigns. The study provides important insights into the effectiveness of current campaigns and offers recommendations for improving future strategies.

ARTICLE HISTORY
Received 14 August 2024
Accepted 17 October 2024

KEYWORDS
Strategic communications; counter-messaging; terrorism; violent extremism

Introduction

In the wake of the rise of the so-called Islamic state and their seemingly effective application of propaganda to recruit and radicalise followers, countering the communication strategies of a range of terrorist and violent extremist actors has become a major challenge for policy makers and practitioners. One key response has been the use of communication campaigns to counter or prevent violent extremism, which has been widely promoted by national governments, international organisations and multilateral forums, civil society organisations and the communications technology industry, backed up by significant financial investment (Briggs & Feve, 2013; Reed et al., 2017). Often designated as 'counter-messaging', 'counter-narrative' or 'alternative narrative' programmes, P/CVE communications interventions typically seek to challenge, undermine or crowd out violent extremist propaganda (counter-narrative), and sometimes seek to

CONTACT Alastair Reed alastair.reed@swansea.ac.uk
Supplemental data for this article can be accessed online at https://doi.org/10.1080/19434472.2024.2419390.

provide non-violent, pro-social content in its place (alternative narrative) (Braddock & Horgan, 2016; Briggs & Feve, 2013.)

Over the past decade counter-messaging campaigns have become a central plank of CVE programming globally, with support from policy makers and practitioners for further funding and implementation (Reed et al., 2017). However, the evidence base for such campaigns remains limited (Ferguson, 2016). This paper adds to the insights gained by a previous systematic review of counter-narratives by Carthy et al. (2020), which found little evidence of their effectiveness in preventing violent radicalisation. However, the studies under review were largely theoretical and experimental contributions, and it is the aim of this review to address the knowledge gap by focusing on 'live' CVE counter-messaging campaigns, or those that were intended to be. Through reviewing all publicly available campaign evaluations that we were able to identify, this study will provide an overview of the types of campaigns implemented, how they were evaluated, and what we can learn from the findings of these evaluations. This review is as far as we are aware, the first review to address 'live' CVE counter-messaging campaigns, and as such makes a unique contribution to the field.

Research questions

To meet the aims of this review, the following questions were formulated:

(1) What types of communications campaigns have been designed and developed that make use of counter-narratives or alternative narratives in countering or preventing violent extremism?
(2) What counter-narratives or alternative narratives were described by the research?
(3) How were the communications campaigns evaluated and what were the findings of these evaluations?
(4) What was the quality of the research conducted?

Methods

It was clear that there were a range of communications campaigns being developed by a variety of state and non-state actors, using counter-messaging narratives. However, it was unclear how much of this had been evaluated, and how much would appear in the peer-reviewed literature. Therefore, it was decided that a systematic integrative review methodology was the most appropriate to use. This approach accommodates a wide variety of study methods, giving a 'broader, more inclusive view of a topic' than a traditional systematic review (Lubbe et al., 2020, p. 308), but also applies a systematic means of conducting the literature review to ensure rigour.

The particular methodology adopted for this review was an adaptation of Whittemore and Knafl's (2005) updated integrative review method, with the addition of a validated quality appraisal tool to facilitate data evaluation – MetaQAT (Rosella et al., 2016). Reporting of the review also adhered to PRISMA systematic review guidelines (Page et al., 2021), including completion of the PRISMA 2020 checklist, see Appendix 1. Using this combination of approaches, the following four steps were followed during the course of this review:

Table 1. Review search terms.

Intervention	Field of study
counter-narrative* OR counter-messag* OR counter fram* OR alternative narrative* OR alternative messag* OR alternative fram* OR strategic communication* OR anti-messag* OR anti-radicalis* OR contested narrative* OR disrupt*	terror* OR counter-terror* OR CT OR anti-terror* OR extremis* OR countering violent extremism OR CVE OR PVE OR P/CVE OR P-CVE OR radicalization OR radicalisation or de-radicalisation OR de-radicalization
Additional for updated search	
Communication* campaign OR awareness campaign OR advertising campaign OR propaganda OR communication* intervention OR information campaign	

Literature search

The review search strategy was devised by the research team, and further refined by a series of pilot searches. This was followed by systematic searching of nine electronic databases related to the social sciences and criminal justice, which were: Applied Social Sciences Index and Abstracts, Campbell Collaboration, Criminal Justice Database, Digital National Security Archive, European Sources Online, International Bibliography of Social Sciences, Proquest Central, Scopus and Web of Science. Searches were completed on all databases by 15th May 2023. It is of note that the searches took place at two main time points, the first being 29th October 2019, and the second search was completed in two parts in May 2023 (5th and 15th), which updated the first search, but also included some additional search terms, as the research questions were refined. The search terms were split into two categories: type of intervention (original search and additional one) and field of study, as presented in Table 1.

Whilst no filters were applied to seven of the databases searched, there were filters applied to both ProQuest Central and Web of Science, which returned a very large number of articles initially. These were then filtered by 'language' (English), 'article type' for ProQuest (excluding article types that were excluded by our inclusion criteria, such as newspapers, magazines and podcasts), and 'topic area' for Web of Science (excluding science, engineering and unrelated social science topics such as radiology, material sciences and archaeology). Appendices 2a, 2b, and 2c give an example of the search strategy at the different time points, and Appendix 3 has the list of filters applied to each database.

In addition to the electronic database search, an extensive hand search strategy was also employed which involved reference mining, internet searches with Google and Google Scholar and searching the websites of 62 organisations and institutions, including governments, academic departments, think-tanks, 3rd sector organisations, and social media companies and groups. These covered the full range of publications up to June 2023, and the specific websites searched are listed in Appendix 4.

Data evaluation

Papers returned by the search were screened by title, abstract and then full-text by two independent reviewers. These were considered for inclusion in the review if they met the following criteria:

(a) Involved a 'live' communication campaign, or aspects of one, run by any group. For the purposes of this review, a communication campaign was defined as: 'purposive

attempts to inform or influence behaviors in large audiences … using an organised set of communication activities and featuring an array of mediated messages in multiple channels generally to produce non-commercial benefits to individuals and society' (Atkin & Rice, 2013, p. 3). In this context, 'live' means that the campaign has been run, or has appeared, in or on various forms of media or social media;

(b) Is focused on preventing or countering terrorism or violent extremism – by any terrorist or violent extremist group;
(c) The campaign explicitly uses counter-narratives or alternative narratives to those posited by the targeted terrorist or violent extremist group;
(d) Any article, paper or report that includes a description of some form of evaluation of the communication campaign, or aspects of a campaign, with an identifiable methods section;
(e) That it is published in English.

Articles and reports were therefore included if they detailed an evaluation of a live communication campaign which used counter-narratives and alternative narratives in targeting the narratives promoted by any terrorist or violent extremist group.

Papers were excluded from the review if they: (i) involved counter-messaging, counter-narratives or alternative narratives which were only devised for use in a study, which had not been run or posted live nor was explicitly intended to be; (ii) targeted non-violent groups or hate speech more broadly; (iii) detailed educational, deradicalisation or 'take-down' approaches or one-to-one interactive messaging only; (iv) did not explicitly evaluate a communication campaign but nevertheless provided recommendations for them in the form of toolkits, handbooks or guidance documents; (v) papers which presented an analysis of a campaign, but not an evaluation, such as a discourse analysis of an advert; (vi) papers which detailed narrative or counter-narrative theory, rather than evaluating a live campaign; and (vii) newspapers, magazines, blogs, podcasts, books, working papers.

Initial extraction of data used a structured summary approach, following Tricco et al. (2018), which described the aims, campaign description, methods, results, and conclusions for each paper. This was combined with the ACME (Audience, Channel, Message and Evaluation) framework (Noar, 2012) to guide data extraction regarding the communication campaign description, and was folded into a standardised extraction form based on the Population, Intervention, Comparator and Outcomes (PICO) formula (Richardson et al., 1995). The data extracted were: author; article data; campaign: country, name, developer, goals, activities, theory, messages, channels, target audience; evaluation: aim, design, sample size and description, outcome measures, intervention group, control group, main outcomes, conclusions, recommendations and limitations. Half of the studies were extracted by two reviewers to check for consistency.

Quality appraisal

The search returned papers that used a variety of methodologies. A validated tool designed to evaluate the quality of studies with a range of methods was considered to be the most appropriate way to allow for both an assessment, and comparison, of quality across the studies. The Meta Quality Appraisal Tool (MetaQAT) was selected for

this purpose (Rosella et al., 2016), and the Revised Cochrane risk-of-bias tool for randomised trials (RoB 2) (Higgins et al., 2019) was also incorporated.

As the MetaQAT does not have a particular categorisation system, the GRADE criteria (Ryan & Hill, 2016) was adapted, and High, Moderate and Low ratings were applied to each study to indicate whether there were no concerns, serious concerns, or very serious concerns in each quality appraisal domain: relevancy, reliability, validity and applicability. The overall quality score was an average of each of the domain scores. Given the inclusive and exploratory nature of this review, no papers were excluded from the review on the basis of quality alone.

Data analysis

It was not possible to conduct a meta-analysis to synthesise the findings of this review, given the variety of methodologies used in the evaluations, and the type of data generated (Harden & Thomas, 2005). Instead, a narrative synthesis approach was undertaken which, together with the data extraction tool employed, enabled each research question to be addressed (Popay et al., 2006).

Results

Thirty reports met the inclusion criteria for the review, representing twenty-five communication campaigns – seven of the studies emanated from the Breaking the ISIS Brand counter-narratives project. Despite there being no date restriction on the search, the studies meeting the inclusion criteria for this review, are from the last decade only, 2013–2023. Figure 1 summarises the stages of the search and screening process in PRISMA format below.

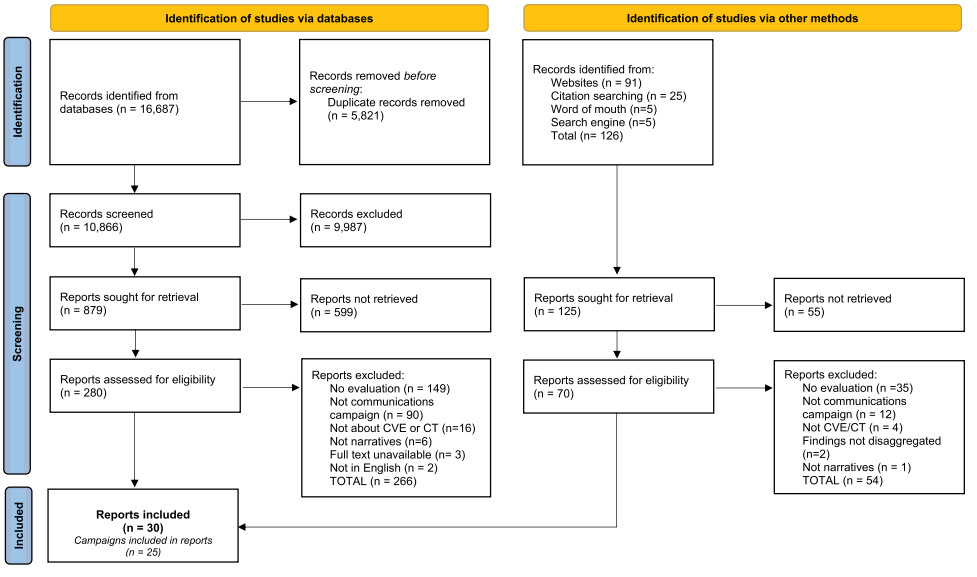

Figure 1. PRISMA 2020 flow diagram depicting the stages of the search and screening process undertaken. From: Page et al. (2021).

A description of the included papers, and of the twenty-five communication campaigns they evaluated, are presented in Table 2, which will be used to describe campaign characteristics, and to address the first two of the review's research questions. This will be followed by the presentation of Tables 3 and 4, and consideration of the third and fourth research questions.

The communication campaigns predominantly targeted Islamist or jihadist extremism ($n = 21$), with two at the far-right, one at extremism generally, and one which targeted both Islamist and far-right extremism. The studies were conducted on campaigns that were based or run in North America ($n = 9$, seven situated in the USA, and three in Canada); nine in Europe (including four in the UK); seven in countries of the Middle East and North Africa (MENA) region; five in Asian countries (with three in Indonesia); five in sub-Saharan African countries, mostly in West Africa; two in Australia, and one in Trinidad and Tobago. It is of note that a number of the campaigns were run in multiple countries. There was no clear funding or sponsor information for eight of the campaigns. For the rest, twelve were government- or EU-funded (mostly by the US government, $n = 7$, although there were others involved in funding some campaigns), three by non-profit organisations (one of which stated that they were also funded by state grants), and two by Facebook and Edventure's Peer-to-Peer challenge, one of which was also partly funded by the US government.

The audience targeted by the campaigns mostly involved young people ($n = 14$), including students, from as young as 13 or 14 years old (Search for Common Ground, 2020; SecDev Foundation, 2016; Silverman et al., 2016; Speckhard et al., 2018a); described as potentially 'at risk' of either exposure to extremist content or recruitment to an extremist cause (Al-Rawi, 2013; Ali et al., 2020; Ihidero, 2019; Monaci, 2020; Search for Common Ground, 2020; Speckhard, Ellenberg, Shaghati, et al., 2020; Voogt, 2017), or sympathetic to violent extremism (Ali et al., 2020; Speckhard et al., 2018a). Others were targeted at broader swathes of selected communities: such as Sunni Muslims (Al-Rawi, 2013; Bélanger et al., 2023), and Somali-Americans (Silverman et al., 2016; Speckhard et al., 2019). Two were targeted at the network around potentially 'at risk' people – family and friends (Al-Rawi, 2013), described as the halo audience in one study (SecDev Foundation, 2016). The target audience was not stated for eight campaigns.

Review question 1: what types of campaign were included?

