

# Comunicar

Media Education Research Journal, n. 71, vol. XXX  
[www.comunicarjournal.com](http://www.comunicarjournal.com)

**Hate speech in communication:  
Researchs and proposals**



English Edition



© COMUNICAR, 71, XXX

MEDIA EDUCATION RESEARCH JOURNAL  
REVISTA CIENTÍFICA DE COMUNICACIÓN Y EDUCACIÓN

ISSN: 1134-3478 / DL: H-189-93 / e-ISSN: 1988-3293  
n. 71, vol. XXX (2022-2), 2<sup>nd</sup> Quarter, April, 1<sup>st</sup>, 2022

## INDEXED INTERNATIONAL SCIENTIFIC JOURNAL

<https://www.revistacomunicar.com/index.php?contenido=factor-de-impacto&idioma=en>



### JOURNAL CITATION REPORTS (JCR)

**JCR 2020 (2022):** Q1. JIF: 6.013. JCI: 3.08. 5 Year Impact Factor: 5.440. Immediacy Index: 1.100; Eigenfactor Score: 0.000275. Article Influence Score: 1.339; Journal Impact Factor (JIF): Education: Q1 (position 7 from 294; 1st Spanish and Iberoamerican); Communication: Q1 (position 9 from 94, 1st Spanish and Iberoamerican); Journal Citation Indicator (JCI): Education: Q1 (position 11 from 725; 1st Spanish and Iberoamerican); Communication: Q1 (position 9 from 207, 1st Spanish and Iberoamerican).

**SOCIAL SCIENCES CITATION INDEX:** The top journal in Spanish in Communication since 2007.



### SCOPUS

**CITE SCORE 2020 (2022):** (7.30); Q1 in Cultural Studies (position 2nd from 1,037) (percentile 99). Q1 in Communication: position 14th from 426 (percentile 96). Q1 in Education (position 24th from 1,319) (percentile 98). **SCIMAGO JOURNAL RANK: SJR 2020 (2021-22):** 1.217; Q1 in Cultural Studies, Communication and Education (first journal in Spanish language in Education and Cultural Studies and second journal in Communication).



### RECYT (FECYT-MEC)

**Ranking FECYT 2020 (2022):** Education: 1st from 56 journals (99.82 points out of 100) (top 1%); Communication, Information and Scientific Documentation: 1st from 16 journals (99.82 points out 100) (top 1%); **FECYT 2021 Seal of Quality Excellence** (12 indicators).



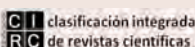
### GOOGLE SCHOLAR

**2022:** Top 100 of Google: Position 2nd (from 100) in the Spanish ranking of all research journals areas. H5: 41. Mediana H5: 55. En 2022-03-02: H: 97; H5: 78 (52,608 accumulated citations). According EC3 Reports, 2018 (UGR): H5 Index (2013-2017): 1st in Education (out of 165; H5: 38); 1st in Communication (out of 51; H5: 38).



### DIALNET METRICS

Q1: EDUCATION 2020 (2022): position 1st (from 230); IF: 4.545 (909 cites).  
Q1: COMMUNICATION 2020 (2022): position 2nd (from 62); IF: 4.545 (909 cites).



### CIRC (INTEGRATED CLASSIFICATION OF SCIENTIFIC JOURNALS) (EC3 Metrics)

2022, Level A+ (highest rated).



### REDIB (CSIC) (IBERO-AMERICAN INNOVATION AND SCIENTIFIC KNOWLEDGE NETWORK)

2021: 2nd position among 1,199 selected journals from all areas. Rating: 58.364.



### ERIH+

Level INT2 (2020).

## PUBLISHED BY: GRUPO COMUNICAR EDICIONES

- [www.comunicarjournal.com](http://www.comunicarjournal.com)
- [www.grupocomunicar.com](http://www.grupocomunicar.com)
- Administration: [info@grupocomunicar.com](mailto:info@grupocomunicar.com)
- Staff: [director@grupocomunicar.com](mailto:director@grupocomunicar.com)
- Mail box 527. 21080 Huelva (Spain)
- COMUNICAR is a member of CEDRO (Spanish Centre for Reprographic Rights).
- COMUNICAR is a quarterly scientific publication, published in January, April, July and October.
- COMUNICAR journal accepts and promotes online institutional exchanges with other scientific journals.

© COMUNICAR is a patented trademark registered at the Spanish Patents and Trademarks Office, n. 1806709,

## INTERNATIONAL CO-EDITIONS

- **ENGLISH EDITION** ([www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#EN](http://www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#EN))
- **PORTUGUESE EDITION** ([www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#BR](http://www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#BR))
- **CHINESE EDITION** ([www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#CN](http://www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#CN))
- **LATIN-AMERICAN EDITION** ([www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#IB](http://www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#IB))
- **RUSSIAN EDITION** ([www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#RU](http://www.revistacomunicar.com/index.php?contenido=coeditores&idioma=en#RU))

PRINTED BY: Estu Graf. Madrid (Spain)

© Reproduction of any text in COMUNICAR requires authorization from CEDRO or the publisher.

E-SHOP: [www.revistacomunicar.com/index.php?contenido=tienda](http://www.revistacomunicar.com/index.php?contenido=tienda)

# CONTENTS

Comunicar, 71, XXX (2022-2)

## Hate speech in communication: Research and proposals

Discursos de odio en comunicación: Investigaciones y propuestas



Comunicar, 71, XXX, 2022

## DOSSIER

### THEMATIC EDITORS

Dr. M<sup>a</sup> Dolores Cáceres-Zapatero, Complutense University of Madrid (Spain)  
 Dr. Mykola Makhortykh, University of Bern (Switzerland)  
 Dr. Francisco Segado-Boj, Complutense University of Madrid (Spain)

01. Adolescents' motivations to perpetrate hate speech and links with social norms ..... 09-19  
 Motivos del discurso de odio en la adolescencia y su relación con las normas sociales  
*Sebastian Wachs, Alexander Wettstein, Ludwig Bilz & Manuel Gámez-Guadix (Germany, Switzerland & Spain)*
02. Hate speech and social acceptance of migrants in Europe: Analysis of tweets with geolocation ..... 21-34  
 Discurso de odio y aceptación social hacia migrantes en Europa: Análisis de tuits con geolocalización  
*Carlos Arcila-Calderón, Patricia Sánchez-Holgado, Cristina Quintana-Moreno, Javier-J. Amores & David Blanco-Herrero (Spain)*
03. Hate speech analysis as a function of ideology: Emotional and cognitive effects ..... 35-45  
 Análisis del discurso de odio en función de la ideología: Efectos emocionales y cognitivos  
*Natalia Abuiñ-Vences, Ubaldo Cuesta-Cambra, José-Ignacio Niño-González & Carolina Bengochea-González (Spain)*
04. A systematic literature review of the representations of migration in Brazil and the United Kingdom ..... 47-58  
 Una revisión sistemática de la literatura de las representaciones de la migración en Brasil y Reino Unido  
*Isabella Gonçalves & Yossi David (Germany)*
05. When negativity is the fuel. Bots and Political Polarization in the COVID-19 debate ..... 59-71  
 Cuando la negatividad es el combustible. Bots y polarización política en el debate sobre el COVID-19  
*José-Manuel Robles, Juan-Antonio Guevara, Belén Casas-Mas & Daniel Gómez (Spain)*

## KALEIDOSCOPE

06. Twitter and human trafficking: Purposes, actors and topics in the Spanish-speaking scene ..... 75-87  
 La trata de personas en Twitter: Finalidades, actores y temas en la escena hispanohablante  
*Alba Sierra-Rodríguez, Wenceslao Arroyo-Machado & Domingo Barroso-Hurtado (Spain)*
07. Special Education Teacher's professional development through digital storytelling ..... 89-99  
 Desarrollo profesional de maestros de educación especial a través de la narración digital  
*Ozgur Yasar-Akyar, C. Rosa-Feliz, Solomon Sunday-Oyelere, Darwin Muñoz & Gyasettin Demirhan (Turkey, Dominican Republic & Sweden)*
08. Detection of traits in students with suicidal tendencies on Internet applying Web Mining ..... 101-112  
 Detección de rasgos en estudiantes con tendencia suicida en Internet aplicando Minería Web  
*Iván Castillo-Zúñiga, Francisco-Javier Luna-Rosas & Jaime-Iván López-Veyna (Mexico)*
09. Booktokers: Generating and sharing book content on TikTok ..... 113-123  
 Booktokers: Generar y compartir contenidos sobre libros a través de TikTok  
*Nataly Guíñez-Cabrera & Katherine Mansilla-Obando (Chile)*
10. The relationship of Twitter with teacher credibility and motivation in university students ..... 125-136  
 Relación de la red social Twitter con la credibilidad docente y la motivación del alumnado universitario  
*Facundo Froment, Alfonso-Javier García-González & Julio Cabero-Almenara (Spain)*



## SUBMISSION GUIDELINES

### GENERAL INFORMATION

'Comunicar', Media Education Research Journal is published by Grupo Comunicar Ediciones (VAT: G21116603). This established non-profit professional group, founded in 1988 in Spain, specialises in the field of media education. The journal has been in print continuously since 1994, published every three months.

Contents are peer reviewed, in accordance with publication standards established in the APA 7 (American Psychological Association) manual. Compliance with these requirements facilitates indexation in the main databases of international journals in this field, which increases the dissemination of published papers and therefore raises the profile of the authors and their centres.

'Comunicar' is indexed in the Social Sciences Citation Index (SSCI), Journal Citation Reports (JCR), Scisearch, Scopus and over 790 databases, catalogues, search engines and international repertoires worldwide.

Each issue of the journal comes in a print (ISSN:134-3478) and electronic format ([www.comunicarjournal.com](http://www.comunicarjournal.com)) (e-ISSN: 1988-3293), identifying each submission with a DOI (Digital Object Identifier System).

### SCOPE AND POLICY

**Subject Matter:** Fundamentally, research papers related to communication and education, and especially the intersection between the two fields: media education, educational media and resources, educational technology, IT and electronic resources, audiovisual, technologies... Reports, studies and experiments relating to these subjects are also accepted.

**Contributions:** 'Comunicar' publishes research results, studies, state-of-the-art articles and bibliographic reviews especially in relation to Latin America and Europe and regarding the convergence between education and communication, preferably written in Spanish although submissions are also accepted in English. The contributions to this journal may be: Research papers, Reports, Studies and Proposals (5,000-6,700 words of text, references included), State-of-the-art articles: (6,000-7,200 words of text, including references).

Unsolicited manuscripts sent in by authors are initially placed in the Miscellaneous section of the journal. The Topics section is organized by an editor through a system of call for papers and specific commissions to experts in the field. If we receive manuscripts within the deadline for a particular topic, the journal editor can refer the manuscript to the Topics editor for assessment and possible publication in this monographic section. The deadline for each Topic section is at least nine months before publication.

### EDITORIAL PROCESS

'Comunicar' confirms receipt of all articles submitted by authors and keeps the authors informed of the acceptance/rejection process, as well as the editing process, in the event of acceptance. The webpage [www.comunicarjournal.com](http://www.comunicarjournal.com) also provides information about manuscript editing procedures, full guidelines for publication, as well as the Annexes: review protocol prior to submission, manuscript structure protocol, external review protocol...

In general, once the external reports have been reviewed, the criteria that justify the editors' decision to accept or reject submissions are as follows: a) Topical and new; b) Relevance: applicability of the results to the resolution of specific problems; c) Originality: valuable information, repetition of known results; d) Significance: advancement of scientific knowledge; e) Reliability and scientific validity: verified methodological quality; f) Organisation (logical coherence and material presentation); g) Presentation: good writing style.

### PRESENTATION AND STRUCTURE OF ORIGINAL PAPERS

Manuscripts must be sent exclusively through the Journal Management Centre ([www.revistacomunicar.com/ojs/](http://www.revistacomunicar.com/ojs/)). These publication guidelines are based on the standards of the American Psychological Association (APA 7): (<https://bit.ly/3cN4XX9>).

**Structure:** The following two files must be sent together: manuscript (main text) and cover letter.

### ETHICAL COMMITMENT AND RESPONSIBILITIES

Each author must submit a statement of authorship and text originality. Previously published material will not be accepted. The cover letter must specify the transfer of copyright ownership of the manuscript for its publication in 'Comunicar'.

[www.comunicarjournal.com](http://www.comunicarjournal.com)

### EDITORIAL BOARD

#### EDITOR IN CHIEF

- Dr. Ignacio Aguaded, University of Huelva, Spain

#### ASSISTANT EDITORS

- Dr. Amor Perez-Rodriguez, University of Huelva, Spain
- Dr. Angel Hernando-Gomez, University of Huelva, Spain
- Ms. Arantxa Vizcaino-Verdu, University of Huelva, Spain
- Dr. Rafael Repiso, UNIR / EC3, University of Granada, Spain
- Dr. Luis-Miguel Romero-Rodriguez, URJC, Spain
- Dr. Rosa Garcia-Ruiz, University of Cantabria, Spain
- Dr. Agueda Delgado-Ponce, University of Huelva, Spain

#### THEMATIC EDITORS

- Dr. M<sup>a</sup> Dolores Cáceres-Zapatero, Complutense Uni. of Madrid, Spain
- Dr. Mykola Makhortyk, University of Bern, Switzerland
- Dr. Francisco Segado-Boj, Complutense University of Madrid, Spain

#### INTERNATIONAL COEDITORS

- **English:** Dr. M. Gant, Univ. Chester and Dr. C. Herrero (MMU), UK
- **Portuguese:** Dr. Vanessa Matos, Univ. Fed. Uberlândia, Brazil
- **Chinese:** Dr. Alice Lee (Hong Kong), Dr. Meng Shen (Spain) and Dr. Yuechuan Ke (USA)
- **Iberoamerican:** Dr. Octavio Islas (Ecuador)
- **Russian:** Dr. Alexander Fedorov (Russia) and Dr. Margarita Bakieva (Spain)

### ADVISORY BOARD

- Dr. Ismar de-Oliveira, University of São Paulo, Brazil
- Dr. Miguel de-Aguilera, University of Malaga, Spain
- Dr. J. Manuel Perez-Tornero, Autonomous University, Barcelona
- Dr. Guillermo Orozco, University of Guadalajara, Mexico
- Dr. Manuel Angel Vazquez-Medel, University of Seville, Spain
- Dr. Cecilia Von-Felitzen, Nordicom, Sweden
- Dr. Joan Ferres, Pompeu Fabra Univer., Barcelona, Spain
- Dr. Agustin Garcia-Matilla, University of Valladolid, Spain
- Dr. Cristina Ponte, Nova University Lisbon, Portugal
- Dr. Pier Cesare Rivoltella, Catholic University of Milan, Italy
- Dr. Javier Marzal, Jaume I University, Castellon, Spain
- Dr. Jesus Arroyave, University of the North, Colombia
- Dr. Francisco Garcia-Garcia, Complutense Univer., Madrid, Spain
- Dr. Alberto Parola, MED, University of Turin, Italia
- Dr. Teresa Quiroz, University of Lima, Peru
- Dr. Concepcion Medrano, University of País Vasco, Spain
- Dr. Maria Luisa Sevillano, National Uni. of Distance Education, Spain
- Dr. Julio Cabero-Almenara, University of Seville, Spain
- Dr. Manuel Cebrian-de-la-Serna, University of Malaga, Spain
- Dr. Ana Garcia-Valcarcel, University of Salamanca, Spain
- Dr. M. Soledad Ramirez-Montoya, Monterrey Institute, Mexico
- Dr. Donaciano Bartolome, Complutense University, Madrid, Spain
- Dr. Samy Tayie, University of Cairo, Egypt
- Dr. Javier Tejedor-Tejedor, University of Salamanca, Spain
- Dr. Sara Pereira, University of Minho, Braga, Portugal
- Dr. Gloria Camarero, Carlos III University, Madrid, Spain
- Dr. Arminda Pinto, University of Coimbra, Portugal
- Dr. Pere Marques, Autonomous University of Barcelona, Spain
- Dr. Xose Soengas, University of Santiago, Spain
- Dr. Octavio Islas, Pontifical Catholic University, Ecuador
- Dr. Don Shin, Chung-Ang University, Seoul, Korea

### ADVISORY BOARD

- Dr. Moises Esteban-Guitert, University of Gerona
- Dr. Patrick Verniers, Conseil Sup. Media Education, Belgium
- Dr. Domingo Gallego, National Uni. of Distance Education, Spain
- Dr. Manuel Area, La Laguna University, Tenerife, Spain
- Dr. Ramon Reig, University of Seville, Spain
- Dr. Gustavo Hernandez, ININCO, Central University, Venezuela
- Dr. Isabel Canton, University of Leon, Spain
- Dr. Juan de Pablos, University of Seville, Spain
- Dr. Gerardo Borroto, CUJAE, La Habana, Cuba
- Dr. Manuel Fandos-Igodo, UNIR, Zaragoza, Spain
- Dr. Jorge Cortes-Montalvo, UACH/REDECA, Mexico
- Dr. Carmen Marta, University of Zaragoza, Spain
- Dr. Silvia Contin, National University of Patagonia, Argentina
- Dr. Begoña Gutierrez, University of Salamanca, Spain
- Dr. Ramón Perez-Perez, University of Oviedo, Spain
- Dr. Carlos Muñoz, Autonomous University of Nuevo Leon, Mexico
- Dr. Carmen Echazarreta, University of Gerona, Spain
- Dr. Evgeny Pashentsev, Lomonosov Moscow University, Russia
- Dr. Fahriye Altinay, Near East University, Turkey
- Dr. Jesus Valverde, University of Extremadura, Spain
- Dr. Yamile Sandoval, University of Santiago of Cali, Colombia
- Dr. Pilar Amaiz, University of Murcia, Spain
- M. Paolo Celot, EAVI, Brussels, Belgium
- Dr. Victoria Tur Viñes, University of Alicante, Spain
- Dr. Jose-Maria Morillas, University of Huelva, Spain
- M. Jordi Torrent, ONU, Alliance of Civilizations, NY, USA
- Ms. Kathleen Tyner, University of Texas, Austin, USA

### INTERNATIONAL REVIEWERS BOARD

- 1,020 Reviewers of 54 countries (2022-2)

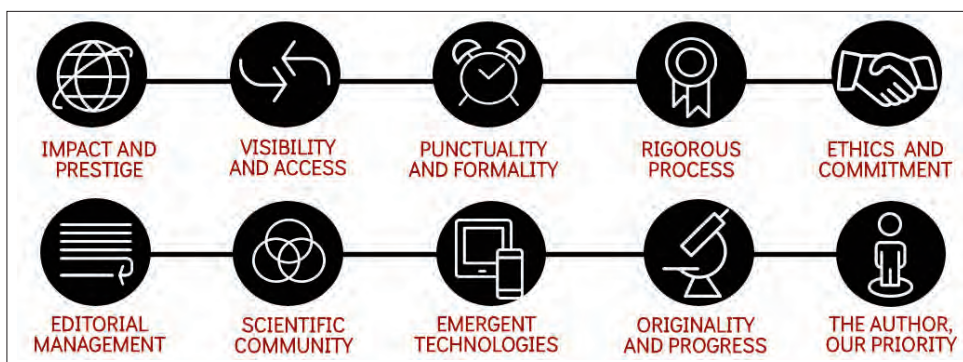
[www.revistacomunicar.com/index.php?contenido=revisores&idioma=en](http://www.revistacomunicar.com/index.php?contenido=revisores&idioma=en)

### BOARD OF MANAGEMENT

- D. Francisco Casado-Mestre, University of Huelva, Spain
- Dr. Patricia De-Casas-Moreno, University of Extremadura, Spain
- Ms. Daniela Jaramillo-Dent, University of Huelva, Spain
- Ms. Mónica Bonilla-Del-Río, University of Huelva, Spain
- Ms. Bárbara Castillo-Abdul, University of Huelva, Spain
- Ms. Sabina Civilá, University of Huelva, Spain
- Dr. Erika-Lucia González-Carrión, Loja National Uni., Ecuador
- Dr. Paloma Contreras-Pulido, UNIR, Spain
- **ICT CONSULTANT:** Alex Ruiz, Spain
- **COMMERCIAL MANAGER:** Belén Pérez, Spain



### QUALITY CRITERIA



The quality criteria of 'Comunicar' Media Education Research Journal are the indicators of our commitment to a rigorous and professional publishing process. They constitute the endorsement for authors, reviewers, collaborators and readers that the final product offered meets the highest standards and academic rigour expected from a publication of national and international reputation.

- **Impact and prestige** are guaranteed by being in the first quartile among the most recognised journals: Q1 in Journal Citation Reports (JCR-SSCI) in Communication and Education; Q1 in Scopus (SJR) in Cultural Studies and Q1 in Scopus (CiteScore) in Communication and Education.

- **Visibility and access** are safeguarded by an open and shared access policy that makes all published manuscripts available to any reader in a bilingual version: Spanish and English.

- **Punctuality and formality** contribute to an efficient flow of manuscripts within established timeframes, facilitating quarterly publication thanks to highly effective schedule compliance.

- A **rigorous process** is supported by an International Reviewers Board of nearly 996 highly qualified researchers in the fields of education and communication from almost 53 countries all over the world.

- **Ethics and commitment** that ensure the upholding of rights and duties that protect the whole academic community: authors, reviewers, readers and editors. Our Code of Ethics is based on the International Committee on Publication Ethics (COPE), ensuring guidance for authors.

- **High quality Editorial Management** is accomplished throughout the OJS Platform, from the Science and Technology Foundation (FECYT), complemented by reviewers' contributions and stylistic department work in English and Spanish languages.

- Our **scientific community** is fostered and increased by an influential academic community of authors (1,957), reviewers (996) and readers all over the world. This Community is supported by a committed editorial group composed of: an Editorial Board, an Academic Committee, an International Reviewers Board, an Advisory Board and a Technical Board.

- Use and development of **emergent technologies** encourages the diffusion and impact of the journal, continuously adjusting the access formats, communication models and academic platforms for academic diffusion (ResearchGate, AcademiaEdu...).

- The **originality and progress** of the outstanding research contributions from the research areas of communication and education are guaranteed by plagiarism controls (CrossCheck), applied to all manuscripts accepted before being published to preserve the originality and progress of research.

- At last but not least the **author is our priority**, since after all, authors uphold and give sense to entire process. Every manuscript is available in the Journal website with accurate information about citation, statistical data, references, impact metrics and interaction in social media.

Quality criteria are, in summary, a set of standards that guarantee the whole process, ensuring a professional treatment for every person involved in the publishing, reviewing, editing and spreading processes of the manuscripts.

### Information on evaluators, acceptance/rejection rates and internationalisation in Comunicar 71

- Number of research works received: 249. Number of research works accepted: 10.
- Percent of manuscripts accepted: 4.02%; Percent of manuscript rejected: 95.98%.
- Received manuscripts internationalisation: 43 countries.
- Numbers of Reviews: 326 (93 internationals and 233 nationals) (update: [www.comunicarjournal.com](http://www.comunicarjournal.com)).
- Scientific Reviewers internationalisation: 26 countries.
- Country of origin: 8 countries (Chile, Dominican Republic, Germany, Mexico, Spain, Sweden, Switzerland & Turkey).
- International databases in COMUNICAR 71: 805 (2022-02) (update: [www.comunicarjournal.com](http://www.comunicarjournal.com)).



# Comunicar 71



## Special issue

Hate speech in communication:  
Research and proposals





**Q1  
2020**

# Comunicar

 **Clarivate Analytics**  
JOURNAL CITATION REPORTS

 **Scopus®**

Journal Impact Factor (JIF)	
7 <sup>th</sup> / 264 Education	2 <sup>nd</sup> / 1,037 Cultural Studies
9 <sup>th</sup> / 94 Communication	14 <sup>th</sup> / 1,319 Communication
Journal Citation Indicator (JCI)	
11 <sup>th</sup> / 725 Education	24 <sup>th</sup> / 426 Education
9 <sup>th</sup> / 207 Communication	





# Adolescents' motivations to perpetrate hate speech and links with social norms

## Motivos del discurso de odio en la adolescencia y su relación con las normas sociales

- id** Dr. Sebastian Wachs. Research Associate, Department of Educational Studies, University of Potsdam (Germany) (wachs@uni-potsdam.de) (<https://orcid.org/0000-0003-2787-6646>)
- id** Dr. Alexander Wettstein. Professor, Institute for Research, Development and Evaluation, Bern University of Teacher Education (Switzerland) (alexander.wettstein@phbern.ch) (<https://orcid.org/0000-0002-5612-3144>)
- id** Dr. Ludwig Bilz. Professor, Department of Health Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Senftenberg (Germany) (ludwig.bilz@b-tu.de) (<https://orcid.org/0000-0003-3017-0492>)
- id** Dr. Manuel Gámez-Guadix. Professor, Department of Biological and Health Psychology, Autonomous University of Madrid (Spain) (manuel.gomez@uam.es) (<https://orcid.org/0000-0002-1575-1662>)

### ABSTRACT

Hate speech has become a widespread phenomenon, however, it remains largely unclear why adolescents engage in it and which factors are associated with their motivations for perpetrating hate speech. To this end, we developed the multidimensional "Motivations for Hate Speech Perpetration Scale" (MHATE) and evaluated the psychometric properties. We also explored the associations between social norms and adolescents' motivations for hate speech perpetration. The sample consisted of 346 adolescents from Switzerland (54.6% boys;  $M_{age}=14$ ;  $SD=0.96$ ) who reported engagement in hate speech as perpetrators. The analyses revealed good psychometric properties for the MHATE, including good internal consistency. The most frequently endorsed subscale was revenge, followed by ideology, group conformity, status enhancement, exhilaration, and power. The results also showed that descriptive norms and peer pressure were related to a wide range of different motivations for perpetrating hate speech. Injunctive norms, however, were only associated with power. In conclusion, findings indicate that hate speech fulfills various functions. We argue that knowing the specific motivations that underlie hate speech could help us derive individually tailored prevention strategies (e.g., anger management, promoting an inclusive classroom climate). Furthermore, we suggest that practitioners working in the field of hate speech prevention give special attention to social norms surrounding adolescents.

### RESUMEN

El discurso de odio se ha convertido en un fenómeno generalizado. Sin embargo, todavía no está claro por qué los adolescentes se involucran en el discurso de odio y qué factores están asociados con las motivaciones para perpetrarlo. Con esta finalidad, desarrollamos una medida multidimensional, la «Escala de Motivaciones para Perpetrar Discurso de Odio» (MHATE), y evaluamos sus propiedades psicométricas. Asimismo, investigamos las asociaciones entre las normas sociales y las motivaciones para participar en el discurso de odio. La muestra estuvo compuesta por 346 adolescentes suizos (54,6% chicos;  $M_{edad}=14$ ;  $DT=0,96$ ) que informaron haber perpetrado discurso de odio. Los análisis revelaron buenas propiedades psicométricas de MHATE, incluyendo adecuada consistencia interna. La subescala con mayor frecuencia fue venganza, seguida de las de ideología, conformidad con el grupo, mejora del estatus, regocijo y poder. Las normas descriptivas y la presión de iguales estuvieron relacionadas con varias motivaciones para perpetrar discurso de odio. Las normas prescriptivas, sin embargo, solo se asociaron con el poder. En conclusión, los hallazgos indican que el discurso de odio cumple varias funciones. Conocer las motivaciones específicas para el discurso de odio ayuda a derivar estrategias de intervención individualmente adaptadas (ej., manejo de la ira, promover un clima escolar inclusivo). Además, sugerimos que los profesionales que trabajan en la prevención del discurso de odio presten especial atención a las normas sociales que rodean a los adolescentes.

### KEYWORDS | PALABRAS CLAVE

Hate speech, cyberhate, motives, social norms, injunctive norms, peer pressure.  
Discurso de odio, ciberodio, motivos, normas sociales, normas cautelares, presión de los pares.



## 1. Introduction

Hate speech can be defined as any communicative form of expression that deliberately promotes, justifies, or disseminates the exclusion, contempt, or devaluation of particular social groups (e.g., due to ethnicity, nationality, religion, sexual orientation, disability, gender etc. Lehman, 2020; Wachs et al., 2020). We know that adolescents are not only the witnesses and victims of hate speech but also that they actively engage in hate speech as perpetrators (Wachs et al., 2021). But there is an important gap in the literature regarding the motivations why adolescents perpetrate hate speech and which factors are associated with the self-reported motivations for devaluing and harassing particular social groups (Ballaschk et al., 2021). Furthermore, not much is known about factors that influence adolescents' motivations for perpetrating hate speech. However, the relevance of social norms for adolescents' deviant behavior and prejudices has been well-established (Allport, 1954; Van-de-Bongardt et al., 2015; Váradi et al., 2021). Whether or not social norms also drive adolescents' motivation to perpetrate hate speech, however, has not yet been investigated. Thus, the present study was conducted to fill these gaps in the literature.

Pinker (2011) developed a taxonomy of motivations for people's violent behavior that includes revenge, power, sadism, ideology, and practical violence. As hate speech can be considered a specific form of violence, it is likely that hate speech has comparable motivations. Ballaschk et al. (2021) found—based on qualitative interviews with students, teachers, and school social workers—evidence for similar causes of hate speech as those proposed by Pinker (2011). Additionally, their analyses highlighted the relevance of peer group conformity and status enhancement among peers as additional motivations that should be considered. We will therefore combine the findings of Pinker (2011) and Ballaschk et al. (2021) as the basis of our investigation into the motivations for hate speech perpetration among adolescents.

The drive to revenge oneself is a natural human impulse (Pinker, 2011). Some adolescents may use hate speech as a means of revenge when they feel threatened, unfairly treated, frustrated, or angry. Indeed, feelings of frustration and inferiority have been found to be positively related to hate speech perpetration among adolescents (Ballaschk et al., 2021). In addition to this, research has shown that hate speech victimization and perpetration are correlated, which might also indicate that some victims intend to take revenge on their perpetrators or members of other social groups (Wachs & Wright, 2021). The notion that revenge is a cause of hate speech is also supported by research that has highlighted that revenge is particularly relevant in traditional bullying and cyberbullying perpetration (Fluck, 2017; Runions et al., 2018; Tanrikulu & Erdur-Baker, 2021). Power can be understood as an individual's ability to influence others, to experience themselves as strong, and desire for dominance and control over others (McClelland, 1975; Pinker, 2011).

The desire for power is centered on experiencing strength and self-confidence and is countered by the fear of losing this power (Sokolowski & Heckhausen, 2010). It is well known that violence can be traced back to a need to exercise power (Gradinger et al., 2012; Tanrikulu & Erdur-Baker, 2021; Wettstein, 2008). And yet, there is limited evidence about the role of power in hate speech perpetration. Only one study indicates that a desire for power is related to hate speech perpetration among adolescents (Ballaschk et al., 2021). Pinker (2011) cited sadism as a motivation for perpetrating violence, describing it as a feeling of pleasure from watching another person's suffering, discomfort, or pain. To avoid any confusion with the psychological disorder of sadism, we use the term exhilaration in the present study. There is some empirical evidence that exhilaration can explain why people engage in hate speech (Ballaschk et al., 2021; Erjavec & Kovač, 2012). Exhilaration has also been shown to be a relevant cause of violence in research on aggression and bullying (Gradinger et al., 2012; Wettstein, 2008).

Hate speech is rooted in in- and out-group processes (Ballaschk et al., 2021). The social groups that people feel part of and identify with are the in-groups, whereas social groups people do not feel a part of or cannot identify with are the out-groups. The positive evaluation of the in-group often goes hand-in-hand with a tendency to devalue the out-group (Tajfel & Turner, 2004). Members of out-groups are generally assigned negative characteristics and members of the in-group feel contempt towards those not included in it. This rejection is often justified by inhumane and discriminatory positions and threatening behaviors. This in-group favoritism can lead to violence, discriminatory behavior, and hate speech between members of in- and out-groups (Tajfel & Turner, 2004). Inhumane worldviews are also prevalent among adolescents

(Möller et al., 2016). Hate speech perpetrators normally target people who are assigned to the out-group and label these groups as enemies while trying to defend their political and ideological values against others through hatred (Erjavec & Kovač, 2012). More recent research on online hate speech has identified ideological beliefs as the driving force behind such behavior (Ballaschk et al., 2021). Whereas Fluck (2017) found that ideology was not a relevant cause for perpetrating bullying, ideology might play a major role for hate speech perpetration due to the reasons outlined here. Adolescents seek a sense of belonging (Baumeister & Leary, 1995) and acceptance from their peer groups, leading to conforming behavior with existing group norms, even if those norms are anti-social and discriminatory against certain groups (Allport, 1954).

Hence, it can be assumed that group conformity might also drive adolescents to perpetrate hate speech. This assumption is supported by initial research on hate speech (Ballaschk et al., 2021) and cyberbullying (Gradinger et al., 2012), but clearly warrants more empirical evidence. Pinker (2011) noted that people use violence to achieve a goal they cannot reach through non-violent means and labelled this practical or predatory violence. Based on Ballaschk et al.'s (2021) findings, we will operationalize this motivation as a form of status enhancement among peers, as this study revealed that hate speech is used instrumentally to enhance one's own status within the peer group. Peer groups are hierarchically organized systems in which adolescents strive for high social status (Salmivalli & Peets, 2009). Studies on bullying also show that the pursuit of high-status positions within the peer group is an essential motivation for bullying (Olthoff et al., 2011; Salmivalli & Peets, 2009).

In sum, it can be assumed that there are various motivations for hate speech perpetration. Only one qualitative study, however, has investigated the motivations for perpetrating hate speech among adolescents. Consequently, no validated scale exists that measures motivations for hate speech perpetration as a multidimensional construct. And in the only existing study in which the participants reported why they believe adolescents perpetrate hate speech, no hate speech perpetrators were interviewed. We are thus missing research that investigates self-reported motivations for engaging in hate speech. Our first research aim was consequently to test the psychometric properties of a newly developed multidimensional scale for measuring motivations for hate speech perpetration.

Three types of social norms can be distinguished. Firstly, injunctive norms, which describe those that inform us about people's perception of attitudes, i.e., what behaviors are typically approved or disapproved by others. Then descriptive norms, which reflect people's perceptions of behaviors typically performed by others. And finally, peer pressure, which can be defined as peers' active encouragement to exhibit a certain behavior (Cialdini & Trost, 1998; Van-de-Bongardt et al., 2015). In his seminal work on the nature of prejudice, Allport (1954) argued that the majority of negative attitudes and feelings towards social groups arise from social norms and attempts to conform to them. What has not yet been the subject of empirical research is the question of whether social norms might also be relevant variables in explaining the motivations for hate speech perpetration. However, initial research revealed a positive relation between social norms and hate speech perpetration (Ballaschk et al., 2021; Wachs et al., in press). In addition, related research showed a positive association between social norms, prejudices, and aggression (Bastiaensens et al., 2016; Cook et al., 2010; Lee & Wong, 2009; Váradi et al., 2021). To add to the current knowledge on hate speech perpetration, the second aim of the present study was to investigate associations between social norms and adolescents' motivations for perpetrating hate speech.

## 2. Material and methods

The total sample consisted of 1,381 participants (44.5% boys, 54% girls, 1.4% gender diverse) between 11 and 18 years old ( $M_{age}=13.92$ ;  $SD=0.98$ ) from 7<sup>th</sup> to 9<sup>th</sup> grade (7<sup>th</sup> grade: 29.9%,  $n=413$ ; 8<sup>th</sup> grade: 27.1%,  $n=374$ ; 9<sup>th</sup> grade: 27.8%,  $n=384$ ; age-mixed grade: 15.2%,  $n=210$ ) from 119 classes of 22 German-speaking schools across two cantons in Switzerland. Based on prior qualitative research (Ballaschk et al., 2021; Krause et al., 2021), an item was created to measure the frequency of hate speech perpetration in schools. The participants were first shown a short video presenting a definition of hate speech. Afterwards they were asked to rate the following item, "In the past 12 months how often have you perpetrated hate speech at your school?", on a five-point scale ("never", "1 or 2 times within the last



month”, “2 or 3 times per month”, “about once a week”, “several times a week”). If students answered that they had perpetrated hate speech at least once, they were asked for their motivations. The obtained subsample included 346 adolescents (54.6% boys, 43.6% girls, 1.7% gender diverse) between the ages of 12 and 18 ( $M_{age}=14$ ;  $SD=0.96$ ) from 106 classes in 19 schools. Regarding migration background, 52.3% ( $n=165$ ) reported that they had a migration background. Concerning socio-economic status (SES), 31.2% ( $n=108$ ) of participants reported living in families of low affluence, 34.4% ( $n=119$ ) in families of medium affluence, and 34.4% ( $n=119$ ) in families of high affluence.

As no existing scale to investigate the motivations for hate speech perpetration was available, we developed the Motivations for Hate Speech Perpetration Scale (MHATE). To develop the MHATE, qualitative interviews with students and school staff were conducted focused on the motivations why students engaged in hate speech perpetration at school (Ballaschk et al., 2021). We then reviewed the literature on motives for perpetrating aggression, bullying, and cyberbullying. Based on these findings, a set of 12 items was developed, which were assumed to reflect six different motivations, namely revenge, power, exhilaration, ideology, group conformity, and status enhancement. Table 1 gives a full overview of all items. Each item could be rated on a five-point scale from “absolutely disagree” (1) to “absolutely agree” (5).

Descriptive hate speech norms were measured by asking participants if they had witnessed hate speech. The scale was developed based on qualitative research (Ballaschk et al., 2021; Krause et al., 2021) and in reference to existing instruments (Reichelmann et al., 2020). Items included, for example: “Someone distributed discriminatory symbols, stickers, pictures, memes, or videos against a specific group of people,” and: “Someone made insulting jokes about a certain group of people.” Possible answers ranged from “never” (1) to “often” (4). The McDonald’s  $\omega$  was good: .83. Perceived injunctive anti-hate speech norms were measured using three items, which the participants were asked to rate after reading a text-based vignette describing an online or offline hate speech scenario. The three items were: “My close friends don’t like it when you say that about other people,” “My family doesn’t like it when you say this about other people,” and “The teachers at my school don’t like it when you say that about other people.” All items could be answered on a five-point scale from “absolutely disagree” (1) to “absolutely agree” (5). The McDonald’s  $\omega$  was acceptable: .74. To measure deviant peer pressure, the following four items were modified from Santor et al. (2000): “At times, I’ve broken rules because classmates have urged me to,” “At times, I’ve done dangerous or foolish things because classmates dared me to,” “I’ve skipped classes, when classmates have urged me to,” and “If a group of classmates at school asked me to do something forbidden, it would be difficult to say no.” All items could be answered on a five-point scale from “absolutely disagree” (1) to “absolutely agree” (5). The McDonald’s  $\omega$  was acceptable: .78.

Participants were asked for their age and gender (male, female, diverse). Grade was added afterwards according to class codes that were used to fill out the questionnaires. Migration background (the standard measure of ethnic diversity in German-speaking countries) was assessed by asking whether the participants themselves, one, or both parents were born in a country other than Switzerland. The socio-economic status (SES) was measured using the Family Affluence Scale (FAS; Hartley et al., 2016). Based on a composite FAS score, an individual FAS category was calculated for each participant (low, medium, and high socioeconomic status). After obtaining ethical clearance from the University of Potsdam Ethics Committee (UP65/2018), 22 Swiss schools were invited to take part in this study. From these 22 schools, 20 agreed to participate in the study (participation rate at the school level: 91%). From 2,593 invited students, 1,381 participated (participation rate at the individual level: 53%). The students and their parent or legal guardian had to sign a written consent form to participate in the study. Data collection took place via an online survey during a school lesson between December 2020 and April 2021. Participants received an access code to the survey via e-mail and subsequently completed the online questionnaire. Participants were told that partaking in the study was optional and participation in the survey could be stopped at any time, without needing to give a reason and without fear of negative consequences. Completion time was 37 minutes on average.