Most of the campaigns ($n = 19$) presented goals which sought to challenge violent extremism and its associated ideology (Al-Rawi, 2013; Aly et al., 2014; Bean & Edgar, 2017; Bélanger et al., 2023; Bilali, 2022; Bodine-Baron et al., 2020; El Damanhoury, 2020, 2023; Ihidero, 2019; Marrone et al., 2020; McDowell-Smith et al., 2017; Monaci, 2020; Octavia & Wahyuni, 2014; Schmitt et al., 2018; Search for Common Ground, 2020; SecDevFoundation, 2016; Silverman et al., 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, 2019, Speckhard, Ellenberg, Shaghati, et al. 2020, Speckhard, Ellenberg, & Ahmed, 2020; Voogt, 2017; Wilner & Rigato, 2017). For some campaigns this challenge also included the presentation of 'real' Islam as a corrective to that presented in extremist messaging, exposing the workings of extremist recruitment strategies, offering a space and support for disengagement from groups including highlighting the negative impact of group membership, and seeking to educate the audience regarding critical

Table 2. Summary of included campaigns.

Study No	Author, Year, Country	Campaign name, targeted extremists; developers	Theory	Goals	Activities	Messages	Channels	Audience
(i) Social media campaigns (full or part)								
1	Ali et al., 2020, Austria	Jamal al-Khatib – My Path!; jihadist or neo-Salafist; social workers, Islamic experts and ex-members, part funded by government and EU	None stated	To develop and disseminate alternatives to jihadist messages using similar high production values	Videos based on biographies; 'online streetwork' – start discussions publicly & one-to-one	Alternative messages: deconstruct us v them narratives; denounce racism; serious issues can be addressed without violence	Social media – Facebook	Adolescents at risk of exposure; jihadist sympathisers
2	Al-Rawi, 2013; Iraq; Saudi Arabia & MENA region	Terror has no Religion, Islamist; unknown sponsors	None stated	Challenge the threat of Al-Qaeda, extremist ideology, & sectarianism; to present the real (more moderate) Islam	Video and print adverts and websites; used shocking images such as a suicide bombs and child victims, and kidnapping	Ideology leads to terrorism; terrorist distortion of Qu'ran; extremists invite criticism of Muslims; pro-peace, anti-violence; terrorist as criminals to be reported; Iraqi unity	State-run TV and newspapers, website	Seems Sunni Salafists; potential 'terrorists', family & neighbours
2 & 3	2. & Aly et al., 2014; population of MENA region	Say No to Terror, Islamist; unknown source	None stated	To uphold and defend true Islam; to expose how terrorists function, challenge the threat	>20 short videos and posters	Counter-narrative: terrorism as crime; terrorists as enemies – kill Muslims, brain-wash children; leaders liars; protect against terror, not support it	Website, Saudi-owned TV & social media (Face-book, You-Tube, Twitter)	Mostly Saudi society; Muslim Arabic
4	Bean & Edgar, 2017, USA	None, Islamist; US State Department – Center for Strategic Counterterrorism Communications.	None stated	To counter extremist discourse	Production of videos.	Not emphasising positives about US, but negatives of the enemy – offence not defence	Social media – YouTube	Not stated
5	Bélanger et al., 2023; Tunisia, Jor-dan, Egypt, Saudi Arabia	None, Islamist, Sawab Center – funded by US and UAE governments	None stated	To challenge Islamic extremism, and the religious misinterpretations they promote	Social media messaging	Message 1: 'school' – hope, ISIS opposition to education and prosperity; Message 2: 'Young Mind' – extremism affecting young minds,	Social media – Facebook, Twitter, YouTube	Sunni Muslims

(*Continued*)

Table 2. Continued.

Study No	Author, Year, Country	Campaign name, targeted extremists; developers	Theory	Goals	Activities	Messages	Channels	Audience
						importance of family values		
6	Berman, 2019, UK	None, Islamist, Quilliam, unclear	None stated	Not stated	Social media messaging	The destructiveness of terrorism; promotion of moderate religious practice	Social Media – Facebook	Not stated
7	Bodine-Baron et al., 2020; Indonesia	'#AkuTemanmu ('I am your friend'); #CapekGakSih ('Aren't you tired?'); Islamist; Search for Common Ground, funded by US government	A logic model – measurable goals.	To increase awareness of fake news, support community diversity, freedom of speech and religion; decrease support for violent solutions to problems	Videos and images with hashtags as branding. Formative research with focus groups, interviews. Use of personal stories	Encounters with different people leads to growth; unity and middle ground can end divisiveness	Social media – Facebook, Twitter, YouTube	18–35 years old; Indonesian, active on social media
8	Effendi et al., 2022, Indonesia	None, Islamist, Duta Damai – national youth community, established by government	None stated	Development of counter-radicalisation narratives to advocate for peace	Asked people to send in writing; infographics and visuals; offline activities (visits to schools) promoted online activity	Promotion of tolerance and diversity; awareness raising on radicalisation and how to respond; nationalism; news; community activity	Social media – Instagram	Youth – age range not specified
9 & 10	El Damanhoury, 2020, 2023; Egypt, Emirates	al-Siham al-Marika ('The Piercing Arrows'); Islamist, televangelist on al-Nahar TV station; unclear	Social cognitive/ identity theory	Entertainment-education – prosocial messages or anti-extremist narratives to address radicalisation	Drama series about life under ISIS control.	Positive role models reject extremism; negative role models are ISIS (manipulate religion to justify crimes); transitional ones: ISIS that change over time. Over-arching spiritual theme	TV; YouTube channel	Not stated
11	Ihidero, 2019; Nigeria	Speak Out, Act Right (SOAR) project, Islamist, developed by universities an NGO, part of P2P (Facebook, EdVenture & US state)	Public health approach	To challenge extremist narratives and educate; address drivers of extremism: bullying, gender extremism & hate speech	Use of credible community voices to deliver messages; short films, concert, theatre, sports/ games, debate competition, lectures, vox pops	Undermine extremist lure; promote peace & humanity; present warning signs and pathways to extremism; raise awareness	University campus, Facebook	At-risk students

(*Continued*)

Table 2. Continued.

Study No	Author, Year, Country	Campaign name, targeted extremists; developers	Theory	Goals	Activities	Messages	Channels	Audience
12–18	(i) McDowell-Smithet al., 2017, USA; (ii) Speckhard et al., 2018a, USA, UK, Canada, Australia; (iii) 2018b, Iraq; (iv) 2019, USA; (v) Speckhard & Ellenberg, 2020, 8 EU coun-tries:, (vi) 2020, Europe, Asia, Middle East, Africa, Trinidad & Tobago; (vii) Speckhard, Ellenberg, & Ahmed, 2020, USA	Breaking the ISIS Brand Counter-Narratives Project; Islamist; developed by the International Centre for the Study of Violent Extremism; no funder information	(i), (iv), (v), (vi): Braddock & Horgan's guides (2015/6) & Davies et al. (2016)	To de-legitimise ISIS, to refute, and provoke questioning, of their narratives, and to seek help	Nineteen videos used across these studies of ex-ISIS members, using ISIS imagery, labelled with pro-ISIS names, created by Hollywood-documentary film producers. Study (v) Facebook inserted their videos in feeds; (vi): videos 'hyper-targeted' at specific profiles with Facebook. Interviews edited to their most 'damaging, denouncing, and derisive content' (Study iii, p. 51)	Counter-narratives. Stories of traumatic experiences to try to dissuade joining ISIS; ISIS hypocrisy, brutality, corruption, inescapability, criminal, present twisted version of Islam, exploiting vulnerability. Terrorists humanised, complex, not us v them binary. Multiple versions of Islam, and 'good' Muslim criteria. No call to action, directed to campaign website; the futility of ISIS promises regarding Caliphate, fails to deliver ultimately	Facebook primarily; YouTube and campaign website. Unclear for study (vii). (i)&(iii) focus groups …	(i) unclear, but not the sample in the study; (ii) young, disillusioned Muslims interested in extremist groups, aged 13+ or 18+ years, (iii): 18–50 years in 10 areas; (iv) Somali-Americans; (v) Males, 18–24 years; (vi): at risk of exposure to ISIS content – mostly male, aged 18–34 years; Study (vii): Unclear
19	Monaci, 2020, France, Italy and Romania	#heartof darkness (#hod); Islamist; EU funded, SAFFRON (private ICT company, military academy & a university)	None stated	To develop alternative narratives; to raise awareness of the motivations ISIS exploit in recruiting	Social media campaign with seven short videos	Violence motivation: society, discrimination, media; peer-expert perspectives; alternative narratives: democracy, integrated, friendly, free expression	Vimeo for videos; Twitter, Facebook, YouTube and website	'at-risk' Muslims: converts, immigrants (2nd/3rd gen), refugees
20	Schmitt et al., 2018, Germany	#WhatIS, Islamist, German government	None stated	counter 'distorted' views of Muslims in Germany	Eight videos with popular German YouTubers	Explains extremism terms – caliphate, haram, jihad	YouTube	Unclear

(*Continued*)

Table 2. Continued.

Study No	Author, Year, Country	Campaign name, targeted extremists; developers	Theory	Goals	Activities	Messages	Channels	Audience
20 & 23	20 & Silverman et al., 2016; USA	ExitUSA, Far-Right, non-profit organisation 'Life after Hate'	None stated	Discredit, discourage recruitment, help to leave or disengage	Four videos using personal stories; run offline exit programme	Discredit groups, inculcate doubt, impact of violence, forgiveness, promote exit	20/23: YouTube; 23: Facebook, Twitter	20: Unclear; 23: 13–60+ across US
21	Search for Common Ground, 2021; Kyrgyzstan	Taasirlink: Citizen Narrative Campaign; Islamist; Search for Common Ground – unclear funding	Positive deviance; Theory of change – not detailed	To increase resilience of 'at-risk' youth; disseminate campaign messages; change knowledge, minds and practice; create a positive environment	Five online campaigns using videos, and linked offline events (training, contests, excursions, sports) using influencers	Indirect messaging, rather than directly addressing radicalisation. Focused on gender empowerment, critical thinking skills, fact-checking, social cohesion, proactiveness, and belonging	Mix of Instagram, YouTube, Facebook, Telegram, WhatsApp	14–36 years, at-risk in three target areas, but ultimately nationwide
22	SecDev Foundation, 2016, Canada	Extreme Dialogue; Islamist & Far-Right; Institute of Strategic Dialogue, Tim Parry Johnathan Ball Foundation; SecDev Foundation. Funded by Public Safety Canada	Target Audience Analysis	To reduce appeal of extremism and offer an alternative: 10% views from target audience; wide sharing of content; positive media coverage; evaluation	Educational, awareness-raising, short documentary films with personal stories disseminated online using paid and unpaid methods; engage media	Tells stories of extremism-affected Canadians to counter extremist messaging – including family of deceased ISIS members, and an ex-far-right extremist	Website, Facebook, Twitter & YouTube, Offline media.	Core: 14–24 years; Halo: teachers & parents. Canada-wide, but four areas specifically
23	Silverman et al., 2016; USA & UK	Average Mohamed, Islamist; non-profit	None stated	Increase capacity of campaigns and encourage critical thinking	Five videos, used animation; have an offline component not clearly described	Alternative narratives, democracy, promote peace & multi-culturalism, challenge ideology; discourage recruitment, anti-slavery;	Facebook, Twitter and YouTube	Somali-Americans 14–25 years, in US & UK
23	Silverman et al., 2016; Pakistan & UK	Harakat-ut-Taleem, Islamist; Institute for Strategic Dialogue, Jigsaw-Google, Twitter & Facebook, unclear funder	None stated	Counter recruitment	Six videos with personal stories – some pre-existing material, used a YouTube creative to edit videos to appeal more	Negative impact of joining; importance of education, peace, repentance & redemption; Taliban-criminal	Facebook, Twitter and YouTube	14–25 years, locations include rural Pakistan

(*Continued*)

Table 2. Continued.

Study No	Author, Year, Country	Campaign name, targeted extremists; developers	Theory	Goals	Activities	Messages	Channels	Audience
24	Voogt, 2017; Australia	Community Action for Preventing Extremism (CAPE), Far-Right; All Together Now, not-for-profit organisation, funded by state grants	Informed by the 'Exit' disengagement approach	To challenge extremist ideology, prevent recruitment & engagement; encourage critical thinking, and expose downsides	Google advertising for high search entry; website & discussion forum; adverts, direct engagement via social media and tours with ex-members	'Evidence-based' counter-narratives: debunk ideology & misinformation with facts; challenge sense of belonging and esteem with far-right; hypocritical leaders; negative impact on life & employment	Website, Google advertising, forum on Facebook	At-risk individuals, searching for far-right keywords in Google
25	Wilner & Rigato, 2017; Canada	60 Days of PVE, general extremism, publicly funded; developed by graduate students as part of the Facebook Global Digital	Public health: including Braddock & Horgan	Preventing violent extremism; fact-based counter-extremist content, use credible local and national people including former extremists	Online peer-to-peer campaign – brief Facebook posts & infographics; pre-campaign focus groups on material & strategy; professional logos	Assumptions; health; recruit & report; media literacy; circumstances of terrorists; gender & sexuality; right-wing nationalism; radical thought-action; deradicalisation; inclusion	Facebook	16–25 years, male and female
(ii) Radio and other campaign types								
26	Aldrich, 2014, Mali	Trans-Sahara Counter Terrorism Partnership, with Radio for Peace Building Northern Mali program; Islamist; Funded by USAID, with a range of local programmers	None stated	Strengthening 'economic and social resilience of local communities' (p. 526) to recruitment to violent extremist groups, and civic participation	Building of two new radio stations, training and technical assistance. Radio programming focused on peace and tolerance, across five years	Pro-peace, anti-Al-Qaida, pro-US, pro-civic engagement	Radio	Town in Northern Mali. No further targeting stated
27	Bilali, 2022, Burkina Faso	'Voices for Peace' project; Islamist, Equal Access International (funded by USAID)	None stated	To counter violent extremism by influencing worldviews and storytelling	A fictional radio drama (52 episodes)	Countering extremism – negative community consequences, need for community-police relationships; acknowledge police corruption	Radio	Not stated
28			Theory of change				Radio	

(*Continued*)

Table 2. Continued.