To investigate the construct validity of the MHATE, we conducted confirmatory factor analyses (CFA) and analyzed composite reliability (CR). We then conducted a series of multi-group confirmatory

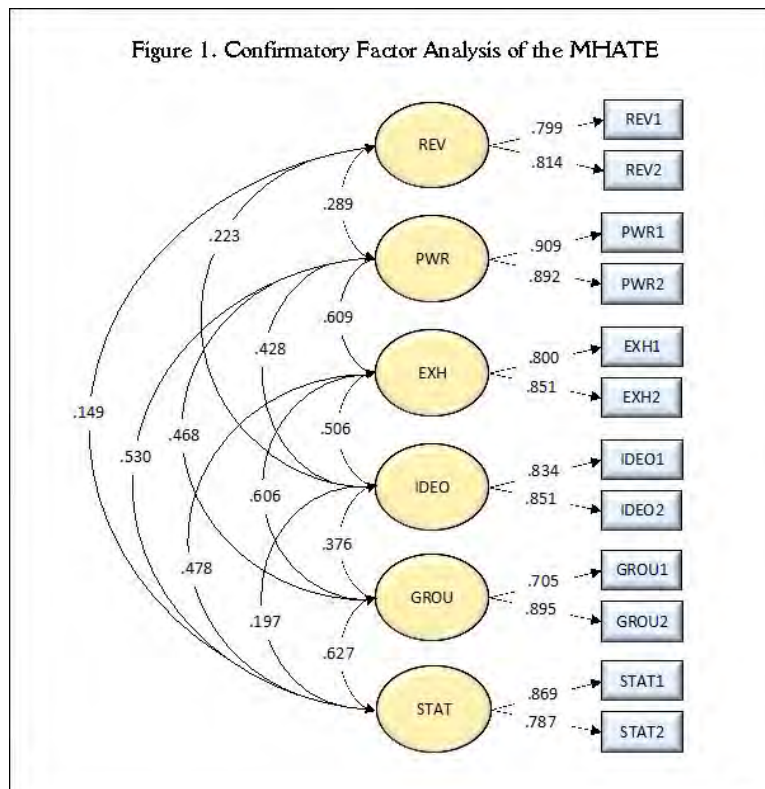
factor analyses (MGCFA) maximum likelihood parameter estimates with standard errors (MLR) in Mplus version 8.3 (Muthén & Muthén, 2017) to analyze the measurement invariance of the MHATE among different demographic groups. To evaluate the measurement invariance assumptions, Chen's (2007) recommendations were followed, according to which decreases in  $\Delta CFI > 0.010$  and increases in  $\Delta RMSEA > 0.015$  indicate that the assumption of measurement invariance is not met. If scalar measurement invariance is established, the meaning of the items can be considered as equal across the groups and latent mean differences can be compared (Van-de-Schoot et al., 2012). Cohen's *d* was used as a measure of effect sizes for latent factor means. Cohen's *d* was calculated by dividing the difference between factor means by the pooled factor standard deviation.

Finally, structural equation modelling was used to test the associations between social norms and the MHATE. The model fit was examined by considering the following fit indices: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). The quality of each model was evaluated using typical cut-off scores reflecting good and adequate fit of the data, respectively: CFI and TLI  $> .95$  and  $.90$ ; RMSEA  $< .06$  and  $.08$ , and SRMR  $< .10$  and  $.05$  (Hu & Bentler, 1999). To account for the multilevel structure of the data (i.e., adolescents nested within schools), standard errors were corrected by using the complex sampling option () in Mplus (Muthén & Muthén, 2017).

### 3. Results

#### 3.1. Construct validity of the motivations for perpetrating hate speech scale

Regarding hate speech perpetration, 74.9% ( $n=1,035$ ) of participants answered that they had never engaged in this behavior; 16.4% ( $n=227$ ) reported this behavior one or two times within the last month; 4.6% ( $n=63$ ) two or three times per month, 1.7% ( $n=23$ ); about once a week; and 2.4% ( $n=33$ ) several times a week.



Note. Only significant correlations are shown. REV=revenge, PWR=power, EXH=exhilaration, IDEO=ideology, GROU=group conformity, STAT=status enhancement.  $N=346$ .

For the following analyses, only the subsample of hate speech perpetrators ( $n=346$ ) was used. First, we tested a model in which all 12 items of the MHATE loaded on one single factor. Results showed that the data did not fit well,  $\chi^2=853.64$ ,  $df=54$ ,  $\chi^2/df=15.52$ ,  $p<.001$ ,  $CFI=.564$ ,  $TLI=0.467$ ,  $RMSEA=0.207$ ,  $SRMR=0.115$ , with standardized factor loadings between 0.18 and 0.70. Then we tested a second model with the proposed six-factor structure of the MHATE. The second model showed a good model fit,  $\chi^2=90.49$ ,  $df=40$ ,  $\chi^2/df=2.26$ ,  $p<.001$ ,  $CFI=.972$ ,  $TLI=0.955$ ,  $RMSEA=0.060$ ,  $SRMR=0.034$ . All standardized factor loadings were significant ( $p<.001$ ) and ranged between 0.78 and 0.90 (Figure 1). The CR was .79 for revenge, .90 for power, .81 for exhilaration, .76 for ideology, .79 for group conformity, and .82 for status enhancement, indicating evidence for factor validity of the MHATE.

### 3.2. Frequency rates of motivations for hate speech perpetration

Table 1 gives a full summary of frequencies, M, and SD of all items and the subscales of the MHATE. The three most frequently reported motivations were: "Because I was made angry by others" (38.4%; REV1), "Because I was hurt or annoyed by others" (33.8%; REV2), and "Because that is the way we talk to each other in my class" (25.4%; GROU1). The three least frequently reported motivations were: "To show who is the boss" (5.7%; PWR2), "To be respected and belong to the cool kids" (7%; STAT1), and "Because it feels good" (7.8%; EXH2). The most frequently endorsed subscale was revenge ( $M=2.78$ ;  $SD=1.34$ ), followed by ideology ( $M=2.17$ ;  $SD=1.15$ ), group conformity ( $M=2.02$ ;  $SD=1.12$ ), status enhancement ( $M=1.73$ ;  $SD=1.03$ ), exhilaration ( $M=1.73$ ;  $SD=1.05$ ), and power ( $M=1.59$ ;  $SD=0.94$ ).

**Table 1. Percentage, means and standard deviation of each item of the MHATE**

Items	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Item	Subscale
	%	%	%	%	%	M (SD)	M (SD)
Because I was made angry by others. (REV1)	32	9.8	19.7	19.9	18.5	2.82 (1.51)	2.78 (1.34)
Because I was hurt or annoyed by others. (REV2)	30.3	15.3	20.5	17.9	15.9	2.73 (1.45)	
To show who is in charge. (PWR1)	64.2	16.5	11.3	5.2	2.9	1.66 (1.05)	1.59 (0.94)
To show who is the boss. (PWR2)	68.2	17.9	8.1	4	1.7	1.53 (0.92)	
Because I think it was funny. (EXH1)	59.8	15.9	11	5.8	7.5	1.85 (1.26)	1.73 (1.05)
Because it feels good. (EXH2)	65.9	15.5	9.8	4.9	2.9	1.62 (1.03)	
Because it corresponds to my beliefs. (IDEO1)	58.4	18.2	11.6	8.7	3.2	1.80 (1.13)	2.17 (1.15)
Because it is my opinion. (IDEO2)	35.8	15	18.2	19.4	11.6	2.55 (1.43)	
Because that is the way we talk to each other in my class. (GROU1)	43.9	13	17.6	15.9	9.5	2.34 (1.41)	2.02 (1.12)
Because this behavior is expected in my class. (GROU2)	60.4	20.5	10.4	5.8	2.9	1.70 (1.05)	
To be respected and belong to the cool kids. (STAT1)	66.8	15.6	10.7	3.5	3.5	1.61 (1.03)	1.73 (1.03)
To belong and not become a victim myself. (STAT2)	56.4	18.5	13.6	5.8	5.8	1.86 (1.19)	

### 3.3. Measurement invariance testing and factor mean differences

Overall, for all group comparisons (i.e., by gender, grade, migration background, and SES) no substantial reduction in model fit (i.e.,  $\Delta CFI>0.010$  and  $\Delta RMSEA>0.015$ , Chen 2007) between the configural and metric invariance model and between the metric and scalar invariance model was found (Table 2). These findings indicate scalar measurement invariance of the MHATE scale across these groups and imply that latent means can be compared across groups. The scalar measurement invariance model (Model 1) was used to compare latent means between boys and girls. Results showed that boys showed lower means in ideology ( $p=.003$ ,  $d=0.40$ ) compared with girls. The scalar measurement invariance model (Model 2) was used to compare latent means between 7<sup>th</sup>/8<sup>th</sup> graders vs. 9<sup>th</sup> graders.

Results revealed that 9<sup>th</sup> graders showed lower means in group conformity ( $p=.013$ ,  $d=0.31$ ) compared with 7/8<sup>th</sup> graders. Next, the scalar measurement invariance model (Model 3) was used to



compare latent means between adolescents with and without migration background. Results showed that adolescents without migration background showed higher means in exhilaration ( $p < .001$ ,  $d = 0.55$ ) and group conformity ( $p < .001$ ,  $d = 0.46$ ) compared with participants with migration background. Finally, the scalar measurement invariance model was used to compare latent means between adolescents of high vs. low SES. Results showed that adolescents with high SES showed higher means in revenge ( $p = .006$ ,  $d = 0.41$ ) and power ( $p = .019$ ,  $d = 0.28$ ) compared to adolescents with low SES.

Table 2. Measurement invariance testing across gender, grade, migration background, and SES							
Models	$\chi^2$ (df)	p	RMSEA	$\Delta$ RMSEA	CFI	$\Delta$ CFI	Invariance rule accepted
<b>Model 1: Gender (boys vs. girls)</b>							
Configural	153.95 (80)	<.001	.073		.961		
Metric	155.96 (85)	<.001	.069	-0.004	.963	0.002	Yes
Scalar	168.51 (91)	<.001	.070	0.001	.959	-0.004	Yes
<b>Model 2: Grade (7<sup>th</sup>/8<sup>th</sup> graders vs. 9<sup>th</sup> graders)</b>							
Configural	152.82 (80)	<.001	.058		.976		
Metric	132.19 (85)	<.001	.057	-0.001	.975	-0.001	Yes
Scalar	144.34 (91)	<.001	.058	0.001	.972	-0.003	Yes
<b>Model 3: Migration background (with vs. without migration background)</b>							
Configural	123.98 (78)	<.001	.058		.975		
Metric	125.73 (84)	.002	.054	-0.004	.977	0.002	Yes
Scalar	129.33 (90)	.004	.050	-0.004	.978	0.001	Yes
<b>Model 4: SES (high vs. low SES)</b>							
Configural	133.79 (78)	<.001	.079		.957		
Metric	145.78 (84)	<.001	.080	0.001	.952	-0.005	Yes
Scalar	155.24 (90)	<.001	.080	0.000	.950	-0.002	Yes

Note.  $\chi^2$ =chi square test of model fit; CFI=comparative fit index;  $\Delta$ CFI=change in CFI compared to the weaker measurement invariance model above; RMSEA=root mean square error of approximation;  $\Delta$ RMSEA=change in RMSEA compared to the weaker measurement invariance model above.

### 3.4. Associations between social norms and motivations for perpetrating hate speech

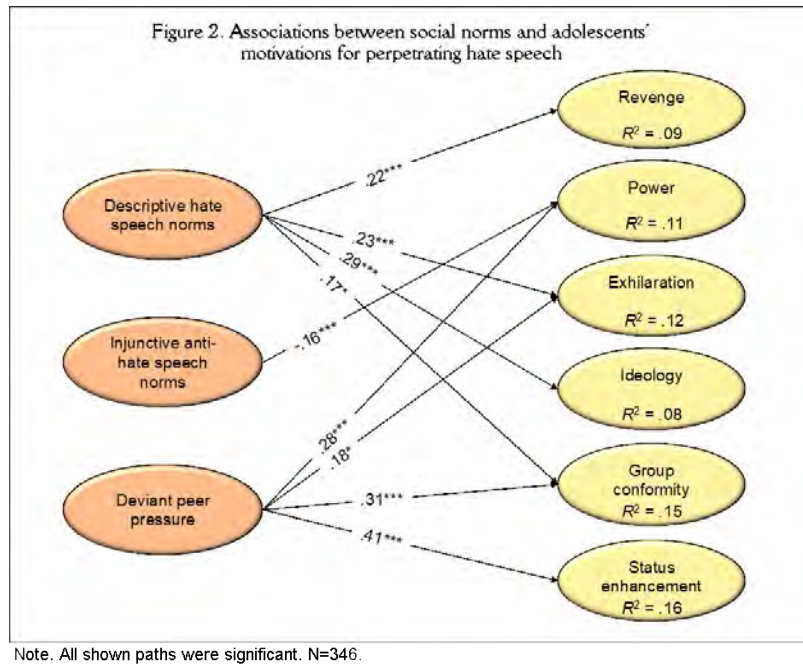
Bivariate correlations among the main variables are shown in Table 3.

Table 3. Bivariate correlations among main study variables									
	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Descriptive hate speech norms	—								
2. Injunctive anti-hate speech norms	.03	—							
3. Deviant peer pressure	.12*	-.13*	—						
4. Revenge	.14**	-.12**	.11*	—					
5. Power	.02	-.13**	.27**	.22**	—				
6. Exhilaration	.13**	-.13**	.20**	.04	.50**	—			
7. Ideology	.18**	-.02	-.01	.13*	.29**	.35**	—		
8. Group conformity	.28**	-.06	.29**	.09	.40**	.47**	.23**	—	
9. Status enhancement	.11*	-.06	.30**	.13*	.42**	.37**	.11**	.43**	—

Note.  $N = 346$  \* $p < .05$  \*\* $p < .01$ .

The SEM to investigate the relationships between social norms and motivations for perpetrating hate speech is shown in Figure 2. The model fit was acceptable,  $\chi^2 = 420.45$ ,  $df = 263$ ,  $p < .001$ , CFI = .946, TLI = 0.934, RMSEA = 0.042, SRMR = 0.047, standardized factor loadings ranged between 0.51 to 0.91. The analyses showed that descriptive hate speech norms had a positive effect on revenge ( $.22$  SE = .06,  $p < .001$ ), exhilaration ( $.23$ , SE = .08,  $p < .001$ ), ideology ( $.29$ , SE = .09,  $p < .001$ ), and group conformity ( $.17$ , SE = .09,  $p = .041$ ).

Injunctive anti-hate speech norms had only a negative effect on power ( $-.16$ , SE = .03,  $p < .001$ ). Finally, deviant peer pressure had a positive effect on power ( $.28$ , SE = .06,  $p < .001$ ), exhilaration ( $.18$ , SE = .08,  $p = .030$ ), group conformity ( $.31$ , SE = .10,  $p < .001$ ), and status enhancement ( $.41$ , SE = .06,  $p < .001$ ). The proposed model explained 9% of variance in the subscale revenge ( $R^2 = .088$ ), 11% of variance in the subscale power ( $R^2 = .111$ ), 12% of variance in the subscale exhilaration ( $R^2 = .124$ ), 8% of variance in the subscale ideology ( $R^2 = .084$ ), 15% of variance in the subscale group conformity ( $R^2 = .153$ ), and 16% of variance in the subscale status enhancement ( $R^2 = .162$ ).



#### 4. Discussion and conclusion

The aims of the present study were twofold. First, to investigate the psychometric properties of the multidimensional MHATE. Second, to investigate associations between social norms and adolescents' motivations for perpetrating hate speech. Regarding our first aim, our analyses supported a six-factor structure of the MHATE. Testing multi-group measurement invariance of the MHATE revealed strong evidence for scalar measurement invariance across different groups (i.e., boys vs. girls, younger vs. older adolescents, adolescents with vs. without migration background, and higher vs. lower SES), indicating an agreement on how motivations of hate speech perpetration manifest between these groups. Our findings support the idea that hate speech fulfills various functions, serving as a defense against or reaction to a perceived threat, conveying a feeling of power, or instrumentally improving one's position in the social group. The motivations to perpetrate hate speech found in this study align with qualitative research on adolescents' motivation (Ballaschk et al., 2021) and adults' motivations (Erjavec & Kovač, 2012) for hate speech perpetration and with research on aggressive behavior (Fluck, 2017; Gradinger et al., 2012; Pinker 2011; Tanrikulu & Erdur-Baker, 2021; Wettstein, 2008).

The most frequently endorsed subscale was revenge, followed by ideology, group conformity, status enhancement and exhilaration (with both having the same mean), and then power. The high prevalence of the revenge and ideology subscales may be attributed to possible justification strategies of the perpetrators. Related research showed that revenge was also one of the most frequently reported motivation for traditional bullying and cyberbullying (Gradinger et al., 2012; Fluck, 2017; Tanrikulu & Erdur-Baker, 2021). Whereas Fluck's (2017) study showed that ideology was not a relevant motivation for cyberbullying, we found that ideology was an important motivation for hate speech perpetration. An explanation might be that ideological motives may be more relevant for hate speech than for cyberbullying. In addition, we found that power was the least frequently reported motivation for hate speech perpetration. This finding corresponds to Gradinger et al.'s (2012) study on cyberbullying but contradicts findings from Tanrikulu and Erdur-Baker (2021), who found that power was the second most often reported motivation for cyberbullying perpetration.

We also found that the various subscales of the MHATE showed differential intercorrelation patterns. Exhilaration and power, as well as exhilaration and group conformity, were most strongly correlated. These findings indicate that hate speech is often ostensibly expressed in the form of jokes to gain power

and to conform to groups within adolescents' peer networks. In contrast, the revenge subscale showed rather weak correlations, or even non-significant correlations with the other subscales. An explanation might be that revenge is a relatively independent motivation and an expression of perceived threat, while group-related hate speech represents controlled attempts to instrumentally attain goals or increase one's status within a given group.

Based on these findings, it appears to be imperative to educate adolescents on strategies of conflict management, the ability to deal with negative emotions and handle frustration to prevent adolescents from using hate speech as a means of revenge. The high endorsement of the ideology subscale highlights the need for democratic education in schools, thereby preventing hate speech perpetration. The approval ratings for group-motivated motivations (i.e., group conformity, status enhancement) indicate that promoting a positive and inclusive classroom climate and encouraging students to reflect on rules and norms are also important starting points for preventing hate speech among adolescents.

Our second aim was to investigate associations between social norms and adolescents' motivations for perpetrating hate speech. Generally, our results confirmed that social norms are associated with adolescents' motivations for hate speech perpetration. This finding is in line with previous research on prejudices and aggressive behavior (Bastiaensens et al., 2016; Cook et al., 2010; Lee & Wong, 2009; Váradi et al., 2021) and initial research on hate speech (Ballaschk et al., 2021). The relationship between descriptive hate speech norms and adolescents' motivation to engage in hate speech might be explained by observational learning (Wachs, in press). An explanation for the association between adolescents' susceptibility to deviant peer pressure and their motivations to perpetrate hate speech might be that these adolescents want to avoid rejection or punishment by their peers. More specifically, the results indicated that motivations for hate speech perpetration are not associated with injunctive norms as much as it is with descriptive norms, as well as perceived peer pressure. It seems plausible that adolescents' motivations are more strongly influenced by their perception of behaviors that are normally performed in their social environment (descriptive norms) or to which they are actively encouraged to (peer pressure) compared to their perception of behaviors that are perceived as typically approved or disapproved (injunctive norms). However, these findings should be interpreted with caution as the three scales for measuring social norms differed from each other, which might limit the comparability. According to this study, the motivation to perpetrate hate speech arises at least in part from the social dynamics of a given social environment, where reciprocal reactive processes, social recognition, social pressure, and imitation, may play an important role. These social dynamics should therefore also be the starting point for prevention measures.

This study has a few limitations which should be addressed by future research. Firstly, the present study tested important psychometric properties of the MHATE, namely construct validity and reliability. Future research should aim to test additional psychometric properties (e.g., predictive validity). In addition to this, the six factors included only two items each, which is often considered problematic for reliability. However, if the two indicators are highly correlated, two items are often sufficient (Hayduk & Littvay, 2012). Secondly, although the sample included a relatively large number of hate speech perpetrators, it cannot be considered representative for Swiss adolescents. Follow-up research should include a representative sample and try to replicate the findings of the present study. Thirdly, due to the cross-sectional study design, we were not able to determine temporal relationships between social norms and motivations for hate speech perpetration. Longitudinal studies are needed to overcome these methodological limitations. Fourthly, we included only interpersonal correlates of adolescents' motivations for hate speech perpetration. Future research should also investigate intrapersonal factors (e.g., moral disengagement). Finally, we measured hate speech perpetration by using self-reports which might have affected the identification of perpetrators, as some participants might have been reluctant to report a socially undesirable behavior such as perpetration. Follow-up research should also use peer reports to overcome such methodological problems.

In conclusion, the present study contributes to the literature by presenting a new instrument for measuring motivations for hate speech perpetration among adolescents and investigating the relationships with social norms. The findings highlight the need to understand hate speech perpetration by taking several motivations and relations to social norms into consideration. The most frequently reported motivations



were revenge, ideology, and group conformity. Prevention programs that aim to tackle hate speech need to consider various strategies, including educating adolescents on social skills and democratic values and creating social environments in which adolescents are not exposed to hate speech or actively encouraged to engage in hate speech.

### Authors' Contribution

Idea, S.W.; Literature review (state of the art), S.W., A.W.; Methodology, S.W. M.G.-G.; Data analysis, S.W. M.G.-G.; Results, S.W.; Discussion and conclusions, S.W., A.W.; Writing (original draft), S.W.; Final revisions, S.W.; Project design and funding agency, S.W., L.B., A.W.