Study No	Author, Year, Country	Campaign name, targeted extremists; developers	Theory	Goals	Activities	Messages	Channels	Audience
	Marrone et al., 2020; Nigeria	Ina Mafita (The Way Forward); Islamist, Equal Access International, funded by USAID		Portfolio of shows countering violent extremism	Radio programme aired every week for a year, with pre-recorded stories and a live call-in show to discuss them	1. Role model and local committee roles; 2. Challenge stigma of kidnap victims – presented as innocent, traumatised		18–35 year olds in 13 states in North Nigeria
29	Octavia & Wahyuni, 2014, Indonesia	Countering & Preventing Radicalization in Indonesian Pesantren; Islamist; Search for Common Ground, NGOs, Netherlands' government funded	None stated	To raise awareness of, and counter, terrorism, radicalisation & promote religious understanding and prevent 'communal' conflict	Over two years: To establish community radio stations in pesantrens. Programmes included: youth bulletins, drama, talk shows, moderate Islamic study; produced a video documentary	Counter extremist messages: present the theological basis for religious freedom, and anti-violence; to reflect the cultural and religious diversity in the 'real' world.	Radio and video	Targeted ten pesantrens in conflict-prone areas (affiliated with the Nahdlatul Ulama)
30	Search for Common Ground, 2020; Tunisia	The Adventures of Daly; Islamist; Search for Common Ground – Tunisia; unclear funding – worked with youth groups	None stated	To reduce support for violent extremist messages; to increase youth organisations' discussions of counter-extremism and – narratives in person and online	Undertook research in targeted areas; produced five comic books; discussions of these in targeted areas; produced digital media based on comics – videos, animation	Localised counter and alternative narratives using local, credible voices, addressing drivers of extremism	Comic books, discussion groups	People aged 15–25 years in five target areas

thinking and fake news. Ten campaigns aimed to present alternatives to extremist messages and violence as a solution to social problems, as a means to prevent or counter terrorism or extremism, by explicitly advocating peace, inclusion, diversity, tolerance and civic engagement (Aldrich, 2014; Ali et al., 2020; Bodine-Baron et al., 2020; Effendi et al., 2022; El Damanhoury, 2020, 2023; Monaci, 2020; Octavia & Wahyuni, 2014; Search for Common Ground, 2021; SecDev Foundation, 2016).

Most of the campaigns were run online using social media platforms ($n = 20$, studies 1–25), and the remaining five used largely radio programmes, with one also using videos, and another a comic book (studies 26–30). Of the online campaigns included, fifteen were run on Facebook, thirteen on YouTube, eight on Twitter, six campaigns had their own websites, two used Instagram, one used Vimeo, and one also used Telegram and WhatsApp. These online campaigns were augmented by offline media ones in four instances, with three in the MENA region using TV channels (Al-Rawi, 2013; Aly et al., 2014; El Damanhoury, 2020, 2023). One study also ran campaign materials at a university campus (Ihidero, 2019) through concerts, theatre and sports, and three within the community (Effendi et al., 2022; Search for Common Ground, 2021; Voogt, 2017). Most of the campaigns did not report the use of any theory or guideline driving its creation or development ($n = 16$). Of the rest, the public health approach (Ihidero, 2019; Wilner & Rigato, 2017) and Braddock & Horgan's guidelines (studies 12–18 in Table 2; Wilner & Rigato, 2017) were mentioned by two campaigns each, and the development of a Theory of Change was detailed in one study (Marrone et al., 2020), and stated, but not described, in another (SFCG, 2021)

The activities of the campaigns were closely linked with the channels of the campaign. Three of the campaigns described conducting formative research as a part of the campaigns' development (Bodine-Baron et al., 2020; Search for Common Ground, 2020; Wilner & Rigato, 2017). Many of the online campaigns ($n = 14$) produced videos, although one of the radio campaigns also filmed videos for use in the community (Octavia and Wahyuni 2014), and a comic book campaign also used videos (Search for Common Ground, 2020). Of the videos, seven campaigns featured ex-members of terrorist or violent extremist groups, or their family members, telling stories of their experience within the organisation and also of leaving it (Ali et al., 2020; Bodine-Baron et al., 2020; McDowell-Smith et al., 2017; SecDev Foundation, 2016; Silverman et al., 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, 2019, Speckhard, Ellenberg, Shaghati,2020; Speckhard, Ellenberg, & Ahmed, 2020; Voogt, 2017). Three campaigns also used influencers, or 'credible' people locally and nationally, to deliver their messages (Ihidero, 2019; Schmitt et al., 2018; Search for Common Ground, 2021). A number of campaigns ($n = 4$) described developing messages to post on social media, which included designing images and infographics (Berman, 2019; Bélanger et al., 2023; Effendi et al., 2022; Wilner & Rigato, 2017), and adverts and posters for their content (Al-Rawi, 2013; Aly et al., 2014; Voogt, 2017). There were two campaigns which were centred around drama series, one for the radio (Bilali, 2022) and one for the TV (El Damanhoury, 2020, 2023), and there were other radio programming, and indeed radio stations, created for the offline campaigns (Aldrich, 2014; Bilali, 2022; Marrone et al., 2020; Octavia & Wahyuni, 2014).

Review question 2: what counter-messaging was described by the studies?

The messages of the campaigns cleaved into two main areas, (i) counter and (ii) alternative messages, consistent with the types of messaging described in the literature, and explored further in this section. Sixteen campaigns presented alternative narratives, and twenty-one used counter-narratives. However, these forms of messaging were not mutually exclusive, with thirteen campaigns using elements of both.

Counter-messages

There were a variety of counter-messages or narratives deployed across the campaigns. A number of these focused on undermining extremist groups and their members, by presenting them as criminals, but also ascribing to them a range of other negative characteristics: corruption, lying, hypocrisy, failure ($n = 9$; Aldrich, 2014; Al-Rawi, 2013; Aly et al., 2014; Bean & Edgar, 2017; Berman, 2019; El Damanhoury, 2020, 2023; McDowell-Smith et al., 2017; Schmitt et al., 2018; Silverman et al., 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, 2019, Speckhard, Ellenberg, Shaghati, et al., 2020; Speckhard, Ellenberg, & Ahmed, 2020). One campaign also suggested that extremist groups encouraged criticism of Muslims, and were presented as enemies of Muslims (Al-Rawi, 2013; Aly et al., 2014).

Eight campaigns also sought to counter extremist messaging in general, targeting us v them binary depictions of the conflict, and seeking to cast doubt on their narratives by using personal stories of trauma experienced by former members, and with 'facts' (Ali et al., 2020; Ihidero, 2019; McDowell-Smith et al., 2017; Octavia & Wahyuni, 2014; Schmitt et al., 2018; SecDev Foundation, 2016; Silverman et al., 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, 2019, Speckhard, Ellenberg, Shaghati, et al., 2020, Speckhard, Ellenberg, & Ahmed, 2020; Voogt, 2017). Extremist use of the Qu'ran and of the practice of Islam more broadly was also presented as a distortion, not 'true' or 'real' Islam (Al-Rawi, 2013; Berman, 2019; McDowell-Smith et al., 2017; Schmitt et al., 2018; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, 2019, Speckhard, Ellenberg, Shaghati, et al., 2020; Speckhard, Ellenberg, & Ahmed, 2020; Voogt, 2017).

Some of the campaigns focused on the machinations of extremist groups' recruitment techniques and strategies, including the exploitation of vulnerable people and children, and presented warning signs and information about what might draw people to such groups ($n = 8$; Al-Rawi, 2013; Aly et al., 2014; Bélanger et al., 2023; Ihidero, 2019; McDowell-Smith et al., 2017; Monaci, 2020; Search for Common Ground, 2020; Silverman et al., 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, 2019, Speckhard, Ellenberg, Shaghati, et al., 2020; Speckhard, Ellenberg, & Ahmed, 2020; Wilner & Rigato, 2017). The negative impact of joining extremist groups on individuals, particularly their future employment prospects, and the wider community, was also highlighted in three campaigns (Bilali, 2022; Silverman et al., 2016; Voogt, 2017).

Although some campaigns included a 'call to action' as part of their narrative, many did not. The actions that were called for were: to report extremists (Al-Rawi, 2013; Bilali, 2022; Wilner & Rigato, 2017); to learn how best to respond to signs of radicalisation (Effendi et al., 2022); to not support terrorist groups, and protect against them (Aly et al., 2014), and to engage in forms of civic and community action (Aldrich, 2014; Marrone et al.,

2020). Two campaigns also appealed more directly to group members in promoting their exit from the groups (Schmitt et al., 2018; Wilner & Rigato, 2017).

Alternative messages

Much of the alternative messaging or narratives developed for the campaigns focused on the promotion of values which were seen as divergent or missing from that of extremist groups. These included diversity, religious freedom, unity, inclusion, integration, community and tolerance, as well as family values and education ($n = 10$; Aldrich, 2014; Al-Rawi, 2013; Bélanger et al., 2023; Bodine-Baron et al., 2020; Effendi et al., 2022; Marrone et al., 2020; Octavia & Wahyuni, 2014; Search for Common Ground., 2021; Silverman et al., 2016; Wilner & Rigato, 2017). Three campaigns also focused explicitly on the promotion of 'democracy' and of the United States more generally (Aldrich, 2014; Monaci, 2020; Silverman et al., 2016).

Two campaigns sought to address grievances articulated by extremist groups, such as denouncing racism and police corruption (Ali et al., 2020; Bilali, 2022). In addition, some of the campaigns were anti-violence as a solution to extremist group grievances, and promoted peace more broadly ($n = 5$; Aldrich, 2014; Ali et al., 2020; Ihidero, 2019; Octavia & Wahyuni, 2014; Silverman et al., 2016).

Review question three: how were campaigns evaluated and what were the findings?

This question will be answered by outlining the types of evaluation undertaken by the studies in the first section, and will summarise their findings and outcomes in the second section.

Section 1: evaluation

There were a variety of methods used to evaluate campaigns, as detailed in Table 3, which gives an overview of the aims, designs and outcome measures used in each study. The social media campaigns predominantly used some form of descriptive quantitative social media metrics ($n = 17$), however, only three studies used metrics alone (Berman, 2019; El Damanhoury, 2023; Wilner & Rigato, 2017). Eight studies used mixed methods which involved an analysis of the comments made in response to campaign social media posts alongside metrics (Al-Rawi, 2013; El Damanhoury, 2020; SecDev Foundation, 2016; Silverman et al., 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, Speckhard, Ellenberg, Shaghati, et al., 2020). The qualitative analytic method used was unclear in a number of these studies, but was stated as either a sentiment analysis (Speckhard & Ellenberg, 2020) or content analysis (El Damanhoury, 2020) in two papers. Four other studies used a mix of social media metrics alongside semi-structured interviews, focus groups and surveys (Effendi et al., 2022; Ihidero, 2019; Search for Common Ground, 2021; Voogt, 2017). Two studies conducted a network analysis alongside metrics to explore the likes of viewers who had 'liked' the campaign (Ali et al., 2020), or the connectedness of the campaign – the videos that were connected to the videos of the campaign (Schmitt et al., 2018).

Of the studies that did not use social media metrics, there were six mixed methods studies which combined surveys with focus groups and/or interviews (McDowell-Smith

Table 3. Study description.

Author, year	Study aims	Study design, length	Sample: (a)size, (b) age, (c) gender, (d) ethnicity	Outcome measures	Intervention group/control group
(i) Social media campaigns					
Ali et al. (2020)	To evaluate the 'success' of using jihadi-style audio-visuals, and reach	Quantitative: social media metrics, social network analysis; April 2019–2020	(a) $n = 500$ Facebook accounts for network analysis; (b–d) not given	View and likes; analysis of likes – of the campaign, and followers	n/a
Al-Rawi (2013)	To examine effectiveness of the campaigns	Mixed methods: social media metrics & comments; August 2010-June 2013	n/a	Comments mined with a 'webometric' tool (Thelwall, 2009)	(a) analysis pertains to Say No To Terror only – 17 videos, the two with most comments; (b) n/a
Aly et al. (2014)	To assess campaign 'noise' and adherence to criteria for effectiveness	Qualitative – descriptive; February – August 2013	n/a	Noise: 5 criteria: credibility; terminology; traditions; partners; act local, think global	n/a
Bean and Edgar (2017)	To compare sonic patterns of ISIS & campaign videos	Genosonic analysis; n/a	n/a	Sonorous communality (chanting together); sonic unmaking (sounds of pain & violence)	(i) 5 most popular US State departments' CVE videos; (ii) 3 high-profile ISIS videos
Bélanger et al. (2023)	To examineeffectiveness of messages in reducing extremism & psychological impact	Randomised controlled trial with surveys; no follow-up	(a) $n = 2009$, (b) mean: 34.22 years, (c) 38% female (d) 89.9% Arab (all Sunni Muslim)	Survey items: psychological reactance; Islamic extremism; Militant Extremist Mindset, Interpersonal Tolerance scale; Pew Global Attitudes Survey; Positive Negative Affect Schedule	(a) Group 1: message one (school); Group 2: message two (young minds); (b) Group 3: brief passage on political systems ($n = 666$)
Berman (2019)	To examine the types of information, content, and geo-political region that appeal to users	Non-experimental quantitative, descriptive – one year 2018	n/a	Modified 'contagious index' – social media likes and shares	(a) all posts from one year ($n = 426$); (b) n/a
Bodine-Baron et al. (2020)	To assess impact of campaign on attitudes	Randomised encouragement design; 2-month intervention: survey at start, midway, end, & 5 week follow-up;	(a) $n = 1579$ at baseline; $n = 940$ at end; (b) 50% 18–24 years; (c) 32.9% female; (d) 80.1% Java	Study-specific survey: reach, resonance, message, attitudes & behaviour: inclusivity, freedom of speech, social media, justification for violence	(i) CVE- content weekly for 2 months; (ii) control content (adverts, public service campaigns) weekly for 2 months
Effendi et al. (2022)	(i) how is the campaign implemented? (ii) how strong is the social media engagement?	Mixed: semi-structured interviews, social media metrics; Jan–June 2020	(a) interviews: $n = 5$; (b–d) not given	Semi-structured interview schedule; metrics – engagement/followers rate	n/a
El Damanhoury (2020)	To assess response to campaign; to explore parasocial interaction & persuasiveness link	Mixed methods: qualitative and quantitative content analysis across 2019	n/a	YouTube comments; parasocial interaction: addressing characters as part of social milieu	(a) 30 episodes of the drama; (b) n/a
El Damanhoury (2023)	To gauge the level of audience engagement	Quantitative: social media metrics; April–Nov 2019	n/a	Number of views, likes, dislikes and comments	(a) 30 episodes of the drama; (b) n/a
Ihidero (2019)				Survey: online questionnaire	n/a

(*Continued*)

Table 3. Continued.