### Funding Agency

This research was funded by the German Research Foundation (DFG), grant number WA 4275/2-1, BI 1046/9-1 and the University of Teacher Education Bern, grant number 19s 0008 01.

### References

- Allport, G.W. (1954). *The nature of prejudice*. Addison-Wesley. <https://bit.ly/3s5s35M>
- Ballaschk, C., Wachs, S., Krause, N., Schulze-Reichelt, F., Kansok-Dusche, J., Bilz, L., & Schubarth, W. (2021). Dann machen halt alle mit. "Eine qualitative Studie zu Beweggründen und Motiven für Hatespeech unter Schüler\*innen. ["Then everyone just goes along with it." A qualitative study on reasons and motives of hate speech among students]. *Diskurs Kindheits- und Jugendforschung / Discourse. Journal of Childhood and Adolescence Research*, 16, 1-18. <https://doi.org/10.3224/diskurs.v16i4.01>
- Bastiaensens, S., Pabian, S., Vandebosch, H., Poels, K., Van Cleemput, K., Desmet, A., & Bourdeaudhuij, I.D. (2016). From normative influence to social pressure: How relevant others affect whether bystanders join in cyberbullying. *Social Development*, 25(1), 193-211. <https://doi.org/10.1111/sode.12134>
- Baumeister, R.F., & Leary, M.R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497-529. <https://doi.org/10.1037/0033-2909.117.3.497>
- Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
- Cialdini, R.B., & Trost, M.R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 151-192). McGraw-Hill.
- Cook, C.R., Williams, K.R., Guerra, N.G., Kim, T.E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly*, 25(2), 65-83. <https://doi.org/10.1037/a0020149>
- Erjavec, K., & Kovari, M.P. (2012). You don't understand, this is a new war!" Analysis of hate speech in news web sites' comments. *Mass Communication and Society*, 15, 899-920. <https://doi.org/10.1080/15205436.2011.619679>
- Fluck, J. (2017). Why do students bully? An analysis of motives behind violence in schools. *Youth & Society*, 49(5), 567-587. <https://doi.org/10.1177/0044118X14547876>
- Grading, P., Strohmeier, D., & Spiel, C. (2012). Motives for bullying others in cyberspace: A study on bullies and bully-victims in Austria. In Q. Li, D. Cross, & P. K. Smith (Eds.), *Cyberbullying in the global playground: Research from international perspectives* (pp. 263-284). Wiley Blackwell. <https://doi.org/10.1002/9781119954484.ch13>
- Hartley, J.E.K., Levin, K., & Currie, C. (2016). A new version of the HBSC Family Affluence Scale-FAS III: Scottish qualitative findings from the international FAS development study. *Child Indicators Research*, 9(1), 233-245. <https://doi.org/10.1007/s12187-015-9325-3>
- Hayduk, L.A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, 12(1), 1-17. <https://doi.org/10.1186/1471-2288-12-159>
- Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Krause, N., Ballaschk, C., Schulze-Reichelt, F., Kansok-Dusche, J., Wachs, S., Schubarth, W., & Bilz, L. (2021). Ich lass mich da nicht klein machen! Eine qualitative Studie zur Bewältigung von Hatespeech durch Schüler/innen. ["I don't let them get me down!"]—A qualitative study on students' coping with hate speech]. *Zeitschrift für Bildungsforschung*, 11(1), 169-185. <https://doi.org/10.1007/s35834-021-00291-w>
- Lee, S.S.T., & Wong, D.S.W. (2009). School, parents, and peer factors in relation to Hong Kong students' bullying. *International Journal of Adolescence and Youth*, 15(3), 217-233. <https://doi.org/10.1080/02673843.2009.9748030>
- Lehman, B. (2020). Hate at school: Victimization and disorder associated with school avoidance. *Sociological Spectrum*, 40(3), 172-190. <https://doi.org/10.1080/02732173.2020.1734890>
- McClelland, D.C. (1975). *Power: The inner experience*. Irvington Publishers.
- Möller, K., Grote, J., Nolde, K., & Schuhmacher, N. (2016). *Die kann ich nicht ab! Ablehnung, Diskriminierung und Gewalt bei Jugendlichen in der (Post-) Migrationsgesellschaft. Analysen zu gesellschaftlicher Integration und Desintegration*. Springer. <https://doi.org/10.1007/978-3-658-02302-7>
- Muthén, L.K., & Muthén, B.O. (2017). *Mplus user's guide*. Muthén & Muthén. <https://bit.ly/3lJorwi>

- Olthof, T., Goossens, F.A., Vermande, M.M., Aleva, E.A., & Van-Der-Meulen, M. (2011). Bullying as strategic behavior: Relations with desired and acquired dominance in the peer group. *Journal of School Psychology, 49*(3), 339-359. <https://doi.org/10.1016/j.jsp.2011.03.003>
- Pinker, S. (2011). *The better angels of our nature: Why violence has declined*. Viking.
- Reichelmann, A., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Rasanen, P., & Zych, I. (2020). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior, 42*(9), 1100-1111. <https://doi.org/10.1080/01639625.2020.1722337>
- Runions, K.C., Salmivalli, C., Shaw, T., Burns, S., & Cross, D. (2018). Beyond the reactive-proactive dichotomy: Rage, revenge, reward, and recreational aggression predict early high school bully and bully/victim status. *Aggressive Behavior, 44*(5), 501-511. <https://doi.org/10.1002/ab.21770>
- Salmivalli, C., & Peets, K. (2009). Bullies, victims, and bully-victim relationships in middle childhood and early adolescence. In K. H. Rubin, W. M. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships, and groups* (pp. 322-340). The Guilford Press.
- Santor, D.A., Messervey, D., & Kusumakar, V. (2000). Measuring peer pressure, popularity, and conformity in adolescent boys and girls: Predicting school performance, sexual attitudes, and substance abuse. *Journal of Youth and Adolescence, 29*(2), 163-182. <https://doi.org/10.1023/A:1005152515264>
- Sokolowski, K., & Heckhausen, H. (2010). Soziale Bindung: Anschlussmotivation und Intimitätsmotivation. In J. Heckhausen, & H. Heckhausen (Eds.), *Motivation und Handeln [Motivation and Action]* (pp. 193-210). Springer. [https://doi.org/10.1007/978-3-642-12693-2\\_7](https://doi.org/10.1007/978-3-642-12693-2_7)
- Tajfel, H., & Turner, J.C. (2004). The social identity theory of intergroup behavior. In J. T. Jost, & J. Sidanius (Eds.), *Political psychology: Key readings* (pp. 276-293). Psychology Press. <https://doi.org/10.4324/9780203505984-16>
- Tanrikulu, I., & Erdur-Baker, O. (2021). Motives behind cyberbullying perpetration: A test of uses and gratifications theory. *Journal of Interpersonal Violence, 36*, 6699-6724. <https://doi.org/10.1177/0886260518819882>
- Van-De-Bongardt, D., Reitz, E., Sandfort, T., & Dekovi, M. (2015). A meta-analysis of the relations between three types of peer norms and adolescent sexual behavior. *Personality and Social Psychology Review, 19*(3), 203-234. <https://doi.org/10.1177/1088868314544223>
- Van-De-Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486-492. <https://doi.org/10.1080/17405629.2012.686740>
- Váradi, L., Barna, I., & Németh, R. (2021). Whose norms, whose prejudice? The dynamics of perceived group norms and prejudice in new secondary school classes. *Frontiers in Psychology, 11*, 3621-3621. <https://doi.org/10.3389/fpsyg.2020.524547>
- Wachs, S., Mazzone, A., Milosevic, T., Wright, M.F., Blaya, C., Gámez-Guadix, M., & Norman, J.H. (2021). Online correlates of cyberhate involvement among young people from ten European countries: An application of the routine activity and problem behaviour theory. *Computers in Human Behavior, 123*, 106872. <https://doi.org/10.1016/j.chb.2021.106872>
- Wachs, S., Schubarth, W., & Bilz, L. (2020). Hate speech als Schulproblem? Erziehungswissenschaftliche Perspektiven auf ein aktuelles Phänomen. In I. van Ackeren, H. Bremer, F. Kessl, H. Koller, N. Pfaff, ... U. Salaschek (Eds.), *Bewegungen – Beiträge aus dem 26. Kongress der Deutschen Gesellschaft für Erziehungswissenschaft* (pp. 223-236). Verlag Barbara Budrich. <https://doi.org/10.2307/j.ctv10h9fjc.19>
- Wachs, S., Wettstein, A., Bilz, L., Krause, N., Ballaschk, C., Kansok-Dusche, J., & Wright, M.F. Playing by the Rules? An investigation of the relationship between social norms and adolescents' hate speech perpetration in schools. *Journal of Interpersonal Violence, 36*, 1056032. <https://doi.org/10.1177/08862605211056032>
- Wachs, S., & Wright, M.F. (2021). Associations between online hate victimization and perpetration: The buffering effects of technical and assertive coping. *Medienpädagogik, 16*, 109-128. <https://doi.org/10.21240/mpaed/jb16/2021.01.14.X>
- Wettstein, A. (2008). *Beobachtungssystem zur Analyse aggressiven Verhaltens in schulischen Settings (BASYS)*. [Observation system for the analysis of aggressive behavior in school settings]. Huber.



Comunicar  
*authors*



## SCHOOL OF AUTHORS

Blog designed by the Editors of Comunicar, which aims to provide resources to authors for the presentation of manuscripts to scientific journals in a planned and strategic way.

[revistacomunicar.com/wp/school-of-authors](http://revistacomunicar.com/wp/school-of-authors)



# Hate speech and social acceptance of migrants in Europe: Analysis of tweets with geolocation

Discurso de odio y aceptación social hacia migrantes en Europa:  
Análisis de tuits con geolocalización

- Dr. Carlos Arcila-Calderón. Associate Professor, Department of Sociology and Communication, University of Salamanca (Spain) (carcala@usal.es) (<https://orcid.org/0000-0002-2636-2849>)
- Dr. Patricia Sánchez-Holgado. Postdoctoral Researcher, Department of Sociology and Communication, University of Salamanca (Spain) (patriciasanc@usal.es) (<https://orcid.org/0000-0002-6253-7087>)
- Cristina Quintana-Moreno. Magister, Department of Sociology and Communication, University of Salamanca (Spain) (crisquimo@usal.es) (<https://orcid.org/0000-0001-6303-6332>)
- Javier-J. Amores. Predoctoral Researcher, Department of Sociology and Communication, University of Salamanca (Spain) (javieramores@usal.es) (<https://orcid.org/0000-0001-7856-5392>)
- David Blanco-Herrero. Predoctoral Researcher, Department of Sociology and Communication, University of Salamanca (Spain) (david.blanco.herrero@usal.es) (<https://orcid.org/0000-0002-7414-2998>)

## ABSTRACT

Hate speech against vulnerable groups is acknowledged as a serious problem for integration and respect for the social diversity existing within the territory of the European Union. The growth of this type of discourse has been supported by the expansion of social media, which have been proven to act as a mechanism for the propagation of crimes against targets such as migrants and refugees, one of the main affected groups. That is why we have conducted the first European study of the social acceptance of migrants and refugees by studying the presence of hate speech. The research is based on the perspective of the theories of intergroup contact and mediated intergroup contact. The methodology includes large-scale longitudinal analysis (2015-2020) of online hate speech on Twitter (N=847,978) and its contrast with existing official indicators. The results suggest that personal intergroup contact is positively correlated with the support of the local population towards migrants and refugees but mediated intergroup contact is not correlated with hate speech on Twitter. We found evidence that those regions where the support for foreigners was higher, there was a lower level of hate speech on Twitter. This is an advance in the study of hate speech by territories and can help in the formulation of action strategies.

## RESUMEN

El discurso de odio contra públicos vulnerables es reconocido como un grave problema para la integración y el respeto a la diversidad social dentro de la Unión Europea. El aumento de este tipo de discurso se ha visto reforzado con la expansión de las redes sociales, donde se ha demostrado que actúan como mecanismo de propagación de delitos contra colectivos como los migrantes y refugiados, uno de los principales afectados. Por ello se aborda el desarrollo del primer estudio europeo de la aceptación social de migrantes y refugiados mediante el estudio de la presencia de discurso de odio. La investigación se basa en la perspectiva de la teoría del contacto intergrupar y el contacto intergrupar mediado. La metodología incluye el análisis longitudinal (2015-2020) a gran escala del discurso de odio en línea en Twitter (N=847.978) y el contraste con indicadores oficiales existentes. Los resultados apuntan a que el contacto intergrupar personal está correlacionado positivamente con el apoyo de la población hacia migrantes y refugiados, pero el contacto intergrupar mediado no está correlacionado con la disminución del discurso de odio. Encontramos evidencia que muestra que en aquellas regiones en las que el apoyo al colectivo era mayor existía un menor nivel de discurso de odio en Twitter. Esto supone un avance en el estudio del discurso de odio por territorios y puede ayudar en el planteamiento de estrategias de actuación.

## KEYWORDS | PALABRAS CLAVE

Immigration, refugees, attitude, hate speech, big data, Twitter.  
Inmigración, refugiados, actitud, discurso del odio, big data, Twitter.





## 1. Introduction and background

Hate speech has alarmingly permeated our society and social media have become the most suitable means of propagation. The number of potentially offensive and dangerous messages against stigmatized groups has increased as a result of the COVID-19 health and economic crisis. The pandemic “not only took human lives, but it reinforced existing problems, and hit harder on otherwise vulnerable minorities” (Bayer & Bárd, 2020: 16-17). So far, there is no single definition for hate speech due to (i) the legal and ethical considerations and implications involved; and (ii) the subjective nature of “hate,” i.e., a sentiment opens to interpretation with blurred boundaries (Cabo-Isasi & García-Juanatey, 2017). United Nations (UN) (2019) defines hate speech as “any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group [...] based on their religion, ethnicity, nationality, race, color, descent, gender or other identity factors.”

The definition of migrant encompasses all those people who move from one place to another, either within their own country or crossing international borders, for multiple reasons. Countries have different processes to determine whether migration is legal or not. Generally, people who have applied for permission to reside in another nation through established procedures and are accepted are considered “legal migrants.” On the other hand, there are undocumented immigrants, also referred as “illegal immigrants”, who cross to another country without knowledge or acceptance of their movement by the host country. There are many reasons why people make these dangerous decisions, such as longing for a better life for themselves or their children, fear for their lives, fleeing extreme poverty in their country, or reuniting with family members who have already resettled in the country. Differently, a “refugee” is defined as someone who, due to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership to a particular social group or political opinion, is outside their country of nationality and is unable, or is unwilling out of fear, to avail themselves to the protection of that country; or who, without having a nationality and being outside the country of their former habitual residence as a result of such events, cannot or, due to such fear, is not willing to return to it (United Nations High Commissioner for Refugees UNHCR, 1951). In practice, hate speech based on xenophobia or racism is similarly directed towards migrants or refugees.

Immigration ranks third among European citizens’ concerns, right after their personal economic situation and government finance. Immigrants and refugees are some of the most affected groups by hate speech. In fact, hate speech towards these vulnerable groups is relevant to the point that some scholars like Müller and Schwarz (2020) have found a direct correlation between the expansion of hate speech and the increase in hate crimes within a given territory. So, detecting hate speech and being able to predict its implications opens new lines of research on its potential effects on citizens. As stated in the Global Compact for Migration (UN, 2018), a better understanding of migration dynamics and the causes thereof is helpful for implementing measures seeking safe, orderly and regular migration. When studying the aspects affecting the flow of people within the European Union, predicting the social acceptance of migrants and refugees in various contexts is quite a challenge. Therefore, we must rely on all the available sources. There have already been some efforts in this direction, e.g., Arcila-Calderón et al. (2022b). This study relies on the Eurobarometer public opinion survey and uses computational methods to estimate the probability of acceptance of migrants and refugees in Europe.

This innovative approach is justified in international projects such as HumMingBird,<sup>1</sup> aimed at predicting the social integration of migrants and refugees in every European city. One of the limitations encountered is the scarce regional-level information provided by European polls and surveys on the social support towards these groups. Consequently, here we present the first large-scale study with a European Union scope. The study detects, simultaneously and longitudinally (including a 6-year period), (i) hate speech on Twitter; and (ii) how it affects citizens’ perception and acceptance of immigrants and refugees.

Hate speech has a major impact on the social environment and public discourse, particularly due to the tension and strain it creates, and also because it has a significant presence in user-generated media. It is difficult to factor this in when studying the probability of acceptance of migrants and refugees. However, incorporating this factor allows for an innovative approach and can offer a new perspective within these lines of work. Furthermore, since social media are currently the main channels for hate speech spread,

it is worth assessing their potential to provide meaningful data on top of existing official indicators. Social media reflect public opinion in certain contexts. Twitter, in particular, with over 500 million posts per day, has a strong impact on a large share of the population (Sayce, 2020). During the so-called refugee crisis in 2015, almost 7.5 million tweets were collected through hashtags such as #refugee or #refugeecrisis (Siapera et al, 2018), and the widespread negative sentiment towards foreigners increased, as well as the hostility leading to the exclusion of migrants from jobs and welfare benefits (Inter-Parliamentary Union, 2015). This also allowed to assess the strategies of anti-immigration actors and xenophobic groups on social media (Ekman, 2019). In the context of this heated social debate, European decision-makers and political leaders try to achieve “fair migration,” benefitting all parties and fulfilling the human rights at stake whilst creating friendly environments and encouraging social cohesion in line with European diversity standards. On this basis, the goals laid down in the 2030 Agenda for Sustainable Development (UN, 2015) include working together to foster the social inclusion of migrants (Canelón & Almansa, 2018).

The origin of xenophobic and hate opinions, attitudes and behaviors can be explained, at least partly, relying on Intergroup Contact Theory (ICT) posed by Allport in *The Nature of Prejudice*, since it remains “a widely used framework in the study of intergroup relations and intergroup prejudice” (Broad et al., 2014: 49). Given the strong social presence of online media, we address a second dimension of ICT, taking the perspective of Mediated Intergroup Contact, particularly through user-generated media like social networking sites. This research thus relies on ICT and the theory of Mediated Intergroup Contact in order to (i) assess individuals’ relationships with migrants; and (ii) analyze such contacts through social media including hate speech as a determinant. ICT’s basic premise is that increased intergroup contact leads to more positive attitudes towards outgroup members (Abrams & Hogg, 2017), there being a relationship between the effects of contact and reduced levels of prejudice. Accordingly, contact between non-migrants and migrants encourages acceptance and a positive attitude towards the latter. ICT or contact hypothesis requires four essential factors to be effective, i.e., to reduce prejudice: equal status between the groups, intergroup cooperation, common goals, and the support of social and institutional authorities. First, equal status entails that there be no unequal hierarchical relationship, e.g., employed-unemployed, since there are negative effects from contact with outgroup members of lower status (Pettigrew, 1998). Second, intergroup cooperation entails that outgroup members work or operate within a non-competitive environment, which is directly related to the third factor, i.e., common goals, since the attainment of common goals must be an interdependent effort without intergroup competition. Finally, no institutional or social authority should prevent or otherwise undermine intergroup contact.

Linking the above conditions with the issue addressed herein, i.e., the acceptance of migrants and refugees, we should first inquire about whether, in places with a significant percentage of the immigrant population, (i) citizens will be more likely to have direct intergroup contact with migrants or refugees; and if, as a result, (ii) they will develop a more positive acceptance than other individuals not engaging in this contact (Abrams et al, 2018). Based on this, we state the first hypothesis to be tested:

- H1: The share of immigrant population in European regions is positively correlated with citizens’ support towards that group, so that the higher the percentage of immigrants, the greater the support towards them.