Author, year	Study aims	Study design, length	Sample: (a)size, (b) age, (c) gender, (d) ethnicity	Outcome measures	Intervention group/control group
	To assess the impact of the campaign on Facebook and the university community	Mixed: survey, focus-groups and interview; Jan 2015–Nov 2017	Survey: (a) $n = 309$, (b,d) not given, (c) 74.1% female; 3 focus groups (a) $n = 12$ each, (b–d) not given; one interview – team lead		
McDowell-Smith et al. (2017)	Focus-testing the effects of the videos	Mixed: 'survey-style focus group'; n/a	(a) $n = 75$; (b) 20 years average, range 18–23 years; (c) 46.7% female; (d) not given	Survey: open and closed-questions; study-specific	(a) two ex-ISIS member videos; (b) n/a
Monaci (2020)	To compare audiovisual features of campaign with other campaigns regarding sense, affect & arousal	Mixed methods: a quantitative index, and focus groups; Feb–June 2018	(a) $n = 26$ (18 'at-risk'; 8 experts); at-risk: (b) 17–32 years; (c) 16.7% female; (d) 2nd/3rd gen. N. African 61%, Chechen 22%, converts 11%, student 6%	Message Sensation Index (Paek et al., 2010): higher scores = >viewer involvement	(i) seven campaign videos; (ii) Abdullah X (17 videos); Average Mohamed (7 videos)
Schmitt et al. (2018)	How are campaign videos linked to: (i) those with extremist content?; and (ii) other counter-messaging videos?	Quantitative: metrics & information net-work analysis; videos from Oct 2015–Jan 2016; data collected March 2017	n/a	Views, likes/dislikes, Eigenvector centrality scores. YTDT Video Network collected 'related videos' and metadata for networks	(a) What IS ($n = 8$ videos) and Exit USA ($n = 4$ videos); (b) n/a
Search for Common Ground (2021)	To assess understanding of messages, change in 'mindset', and lessons learned	Mixed methods: surveys, interviews & focus groups; campaign March-Dec 2020; study Dec 2020–Jan 2021	Survey: (a) $n = 314$, (b) 59% (14–18 years); (c) 71% female; (d) 72% Kyrgyz, 26% Uzbek; Interviews: (a) $n = 18$; (b–d) none; Focus groups: (a) $n = 50$, (b) 52% (14–18 years); (c) 54% female; (d) none	Not presented	n/a
SecDev Foundation (2016)	To analyse the effectiveness of platforms, dissemination, & impact	Mixed methods: metrics, comment analysis; launch period – first week only	n/a	Social media metrics: reach, awareness & engagement; impact indicators	n/a
Silverman et al. (2016)	To monitor the impact of campaigns to analyse campaign 'success metrics'	Mixed methods: social media metrics; analysis of comments; year-long, 'deployed' Oct 2015,	n/a	Metrics: awareness engagement, impact, comments: (i) engagement; (ii) ratio of engagement to impressions; (iii) retweets, (iv) Facebook likes, (v) Twitter followers, (vi) YouTube subscribers. Analysis of comments	Twelve videos, three campaigns: 1. ExitUSA ($n = 4$), 2. Average Mohamed ($n = 5$), 3. Harakat-ut-Taleem ($n = 3$)
Speckhard et al. (2018a)	Unclear – the purpose of the campaign was to reach as many people as possible	Mixed methods: social media metrics, qualitative analysis; 24 days in Dec 2017	n/a	Metrics: awareness, engagement and impact; analysis of comments	(a) two videos; (b) n/a
Speckhard et al. (2018b)	Unclear. The purpose of the campaign was to reach as many people as possible	Mixed methods: social media metrics, qualitative	n/a	Facebook metrics: awareness, engagement, impact; comments analyses	(a) one video; (b) n/a

(*Continued*)

Table 3. Continued.

Author, year	Study aims	Study design, length	Sample: (a)size, (b) age, (c) gender, (d) ethnicity	Outcome measures	Intervention group/control group
		analysis; 24 days in Dec 2017			
Speckhard et al. (2019)	To understand views of ISIS and campaign content & role in creating 'resilience' to extremism	Mixed – quantitative and qualitative; n/a	2 groups: (a) $n = 10–12$; (b) average: 26 years (21–30 years), (c) all male; (d) all Sunni, Somali-American	Semi-structured survey and focus group (class-like discussion forum)	(i) Two videos; (ii) n/a
Speckhard and Ellenberg (2020)	To consider how to increase viewership in Europe	Mixed: social media metrics, comments – sentiment analysis	n/a	Facebook metrics: reach, ThruPlay (>15 s), view (>3 s), complete views, reactions, shares, saves. Comments.	(a) 20 campaign videos; (b) n/a
Speckhard, Ellenberg, Shaghati, et al. (2020)	To assess engagement and impact of campaign, and 'behavioural shifts'	Mixed methods: metrics & comments; 13 days in Apr 2020	n/a	Awareness, engagement and impact metrics	(a) $n = 16$ campaigns (10 narratives); (b) n/a
Speckhard, Ellenberg, & Ahmed (2020)	To evaluate video quality, and peoples' views and experiences of extremism	Mixed methods: surveys and focus groups; n/a	(a) $n = 10$; (b) 25.8 years (male), 23.4 years (female); (c) 50% female; (d) all Sunni Somali-American	Pre-video: questionnaire on violence-extremism beliefs; post-video: feelings on video, and questionnaire as above	(i) All shown campaign video; (ii) n/a
Voogt (2017)	Outline of campaign & 'the advantages of using this method of engagement' (p. 34)	Quantitative: survey, metrics; since 2012 launch, but differs by metric & methos	Not given	Page views; evaluation-specific survey	n/a
Wilner and Rigato (2017)	To analyse the campaign's 'digital footprint'	Quantitative: social media metrics; 60 days	n/a	Number of likes, follows, shares, people reached, website visitors, location	(a) 150 posts; (b) n/a
(ii) Offline campaigns					
Aldrich (2014)	To explore the effect of radio programming campaign	Quasi-experimental paired comparison design; programmes: 5 years (2005–2010); survey Dec 2010	(a) $n = 200$; (b) mid-30s, estimated average; (c) not given; (d) majority were Sonrai, all Muslim	Survey items: USAID PDEV Baseline survey 2009, Afrobarometer AFRICOM (2007–2008). Outcomes: listen to programmes; civic engagement; pro-US/anti-Al-Qaeda views	(i) residents in Timbuktu exposed to radio programmes; (ii) residents in Dire not exposed to programmes
Bilali (2022)	To assess whether the drama alters violent extremist attitudes, beliefs and intended behaviours	Cluster (by village) randomised controlled trial. Surveys (1–4 weeks after end); one facilitated discussion during intervention	(a) $n = 2904$ (22 people from 132 villages); (b) mean 39–40 years (range 16–89 years); 43.6% female; 94% Fulani (>96% Muslim)	Measures study-specific and existing, taken from: Afrobarometer; Finkel et al., 2021; Buchanan-Clarke & Lekalake, 2016; Gau, 2014	(i) listened to radio drama 1 hour per week for 12 weeks (six months' content), before broadcast; (ii) business-as-usual for the remaining villages

(*Continued*)

Table 3. Continued.

Author, year	Study aims	Study design, length	Sample: (a)size, (b) age, (c) gender, (d) ethnicity	Outcome measures	Intervention group/control group
Marrone et al. (2020)	To evaluate the effect of the show	Randomised encouragement design: SMS-based longitudinal panel survey at weeks 5 & 10; with 3 & 20 week follow-up	(a) $n = 2064$; (b) 18–24 years (39.2% intervention, 38.9% control); (c) Male (81.6%, 81.4%); (d) all north Nigerian	Study-specific questions: role models, local committees, kidnap, diversity linked with theory of change. One item from Pew Research Center survey	(i) $n = 1032$, listened to show weekly for 12 weeks; (ii) $n = 1032$ – Premier League match weekly for 12 weeks
Octavia and Wahyuni (2014)	To evaluate programme implementation, and campaign relevance, effectiveness & sustainability	Mixed methods: qualitative focus groups & interviews; surveys; campaign from Sep 2011 to August 2013	(a) 10 focus groups $n = 10–15$ (students, community); interviews: $n = 21$ teachers & community leaders; surveys $n = 379$, 218 students, 161 community; (b–d) none	Some survey questions taken from baseline survey; interview and focus group schedules	n/a
Search for Common Ground (2020)	To assess effectiveness of alternative narratives; assess knowledge & attitude change	Mixed methods: surveys, interviews & focus groups; campaign from Aug 2017 to Nov 2018	Survey (a) $n = 200$, (c) 42% female; Inter-views: (a) $n = 10$ Imams, youth, experts, (c) not given; Focus groups: (a) $n = 176$, (c) 40.9% female. (b) & (d) none	Survey used OECD DAC evaluation quality standards and 3Rs framework. Semi-structured interview and focus group guides	n/a, but focus groups with (a) readers and (b) non-readers

et al., 2017; Octavia & Wahyuni, 2014; Search for Common Ground, 2020; Speckhard et al., 2019, 2020). A further study also used mixed methods with focus groups, but paired this with the application of a Message Sensation Index (Monaci, 2020).

Four of the studies (all post-2020) used some form of randomisation in their study design with a randomised controlled trial (Bélanger et al., 2023), a cluster randomised controlled trial (Bilali, 2022) and two using a 'randomised encouragement design' (Bodine-Baron et al., 2020; Marrone et al., 2020). All of these used surveys as part of the evaluation, with one administering these using SMS messaging in a harder to reach area (Marrone et al., 2020), and another also using facilitated discussions (Bilali, 2022). The sample sizes for these studies was between $n = 1579$ to $n = 2904$, and the follow-up period ranged from none to 20 weeks post-intervention. A further study also used a quasi-experimental non-randomised design, which used surveys (Aldrich, 2014).

The remaining two studies sought to assess campaign videos for their sonic attributes when compared to those of ISIS (Bean & Edgar, 2017), and another assessed the Say No to Terror campaign against a set of pre-existing criteria of effectiveness (Aly et al., 2014).

Section two: findings

Table 4 gives an overview of the main findings, limitations, and recommendations generated by each study, which will be explored further in this section which will present: (i) the impact or effects of campaigns; (ii) the qualities of the campaigns themselves; and (iii) recommendations made by studies regarding campaigns and campaign evaluations.

Impacts and effects. This section will explore the impacts and effects of the communications campaigns as reported in the included studies. These will include: (a) the number of views, likes, reach and comments made as measured by social media metrics; (b) attitudes and beliefs regarding extremism and extremist groups; (c) views of the US and its allies; (d) civic or community engagement; and (e) emotional effects.

Social media metrics

A number of metrics were used in the studies, which differed by social media platform, and also in the time frame in which they were under assessment, from 13 days (Speckhard, Ellenberg, Shaghati, et al., 2020) to two years and 10 months (Al-Rawi, 2013; Ihidero, 2019).

The number of campaign views were reported by ten studies, ranging from 18,132 views for four videos of the Exit USA campaign (Schmitt et al., 2018) to 11.5 million for 30 episodes of the TV drama al-Siham al-Marika (El Damanhoury, 2020), both across the course of a year. It is of note that one of the Breaking the ISIS Brand videos had 'close' to 1.7 million views in a 24-day period (Speckhard et al., 2018b).

Nine studies reported the number of Facebook likes received by campaigns, with the most liked being Say No to Terror ($n = 954{,}000$; Al-Rawi, 2013), and the least number of campaign likes for the Harakut-ut-Taleem page ($n = 116$; Silverman et al., 2016).

The number of comments made on campaigns were far lower than the number of views across all campaigns. The highest number of comments received were on al-Siham al-Marika's YouTube channel (8615 comments), which represents around 0.007% of views (El Damanhoury, 2020). The lowest number of comments received was for Harakut-ut-Taleem (Silverman et al., 2016) with 40 comments. The comments were

Table 4. Study findings.

Author, year	Main outcomes – summary of findings	Limitations	Recommendations: (i) campaign; (ii) evaluation	Quality
(i) Social media campaigns				
Ali et al. (2020)	Views: first season 200,000; second season 450,000. Some followers also liked neo-Salafist accounts and there were extremist spam 'attacks', indicating target audience reached to an extent. Concluded the campaign was a success as it gave space to air issues	Basic design, imprecise figures, demographics missing, vague reporting, unclear sample, analysis lacked depth, no limitations, conclusion overstated	(i) real narratives including former extremists; 'right' keywords & hashtags; similar audiovisual 'codes' as extremists – images, music, narrator, language; create space to reconsider ideas, that combine familiar and new; (ii) none	LOW
Al-Rawi (2013)	Facebook: 954,000 likes, no negative comments; Twitter 224 followers; YouTube: 1,348,791 views, 281 comments, 60% negative (anger-mistrust regarding sponsor, channels & conflating Islam & terror; anti-US & -Israel), 20% positive. Motivated 359 parody videos with 397,255 views	Missing method detail; some lack of analytic depth; conclusion somewhat overstated; no limitations; not conventionally presented	(i) avoid: concealing sponsor; conflating Islam & terrorism; propaganda-like messages; framing insurgents as terrorists. Should be interactive; (ii) none	LOW
Aly et al. (2014)	Significant problem regarding 'credibility', (astroturfing; religious authority), and campaign 'partners'. Positive use & challenge of 'terminology' and 'traditions'	Presentation and results somewhat unclear; lack of analytical framework; difficult to know how criteria were applied; lack of depth and rigour of analysis	(i) credibility key; offline campaign to support online one; (ii) long-term assessment of audience attitude and behaviour to the campaign, as well as the contexts from which they engage with it	LOW
Bean and Edgar (2017)	Campaign as 'simplistic and sonically sterile' (p. 339), lacked the 'sonorous communality' of nasheeds, and 'sonic unmaking' (sounds of pain, violence, torture), that strengthen the message and appeal of ISIS videos. Sonic experience may be more important than content	ISIS video selection not well-justified; measures not well defined; not systematically presented, more depth in ISIS video analysis	(i) need to consider how sound impacts emotion in CVE – the 'experiential' quality; (ii) none	MODERATE
Bélanger et al. (2023)	Only significant differences: (i) intervention higher 'agreement with message' scores than control (group 1, $p = 0.47$; group 2, $p < 0.001$); (ii) psychological reactance – group 2 significantly lower than control ($p = 0.005$); (iii) Islamic extremism – group 1 lower than control ($p = 0.01$); (iv) group 2 higher than control ($p = 0.02$). Effect	Unclear aim, sample choice, & procedures: recruitment and randomisation; no baseline measures, some demographic information & analysis description missing; conclusions fairly overstated; simplistic, possibly inappropriate design	(i) use evidence-based strategies supported by a theory of change – suggest inoculation theory/ messaging; fit content to context; (ii) none	LOW

(*Continued*)

Table 4. Continued.