Focusing on mediated intergroup contact, the boom of social media has raised significant new research questions, challenges, and opportunities regarding online social behavior and its impact on real life. Considering that interpersonal contact via social media is both unlimited and cheap. Online intergroup contact, whether positive or negative, can lead to attitudes, beliefs or thoughts that are expressed outside of social networks. Hate speech emerges as a determinant of mediated intergroup contact, so that if there is negative contact through social media triggered by hate speech, the offline attitude and acceptance will also be negative. Previous studies tried to explain if online engagement between persons with different backgrounds, experiences or opinions represents real life, offline intergroup relationships (Gallacher et al, 2021). The findings indicate that antagonistic or opposing groups took their online behavior to their real-world encounters. On top of that, their real-life behavior derived from negative intergroup contact, which increases concerns about polarization, because we could expect individuals to identify themselves as members of a social tribe thus expressing hostile attitudes towards outgroup members (Zhang et al,

2019). These concerns have given rise to the so-called “echo chamber effect” on social media, aimed at explaining how social media users seek or avoid information based on their ideological background whilst seeking like-minded users framing and reinforcing a shared narrative and thus affecting their offline behavior (Cinelli et al, 2021). Platforms like Twitter, with more than 400 million active accounts (We Are Social/Hootsuite, 2021), may have radicalized citizens’ beliefs, thereby making them reject or dismiss outgroup members outside their cultural, social, and economic circle. Therefore, we can assume that “it is increasingly important to understand intergroup contact as a straightforward yet potentially powerful strategy to reduce prejudice between groups” (Zhang et al, 2019). See next a few examples of messages containing hate speech on Twitter (translated from Spanish): “While they hinder Spaniards’ mobility, more than 700 illegal immigrants have assaulted our borders in the last few days. The Government’s call effect continues and the resulting migratory avalanche that dooms the future of our neighborhoods. STOP THE INVASION!” (Vox, 2021). “Here is the beautiful Christmas present from those who hate us. The perpetrator of this disgusting massacre, a Palestinian ‘holder of a REFUGEE travel document issued by Belgium!’ For certain offenders, no red zones, no travel bans....” (Salvini, 2020).

Accordingly, if applied to the current times where new technologies reinforce communication on social media, “any strategy that seeks to understand and fight hate speech must include a communication approach” (Arcila-Calderón et al, 2021). Consequently, even if there is a significant share of migrants and refugees in a given region with positive direct contact, we can analyze the existing relationship with the hate level online in that territory. We thus state a second hypothesis:

- H2: The share of immigrant population in European regions is correlated with the level of hate on Twitter, so that the greater the percentage of immigrants, the lesser the hate.

The foregoing suggests that it is necessary to inquire about the relationship between social media and their role representing social reality. The research conducted by Bollen et al (2011) found a statistically significant correlation between the mood states of tweets (tension, depression, anger, vigor, fatigue, confusion) and daily events gathered from media and other sources, attributing it a predictive value. To address this, we must use the available macrodata, providing information that can supplement other existing sources like the Eurobarometer (Eurobarometer, n.d.), the European Social Survey (ESS-ERIC Consortium, 2021), Gallup’s Migrant Acceptance Index (Esipova et al, 2020), and other national sources, adding value to the predictions on the acceptance of migrants and refugees in their host territories. These sources and indicators have certain limitations, e.g., the social desirability bias, leading respondents to underreport socially undesirable attitudes and behaviors and to give more likeable answers that others want to hear, making it difficult to detect openly hateful content, like xenophobic or racist statements (Arcila-Calderón et al, 2020). Another meaningful limitation is that there is little information on the location of opinions because they are presented at a national or regional level. At this point, Twitter data can provide geolocation information, which would allow to cross-reference and contrast variables. On this basis, we expect Twitter to help model a depiction or representation of society relying on the posted messages, analyzing hate speech as a predictor of social acceptance of migrants and refugees in Europe and adding value to the existing indicators. Therefore, we state a final hypothesis:

- H3: The level of hate speech on Twitter towards migrants and refugees in European regions is correlated with the degree of citizens’ support for these groups, so that the lesser the level of hate speech, the greater the acceptance of migrants and refugees.

## 2. Materials and methods

### 2.1. Sample and process

For this research we have used primary sources (databases with tweets and their hate level from all Europe between 2015 and 2020) and secondary sources (information provided by other studies and existing databases). A tweet database was generated: collected tweets using Twitter’s Application Programming Interface (API) (v2 full-archive search endpoint, using Academic research product track), which provides access to the historical archive of messages since Twitter was created in 2006 (the downloaded data from each tweet can be consulted in <https://doi.org/10.6084/m9.figshare.16708942.v3>). To download the tweets, we first defined the search filter by keyword and geographic zones using the

Python programming language and the NLTK, Tensorflow, Keras and Numpy libraries. We established generic words directly related with the topic, taking into account linguistic agreement in Spanish (i.e., gender and number inflections) but without considering adjectives, for instance: migrant, migrants, immigrant, immigrants, refugee (both in masculine and feminine forms in Spanish), refugees (both in masculine and feminine forms in Spanish), asylum seeker, asylum seekers (the keywords are available as supplementary materials at: <https://doi.org/10.6084/m9.figshare.16708945.v3>). Then, we selected only those messages that had geolocation coordinates (coordinates tag in the Twitter object), so that each tweet provided the exact location using a longitude and latitude matrix, and only those messages that were not retweets or answers. Next, for categorizing the downloaded tweets into the 27 EU Member States (adding Switzerland, the United Kingdom and Norway in aggregated form), we used the Nomenclature of Territorial Units for Statistics (NUTS), which follows a hierarchical system for dividing up the economic territory of the EU and the UK into the major socio-economic regions (NUTS 1), basic regions for the application of regional policies (NUTS 2) and small regions for specific diagnoses (NUTS 3), relying on the Nominatim geocoder<sup>2</sup> and its association with the codes through Nuts Finder.<sup>3</sup> We focus on the NUTS 2 level, including regions that are able to implement migration-related policies. This work was supported by the infrastructure of the Supercomputing Center of Castille and Leon (Scayle).

Finally, for the process of hate speech detection in tweets, we used as a basis a tool (<http://pharm-interface.usal.es>) created and validated by Vrysis et al. (2021). For this research, the tool has been retrained with (i) supervised dictionary-based term detection; and (ii) also taking an unsupervised approach (machine learning with neural networks) using a corpus of 90,977 short messages, from which 15,761 were in Greek (5,848 with hate toward immigrants), 46,012 were in Spanish (11,117 with hate toward immigrants) and 29,204 in Italian (5,848 with hate toward immigrants). This corpus comes from two sources: one, the import of already classified messages in other databases ( $n=57,328$ , of which 5,362 are generic messages in Greek, 23,787 are generic messages and 9,727 are messages with hate toward immigrants in Spanish, and 18,452 are generic messages in Italian), and the other from messages manually coded by local trained analysts (in Spain, Greece and Italy), using at least 2 coders with total agreement between them (the level of agreement in the tests was 94%), dismissing those without a 100% intercoder agreement ( $n=33,649$ , of which 6,040 are messages about immigration without hate and 4,359 are messages with hate toward immigrants in Greek; 11,108 are messages about immigration without hate and 1,390 are messages with hate toward immigrants in Spanish; and 4,904 are messages about immigration without hate and 5,848 are messages with hate toward immigrants in Italian). The corpus was divided into 80% training and 20% test. In the models, embeddings were used for the representation of language and Recurrent Neural Networks (RNN) for the supervised text classification. Specifically, the embeddings were created with the 1,000 most repeated words with 8 dimensions (first input layer), two hidden layers' type Gated Recurrent Unit (GRU) with 64 neurons each, and a dense output layer with one neuron and softmax activation (the model is compiled with Adam optimizing and the Sparse Categorical Crossentropy loss).

The evaluation of the detection model offered good results: in Greek, hate (accuracy=0.78, precision=0.79, recall=0.75, F-score=0.77), no hate (accuracy=0.78, precision=0.76, recall=0.80, F-score=0.78); in Spanish, hate (accuracy=0.87, precision=0.87, recall=0.86, F-score=0.87), no hate (accuracy=0.87, precision=0.86, recall=0.87, F-score=0.87); and in Italian, hate (accuracy=0.90, precision=0.93, recall=0.87, F-score=0.90), no hate (accuracy=0.90, precision=0.88, recall=0.94, F-score=0.91). Additionally, we developed an external validation phase –with 10,285 new tweets collected in a later period, between November and December 2020, classified by the model and by one of the human trained coders–. This validation offered acceptable quality evaluation metrics for the hate category: accuracy=0.85, AUC-ROC=0.88, F-score=0.74, Loss=0.46. The specific results of the tests of previous models with different algorithms, parameters and training corpus, as well as the validation processes, are completely reported in Arcila-Calderón et al. (2022a).

For this study, we have managed to download and store a database with a total sample of 847,978 Twitter messages from the last six years (193,676 in 2015, 182,634 in 2016, 121,465 in 2017, 140,293 in 2018, 112,552 in 2019 and 97,358 in 2020) from 30 European countries with messages referring to migrants or refugees and including geolocalization data (in the supplementary materials, the number of



tweets per year and country can be found: <https://doi.org/10.6084/m9.figshare.17186108.v2>). These messages have been run over the detector to obtain the results that will be explained next.

To make up for the language diversity across the European Union, the Google Translate<sup>4</sup> and Python texblob<sup>5</sup> APIs provide a translation, even if the language has not been declared. Relying on these APIs, all messages have been automatically translated into Spanish, as it was the language in which the external validation was conducted. Furthermore, in order to correlate and contrast these Twitter data with existing data and study migrant and refugee acceptance scenarios, we relied on various secondary sources. First, the so-called Census Hub, within the European Statistical System (ESS), which provides detailed data on the size, characteristics of the population and housing in Europe. These data are updated every 10 years, so the relevant census for this research is the 2011 census (European Commission, 2016).

Second, we use the results of the study carried out by Arcila-Calderón et al. (2022b), which estimates the probability of acceptance of refugees in Europe at a regional level (NUTS 2) based on Eurobarometer (2015-2017) data and applying computational methods that combine machine learning and synthetic populations. The study creates “acceptance probabilities” toward migrants for each NUTS 2, extrapolating demographic data in each region and comparing them with the models produced by national surveys (using the algorithms: logistic regression, decision trees, random forest, vector machines support k-nearest neighbour), so that it created artificial or synthetic populations with 10,000 inhabitants in each of the 271 European regions to estimate the probability of each individual and then the average of all the individuals of the region.

## 2.2. Measures and analysis

Three variables were used in this study:

1. The level of hate speech online towards immigrants and refugees on Twitter: the average hate level of posted tweets broken down by European regions, measured using the hate detector, which is based on the one developed by Vrysis et al. (2021), retrained for this study. The values range between 0 and 1, 0 being the absence of hate speech and 1 being the presence of hateful content. These data were obtained by applying the detector to the Twitter archive created for this research and filtered by content, date and geolocation.
2. Level of support towards migrants and refugees: this indicator is based on average public opinion regarding the support for measures aimed at integrating refugees. The indicator was created using the Eurobarometer survey at various stages (second semester of 2015, first semester of 2016, second semester of 2016, first semester of 2017 and second semester of 2017). The values are continuous and range between 0 (less support) and 1 (greater support). These data were obtained from the simulations conducted in the study by Arcila-Calderón et al. (2022b) based on the Eurobarometer public opinion survey (2020).
3. Share of foreign population: this measure was obtained from the number of foreign persons in each European region (NUTS 2) overall. The values range between 0 and 1, 0 being the absence of foreign population and 1 meaning that the foreign population is 100% overall. These figures were obtained from the Census Hub database.

Visual,<sup>6</sup> descriptive and correlational analyses are used for analyzing the data and indicators and testing the hypotheses. Regarding the visual analyses, we prepared comparative maps for every year using the Tableau software, designed for data processing, analysis and visualization. We used other previously created maps showing the level of support towards the analyzed groups (Arcila-Calderón et al., 2022b) and the share of immigrants and native inhabitants per region. The maps represent descriptive data for every measure (averages) using Excel and SPSS as statistical software. SPSS is also used to test statistical correlations in order to analyze if there are statistically significant relationships between variables.

## 3. Analysis and results

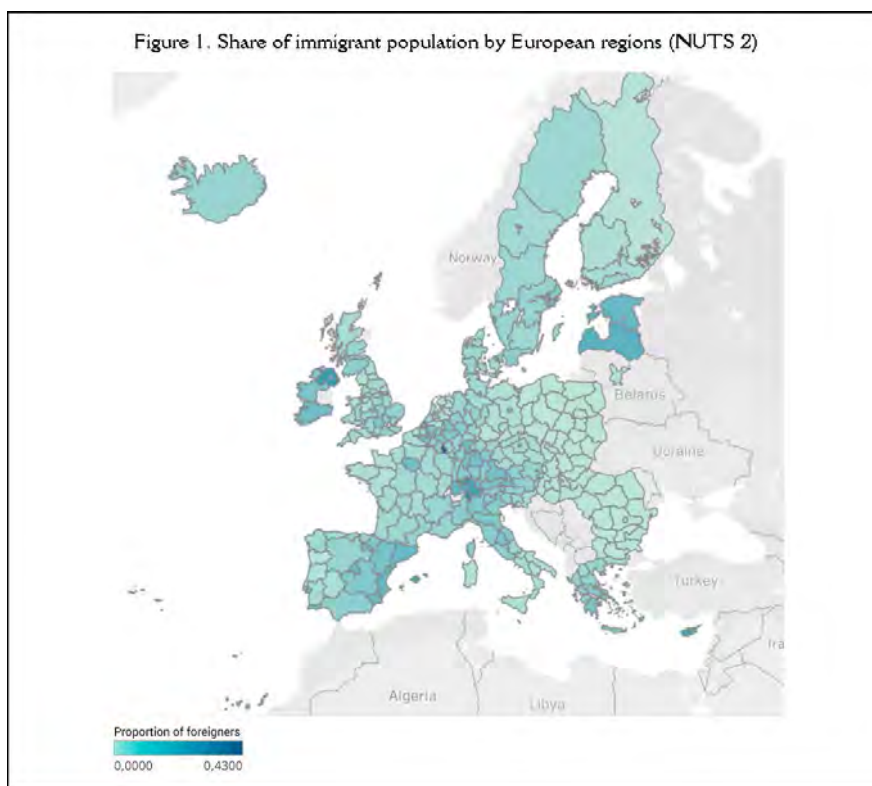
The results of using the hate detector to analyze the database comprising 847,978 messages per year in EU countries are presented in Table 1.

**Table 1. Average Presence of Hate Speech in Twitter by Country and Year**

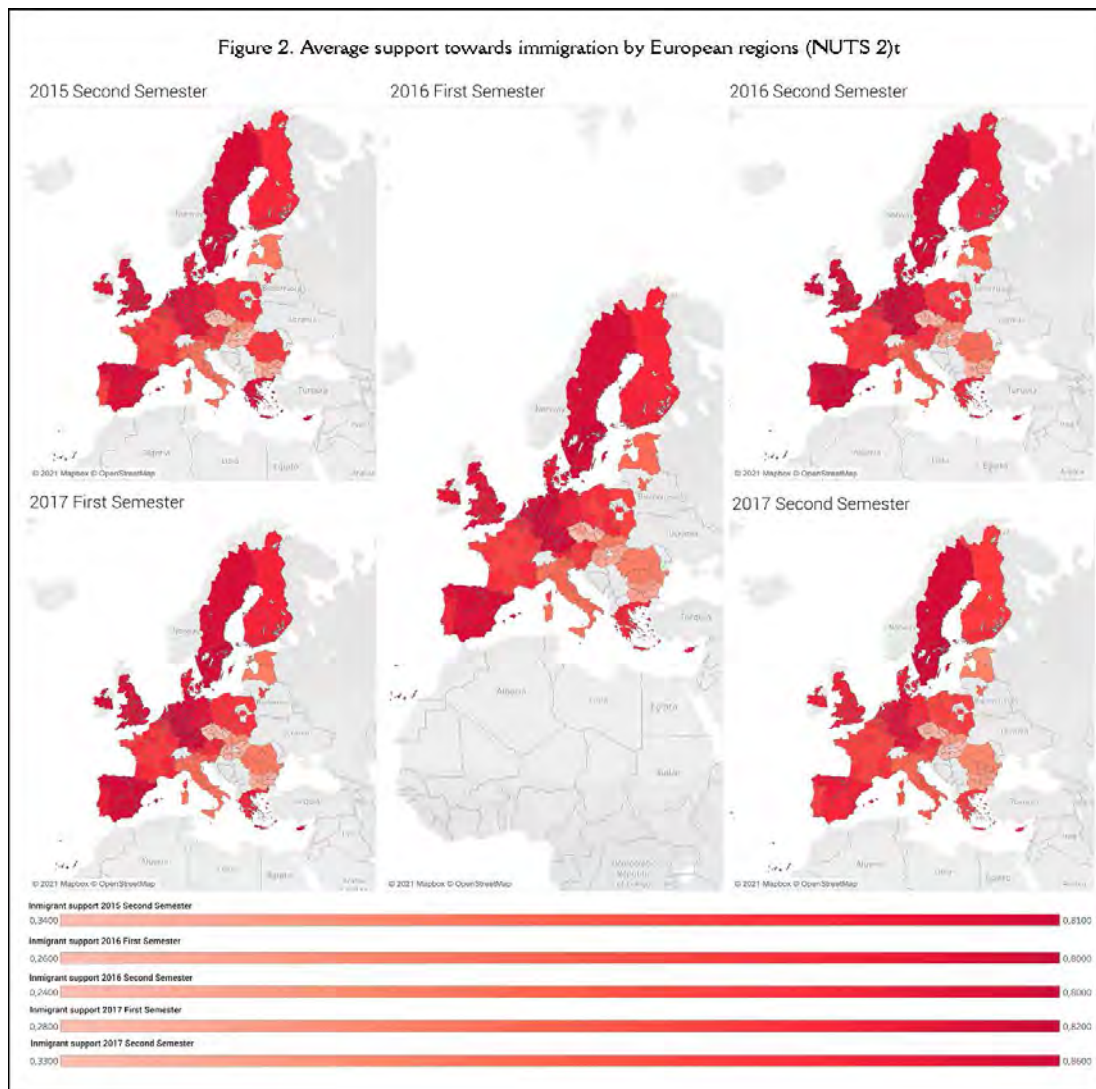
COUNTRY	2015	2016	2017	2018	2019	2020
Austria	0.2343	0.2326	0.2275	0.1504	0.1495	0.1490
Belgium	0.1871	0.1898	0.1695	0.1372	0.1280	0.1354
Bulgaria	0.2400	0.1747	0.1416	0.1678	0.3540	0.0400
Cyprus	0.5172	0.4173	0.3260	0.5604	0.5513	0.5329
Czech Republic	0.2104	0.2076	0.1317	0.1520	0.1086	0.1465
Germany	0.2433	0.2395	0.1973	0.1654	0.1558	0.1525
Denmark	0.2256	0.2137	0.2118	0.1572	0.1245	0.1244
Estonia	0.1804	0.1622	0.1600	0.2025	0.1217	0.3650
Spain	0.2604	0.2514	0.2173	0.1899	0.2088	0.1783
Finland	0.1923	0.1894	0.1834	0.1669	0.1429	0.1499
France	0.1773	0.1634	0.1563	0.1300	0.1311	0.1193
Great Britain	0.2204	0.2126	0.1918	0.1751	0.1706	0.1717
Greece	0.6509	0.7069	0.6117	0.6490	0.5812	0.6792
Croatia	0.1814	0.2125	0.2371	0.2580	0.1322	0.1458
Hungary	0.3037	0.1994	0.0895	0.0679	0.1415	0.0868
Ireland	0.1888	0.1964	0.1877	0.1637	0.1534	0.1243
Italy	0.6637	0.5438	0.5063	0.5691	0.5332	0.5260
Lithuania	0.1769	0.4200	0.0500	0.8800	0.2600	0.5300
Luxembourg	0.1612	0.1540	0.1738	0.1252	0.2176	0.1845
Latvia	0.2521	0.2512	0.1853	0.1988	0.1849	0.1871
Malta	0.1679	0.1666	0.1963	0.1425	0.1536	0.1605
Netherlands	0.1983	0.1895	0.1944	0.1603	0.1536	0.1467
Poland	0.2129	0.2238	0.1755	0.1661	0.1874	0.1260
Portugal	0.2679	0.2420	0.2274	0.1963	0.1681	0.1723
Romania	0.2201	0.2834	0.2862	0.3514	0.2386	0.0744
Sweden	0.2274	0.2094	0.1815	0.1599	0.1630	0.1737
Slovakia	0.1833	0.2221	0.4600	0.1025	0.3413	0.1567

Note. continuous variable from 0=no hate to 1=presence of hate.

We found that some countries are noteworthy for their peaks, e.g., Cyprus, Greece, Italy or Lithuania. Our findings show some major variations depending on the year.



Sometimes it is a decrease in hateful content (see Lithuania's downward variation from 0.88 in 2018 to 0.53 in 2020) and in other cases there is an increase in hate speech (e.g., Cyprus went from a value of 0.32 in 2017 to a 0.53 score in 2020). The supplementary materials include the yearly average broken down by regions (NUTS 2: <https://doi.org/10.6084/m9.figshare.16708969.v1>) and cities (NUTS 3: <https://doi.org/10.6084/m9.figshare.16708954.v1>).