Author, year	Main outcomes – summary of findings	Limitations	Recommendations: (i) campaign; (ii) evaluation	Quality
	sizes small. Concluded: messages had positive but limited impact, especially on extremism			
Berman (2019)	No relationship between region & likes/shares. Personal stories > likes than research/political analysis ($p = 0.035$), but not news, doctrine or military defeat stories. No relationship with shares. Written posts with video > shares than written & link ($p = 0.000$). No relationship with likes	Highly simplistic design; unclear justification for campaign selection, limited potential for impact, abstract conclusion overstates results	(i) make terrorism seem a minority endeavour unsupported by friends & family; disseminate resources to parents & teachers to help if issues arise; tackle religious discrimination; (ii) none	MODERATE
Bodine-Baron et al. (2020)	Understood message one, confused by message two. Few significant differences except: 'promoting inclusivity on social media' at 5-week follow-up, potentially due to external factors; significant negative effect on 'living in separated communities' & 'judging by appearance'	Follow-up not representative of baseline; campaign on social media before research, so control group could have been exposed. Some falsified & questionable data referenced, affecting results interpretation	(i) random assignation of content direct to Facebook feeds; distribute surveys at the same time as live campaign; use resonance analysis (Marcellino et al., 2017); (ii) none	LOW
Effendi et al. (2022)	Average engagement rate was 8.31% over six months. Peaked in month two (15.43%) dropped off thereafter	Small sample size, analysis not well explained, very basic design, qualitative analysis brief and lacked depth	(i) need for both online & offline campaign components; adequate funding, staffing & scheduling consistency; (ii) none	LOW
El Damanhoury (2020)	11.5 million viewers year after aired; 8615 comments: mostly good (83%), ⅙ comments evidencing 'parasocial interaction'. Negative comments: anti-Islam propaganda, poor production – reduced over time	Findings hard to follow, not fully reported & fairly unsynthesised; unclear use of grounded theory, need fuller discussion of limits of method	(i) TV dramas show promise as longer time frames increase chances for persuasion; (i) parasocial identification may be useful to monitor	MODERATE
El Damanhoury (2023)	Total (and average per episode) of: 10,173,729 views (339,124); 50,000 likes (1667); 5974 dislikes (199); 8482 comments (283). Levels fluctuated, peaking in the first episode	Simplistic design, figures non-specific, raw data; views not unique users; lack of depth in general	(i) mainstream TV may be more credible than state production; mainstream TV, with good budget & marketing, can reach a large audience especially when also on YouTube; (ii) none	LOW
Ihidero (2019)	Facebook: reach of 5,225,300, >700 shares, >1210 comments and reactions. 88.6% reported campaign had impact: > knowledge of using social media, countering extremism, and identification of extremist content	Unclear method; incomplete sample description and results: no regressions; repeated results; lack of study limitations; no ethics described; not easy to follow	(i) use of range of communication methods; use of small, grassroots community groups; adequate campaign finance and sustainability; (ii) none	LOW

(*Continued*)

Table 4. Continued.

Author, year	Main outcomes – summary of findings	Limitations	Recommendations: (i) campaign; (ii) evaluation	Quality
McDowell-Smith et al. (2017)	90%+: believed 'defector', reported ISIS: not real Islam, is a terror group, behaves badly, abuses Qu'ran; felt video could dissuade potential recruits; 64% reported video altered feelings for ISIS. Frequent adjectives for video: information ($n = 9$), crazy, saddening ($n = 5$); feelings evoked: sadness ($n = 14$), anger, scared, upset ($n = 9$)	Unclear aims; descriptive statistics; convenience sample; no ethnicity data; no pre-test data; focus group element unclear; analysis lacked depth, overstated conclusions; lengthy	(i) use former ISIS members; (ii) need to test materials with other cultures, languages and vulnerable populations	LOW
Monaci (2020)	Campaign had higher Index scores (8.43) than Abdullah X (3.82) & Average Mohamed (5.14). Qualitative: positive on sound & visuals; introduction less engaging; identified with characters, real stories & young people's involvement. Concluded that the 'dialectical storytelling' used delivers effective narratives	No qualitative method described; results lack depth; no ethics described; not presented conventionally	(i) could use very short web interstitials, such as at the start of YouTube videos, to increase views; (ii) to develop a more nuanced, verified measure than the MSV for storytelling strategy	MODERATE
Schmitt et al. (2018)	#WhatIS more popular (452,488 views), connected & influential (Eigenvector centrality 0.031–0.1787) than ExitUSA (18,132 views, very low Eigenvector scores 0.031–0.1787). Campaign videos 'might' be closely linked to extremist content – similar keywords & topics – could be exposed to extremist content within two clicks via YouTube recommendations. Campaigns less linked with other counter-messaging videos.	Data in method section; findings speculative – rely on algorithms & a YouTube tool that defines related videos; networks are built from blank user profile, with no browser history nor cookies – not the same as a regular user	(i) use credible voices; consider the impact of algorithms, and of recommendations & related content – gain a fuller understanding of social media principles; on & offline CVE and PVE should be combined; (ii) none	MODERATE
Search for Common Ground (2021)	9570 Instagram subscribers, only 3% 'active'. Offline events better disseminators than social media (62–24%), influencers had greatest impact (86%). Social cohesion & proactiveness as most effective messaging areas, critical thinking and gender empowerment the least, with males particularly. Challenges: older people lack interest, gender-age roles & customs, parental restrict and lack of money	Large survey dropout rate; presentation of results difficult to follow and partly un-synthesised; results & summary did not match; aims presented in results; no ethics described, conclusions somewhat overstated	(i) use of indirect messaging and influencers across platforms; target advertising; engage audience with activities (i.e. contests); (ii) none	LOW
SecDev Foundation (2016)	3707 visitors, 51% >35 years, 2.5 minutes average stay; YouTube: 73 subscribers, 50,673 views, 3.30 mins average, 78% 18–24 years, 68% male. Facebook: 946 likes, 35–44 years (20%), 66% likers female; paid ads had higher click-through rate than boosted posts at lower cost. Twitter: 22 followers from 29 tweets, most from CVE field, almost all positive. News media: > 240 stories, generating 9.8% video views. Comments mostly positive, discussed radicalisation, changing views & empathy; ⅓ – ¼ negative: futile, Islamophobic	Aims unclear, simplistic design, qualitative analysis method absent, few quotes presented, YouTube comments vetted – discussion likely limited; campaign not described in full,	(i) memorable brand & content; defined strategy & goals; best platform for audience; spread ads over time – Facebook best; avoid over-exposure; regular new content to keep media interest; embed videos in media stories to click on; use 'organic' outreach tactics; (ii) build-in evaluation; use viewer metrics & retention rates	LOW

(*Continued*)

Table 4. Continued.

Author, year	Main outcomes – summary of findings	Limitations	Recommendations: (i) campaign; (ii) evaluation	Quality
Silverman et al. (2016)	'Dramatic' improvement. 378,694 views, higher on Facebook & Twitter. Average Mohamed most views, ExitUSA higher retention. Metrics: Average Mohamed: (i) 10,810, (ii) 1:14, (iii) 275, (iv) 859, (v) 183, (vi) 274; Harakut-ut-Taleem: (i) 5463, (ii) 1:21, (iii) 40, (iv) 116, (v) 62, (vi) 6; ExitUSA (i) 4421, (ii) 1:123, (iii) 141, (iv) 286, (v) 155, (vi) 32. Comments: Average Mohamed, $n = 305$, 66% supportive; $n = 139$ for ExitUSA, 32% support; $n = 40$ for Harakat-ut-Taleem, equally positive-negative. ExitUS more engagement – responded more, videos provoke; Harakat-ut-Taleem – minimal engagement	Not all data given; analysis lacked depth, tagging categorised as supportive; 'misunderstood' is problematic. No Urdu speaker for Harakat-ut-Taleem, limited engagement online. Conclusions slightly overstated, although limitations acknowledged, and the challenge of measuring impact	(i) test topics & tone; interactivity; variety of media & platforms; combine on & offline campaigns; spread budgets; (ii) identify hypotheses & demonstrable results; ***Other:*** platforms to: train & support NGOs with ads, messaging & audiences; not block content; supply accurate analytics. Governments: fund & support organisations – train, capacity build or source private sector partners	MODERATE
Speckhard et al. (2018a)	Reach: 1,048,133, views: 604,000. 12.92% of video watched if auto-played, 31.38% if clicked. Most 'reached' 25–34 years. 296,389 engagements, 3302 shares, 3297 reactions (1663 likes). 1079 comments in three categories: anti-ISIS, anti-Islam, defend Islam. No pro-ISIS, many felt videos were propaganda	No clear study aim, not all results presented; analysis lacked depth; conclusion not wholly consistent with results presented	(i) expand this internet ad campaign; use influencers to magnify impact; (ii) none	LOW
Speckhard et al. (2018b)	Reach: 1,287,557 & 2,339,453 impressions – more males (82%), aged 25–34 years. Almost 1.7 million views (266,857 click to play). 8.34% of video watched on average. 787,7413 'engagements' – 338 shares, 4991 reactions (4500 likes). 338 comments, mostly from Shia Muslims regarding the ISIS defector & ISIS (both negative); some mocked; some support Iraq against ISIS. Some conspiracy theories on ISIS roots. Concluded that they challenged extremism & channelled anger	Method, findings and campaign description overlapped; unclear qualitative analysis method; impact measure results missing, no limitations of method discussed; conclusions not consistent with findings	(i) pre-test narratives; use ISIS defectors; model videos on ISIS propaganda – emotionally evocative; depict Muslim victims; make source of video clear; respect the viewer – don't insult Islam; (ii) none	LOW
Speckhard et al. (2019)	4/5 found ISIS defectors credible & truthful, others questioned authenticity. Survey: unfavourable view of ISIS, some frustrated with Iraqi politics, & US in the Middle East. All understood the messages. Concluded participants may have become deradicalised or 'even further turned away from ISIS … bolstered by the counter-narratives' (p. 60)	Very small sample size; unclear: sampling, why sample were 'vulnerable', recruitment, analytic method, results in parts; no baseline measures; presented raw data; lack of analytic depth; conclusions over-stated	(i) use of focus group members in formative research to act as influencers; (i) use of more experimental designs in evaluating	LOW
Speckhard and Ellenberg (2020)	'High' reach numbers, mostly male, older than target (35–64 years). One-minute videos more popular. No particular narrative most popular. Comments: appearance of video on news feed, why being targeted; some pro-ISIS, some anti-ISIS but not jihadism, conspiracy theories; defector as stupid or weak. Concluded: ISIS endorsers can be reached & engaged, especially with their emotive videos	Simplistic design; results not presented clearly – descriptive; lack of females; low number of comments, lack of depth in analysis; conclusions overstated in places – overly optimistic	(i) shorter videos can link to longer ones; need to address women to disrupt their recruitment; focus on emotional evocativeness and 'genuine' speakers; (ii) none	LOW

(*Continued*)

Table 4. Continued.

Author, year	Main outcomes – summary of findings	Limitations	Recommendations: (i) campaign; (ii) evaluation	Quality
Speckhard, Ellenberg, Shaghati, et al. (2020)	Reach: 118,559; views: 51,686; view rate: 3.1%; average watch: 10.5 s; reaction: 5110. Comments: anti-US, anti-Israel, pro-ISIS dominant in some areas; some disbelief & negativity regarding defectors; some positive comments: anti-ISIS, -terror, -violence. No direct contacts with campaign. Conclude campaign better at engaging target audience than general population, so, campaign effective	Simplistic design; method in Results; unclear sample selection; results not synthesised – some missing; analysis method vague, analysis lacked depth, some quotes not apposite; partly overstated conclusion	(i) develop relationship with social media platform to assist campaign, including targeting; cross-sector partnerships; short videos – around 1 min; (ii) track people's online behaviour after viewing campaign	LOW
Speckhard, Ellenberg, and Ahmed (2020)	No significant differences in pre and post questionnaires. No participants found terror groups appealing. More males endorsed 'pro-extremist ideology'. ISIS defector seen as authentic, some doubted remorse, saw ISIS negatively. Claimed this a success and effective at protecting against recruitment	Small sample size; unclear why sample is 'at-risk'; vague on recruitment, ethics, measures, type of analysis; some results missing; items seem mis-labelled pro-extremist; lack of analytic depth, conclusions over-stated in parts; no limitations	(i) use of former extremists as key; (ii) none	LOW
Voogt (2017)	Site near top of returns by Google for far-right keywords. Website: >22,000 users between 2013 and 2015; Facebook: 2000 likes in the first year. Popular articles: personal motivations for involvement rather than de-bunking ideological claims. Around half found site helpful in reconsidering further involvement	Limited description of outcomes; full results not presented; small sample; 'finding more sophisticated methods of measuring the impact of … engagement is required' (p. 39)	(i) different groups require different approaches; focus messages on: marginalisation as a result of involvement, unacceptability of violence, hypocrisy; involve and support community to challenge extremists. Ensure adequately resourced; (ii) none	LOW
Wilner and Rigato (2017)	Nearly 1500 likes & follows, 550 shares, reached nearly 75,000 people, 700 website visitors. Campaign Canada-focused ($n = 1519$ website views), with some wider reach. Ex-extremist & campaign designed posts were most popular, amplified by events. Conclude campaign was generally successful	Very simplistic design; authors state they were rushed & constrained by P2P guidelines which were not always clear	(i) need time and money to run a campaign well; formative research; decision-making processes to be planned out; (ii) need time and money to evaluate a campaign	LOW
(ii) Offline campaigns				
Aldrich (2014)	Peace & tolerance and civic engagement: significant ($p = 0.017$ and $p = 0.002$) with Chi-squared, and regressions ($p = 0.007$, CI[0.121, 0.775] and $p =$ 0,031, CI[0.034, 0.708]). Using propensity matching, both significant – average treatment effects: 0.43 ($p = 0.02$, CI[0.06, 0.80]) & 0.40 ($p = 0.03$, CI[0.05, 0.76]). No significant difference on pro-US or anti-Al Qaeda outcomes. Concluded programming had some positive, but limited, results	Small sample size; not randomised; no baseline measures; cannot link people & programmes directly; data collectors not transparent; findings in Method; socio-demographics unclear; not all results given; authors stated sample were fairly moderate	(i) long-term investment; understand how to get people to listen to programmes – use 'norm messaging' in local languages, cultures & institutions. Consider social media; (ii) bigger, directly exposed, completely randomised; longitudinal studies with baseline measures	MODERATE

(*Continued*)

Table 4. Continued.