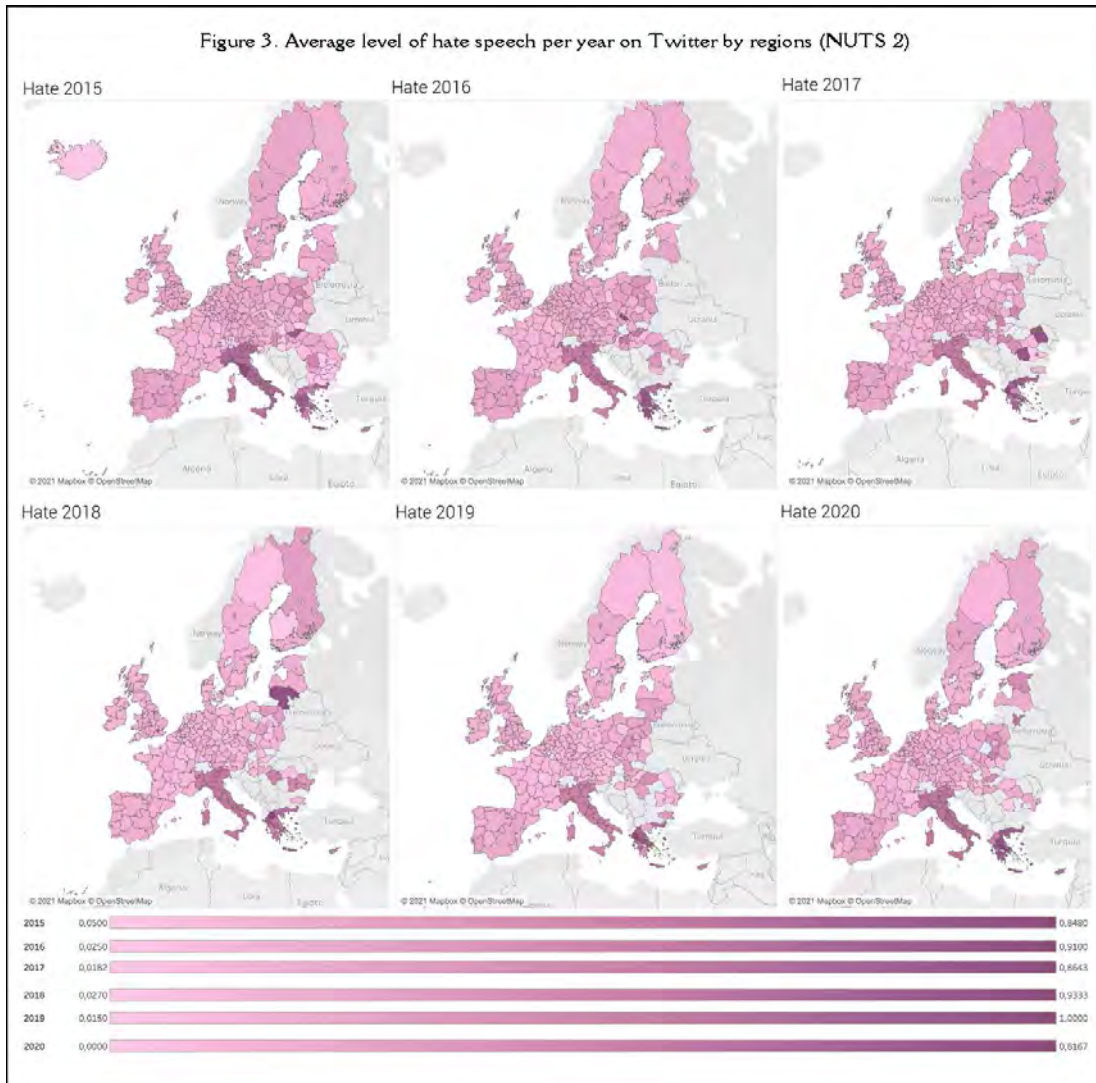


Note. Created by the authors based on the data obtained from the simulations based on the Eurobarometer by Arcila-Calderón et al. (2021).

Considering the first hypothesis (H1), posing that the share of immigrant population in European regions (NUTS 2) is related to citizen support towards immigrants, is depicted in a map (Figure 1) visually presenting the percentage of immigrants by region. As can be noticed, Southern or Mediterranean Europe has a higher percentage of immigrants than other mainland areas. Furthermore, a visual map is also used to show the average support towards immigrants and refugees during each period selected, with data obtained from the Eurobarometer (Figure 2). The mapped average support is mostly positive except in certain regions that drag the average down, like Italy and Eastern Europe.

The correlation test shows a statistically significant correlation between the variable “share of immigrant population by European regions” (NUTS 2) and the variable “support towards immigrants” in the second semester of 2015 ( $r=.254$ ,  $p<.001$ ), in 2016, first semester ( $r=.298$ ,  $p<.001$ ) and second semester

( $r=.308$ ,  $p<.001$ ), as well as in 2017, both in the first ( $r=.300$ ,  $p<.001$ ) and second semesters ( $r=.292$ ,  $p<.001$ ) (Table 2). Relying on the work of Sampieri et al. (2014), it is worth noting that the correlation is positive but weak, since it ranges between .10 and .50. Note that it is a positive relationship, thus confirming the hypothesis (H1) that the higher the percentage of immigrants in a given region, the greater the support towards them.



Note. Created by the authors based on the data obtained from the simulations based on the Eurobarometer by Arcila-Calderón et al. (2021).

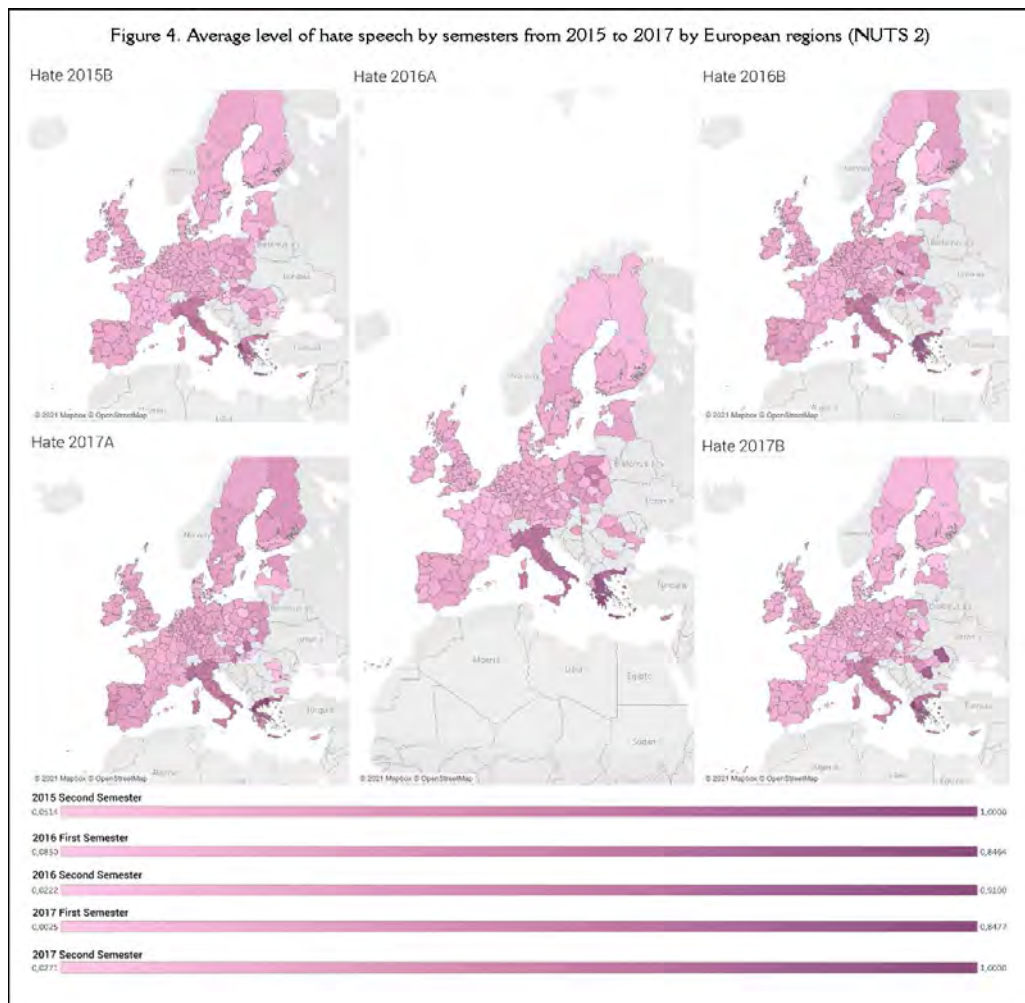
To test the second hypothesis (H2), which suggested that the share of immigrant population in European regions (NUTS 2) is correlated with the hate level on Twitter, first see the maps depicting the average values for both variables (Figure 1 and Figure 3) broken down by regions and years. As for the hate level, Figure 3 shows that in the darker areas there is more hate, and in 2020 we found remarkably high levels of hate in Italy and Greece.

Table 2 below includes the correlation analysis between the share of immigrant population and the average level of hate speech found on Twitter, broken down by the studied years and European regions (NUTS 2). We found that the correlation between the share of immigrants by European region is statistically trending in 2015 ( $r=.089$ ,  $p<.10$ ) and 2018 ( $r=.089$ ,  $p<.10$ ), and not significant in 2016 ( $r=.070$ ,  $p>.10$ ), 2017 ( $r=.045$ ,  $p>.10$ ) and 2019 ( $r=.074$ ,  $p>.10$ ). However, there is a statistically



significant correlation in 2020 ( $r=.147$ ,  $p<.05$ ). There is a positive trend, thus disproving our hypothesis (H2), i.e., that the greater the percentage of immigrants in a region, the lesser the level of hate speech on Twitter towards migrants and refugees, since the only significant result and the two trends show the opposite relationship.

To test hypothesis H3, posing that the level of hate speech on Twitter towards migrants and refugees in European regions (NUTS 2) is correlated with the degree of citizen support for these groups, Figure 2 provides the visual map depicting the latter variable. Also, a new map is added, showing the average level of hate speech over the same time periods (Figure 4), which is very similar to the yearly values of Figure 3, with greater hate speech in countries like Italy or Greece.



Note. Created by the authors based on the data obtained from the simulations based on the Eurobarometer by Arcila-Calderón et al. (2021).

In Table 2, the correlation between the support towards immigration (by semester) and the level of hateful content on Twitter during those time periods shows that the second semester of 2015 ( $r=-.104$ ,  $p<.05$ ), the first semester of 2016 ( $r=-.266$ ,  $p<.001$ ), the second semester of 2016 ( $r=-.234$ ,  $p<.001$ ), and the second semester of 2017 ( $r=-.245$ ,  $p<.001$ ) were statistically significant with a negative trend. Therefore, the hypothesis that regions with a lower level of hate speech on Twitter are more supportive of immigrants was validated. The only non-significant period was the first semester of 2017 ( $r=-.096$ ,  $p>.05$ ).

<b>Table 2. Pearson correlation between the indicated variables by European regions (NUTS 2)</b>						
<b>Correlation between the share of immigrants and the support for immigration by semesters and time periods</b>						
	Level of support for immigration					
	2015 B	2016 A	2016 B	2017 A	2017 B	
Share of immigrant population	.254**	.298**	.308**	.300**	.292**	
<b>Correlation between the share of immigrants and the level of hate on Twitter by year</b>						
	Hate level					
	2015	2016	2017	2018	2019	2020
Share of immigrant population	.089	.070	.045	.089	.074	.147*
<b>Correlation between the average hate level on Twitter and the support for immigration by semester and time period</b>						
	Hate level					
	2015B	2016A	2016B	2017A	2017B	
Level of support for immigration	-.104*	-.266**	-.234**	-.096	-.245**	

Note. \*\* Correlation is significant at the 0.01 level (1-tailed).

\* Correlation is significant at the 0.05 level (1-tailed).

#### 4. Conclusions and discussion

This research has required demanding documentation efforts, collecting large datasets, classifying, and applying advanced computational methods to supplement the common framework of studies on migration flows and dynamics in Europe. Relying both on data generated at large scale and existing indicators, this research was aimed at studying future scenarios of acceptance towards immigrants and refugees, considering the intergroup contact theory (ICT) to attempt an explanation of the hostility towards these social groups, and using hate speech, for the first time, as a predictor of such acceptance. A valuable contribution of this study includes proving the ability to obtain geolocated data from social media allowing for new cross references from innovative perspectives, also overcoming some of the limitations of the existing official indicators based on surveys, e.g., high costs, time, the amount of data or the lack of detail for specific locations.

First, this research tests the long-standing principles of Allport's (1954) ICT, analyzing if regions with higher immigration rates foster positive interpersonal contact, assuming (based on the ICT) that intergroup contact can encourage tolerance and acceptance. This claim, raised in hypothesis H1, is validated by our findings and results. Therefore, it is worth stating that there is greater support for immigrants in European regions with a higher share of immigrants. Along these lines —although assuming the perspective of the mediated ICT and considering hate speech as a factor mediating between the effects of contact with outgroup members and individuals' attitudes— this research analyzed if regions with higher immigration rates show a lower level of hate speech on Twitter, as stated in hypothesis H2. In this case, only a significant result was obtained, but the opposite premise was proven, i.e., that the higher the share of immigrants, the greater the level of hate speech on Twitter. As noted above, this can be due to the social desirability shown by respondents, who give answers that differ from their actual attitudes, values or behaviors, to look better to others or to provide a socially preferred image (Larson, 2019). So, social media and virtual communities can provide users with anonymity for them to express their real ideas and opinions, thereby changing their behavior online from their real-world behavior (Joinson, 1999).

Finally, this research has reviewed the relationship between the level of hate speech towards migrants and refugees on Twitter and the level of support for immigration obtained from the Eurobarometer and stated in hypothesis H3. The result is significant, thus confirming H3 since regions with greater support recorded a lower level of hate speech on Twitter. This is an innovative knowledge scenario for the acceptance of migrants and refugees in European regions relying on the presence of geolocated hate speech. For the first time, we can conduct new studies based on existing indices and indicators or previous results yet enhanced by geolocation and the pulse of anonymized social engagement and conversation in user-generated media. Therefore, this work has accomplished its objective of discovering social acceptance of migrants and refugees in European regions, through a large-scale longitudinal analysis of online hate speech on Twitter and its contrast with existing official indicators. These findings are a great leap for studying the origins of hate speech narrowing it down to specific territories, and they can help to come up with solutions and measures to tackle the expansion of discriminatory, racist, and xenophobic behaviors. Social media pose a true challenge for social scientists seeking to analyze online messages with a purpose of better understanding human interaction and improving the human condition (Felt, 2016).

Future lines of work can also deepen into the influence of other mass media in the acceptance of migrants and refugees to compare with these indicators.

As for the limitations of this research work, first we need to take into account that Twitter is not considered representative of the whole society and its reality, but it can offer very relevant guidelines given the free expression of opinions among its diversity of users. Furthermore, handling and processing geolocation macrodata has been complex, and this complexity has led to reducing the sample size for the sake of data quality. At the same time, updating the data is essential to accurately determine the results, since the census used is updated every 10 years. As for the time scope, the different periods considered for each measure should be acknowledged as a limitation. For future works, it would be advisable to increase uniformity, e.g., for the immigrant support indicators in the Eurobarometer surveys (2020). Other limitation to consider is automatic translation (instead of the creation of *ex profeso* models for each language), given that in a new external validation with 500 translated tweets (including messages in Spanish, Greek or Italian) randomly selected from the sample and compared with the predictions of the model, we found that, although in general terms classes can be detected with some accuracy (accuracy=0.72), the prediction of hate is very weak (precision=0.18, recall=0.28, F-score=0.22), compared to the prediction of non-hate (precision=0.80, recall=0.79, F-score=0.83) for these automatically translated messages. Additionally, it should be mentioned that the results might be influenced by online bot campaigns or by offline events strongly affecting online conversations, such as the “Aquarius” case (Arcila-Calderón et al., 2021). The approach to hate speech is generic given that it is analyzed using thematic filters, but the different types of online hate are not studied in detail. Besides, the keywords are specific, and a more in-depth study could be conducted considering the linguistic idiosyncrasy of each country and the fact that, in each European territory, there are normative frameworks and regulations of speech in social media that can also condition the existence of a larger or smaller degree of hate speech.

Generally, this work raises new questions and opens additional lines of research, allowing (i) to continue inquiring about the use of new technologies and online platforms on top of the existing traditional methodologies; and, after creating more complex models, (ii) to present possible tools and means to fight the detected social issues.

### Authors' Contribution

Conceptualization, C.A-C.; Review of literature (state-of-the-art), C.Q-M, D.B-H, J.J.A.; Methodology, C.A-C, P.S-H; Data analysis, C.Q-M., C.A-C, P.S-H, D.B-H, J.J.A.; Results, C.Q-M., P.S-H, J.J.A, D.B-H; Discussion and conclusions, C.A-C, P.S-H; Writing (original draft preparation), C.A-C, P.S-H, C.Q-M.; Final review and editing, C.A-C, P.S-H, C.Q-M, J.J.A, D.B-H; Project design and sponsorship, C.A-C, P.S-H, J.J.A, D.B-H.

### Notes

<sup>1</sup> European Project HumMingBird. Additional information at <https://hummingbird-h2020.eu/>

<sup>2</sup> Geocoding. Additional information at <https://nominatim.org/>

<sup>3</sup> NutsFinder. Classification of NUTS codes. See the app at <http://www.pypi.org/project/nuts-finder/>

<sup>4</sup> Google Translate APIs. Available at <https://buildmedia.readthedocs.org/media/pdf/textblob/latest/textblob.pdf>

<sup>5</sup> Python textblob code used for the translation. Additional information at <https://textblob.readthedocs.io/en/dev/>

<sup>6</sup> All the interactive maps are available at: <https://doi.org/10.6084/m9.figshare.16708960.v2>

### Funding Agency

This research falls within the scope of the international research project “Enhanced migration measures from a multidimensional perspective (HumMingBird)” funded by the European Union under the Horizon 2020 Research and Innovation Programme, with reference number 870661 (<https://hummingbird-h2020.eu/>). It has also been supported by international research project “Preventing Hate Against Refugees and Migrants (PHARM)” funded by the European Union under the Rights, Equality and Citizenship Programme (2014-2020) with reference number 875217 and by the tool Stop-Hate, funded by the Fundación General de la Universidad de Salamanca as a competitive proof of concept within the TCUE plan, with reference PC-TCUE18- 20\_016. This work has been supported by the Research Group Observatorio de los Contenidos Audiovisuales ([www.ocausal.es](http://www.ocausal.es)) of the University of Salamanca. The authors would also like to explicitly thank the members of the projects by their assistance during the research, specially: Andreas Veglis, Charalampos Dimoulas, Lazaros Vrysis, Nikolaos Vryzas, Sergio Splendore and Martín Oller. The authors want to thank as well the support provided by the operative online infrastructure of the Supercomputing System of Castille and Leon (Scayle). Finally, we would like to thank Guillermo Frutos (Frutos Miranda Traductores) for his translation work.