Author, year	Main outcomes – summary of findings	Limitations	Recommendations: (i) campaign; (ii) evaluation	Quality
Bilali (2022)	Positive results limited: reduction in justification of violence (1.423–1.467, $p = 0.04$); increased intention to collaborate with the police (63.3% to 57.8%); and need for government to deal with extremism (17% v, 14.6%). No difference: attitude to police & collaboration, extremism, community dealing with extremism. Qualitative: approve peace-promoting acts, & collaboration	Some lack of ecological validity in the way drama listened to; not a placebo control – people likely to know what arm of the trial they are in; lack of ethics detail	(i) more research on best narratives in different contexts, and how the impact of these are mediated by both narratives themselves and their interpretation in social interaction; (ii) none	HIGH
Marrone et al. (2020)	Largely no statistically significant results including sympathy for victims and none or possible negative on value of diversity; significant positive effect on beliefs about the importance of being a role model in weeks 5 and 10 of the intervention	Non-representative sample: male, urban, university; SMS texts – very short questions; issues with translations; cross-contamination possible; lack access to show content – differed by area	(i) need for literature regarding best practices; context important; (ii) more rigorous studies; use of CATI surveys and focus groups; assess components of multi-layered intervention together	LOW
Octavia and Wahyuni (2014)	Relevance: 'quite relevant' (p. 5), targeted unevenly. Effectiveness: mixed – little change on tolerance & religious difference; religious freedom more negative in 3 areas. Concluded: campaign raised awareness of tolerance & radicalization; students more able to stop radicalisation amongst peers. Better pesantren-community relationships	Limited sample; raw data, not consistent with text; most pesantren in peaceful areas; students not at-risk, some not allowed to listen; staff too busy to campaign; partly overstated conclusions; unclear reporting; rationale & procedure not fully presented, ethics not described	(i) programs should deal with local issues & diversity, reflect society; more funding, consider succession and sustainability; support grassroots initiatives; broaden out into community TV; (ii) replicate project with other groups, in more radical-'terrorist-prone' areas	LOW
Search for Common Ground (2020)	Relevance: 67% knew people with similar experiences; Effectiveness: better under-standing of extremism (93%) & confidence in responding (92%); effective elements: true stories (34.6%), emotional (22.5%), defensive argument (19.4%). Resonance: 85.5% told someone about comic. Bottom-up development seen as effective & credible leading to sense of belonging & resilience	Unsure where sample was from or how randomised; results not always presented conventionally and analytic method unclear, no ethics mentioned	(i) formative research; combine online & offline; give young people ownership; use characters & messages based in lived experience; context-specific; raise parents' awareness; support young people with experts; soft approach for younger children (6–12 years); teach PVE in school; (ii) none	MODERATE

eported to be a mixture of mostly positive for some studies and campaigns (Say No to Terror on Facebook, Al-Rawi, 2013; El Damanhoury, 2020; Ihidero, 2019; SecDev Foundation, 2016; Average Mohamed in Silverman et al., 2016), mostly negative for a couple of others (Say No to Terror on YouTube, Al-Rawi, 2013; Exit USA in Silverman et al., 2016) and mixed for Harakat-ut-Taleem (Silverman et al., 2016).

There were mixed results regarding whether campaigns reached or were viewed by the audience targeted by campaigns. All of the studies reported that targets had been reached at least in part (Ali et al., 2020; SecDev Foundation, 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b). However, there were also reports that audiences were older, and either mostly male or mostly female contrary to the desired audience (SecDev Foundation, 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a), or Shia Muslims, when Sunni Muslims were the target (Speckhard et al., 2018b). It is of note that a number of studies reported that the audience to their campaigns appeared to be, or were likely to be, anti-extremist violence (Aldrich, 2014; McDowell-Smith et al., 2017; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, 2019, Speckhard, Ellenberg, Shaghati, et al., 2020; Speckhard, Ellenberg, & Ahmed, 2020).

Extremism and extremist groups

Only three studies sought to directly measure changes to attitudes and beliefs around extremism or extremist groups. Of these, either no significant differences were reported (Aldrich, 2014; Bilali, 2022) or there were mixed effects (Bélanger et al., 2023) – some better, some worse. Only Bilali (2022) used baseline measures.

Three studies reported that the campaign had helped viewers to feel more equipped to challenge extremism amongst their peers or community (Ihidero, 2019; Octavia & Wahyuni, 2014; SFCG, 2020). It was also suggested by two campaigns that some respondents felt campaigns could prevent people from joining, or further exploring involvement with, extremism or extremist groups (McDowell-Smith et al., 2017; Speckhard et al., 2019, Speckhard, Ellenberg, & Ahmed, 2020; Voogt, 2017).

Views of the US and allies

Some of the comments made in response to campaigns were considered to be anti-US or anti-Israel, with some conspiracy theories regarding their involvement in the origins of ISIS present in some posts (Al-Rawi, 2013; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, Speckhard, Ellenberg, Shaghati, et al., 2020). Aldrich (2014) reported USAID radio programming had no significant effect regarding pro-US sentiment.

Civic/community engagement

It was in the area of civic/community engagement that the (quasi-) experimental studies found the most effects, both positive and negative. Endorsement of pro-civic engagement items, including intention to collaborate with the police regarding extremism (Bilali, 2022), and to participate in local decision-making processes (Aldrich, 2014) reached statistical significance, as did the importance of being a role model (Marrone et al., 2020). However, Bilali (2022) also found little difference in attitudes regarding police collaboration, and Octavia and Wahyuni (2014) found little impact on tolerance and religious difference.

There were also some significantly negative effects on items such as 'living in separated communities' and 'judging by appearance' in another study (Bodine-Baron et al., 2020), and Octavia and Wahyuni (2014) reported more negative attitudes regarding religious freedom in some areas. Despite this, they also concluded that pesantren-run radio programming resulted in better relationships between the pesantrens or schools and the local community.

Emotional impact

There were few findings reported regarding how campaigns made people feel, or impacted them psychologically. There were mixed results regarding 'psychological reactance' or resistance activated when beliefs are challenged or threatened, with one intervention group having lower scores, and the other higher, than the control group (Bélanger et al., 2023). People described themselves as feeling angry at the campaign sponsor as highlighted by the studies in the next section, and one study claimed that the campaign made them feel sad, angry, scared and upset (McDowell-Smith et al., 2017), finding the videos emotionally evocative which was considered to be a facilitator of the campaign's message (Speckhard, Ellenberg, & Ahmed, 2020). This is consistent with the conclusion of Bean & Edgar's study that the 'simplistic and sonically sterile' US state department campaign videos were less emotionally appealing than those of ISIS (p. 339).

Campaign qualities. A number of studies presented findings that related to the qualities of the campaigns, which were posited as facilitating or frustrating campaign efforts. These included: (a) the origins or foundations of the campaigns; (b) production; (c) messages; (d) 'characters' or messengers; and (e) the use of social media.

Campaign foundations

Responses and comments from viewers to some campaigns suggested that the identity (or lack thereof) of campaign sponsors or channels could have a negative impact. Comments were made in five studies expressing anger or mistrust of campaign sponsors or channels, with some suggesting that they were purveyors of anti-Islamic propaganda or Islamophobia, or that they had conflated Islam with terrorism (Al-Rawi, 2013; Aly et al., 2014; El Damanhoury, 2020; SecDev Foundation, 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a). Funder obscurity and suspicion regarding the origins of campaigns was considered highly problematic by Aly et al.'s (2014) evaluation, to a potentially insurmountable degree in the Say No to Terror campaign in particular, but appears to be a crucial factor in general (Al-Rawi, 2013; Aly et al., 2014).

Campaign production

Campaign production, the features, quality and length, were considered important factors in a number of studies. Ali et al. (2020) and all of the studies under the Breaking the ISIS Brand umbrella, attributed the 'success' of their campaigns to the 'audiovisual mirroring' of their videos with those of ISIS (see studies 12–18 in Table 2). Similarly, Monaci (2020) also reported that viewers were positive about the production of the Heart of Darkness campaign, and their analysis indicated that this was consistent with the campaign's high Sensation scores.

Conversely, Bean and Edgar (2017), in their comparison of US state department CVE videos with ISIS videos, concluded that the sonic sterility of the former cannot compete with the latter, suggesting that the CVE message is thus weaker and less emotionally appealing as a consequence. There was also reported criticism from some viewers of the production quality of campaigns, and general mockery of others (El Damanhoury, 2020; SecDev Foundation, 2016; Speckhard et al., 2018b), with a large amount of parody videos observed in response to the Say No to Terror campaign (Al-Rawi, 2013).

In terms of length, one-minute videos, or shorter, were considered to be more engaging by a couple of studies (Speckhard & Ellenberg, 2020; Speckhard, Ellenberg, Shaghati, et al., 2020). However, it was also suggested that the time and space available in longer-form formats, such as TV drama, to develop narratives and characters allows for greater resonance with viewers (El Damanhoury, 2020)

Campaign messages

In terms of campaign messages, consideration of agreement with, or understanding of, messages was deemed important in a number of studies, as well as the type of messages. Speckhard et al. (2019) found that all of their sample understood campaign messages, and there were also studies that reported that social media comments were mostly consistent with the messages of campaigns (El Damanhoury, 2020; SecDev Foundation, 2016). In contrast, Silverman et al. (2016) suggested that there was a 'persistent issue' (p. 34) of people misunderstanding the content of Exit USA. Interestingly, although it seems reasonable to assume that agreement with, or understanding of, messages would be important in campaigns, Bélanger et al. (2023) found a significantly higher level of agreement with messages in the intervention group, but no significant links with extremism-related outcomes, and Bodine-Baron et al. (2020) concluded that the links between understanding of messages and outcomes were unclear (Bodine-Baron et al., 2020).

There were a number of different 'types' of messages that studies positioned as being received more positively than others, including: posts with videos significantly favoured over posts with links (Berman, 2019); personal stories (Berman, 2019; SFCG, 2020; Voogt, 2017; Wilner & Rigato, 2017); messaging regarding social cohesion and being proactive (SFCG, 2021), messages which were emotionally evocative or provocative (SFCG, 2020; Silverman et al., 2016). Less well received messages included those focused on: research/political analysis, or challenging of ideology (Berman, 2019; Voogt, 2017); critical thinking and gender empowerment – the latter with males in particular (SFCG, 2021).

Campaign 'characters'

The use of former extremist group members or family members telling their stories in campaigns, was largely positively received and they were endorsed as authentic (Berman, 2019; El Damanhoury, 2020, 2017; Monaci, 2020; SFCG, 2020; Speckhard et al., 2019; Voogt, 2017; Wilner & Rigato, 2017). The use of credible messengers or 'influencers' to promote messages on social media was also considered to have had the greatest impact on one campaign (SFCG, 2021).

There were also some negative comments about the former members of extremist groups presented – viewed by some as inauthentic, stupid, weak, or lacking remorse

(Speckhard & Ellenberg, 2020; Speckhard et al., 2018b, 2019, Speckhard, Ellenberg, & Ahmed, 2020). It is unclear what impact this was considered to have.

Use of social media in campaigns

The majority of the campaigns included in this study were conducted on social media, at least in part. A number of studies commented on the possibilities for online campaigns to be made more effective when supported by offline interventions, and recommendations were also made of the same (Ali et al., 2020; Aly et al., 2014; Effendi et al., 2022; Ihidero, 2019; SecDev Foundation, 2016; Silverman et al., 2016). One study reported that offline events were better disseminators of information than online ones (SFCG, 2021). Conversely, SFCG (2020) which reported on a campaign which used comic books, recommended the use of social media to increase awareness of the campaign, alongside other offline events.

It is of note that Schmitt et al. (2018) suggested that there are potential difficulties with social media campaigns. They observed that P/CVE campaigns are likely to be closely linked with extremist campaigns through the use of similar topics and keywords on social media platforms, and that this could then expose people to extremist videos.

Recommendations. The recommendations made within the papers under review could be separated into those made about: (i) the campaigns; and (ii) evaluation of the campaigns. One paper also included recommendations for social media companies and government (Silverman et al., 2016).

Campaigns

A range of recommendations were made regarding campaigns in all of the papers, which included:

- Campaign organisation: two studies recommended that campaign sponsors should not be obscured (Al-Rawi, 2013; Aly et al., 2014); three studies suggested campaigns should be delivered by grassroots community groups (Ihidero, 2019; Octavia & Wahyuni, 2014 ; SFCG, 2020).
- Campaign conception: the importance of formative research was highlighted by some studies (SFCG, 2020; Silverman et al., 2016; Speckhard et al., 2018b; Wilner & Rigato, 2017); others recommended the use of a defined evidence-based strategy or theory of change (Bélanger et al., 2023; Bilali, 2022; Marrone et al., 2020; SecDev Foundation, 2016);
- Appreciation of context: using local languages and the cultural and institutional landscape to shape content was recommended by some (Aldrich, 2014; Bélanger et al., 2023; Bilali, 2022; Marrone et al., 2020; Octavia & Wahyuni, 2017; SFCG, 2020);
- Messages: there was no overall consensus, but 'types' of messaging recommended were: 'norm' messaging (Aldrich, 2014; Berman, 2019; Voogt, 2017); inoculation messaging (Bélanger et al., 2023); focusing on hypocrisy, depicting victims as predominantly Muslims (Speckhard et al., 2018b; Voogt, 2017); not conflating terrorism with Islam or otherwise insulting Islam (Al-Rawi, 2013; Speckhard et al., 2018b);
- Delivery: the use of 'real' stories or narratives delivered by former members of extremist groups or using influencers to amplify impact (Ali et al., 2020; McDowell-Smith et al.,

2017; Schmitt et al., 2018; SFCG, 2020, 2021; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, 2019; Speckhard, Ellenberg, & Ahmed, 2020).