## References

- Abrams, D., & Hogg, M.A. (2017). Twenty years of group processes and intergroup relations research: A review of past progress and future prospects. *Group Processes & Intergroup Relations*, 20, 561-569. <https://doi.org/10.1177/1368430217709536>
- Abrams, J.R., Mcgaughey, K.J., & Haghighat, H. (2018). Attitudes toward Muslims: a test of the parasocial contact hypothesis and contact theory. *Journal of Intercultural Communication Research*, 47(4), 276-292. <https://doi.org/10.1080/17475759.2018.1443968>
- Allport, G.W. (1954). *The nature of prejudice*. Addison-Wesley.
- Alto Comisionado de las Naciones Unidas para los Refugiados (Ed.) (1951). *Convención sobre el Estatuto de los Refugiados*. <https://bit.ly/2Zz2Y70>
- Arcila-Calderón, C., Amores, J.J., Sánchez-Holgado, P., & Blanco-Herrero, D. (2022a). Using text classification to detect online hate towards migrants and refugees. Developing and evaluating a classifier of racist and xenophobic hate speech using shallow and deep learning. [Submitted for publication].
- Arcila-Calderón, C., Amores, J.J., & Stanek, M. (2022b). Using machine learning and synthetic populations to predict support for refugees and asylum seekers in European regions. In S. Korkmaz, & B. Bircan (Eds.), *Data science for migration and mobility*. Oxford University Press.
- Arcila-Calderón, C., Blanco-Herrero, D., Frías-Vázquez, M., & Seoane-Pérez, F. (2022). Refugees welcome? Online hate speech and sentiments in Twitter in Spain during the reception of the boat Aquarius. *Sustainability*, 13(5). <https://doi.org/10.3390/su13052728>
- Arcila-Calderón, C., Blanco-Herrero, D., & Valdez-Apolo, M.B. (2020). Rechazo y discurso de odio en Twitter: Análisis de contenido de los tuits sobre migrantes y refugiados en español. *Revista Española de Investigaciones Sociológicas*, 172, 21-40. <https://doi.org/10.5477/cis/reis.172.21>
- Bayer, J., & Bárd, P. (2020). *Hate speech and hate crime in the EU and the evaluation of online content regulation approaches*. Policy Department for Citizens' Rights and Constitutional Affairs. <https://doi.org/10.2861/28047>
- Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Proceedings of the International AAAI Conference on Web and Social Media*, 5, 450-453. <https://bit.ly/3DO1hRR>
- Broad, G.M., Gonzalez, C., & Ball-Rokeach, S.J. (2014). Intergroup relations in South Los Angeles. Combining communication infrastructure and contact hypothesis approaches. *International Journal of Intercultural Relations*, 38(1), 47-59. <https://doi.org/10.1016/j.ijintrel.2013.06.001>
- Cabo-Isasi, A., & García-Juanatey, A. (2017). *Hate speech in social media: A state-of-the-art-review*. <https://bit.ly/33bfH3>
- Canelón, A.R., & Almansa, A. (2018). Migración: Retos y oportunidades desde la perspectiva de los Objetivos de Desarrollo Sostenible (ODS). *Retos*, 8. <https://doi.org/10.17163/ret.n16.2018.08>
- Cinelli, M., Morales, G.D.F., Galeazzi, A., Quattrociocchi, V., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), 1-8. <https://doi.org/10.1073/PNAS.2023301118>
- Ekman, M. (2019). Anti-immigration and racist discourse in social media. *European Journal of Communication*, 34(6), 606-618. <https://doi.org/10.1177/0267323119886151>
- Esipova, N., Ray, J., & Pugliese, A. (2020). *World grows less accepting of migrants*. Gallup. <https://bit.ly/3uDGL3r>
- ESS-ERIC Consortium (Ed.) (2021). *European Social Survey (ESS)*. <https://bit.ly/3ilifiq>
- Eurobarómetro (Ed.) (2020). *Public opinion in the European Union*. <https://doi.org/10.2775/460239>
- Eurobarómetro (Ed.) (s.f.). *Sondeos periódicos de opinión del Parlamento Europeo*. <https://bit.ly/3up3rnR>
- European Commission (Ed.) (2016). *The Census Hub: Easy and flexible access to European census data*. Publications office of the European Union. <https://doi.org/10.2785/52653>
- Felt, M. (2016). Social media and the social sciences: How researchers employ Big Data analytics. *Big Data & Society*, 3(1). <https://doi.org/10.1177/2053951716645828>
- Gallacher, J.D., Heerdink, M.V., & Hewstone, M. (2021). Online engagement between opposing political protest groups via social media is linked to physical violence of offline encounters. *Social Media + Society*, 7(1). <https://doi.org/10.1177/2056305120984445>
- Inter-Parliamentary Union (Ed.) (2015). *Migration, human rights and governance*. The International Labour Organization/The United Nations. <https://bit.ly/3F0KDjz>
- Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, 31(3), 433-438. <https://doi.org/10.3758/bf03200723>
- Larson, R.B. (2019). Controlling social desirability bias. *International Journal of Market Research*, 61(5), 534-547. <https://doi.org/10.1177/1470785318805305>
- Müller, K., & Schwarz, C. (2020). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, 19(4), 2131-2167. <https://doi.org/10.1093/jea/jvaa045>
- Organización de las Naciones Unidas (Ed.) (2015). *Agenda 2030 - Objetivo 10. Reducción de las desigualdades*. ONU. <https://bit.ly/3ulmcZi>
- Organización de las Naciones Unidas (Ed.) (2018). *Global compact for safe, orderly and regular migration*. ONU. <https://bit.ly/3op7VtL>
- Organización de las Naciones Unidas (Ed.) (2019). *International migration policies. Data booklet*. Statistical Papers - United Nations. <https://doi.org/10.18356/0a2bc93d-en>
- Pettigrew, T.F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49(1), 65-85. <https://doi.org/10.1146/annurev.psych.49.1.65>
- Salvini, M. [@matteosalvinimi] (2020). *Ecco il bel regalo di Natale di chi ci odia. Autore di questo scempio schifoso un palestinese*. [Tweet]. Twitter. <https://bit.ly/3IHdIXK>



- Sampieri, R.H., Collado, C.F., Lucio, P.B., Valencia, S.M., & Torres, C.P.M. (2014). *Metodología de la investigación*. McGraw-Hill Education. <https://bit.ly/3ESaIRq>
- Sayce, D. (2020). *The number of tweets per day in 2020*. <https://bit.ly/3CZbUB2>
- Siapera, E., Boudourides, M., Lenis, S., & Suiter, J. (2018). Refugees and network publics on Twitter: Networked framing, affect, and capture. *Social Media + Society*, 4(1). <https://doi.org/10.1177/2056305118764437>
- Vox [@Vox\_es] (2021). *Mientras impiden la movilidad a los españoles, más de 700 inmigrantes ilegales han asaltado nuestras fronteras en los últimos días*. [Tweet]. Twitter. <https://bit.ly/3GAmGQe>
- Vrysis, L., Vryzas, N., Kotsakis, R., Saridou, T., Matsiola, M., Veglis, A., Arcila-Calderón, C., & Dimoulas, C. (2021). A Web interface for analyzing hate speech. *Future Internet*, 13(3), 80-80. <https://doi.org/10.3390/fi13030080>
- We are social/Hootsuite (Ed.) (2021). *Digital 2021 October Global Statshot Report*. Digital 2021 Global Digital Overview. <https://bit.ly/32c89el>
- Zhang, J.S., Tan, C., & Lv, Q. (2019). Intergroup contact in the wild: Characterizing language differences between intergroup and single-group members in NBA-related discussion forums. In A. Lampinen, D. Gergle, & D. A. Shamma (Eds.), *Proceedings of the ACM on Human-Computer Interaction* (pp. 1-35). CSCW. <https://doi.org/10.1145/3359295>



Audiovisual project for childhood  
media literacy development

GRUPO  
**Comunicar**



Follow us: <http://www.bubuskiski.es/>





# Hate speech analysis as a function of ideology: Emotional and cognitive effects

Análisis del discurso de odio en función de la ideología:  
Efectos emocionales y cognitivos

- ID** Dr. Natalia Abuín-Vences. Professor, Department of Applied Communication, Complutense University of Madrid (Spain) (nabuinve@ucm.es) (<https://orcid.org/0000-0002-4153-9390>)
- ID** Dr. Ubaldo Cuesta-Cambra. Full Professor, Department of Communication Theory and Analysis, Complutense University of Madrid (Spain) (ucuestac@ucm.es) (<https://orcid.org/0000-0001-7023-7132>)
- ID** Dr. José-Ignacio Niño-González. Associate Professor, Department of Communication Theory and Analysis, Complutense University of Madrid (Spain) (josenino@ucm.es) (<https://orcid.org/0000-0001-6940-2399>)
- ID** Carolina Bengochea-González. Assistant Professor, Department of Communication Theory and Analysis, Complutense University of Madrid (Spain) (cbengoch@ucm.es) (<https://orcid.org/0000-0003-2250-160X>)

## ABSTRACT

Hate speech is a major problem, especially in the political environment, where it generates polarization and social conflict. There are no experimental laboratory studies that have analyzed the mechanisms of action of hate speech using biometric records to evaluate the implicit negative emotion caused by these discourses, despite the essential importance of emotion in this problem. This work investigates the interaction between the ideology of the source and the audience (progressive vs. conservative) and the type of message (hate vs. non-hate) on cognitions (perception of the veracity of the message and the source) and negative emotion (biometric and self-reported). Biometric emotion was assessed by recording heart rate variability and electrodermal response. The results highlight the importance of evaluating implicit biometric responses that reveal inaccessible data with other methodologies: conservatives show more emotion in their biometric records than self-reported ones. With these measures, hate desensitization effects are also detected in both audiences. With self-reported measures, more negative emotion is detected in progressive audiences only when the source is contrary to their ideology; on the other hand, conservatives show less negative emotion in all cases. These results are interpreted as a consequence of normative pressure and endogroup cognitive biases.

## RESUMEN

El discurso de odio constituye un importante problema, especialmente en el entorno político, donde genera polarización y conflicto social. No existen trabajos experimentales de laboratorio que hayan analizado sus mecanismos de acción empleando registros biométricos para evaluar la emoción negativa implícita provocada por estos discursos, a pesar de la importancia esencial de la emoción en este problema. El objetivo de este trabajo es investigar, en una muestra universitaria, la interacción entre la ideología del emisor y el receptor (progresista vs. conservador) y el tipo de mensaje (odio vs. no-odio) sobre las cogniciones (percepción de credibilidad del mensaje y del emisor) y la emoción negativa (biométrica y auto-informada). La emoción biométrica se evaluó mediante registros de la variabilidad intercardíaca y la respuesta dermoeléctrica. Los resultados destacan la importancia de evaluar las respuestas biométricas implícitas que revelan datos inaccesibles con otras metodologías: los conservadores muestran mayor emoción en sus registros biométricos que en los auto-informados. También se detectan, con estas medidas, efectos de desensibilización del odio en ambas audiencias. Con medidas auto-informadas, se detecta mayor emoción negativa en las audiencias progresistas solo cuando el emisor es contrario a su ideología. En cambio, los conservadores manifiestan menor emoción negativa en todos los casos. Se interpretan estos resultados como consecuencia de la presión normativa y de sesgos endogrupales de conformidad.

## KEYWORDS | PALABRAS CLAVE

Hate speech, implicit emotion, ideology, neuro communication, credibility, GSR.  
Discurso de odio, emoción implícita, ideología, neurocomunicación, credibilidad, GSR.



## 1. Introduction and state of the art

In recent years, social networking sites have functioned as amplifiers (Colleoni et al., 2014) of more radical views than face-to-face interaction, enabling hate speech to infiltrate political discourse through these digital media. Likewise, it seems that this escalation in hate speech is accompanied by a higher prevalence of this type of crime on the streets: 1,598 in 2018, 1,706 in 2019, and 1,401 in 2020 (the decline during this year was due to home confinement and mobility restrictions resulting from the COVID-19 pandemic (Ministerio del Interior, 2018; 2019; 2020). Hate speech models the ideology of recipients with like-minded ideas, aiming to reaffirm social identity and setting the difference with other social groups resulting in the phenomenon of desensitization, increasing prejudice, and prompting avoidance of the subjects targeted by hate, potentially triggering violent acts.

Socially, the problem seems to be increasing, so research in this field is growing and providing new ways to understand and behave. Thus, research on homophobic behaviors through allocation matrices developed by Fasoli et al. (2014) provided important new avenues of work (e.g. on discrimination and resource allocation). Furthermore, Saha et al. (2018) formulated machine learning models for the detection of gender-based violence on social media platforms that enable the screening of hate messages using algorithms.

In this paper we also try to bring new perspectives to this area of study: we seek to analyze the emotional impact of hate speech in young university students according to their ideology, using new models and tools derived from neuro communication for the evaluation of the emotional impact. More specifically, the study examines whether an ideology aligned with the subject who utters the speech interacts with this type of messages and whether they are perceived as hate or, on the contrary, when the individual who delivers the speech has ideas aligned with those of the receiver, a desensitization phenomenon is produced, which causes the individual to normalize this discourse, increasing its persuasive capacity.

### 1.1. Hate speech, social identity, and ideology

Delgado and Stefancic (1995) defined hate speech as a public, conscious, and deliberate statement aimed at denigrating a group of people. In this sense, in his pioneering work Tajfel (1978) indicated that the configuration of social stereotypes is characterized by the need to highlight the perceived similarities between members of the same group or social category and to emphasize the differences with different groups or categories, frequently denigrating the opposing group.

Billing and Tajfel (1973) concluded that group construction and intergroup behavior occurred as a result of social categorization processes, which activate a social identity that triggers behaviors related to endogroup favoritism. In fact, social identity is the psychological driver of intergroup behavior, which is based on a social categorization mechanism that triggers a social comparison process. As identification with the in-group increases, one moves from the interpersonal to the intergroup extreme (Canto & Moral, 2005). This pushes the individual to search for and accentuate a positive differentiation in favor of his group when compared to others (Tajfel & Turner, 1979). This theory would help to explain that the targets of hate speech are other identity groups since it explains prejudice, discrimination, and intergroup relations by resorting to social categories where certain social identities are reaffirmed. In this line, Leets (2002) indicates that hate speech intends to denigrate a given target based on perceived differences.

Leader-Maynard and Benesch (2016) argue that both this discourse and the dangerous ideology that promotes it constitute a real risk of turning into crimes and attacks, so it is necessary to monitor and combat all hate speech given its dangerousness. As early as 1954, Allport (1954) indicated that verbal expressions of prejudice, which he called antilocutions, can lead to avoidance of the target group, discrimination, physical attack, and extermination. A clear example of this scale is the Nazi extermination, which began by singling out a particular group and facilitated its isolation and subsequent elimination (Bilewicz & Soral, 2020). In this sense, we can state that there is considerable consensus among scholars and international actors that surrounding ideologies and discourse play a fundamental role in the escalation of violence (Leader-Maynard & Benesch, 2016).

Likewise, and as indicated by Atienza-Cerezo and van-Dijk (2010), social identities are not innate, but are acquired from childhood, although it is true that they gradually change and transform through

discourse and other forms of interaction. Moreover, discourse aimed at promoting a given social identity involves underlying ideological components that can create prejudice towards certain groups (Atienza-Cerezo & van-Dijk, 2010). This suggests that different ideologies contribute to the configuration of social identities that coexist in society and do so through discourse. Thus, the relationship between discourse and ideology has long been a subject of theoretical reflection in different scientific disciplines such as linguistics, philosophy, history, political theory, and social theory (Leader-Maynard & Benesh, 2016). However, it generally lacks theoretical clarity in the contemporary literature on mass violence. This is highly relevant, as discourse and ideology are intrinsically related.

Following recent trends in the study of ideology in the social sciences, we conceptualize ideology broadly, defining it as a distinctive system of normative ideas, typically shared by members of groups or societies, that underpin their understanding of the world and guide their political behavior (Leader-Maynard & Benesh, 2016). As systems of ideas, ideologies are stored in the memory, providing cognitive resources for thought processes, including decision-making, and thus shaping individual behavior. Each person's ideology is unique, but social scientists also analyze group ideologies: analytic constructs that describe important similarities between group members' personal but heterogeneous ideologies (Leader-Maynard & Benesh, 2016).

Ideologies can also acquire important social dimensions by being embedded in institutions and recognized in political discourse. There is also evidence that hatred fuels political intolerance, defined as the support or willingness to denounce basic democratic values and equal rights of people belonging to a defined out-group in a particular society (Gibson, 2006), and is considered one of the most problematic phenomena in democratic societies. Moreover, in today's informational contexts, different types of discourse such as anti-vaccine, hate speech, negationism, and polarization, contribute to creating strong social unrest and hindering public health policies, which is especially relevant in cases such as the COVID-19 pandemic (Picazo-Sánchez et al., 2020). To study the emotional effects of hate speech in this context of polarization and political intolerance, the stimuli used in this research involved two political figures that provoke socio-political polarization (Olaz-Capitán & Ortiz-García, 2021): Pablo Iglesias and Santiago Abascal.

## 1.2. Exposure to hate speech, populism, desensitization, and selective perception

When we speak of discourse, we refer to any act of human communication, not only in the form of verbal discourse but also to non-verbal communication such as images, gestures, music, rituals, among others. Ideologies are communicated through discourse and, consequently, they are constructed and altered through discourse. They are also produced through an individual's thinking, but genuinely creative thinking is arduous. Most people avoid it most of the time, and instead appropriate ideas they glean from social discourse (Leader-Maynard & Benesh, 2016).

Individuals' continued exposure to such messages produces the phenomenon of desensitization: after an initial physiological and affective (negative) arousal, individuals gradually learn to ignore these messages and become desensitized. The General Model of Aggression outlined by Carnagey et al. (2007) explains how these physiological responses generate cognitive and affective outcomes that could lead to increased aggressive behavior and a decreased likelihood of helping victims. Desensitization processes observed at lower physiological levels (decreased heart rate and decreased skin conductance) reflect the extinction of negative emotional reactions towards violence associated with: decreased perception of the seriousness of the aggression decreased attention to violent events, decreased sympathy for victims of violence, increased belief that violence is normative, and decreased negative attitudes towards violence. Along the same lines, Soral et al. (2018) conducted several experiments that demonstrated that the greater the desensitization of the individual towards hate speech, the greater the persuasive capacity of the message and the prejudice towards the group targeted by hatred. In short, the individual becomes desensitized, normalizes hate speech, and transforms it into resentment, increasing prejudice and violence towards the subjects of hate.

Likewise, it is important to consider that when speakers delivering hate speech display traits such as aggressiveness, shyness, or fear, this affects their credibility (Cole & McCroskey, 2003), which becomes more pronounced in the case of a political figure delivering the message. Thus, the deployment of



aggressive persuasive strategies negatively affects the credibility of the message, which is perceived more negatively (Nau, 2012).

In this sense, it should be emphasized that hate speech feeds political intolerance, defined as the support or willingness to denounce basic democratic values and the equal rights of people belonging to a defined out-group in a particular society (Gibson, 2006). This is considered one of the most problematic phenomena in democratic societies. Hence the importance of analyzing the emotions it provokes, how individuals perceive this type of message and how ideology influences resulting emotions and perceptions. Desensitization to violence and hatred has been widely studied in the field of videogames (Bermejo-Berros & Soto-Sanfiel, 2015), but there is very little scientific evidence derived from experimental work in the case of political discourse in the media.

On the other hand, “selective perception bias” favors the positive evaluation of the discourse by senders with ideological affinity and the rejection of ideologically opposed discourse, especially in the field of political communication (Paz-García et al., 2020). These phenomena are causing an intense political polarization often accompanied by emotional reactions that are more linked to the sender’s profile than to the discourse itself. Some authors interpret these psycho-social mechanisms within the framework of “post-truth theories” (Waisbord, 2018).

Moreover, the elaboration likelihood model has demonstrated the importance of peripheral aspects in understanding the persuasive effectiveness of a message (Petty et al., 1983; Paredes et al., 2021). In the field of political communication, peripheral processing plays a very important role, given that receivers are not usually motivated to exert the effort to process through the central route (Marañón-Lazcano, 2015). In this sense, this model would confirm the importance of cognitive biases related to selective perception, as well as theories of intergroup biases, conflict, and prejudice (Crawford & Brandt, 2020). These theories analyze related but different mechanisms, such as desensitization, processing type, cognitive biases of selective perception, intergroup biases or polarization, and populism. However, all of them agree in proposing a likely result when analyzing political discourse mediated by a highly polarized and ideologically defined speaker: the cognitive and emotional effects provoked by hate messages will not depend so much on the content of the discourse as on the ideological affiliation of the receiver with that of the sender. Accordingly, this paper proposes the following hypotheses:

- Hypothesis 1: The cognitive and emotional effects triggered by discourse (hate vs. neutral) delivered by highly polarized senders will depend on their ideological affinity with the receiver
- Hypothesis 2: These effects will also display interaction between the main repeated measures factor (message type) and the grouping factor (ideology).