- Design: some suggested taking a similar approach as extremists in audiovisuals, including emotionally evocative sounds and images (Ali et al., 2020; Bean & Edgar, 2017; Speckhard & Ellenberg, 2020; Speckhard et al., 2018b); interactive elements such as chats or competitions were also suggested (Al-Rawi, 2013; SFCG, 2021; Silverman et al., 2016). Regarding length, shorter videos or web interstitials were advocated by three studies (Monaci, 2020; Speckhard & Ellenberg, 2020; Speckhard, Ellenberg, Shaghati, et al., 2020), but also long form TV dramas were also recommended (El Damanhoury, 2020, 2023);
- Channels: the use of multiple platforms or channels to amplify messages or reach more people was suggested by a number of studies, particularly the use of offline interventions to support online ones – and care in selecting the right channel for the audience was seen as important (Aly et al., 2014; Effendi et al., 2022; El Damanhoury, 2023; Ihidero, 2019; Octavia & Wahyuni, 2017; Schmitt et al., 2018; SFCG, 2020, 2021; SecDev Foundation, 2016; Silverman et al., 2016).
- Funding: sustainable funding, time and resources were recommended by a number of studies, with some suggestion that it should be seen as a long-term investment (Aldrich, 2014; Effendi et al., 2022; Ihidero, 2019; Octavia & Wahyuni, 2017; Voogt, 2017; Wilner & Rigato, 2017); it was also recommended that advertising spends should be spread over time, and not just for an initial push (SecDev Foundation, 2016; SFCG, 2020; Silverman et al., 2016)
- Media and marketing: engaging with media channels, producing regular new content, and ensuring easy links to the campaign was recommended, as was a memorable brand and understanding how to reach more people with campaign material (Aldrich, 2014; El Damanhoury, 2023; SecDev Foundation, 2016).
- Social media and government: Silverman et al. (2016) suggested social media organisations could train and support NGOs in developing campaigns, messaging and targeting audiences, and not blocking content, and governments should fund and support NGOs regarding training, building capacity or sourcing private sector partners.

Evaluation

Seventeen studies did not make any recommendations regarding evaluation. Of the rest, there was a wide variety of recommendations made regarding campaign evaluation. A couple of studies advocated that evaluation be built into campaigns from the start, with due time and money allocated (SecDev Foundation, 2016; Wilner & Rigato, 2017). The need for more rigorous studies with more experimental designs, baseline measures, larger samples, control for confounding variables such as exposure, and complete randomisation, was suggested by a number of studies in assessing effectiveness (Aldrich, 2014; Bodine-Baron et al., 2020; Marrone et al., 2020; Silverman et al., 2016; Speckhard et al., 2019). The use of long-term or longitudinal study designs was also recommended by a couple of studies (Aldrich, 2014; Aly et al., 2014). Two studies advocated the testing of materials in different contexts and across different cultures or vulnerable populations (McDowell-Smith et al., 2017; Octavia & Wahyuni, 2014). Marrone et al. (2020) suggested that evaluations 'could assess the effect of a portfolio of CVE interventions simultaneously' (p. 19), where communications campaigns are a part of a multi-component CVE approach.

Regarding social media metrics, whilst they are used widely, some caution was suggested in their use and interpretation (SecDev Foundation, 2016). Some relatively novel suggestions for online study included tracking people's online behaviour after they have viewed campaign material (Speckhard, Ellenberg, Shaghati, et al., 2020), and resonance analysis – analysing whether the language of the campaign is taken up and used by a target populations' social media posts (Bodine-Baron et al., 2020).

Review question four: what is the quality of the research?

The 'quality' of the research was assessed using the MetaQAT tool, with studies largely found to be of low quality ($n = 21$), with some at a moderate level ($n = 8$), and one study being assessed to be of high quality (Bilali, 2022) – as seen in Table Three. For further information, Appendix 5: Quality Appraisal Table sets out the specific criteria assessed, and level assigned, for each study.

There were a number of limitations present across all areas of the studies as reported. Given that almost all of the studies were assessed to be of low or moderate quality, resulting from problems across the full range of possible reporting from design to conclusions, only the most frequent or problematic issues will be discussed: (a) design, (b) sampling, (c) lack of procedural detail, (d) the analysis, (e) conclusions drawn, and (f) reporting overall. Each of these have a negative impact on the interpretation of the findings of studies, the credibility and applicability of the findings.

Design

The use of basic or overly simplistic study designs to evaluate the complex phenomena of communications campaigns, particularly the seeming reliance on social media metrics as the main outcome measure, was highlighted. Another issue was that some of the experimental/quasi-experimental studies possibly exposed the control group to the campaign material – or it was at least difficult to rule this in or out (Aldrich, 2014; Bodine-Baron et al., 2020; Marrone et al., 2020). Baseline measures were also not collected in two of these studies (Aldrich, 2014; Bélanger et al., 2023) and in another, researchers did not have access to radio programme content which varied by area (Marrone et al., 2020).

Sample

Sample descriptions were missing or unclear in ten studies, including justification for the sample choice (Aldrich, 2014; Ali et al., 2020; Bélanger et al., 2023; Ihidero, 2019; Marrone et al., 2020; McDowell-Smith et al., 2017; Octavia & Wahyuni, 2014; Speckhard et al., 2019, Speckhard, Ellenberg, Shaghati, et al., 2020; Speckhard, Ellenberg, & Ahmed, 2020). Six studies included relatively small sample sizes, and the sampling procedure was unclear in four studies, including the randomisation procedures in an RCT (Bélanger et al., 2023). In one study, the follow-up sample was not representative of baseline sample (Bodine-Baron et al., 2020).

Lack of procedural detail

There was a lack of detail or clarity around the method of analysis undertaken (particularly qualitative) and outcome measures in twelve studies (Aly et al., 2014; Bean & Edgar, 2017;

Bélanger et al., 2023; Effendi et al., 2022; El Damanhoury, 2020; SecDev Foundation, 2016; SFCG, 2020; Speckhard et al., 2018b, 2019, Speckhard, Ellenberg, Shaghati, et al., 2020, Speckhard, Ellenberg, & Ahmed, 2020; Voogt, 2017), as well as absent or incomplete details of ethics procedures in a further seven studies (Bilali, 2022; Ihidero, 2019; Monaci, 2020; Octavia & Wahyuni, 2014; SFCG, 2020, 2021; Speckhard, Ellenberg, & Ahmed, 2020)

Analysis

The appraisal found a lack of depth or rigour to the qualitative analysis undertaken, or it was applied unevenly, in fifteen studies (Ali et al., 2020; Al-Rawi, 2013; Aly et al., 2014; Bean & Edgar, 2017; Effendi et al., 2022; El Damanhoury, 2023; McDowell-Smith et al., 2017; Monaci, 2020; SecDev Foundation, 2016; Silverman et al., 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2019, Speckhard, Ellenberg, Shaghati, et al., 2020; Speckhard, Ellenberg, & Ahmed, 2020). There were also issues with imprecise and missing statistics in ten studies (Aldrich, 2014; Ali et al., 2020; El Damanhoury, 2020; Ihidero, 2019; Silverman et al., 2016; Speckhard et al., 2018a, 2018b, Speckhard, Ellenberg, Shaghati, et al., 2020; Speckhard, Ellenberg, & Ahmed, 2020; Voogt, 2017), or the presentation of raw and unsynthesised data in a further eight studies (El Damanhoury, 2020, 2023; McDowell-Smith et al., 2017; Octavia & Wahyuni, 2014; SFCG, 2021; Speckhard & Ellenberg, 2020; Speckhard et al., 2019, Speckhard, Ellenberg, Shaghati, et al., 2020).

Conclusions

The conclusions drawn by a number of studies were judged to be inconsistent with the results presented and/or were somewhat overstated in fourteen of the studies (Ali et al., 2020; Al-Rawi, 2013; Bélanger et al., 2023; Berman, 2019; McDowell-Smith et al., 2017; Octavia & Wahyuni, 2014; SFCG, 2021; Silverman et al., 2016; Speckhard & Ellenberg, 2020; Speckhard et al., 2018a, 2018b, 2019, Speckhard, Ellenberg, Shaghati, et al., 2020; Speckhard, Ellenberg, & Ahmed, 2020).

Reporting

The reporting and presentation of studies was vague in at least some areas, was not systematic, was overly long, or otherwise did not fit the conventions of the field in fifteen of the studies, making them hard to follow (Aldrich, 2014; Ali et al., 2020; Aly et al., 2014; Bean & Edgar, 2017; El Damanhoury, 2020; Ihidero, 2019; McDowell-Smith et al., 2017; Monaci, 2020; Octavia & Wahyuni, 2014; Schmitt et al., 2018; SFCG, 2020, 2021; Speckhard et al., 2018b, 2019, Speckhard, Ellenberg, Shaghati, et al., 2020).

Discussion

A key contribution of this paper is that it is, as far as we are aware, the first review of 'live' counter-messaging campaigns. This review set out to answer four broad questions pertaining to the evaluation of live counter-messaging communications campaigns. In terms of the types of campaigns studied, the majority of these involved social media campaigns run across a range of platforms, in full or part, with the offline campaigns mostly using radio platforms. The campaigns made use of both counter and alternative narratives, and sometimes both. The counter-messages focused on undermining extremist

groups and messaging, and highlighting exploitative recruitment strategy. Alternative messages focused on promoting diversity, inclusion and democracy, and denouncing issues of grievance.

The majority of studies used social media metrics to explore the impact of campaigns, although these were largely combined with a mixture of other methods, predominantly an analysis of comments. It is of note that there were four randomised experimental studies included in the review, all published from 2020 onwards, which perhaps signals a maturing research field.

Whilst there were some positive effects reported for the campaigns included in this study, these results should be interpreted with caution, as they were largely derived from studies of low or moderate quality which were therefore at a high risk of bias. Indeed, the experimental studies typically concluded that whilst there may have been some positive findings, ultimately impact was limited particularly regarding attitudes and beliefs about extremism and extremist groups. Given that none of the studies was designed to measure actual behaviour, this was perhaps the most important outcome under examination. Nevertheless, the review suggests that there may be some merit in pursuing counter-messaging communications campaigns, potentially as part of a broader suite of approaches to preventing or countering violent extremism that would also involve more offline interventions.

One point that the evaluations reflected was the wide variation in counter-messaging campaigns, with diversity across the type of message (counter, alternative), the medium of communication (social media, TV, radio), target audience (young people, wider community, networks around at risk individuals) and the campaign objective (prevention, disengagement, media literacy). Whilst these insights can provide a wealth of inspiration for practitioners designing new campaigns, it was not possible with the data available to start to draw conclusions about what aspects make the most effective campaigns. Not all counter-messaging campaigns are equal, and while there is evidence of their overall positive impact, we still have limited understanding of what makes an effective campaign. As such a key area of future research that needs to be addressed is not so much whether counter-messaging is effective, but what types of campaigns achieve what effects.

The limitations of the studies under review point to the weak empirical foundations of counter-messaging campaigns, both in terms of campaign design and campaign evaluation: 16 of the 25 campaigns did not report the use of any theory or guidelines driving its creation or development. Several of these studies also recommended more research to support the design, implementation and evaluation of counter-messaging campaigns.

Following on from this, perhaps more concerning was that only two campaigns used a theory of change (with only one stating what it was). A theory of change is fundamental to good project implementation. It sets out the campaign objectives, and how implementing the campaign will achieve these objectives. In addition to setting out the causal process through which this is intended to happen, ideally drawing on empirical evidence to substantiate this process. Using a theory of change is not just important for good project management, but also crucial for evaluating impact effectiveness. Unless it is clear what the campaign objectives were, and how they were meant to be achieved, it is hard to measure their intended impact.

A wider point highlighted by this report is the low to moderate quality of the evaluations reviewed: only one of the studies was of high quality, whilst 21 of 30 reports were classed as low quality. This underscores the need for a more methodologically rigorous approach to internal and external evaluations of counter-messaging campaigns, and the need for more effective evaluation methods, particularly in establishing the impact and effectiveness of campaigns in delivering behavioural change. A useful resource for the field would be the development of a database of effective evidence-based methods and tools to evaluate different types and aspects of campaigns, addressing the wide variety of campaigns.

A further and fundamental issue is the limited number of publicly available evaluations of counter-messaging campaigns. This review found 30 papers relating to 25 campaigns, which presumably represents a fraction of the counter-messaging campaigns implemented worldwide. Campaign evaluations provide a crucial step in testing theory and practice, and furthering our knowledge of what makes an effective and impactful campaign. Failing or refusing to make results publicly available restricts the ability of funders and implementing agencies to improve the quality and effectiveness of counter-messaging campaigns.

Recommendations for policy and practice

Recommendation 1: Further research to increase our understanding of how communication campaigns work, and what makes them effective, in the counter terrorism and countering violent extremism space.
Recommendation 2: For all counter-messaging campaigns to be based on and clearly set out their theory of change.
Recommendation 3: To develop more methodologically rigorous evaluations that can more effectively determine the effectiveness and impact of campaigns.
Recommendation 4: Evaluations should be hardwired into CT-CVE Communication Campaigns at the design phase, and not left as an afterthought.
Recommendation 5: Campaigns should endeavour to ensure that where evaluations are carried out that the findings are disseminated and made publicly available. This is something that funders should also consider making a requirement of accepting funding.

Limitations

As with all reviews, there are a number of potential limitations that need to be taken into consideration. First, some evaluations may have been missed in the review due to deficiencies in our search strategy or because of publication bias, such as having a more in-depth knowledge of UK grey literature than non-UK ones. Second, there is a language bias as only studies published in the English language were included – the majority of studies took place in 'the West', although there were a number of non-Western countries also included which is a strength of the study, but it is likely that this may have meant we missed studies. Third, a narrative synthesis approach can introduce some bias, although extraction involved two researchers which hopefully mitigated some of this potential bias. Finally, the heterogeneity of the campaigns and the evaluations that had taken place meant that it was not possible to appraise or synthesise

them using a hierarchy of evidence as this seemed untenable, but this does deviate from traditional systematic review guidelines.

Conclusion

This review described 'live' counter-messaging communications campaigns, and their evaluation. Whilst there were a range of campaigns and evaluations, the majority were run on social media platforms, in full or part, and used social media metrics in their evaluations. The findings highlighted that whilst there is some evidence of positive impacts from the campaigns, the evidence is limited and must be interpreted with caution due to the low and moderate quality of the evaluations. Furthermore, the review called attention to the weak empirical foundations underpinning both the campaigns and their evaluations. Whilst there are good reasons to believe that counter-messaging campaigns can play an important role in CVE, there is much work to be done to identify and develop what makes an effective counter-messaging campaign.

Acknowledgements

The authors wish to thank Angharad Devereux for her assistance in data collection and analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This project was made possible by research funding from the UK Home Office to Swansea University, and financial support provided by Swansea University's Legal Innovation Lab Wales (which is part funded by the European Regional Development Fund through the Welsh Government).

Notes on contributors

Samantha Treacy is currently a Research Officer in the Department of Psychology at Swansea University, and is also a PhD student within the Department of Criminology, Sociology and Social Policy at Swansea University.

Alastair Reed is Professor of Security and Strategic Communications at the Centre for Peace and Security, Coventry University. Previously, he was an Associate Professor at the Hillary Rodham Clinton School of Law, Swansea University (where the research for this paper was conducted).

Andrew Glazzard is Professor of National Security Policy and Practice, at the Centre for Peace and Security, Coventry University.

References

Studies under review

Aldrich, D. P. (2014). First steps towards hearts and minds? USAID's countering violent extremism policies in Africa. *Terrorism and Political Violence*, *26*(3), 523–546. https://doi.org/10.1080/09546553.2012.738263

Ali, R., Šibljaković, D., Lippe, F., Neuberg, U., & Neuberg, F. (2020). You're against Dawla, but you're listening to their Nasheeds?' Appropriating jihadi audiovisualities in the online streetwork project

Jamal al-Khatib – My Path!. In C. Günther & S. Pfeifer (Eds.), *Jihadi audiovisuality and its entanglements: Meanings, aesthetics, appropriations* (pp. 222–246). Edinburgh University Press.

Al-Rawi, A. K. (2013). The anti-terrorist advertising campaigns in the Middle East. *Journal of International Communication*, *19*(2), 182–195. https://doi.org/10.1080/13216597.2013.833534

Aly, A., Weimann-Saks, D., & Weimann, G. (2014). Making 'Noise' online: An analysis of the Say No to Terror online campaign. *Perspectives on Terrorism*, *8*(5), 33–47.

Bean, H., & Edgar, A. N. (2017). A genosonic analysis of ISIL and US counter-extremism video messages. *Media, War & Conflict*, *10*(3), 327–344. https://doi.org/10.1177/1750635217694124

Bélanger, J. J., Snook, D. W., Dzitac, D., & Cheppih, A. (2023). Challenging extremism: A randomized control trial examining the impact of counternarratives in the Middle East and North Africa. *Current Research in Ecological and Social Psychology*, *4*, 100097. https://doi.org/10.1016/j.cresp.2023.100097

Berman, E. R. (2019). *Evaluating the effectiveness of counter-narrative tactics in preventing radicalization* [22624402 Ph.D.]. Walden University. https://www.proquest.com/docview/2303121236?accountid=14680&sourcetype=Dissertations%20&%20Theses

Bilali, R. (2022). Fighting violent extremism with narrative intervention: Evidence from a field experiment in West Africa. *Psychological Science*, *33*(2), 184–195. https://doi.org/10.1177/09567976211031895

Bodine-Baron, E., Marrone, J. V., Helmus, T. C., & Schlang, D. (2020). *Countering violent extremism in Indonesia: Using a online panel survey to assess a social media counter-messaging campaign*. https://www.rand.org/pubs/research_reports/RRA233-1.html

Effendi, R., Sukmayadi, V., Unde, A. A., & Triyanto. (2022). Social media as a medium for preventing radicalization (A case study of an Indonesian youth community's counter-radicalization initiatives on Instagram). *Plaridel*, *19*(2), 1–28. https://doi.org/10.52518/2021-14edut

El Damanhoury, K. (2020). Entertainment-education versus extremism: Examining parasocial interaction among Arab viewers of anti-ISIS TV drama. *Journal for Deradicalization*, *24*, 40–78.

El Damanhoury, K. (2023). Fighting religious extremism with faith-based entertainment-education: The portrayal of Isis in Arab drama. *Journal of Religion, Media and Digital Culture*, *11*(2), 171–197. https://doi.org/10.1163/21659214-bja10070

Ihidero, V. O. (2019). An appraisal of 'Facebook's global peer-to-peer campaign on challenging violent extremism' in Nigeria: The SOAR peer-to-peer project. *Journal of English Literature and Cultural Studies*, *2*(1), 94–109.

Marrone, J. V., Helmus, T. C., Bodine-Baron, E., & Santucci, C. (2020). *Countering violent extremism in Nigeria: Using a text-message survey to assess radio programs*. https://www.rand.org/pubs/research_reports/RR4257.html

McDowell-Smith, A., Speckhard, A., & Yayla, A. S. (2017). Beating ISIS in the digital space: Focus testing ISIS defector counter-narrative videos with American college students. *Journal for Deradicalization*, *10*, 50–76.

Monaci, S. (2020). Social media campaigns against violent extremism: A new approach to evaluating video storytelling. *International Journal of Communication*, *14*(1), 980–1003.

Octavia, L., & Wahyuni, E. (2014). *Final evaluation report for the project: Countering and preventing radicalization in Indonesian Pesantren*. https://www.sfcg.org/wp-content/uploads/2014/08/DUT_Evaluation_Report__FINAL.pdf

Schmitt, J. B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as prevention or promotion of extremism?! The potential role of YouTube. *Journal of Communication*, *68*(4), 780–808. https://doi.org/10.1093/joc/jqy029

Search for Common Ground. (2020). '*The Adventures of Daly' graphic novel campaign: Internal Final Evaluation*. https://www.sfcg.org/wp-content/uploads/2020/02/SFCG-Final-Evaluation-The-Adventures-of-Daly-Graphic-Novel-Campaign-Nov-2018.pdf

Search for Common Ground. (2021). *TaasirLink: Citizens narrative campaign - Final evaluation report*. https://www.sfcg.org/wp-content/uploads/2021/04/SFCG_Final-Evaluation-Report_M-Vector_final.pdf

SecDev Foundation. (2016). '*Extreme dialogue: Social media target audience analysis and impact assessments in support of countering violent extremism: An abridged summary report of findings*

and lessons learned. https://preventviolentextremism.info/sites/default/files/Kanishka-Secdev%20Extreme%20Dialogue-%20Social%20media%20Target%20Audience%20Analysis%20and%20Impact%20Assessments%20in%20support%20of%20countering%20violent%20extremism.pdf

Silverman, T., Stewart, C. T., Amanullah, Z., & Birdwell, J. (2016). *The impact of counter-narratives: Insights from a year-long cross-platform pilot study of counter-narrative curation, targeting, evaluation and impact*. https://www.isdglobal.org/wp-content/uploads/2016/08/Impact-of-Counter-Narratives_ONLINE_1.pdf

Speckhard, A., & Ellenberg, M. (2020). Breaking the ISIS brand counter narrative Facebook campaigns in Europe. *Journal of Strategic Security*, *13*(3), 120–148. https://doi.org/10.5038/1944-0472.13.3.1844

Speckhard, A., Ellenberg, M., & Ahmed, M. (2020). Jihad is our way: Testing a counter narrative video in two Somali American focus groups. *Bildhaan*, *20*, 54–73.

Speckhard, A., Ellenberg, M., Shaghati, H., & Izadi, N. (2020). Hypertargeting Facebook profiles vulnerable to ISIS recruitment with "breaking the ISIS brand counter narrative video clips" in multiple Facebook campaigns. *Journal of Human Security*, *16*(1), 16–29. https://doi.org/10.12924/johs2020.16010016

Speckhard, A., Shajkovci, A., & Ahmed, M. (2019). Intervening in and preventing Somali-American radicalization with counter narratives: Testing the breaking the ISIS brand counter narrative videos in American Somali focus group settings. *Journal of Strategic Security*, *11*(4), 32–71. https://doi.org/10.5038/1944-0472.11.4.1695

Speckhard, A., Shajkovci, A., Wooster, C., & Izadi, N. (2018a). Engaging English speaking Facebook users in an anti-ISIS awareness campaign. *Journal of Strategic Security*, *11*(3), 52–78. https://doi.org/10.5038/1944-0472.11.3.1679

Speckhard, A., Shajkovci, A., Wooster, C., & Izadi, N. (2018b). Mounting a Facebook brand awareness and safety ad campaign to break the ISIS brand in Iraq. *Perspectives on Terrorism*, *12*(3), 50–66.

Voogt, S. (2017). Countering far-right recruitment online: CAPE's practitioner experience. *Journal of Policing, Intelligence and Counter Terrorism*, *12*(1), 34–46. https://doi.org/10.1080/18335330.2016.1215510

Wilner, A., & Rigato, B. (2017). 60 days of PVE campaign: Lessons on organizing an online, peer-to-peer, counter-radicalization program. *Journal for Deradicalization*, *12*, 227–268.

Other references

Atkin, C. K., & Rice, R. E. (2013). Theory and principles of public communication campaigns. In R. E. R. C. K. Atkin (Ed.), *Public communication campaigns* (4th ed., pp. 3–19). Sage.

Braddock, K., & Horgan, J. (2016). Towards a guide for constructing and disseminating counternarratives to reduce support for terrorism. *Studies in Conflict & Terrorism*, *39*(5), 381–404. https://doi.org/10.1080/1057610x.2015.1116277

Briggs, R., & Feve, S. (2013). Review of programs to counter narratives of violent extremism: What works and what are the implications for government? *Institute for Strategic Dialogue*. https://core.ac.uk/reader/30675430

Buchanan-Clarke, S., & Lekalake, R. (2016). Violent extremism in Africa. Public opinion from the Sahel, Chad, and the Horn (Afrobarometer Policy Paper No. 32). Afrobarometer. https://afrobarometer.org/sites/default/files/publications/Policy%20papers/r6-afropaperno32-violent-extremism-inss-africa-en.pdf

Carthy, S. L., Doody, C. B., Cox, K., O'Hora, D., & Sarma, K. M. (2020). Counter-narratives for the prevention of violent radicalisation: A systematic review of targeted interventions. *Campbell Systematic Reviews*, *16*(3), e1106. https://doi.org/10.1002/cl2.1106

Ferguson, K. (2016). Countering violent extremism through media and communication strategies: A review of the evidence. *Partnership for Conflict, Crime and Security Research*. https://bpb-eu-w2.wpmucdn.com/blogs.bristol.ac.uk/dist/c/892/files/2016/03/Countering-Violent-Extremism-Through-Media-and-Communication-Strategies-.pdf

Finkel, S. E., McCauley, J. F., Neureiter, M., & Belasco, C. A. (2021). Community violence and support for violent extremism: Evidence from the Sahel. *Political Psychology*, *42*(1), 143–161. https://doi.org/10.1111/pops.12692

Gau, J. C. (2014). Procedural justice and police legitimacy: A test of measurement and structure. *American Journal of Criminal Justice*, *39*, 187–205. https://doi.org/10.1007/s12103-013-9220-8

Harden, A., & Thomas, J. (2005). Methodological issues in combining diverse study types in systematic reviews. *International Journal of Social Research Methodology*, *8*(3), 257–271. https://doi.org/10.1080/13645570500155078

Higgins, J. P. T., Savović, J., Page, M. J., & Sterne, J. A. C. (2019). *Revised Cochrane risk-of-bias tool for randomized trials (RoB 2): Template for completion*. https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool?authuser=0

Lubbe, W., ten Ham-Baloyi, W., & Smit, K. (2020). The integrative literature review as a research method: A demonstration review of research on neurodevelopmental supportive care in preterm infants. *Journal of Neonatal Nursing*, *26*(6), 308–315. https://doi.org/10.1016/j.jnn.2020.04.006

Marcellino, W. M., Cragin, K., Mendelsohn, J., Cady, A. M., Magnuson, M., & Reedy, K. (2017). Measuring the popular resonance of Daesh's propaganda. *Journal of Strategic Security*, *10*(1), 32–52. http://doi.org/10.5038/1944-0472.10.1.1527

Noar, S. M. (2012). An Audience–Channel–Message–Evaluation (ACME) framework for health communication campaigns. *Health Promotion Practice*, *13*(4), 481–488. https://doi.org/10.1177/1524839910386901

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, E., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*, n160. https://doi.org/10.1136/bmj.n160

Paek, H. J., Kim, K., & Hove, T. (2010). Content analysis of antismoking videos on YouTube: Message sensation value, message appeals, and their relationships with viewer responses. *Health Education Research*, *25*(6), 1085–1099. https://doi.org/10.1093/her/cyq063

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC methods programme*. https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf

Reed, A., Ingram, H., & Whittaker, J. (2017). Countering terrorist narratives. *European Parliament's Committee on Civil Liberties, Justice and Home Affairs/Policy Department for Citizen's Rights and Constitutional Affairs*. https://scholarlypublications.universiteitleiden.nl/access/item%3A2951675/view

Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. (1995). The well-built clinical question: A key to evidence-based decisions. *ACP Journal Club*, *123*(3), A12–A13. https://doi.org/10.7326/ACPJC-1995-123-3-A12

Rosella, L., Bowman, C., Pach, B., Morgan, S., Fitzpatrick, T., & Goel, V. (2016). The development and validation of a meta-tool for quality appraisal of public health evidence: Meta Quality Appraisal Tool (MetaQAT). *Public Health*, *136*, 57–65. https://doi.org/10.1016/j.puhe.2015.10.027

Ryan, H., & Hill, S. (2016). *How to GRADE the quality of the evidence*. http://cccrg.cochrane.org/author-resources

Thelwall, M. (2009). *Introduction to webometrics: Quantitative web research for the social sciences*. Morgan & Claypool.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, *52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x