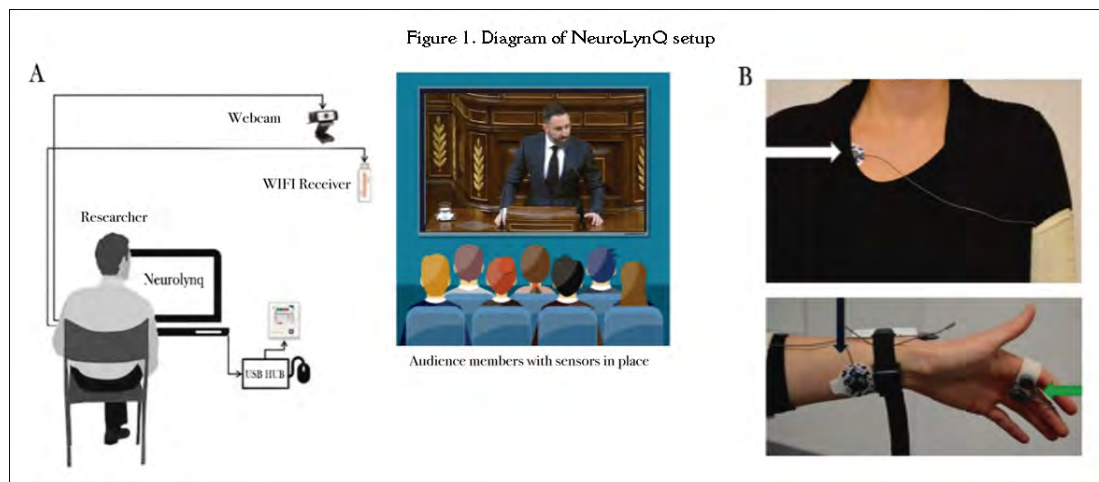
## 2. Material and methods

### 2.1. Material and instruments

This research aims to analyze the cognitive and emotional effects of the message. In this sense, there is strong empirical evidence that shows that the “perceived truthfulness” of both the sender and the message is among the most important cognitive effects (Hovland et al., 1953). The assessment of these variables has been shown to be very effective using traditional Likert-type self-report scales (Woodruff & Cashman, 1993). However, the assessment of emotion elicited by political speech has been controversial, since self-reports are often unreliable. This is due to the emotions that are produced implicitly, and the subjects’ insufficient awareness of the emotion elicited by the situation since it is produced at a pre-conscious level (Smith, 2020). This methodological difficulty has been a major problem for the objective study of emotional processes. However, neuro communication techniques are now available, making it possible to evaluate these emotions using highly accurate and valid biometric measures. In this area, the pioneering work of Cline et al. (1973) analyzed young people’s “desensitization” to broadcasts of violence on television using a methodology based on the measurement of autonomic responses (electrodermal skin response and heart rate variability). However, this line of research has not been developed, due to the enormous technical difficulties that existed to evaluate both variables together, to synchronize them adequately with the stimulus presented, and to carry it out in a non-invasive environment. In this study, we have evaluated the biometric emotional response using the NeuroLynQ tool, which allows the simultaneous collection of

biometric data from participants by measuring emotional arousal through electrodermal activity (EDA) and heart rate variability (HRV). Heart rate variability (HRV) is the result of interactions between the autonomic nervous system (ANS) and the intrinsic heart function mechanism. HRV is based on the balance between the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS), and it is one of the most effective measures of emotion (Fonfría et al., 2011).

Thus, the measurement of cardiac variability not only provides relevant information about emotional changes, but it is also the variable that allows validating the values obtained in the measurement of the electrodermal response avoiding false positives due to internal homeostasis processes (Tarnowski et al., 2018). This occurrence of false positives in the assessment of emotion by EDA is, unfortunately, very frequent, limiting the validity of the measurement. Consequently, the use of both measures simultaneously (HRV and EDA) is a dramatic advance in the objective assessment of subjects' emotions. Figure 1 shows a diagram of setup A, which includes both the computer with NeuroLynQ and the WIFI receiver with a webcam to record the session, and setup B for each participant, which includes both electrodes for heart rate (white arrow) and galvanic response (black and green arrow).



## 2.2. Methods

A multigroup repeated measures design was used. The “within-group” independent variable (repeated measures) was “type of speech/video” with four experimental conditions: Video-Iglesias-Hate, Video-Iglesias-No-Hate, Video-Abascal-Hate, Video-Abascal-No-Hate). The independent “between-group” (grouping) variable was “ideology type”. The dependent variables were measures of message and sender perception, as well as biometric and self-report measures of emotion. A group of 39 volunteer participants (31 females and eight males with a mean age of 21) at the undergraduate level were randomly recruited for the experiment. Previous work has validated that this sample size allows reliable conclusions to be reached (Cuesta et al., 2020). They were informed that they would be watching videos, but no information was provided about the purpose. They were then presented with a series of four 20-second videos featuring two Spanish politicians. Two of the videos featured Pablo Iglesias and the other two featured Santiago Abascal. Each politician was featured in one video delivering a neutral speech (about global politics) and one video delivering a hate speech (about immigration and inter-party power struggles). The videos were practically identical: they lasted 20 seconds and took place in the Spanish Parliament. These videos were designed by three experts in political communication, specifically, three political communication university professors with experience in defining these types of messages and procedures. A pilot study carried out with 15 university students confirmed the validity of the stimuli.

Before watching the videos, participants were given a questionnaire in which they indicated their political ideology on a scale from one (very conservative) to seven (very progressive). The videos were then presented to them, pausing after each one for participants to complete a 7-point Likert-type questionnaire

where they reported on the degree of credibility of the message and of the speaker, as well as on the “anger” (emotion) produced by the video. This type of scale has been previously used in similar research, where they have demonstrated its validity for this type of study (Cuesta et al., 2012). The order of the videos was randomized. The session was recorded with a webcam that sent the feed to a computer that collected the biometric signals, in order to identify the intervals where the videos were shown and synchronized with the biometric signals. The participants were finally informed about the purpose of the study following the ethical protocols validated by the university.

### 2.3. Data analysis

To compare the percentage of emotional response in the biometric variables between videos, a one-factor ANOVA (type of speech) was performed, followed by a Turkey post hoc test and Q-test to find out the effect between different stimuli. The significance level was set at  $p \leq 0.05$ . In addition, the graphs provided by the NeuroLynQ tool where emotion levels are reflected as a function of the type of video and the ideological affiliation of the recipient were analyzed. For this purpose, the political affiliation variable was dichotomized (1=conservative, 2=progressive) based on the extreme values of the scale. The results of the dependent variables, perception of the truthfulness of the sender and the message, as well as self-referred emotion (anger), were also subjected to analysis of variance and post hoc tests.

### 3. Analysis and results

Figure 2 shows how subjects holding a conservative ideology (red line) show a significantly lower level of emotion when confronted with Santiago Abascal’s two videos (hate and non-hate) than subjects holding a progressive ideology (blue line). However, their emotion levels increase during the two statements of Pablo Iglesias, although to a much lesser extent when confronted with Pablo Iglesias’ hate speech. This decrease in emotion in the face of the hate message coincides with the interpretation of desensitization theory. The emotional pattern presented by subjects whose ideology is progressive is very different: they show very similar elevated indices when faced with the four types of discourse, with a slight decrease in emotion when faced with Abascal’s hate condition. As in the previous case, these data coincide with the interpretation of desensitization theory.

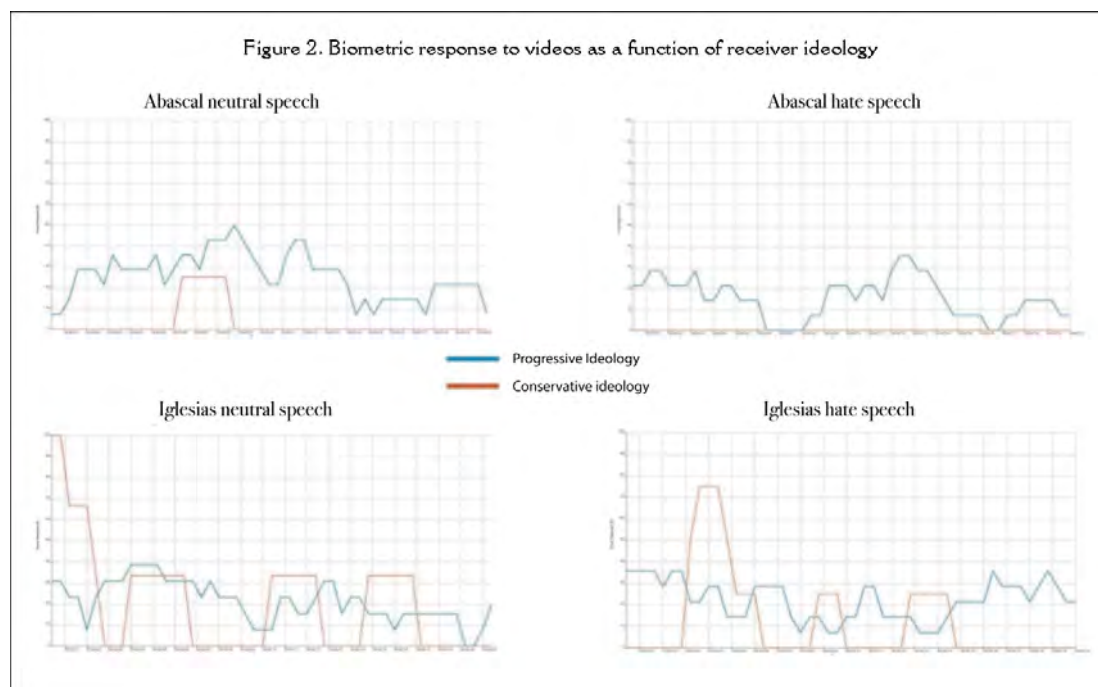
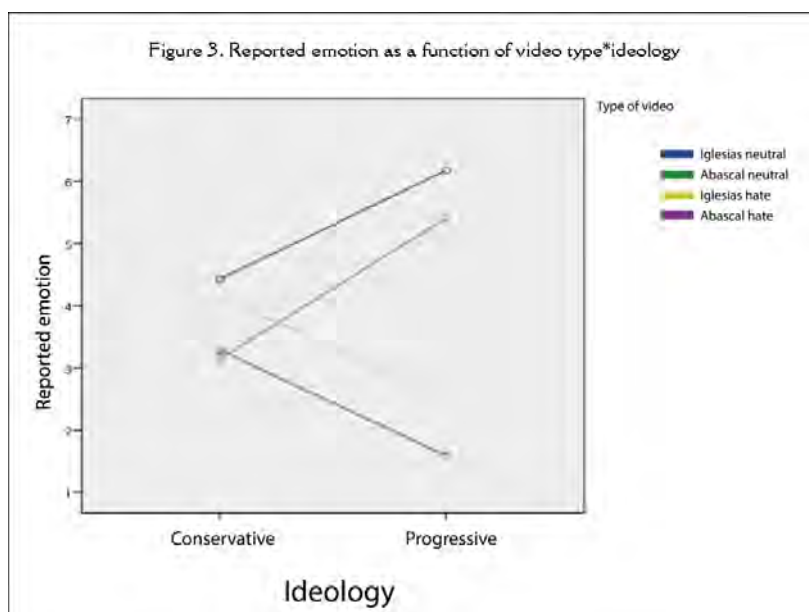


Table 1 shows the mean values and variances for these results. This confirms the data presented in Figure 2: the mean value of biometric emotion for the progressive group, in general (21.17) was higher than the level for the conservative group (9.27). The analysis of variance performed using the repeated measures factor “type of video/message” as an independent variable yielded a significant main effect ( $p < .000$ ). Mean differences post hoc tests showed statistically significant values for the conservative group ( $p < .000$ ): the conservative group shows, in general, less biometric emotion than the progressive group (9.27) but it increases significantly compared to the progressive sender (22.43 Iglesias-neutral and 11.76 Iglesias-hate). The pattern for the progressive group is very clear: their emotional level is much higher in general (21.17), and only shows a slight decrease in the “Abascal-hate” message (15.34). In summary: conservatives generally display less biometric negative emotion except when the sender is contrary to their ideology and delivers a hate message. Progressives show more emotion in general, but especially when it comes to the “Abascal-hate” condition.

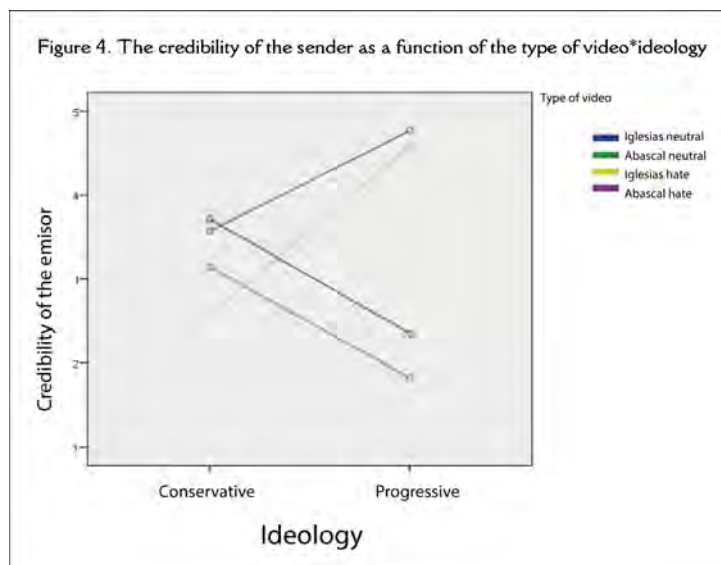
Table 1. Means and variances of biometric emotion for the groups (progressive vs conservative) as a function of video type		
Progressive		
Groups	Mean	Variance
Iglesias neutral	21.28	94.45
Abascal neutral	25.36	123.20
Iglesias hate	22.69	99.37
Abascal hate	15.34	94.06
Total	21.17	
Conservative		
Groups	Mean	Variance
Iglesias neutral	22.43	772.15
Abascal neutral	2.88	65.04
Iglesias hate	11.76	433.82
Abascal hate	0	0
Total	9.27	

An analysis of variance was then performed using the independent variable “type of video” as a repeated-measures factor and the independent variable “ideology” as a grouping factor. In this case, however, the dependent variable was “reported emotion” (dichotomized from the continuous scale). The analysis of variance yielded significant interaction of both main effects of video type ( $p < .001$ ) and video\*ideology interaction effects ( $p < .003$ ). Figure 3 presents the graph where these effects can be seen for the dependent variable “reported emotion”.

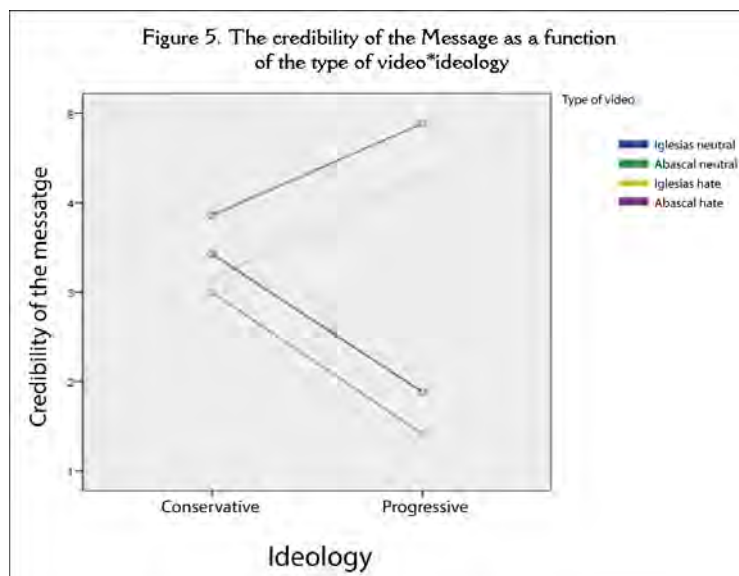




This figure shows that the results are similar, although not identical to those found with the “biometric emotion”: subjects with a conservative ideology tend to present lower and more stable mean values of emotion. Thus, they present very similar values for the four videos: 3.14; 3.29; 4.14; 4.42. These values do not differ from a statistically significant value ( $p > .10$ ), which means that for conservative subjects there is no difference either between the type of message (hate vs. non-hate) or between the type of sender (Abascal vs. Iglesias).



For the progressive group, the pattern is different: the emotion values are low, 1.59 and 2.53 for Iglesias' discourse non-hate and hate, respectively (but the differences are not significant  $p < .10$ ) and very high for Abascal's discourse: 5.41 and 6.18 for Abascal's non-hate and hate respectively (neither have significant differences between them  $p > .10$ ). This means that for progressive subjects what is relevant is not the type of message (hate vs. non-hate) but the type of speaker: Abascal provokes anger, regardless of his message; Iglesias does not provoke anger, also regardless of his message.



Finally, an analysis of variance was performed using the independent variable “type of video” as a repeated-measures factor and the independent variable “ideology” as a grouping factor, but this time using “credibility of the message” and “credibility of the sender” as dependent variables. The analysis of variance yielded significant interaction of both main effects of video type ( $p < .001$ ) and video\*ideology interaction effects ( $p < .003$ ). Figures 4 and 5 below present the results of this analysis of variance showing both main and interaction effects.

As can be seen in the two figures, the pattern of behavior of both dependent variables is identical to that observed for the variable “provoked emotion”: among subjects of conservative ideology, neither the type of discourse nor the type of sender has an influence, while among progressive subjects, when the speaker is Abascal, both variables (credibility of the message and the sender) decrease, and when the speaker is Iglesias, both increase. The statistical significance values are also identical to those found for the dependent variable “reported emotion”. The inclusion of “gender” as a co-variable in the analysis of variance showed that its effect does not affect the conclusions.

#### 4. Discussion and conclusions

One of the most important results of this study demonstrates the importance of the polarization of audiences according to conservative vs. progressive ideologies. In this work, we have analyzed the emotional impact of hate speech in young university students according to their ideology, analyzing if an ideology akin to the speaker interacts with these types of messages and they are perceived as hate or if, on the contrary, when the speaker has ideas that are similar to those of the receiver, a desensitization phenomenon occurs, which causes the individual to normalize the speech, thus increasing its persuasive capacity.

In this sense, the first hypothesis put forward is in line with other similar empirical studies: “The cognitive and emotional effects triggered by discourse (hate vs. neutral) delivered by highly polarized senders will depend on their ideological affinity with the receiver”. The research results have partially confirmed the hypothesis: ideology strongly modulates the cognitive (perceived truthfulness of the sender and the message) and emotional (anger and biometric activity) effects of the message. In this sense, the results of this research provide a substantial contribution since the interaction hypothesis is also partially confirmed: conservative subjects experience lower levels of reported emotion towards both political leaders, regardless of ideological bias. On the other hand, progressive subjects experience higher levels of reported emotion towards the conservative speaker and lower levels of reported emotion towards the speaker of their own ideology.

Our study provides several important contributions to existing research. First, there is an unexpected interaction effect, since “message type” asymmetrically interacts with ideology. In other words, “message type” only interacts with “sender type” among progressive subjects. Moreover, the interaction does not occur due to the variable “hate”, but due to the variable “Iglesias”, meaning that it is at experimental treatment levels 1 (non-hate Iglesias) and 3 (hate Iglesias) of the “message type” variable that the highest levels of reported emotion appear, but only among progressive subjects (Figure 3).

Secondly, the results related to the dependent variables degree of “truthfulness of the message” and “truthfulness of the sender” present a profile in line with the previous ones: among conservative subjects, neither the type of message (hate vs. non-hate) nor the type of sender (Abascal vs. Iglesias) modifies the variables. However, among progressive subjects, the message and the sender are perceived as more credible when it is consistent with their progressive ideology (Figures 4 and 5). These results are of great interest: they confirm previous literature indicating that the higher the negative emotion, the lower the credibility of the sender and the message, thus providing greater robustness and confirming previous data regarding the interaction “type of message\*ideology”.

Thirdly, this research contributes novel data to the scientific literature on hate messages, ideology, and emotion: the biometric study of emotions. Again, the results are of great interest: conservatives show, in general, less biometric negative emotion, except when the sender is contrary to their ideology and delivering a hate message. Progressives show more biometric negative emotion in general, but especially when it comes to the “Abascal-hate” condition.

These data seem to indicate the following. In essence, the patterns of biometric and reported negative emotion essentially agree. In this sense, there is less overall activity in the conservative group and more in the progressive group, especially when the ideology of the sender does not coincide. However, there are some relevant differences between the two types of measures: among conservatives, the progressive speaker elicits some biometric emotion, which is “denied” in self-reported emotion. Both the conservative and progressive groups experience “less” negative emotion with hate messages than with neutral ones, which is consistent with the interpretation of desensitization theory (Soral et al., 2018). Thus, biometric emotion seems to be more reliable when assessing ideology\*message interactions. As previous researchers have highlighted, biometric measures are more reliable when assessing topics where the “politically correct” bias exerts strong pressure (Cuesta et al., 2021).

As a whole, this research confirms that the “selective perception bias” favors the positive evaluation of the discourse by speakers with whom there is ideological affinity and the rejection of discourse by speakers whose ideology is in conflict, especially in the field of political communication (Paz-García et al., 2020) and with a certain independence of the degree of hate transmitted, which would confirm the importance of the cognitive biases of selective perception, as well as the theories on intergroup biases, conflict, and prejudice (Crawford & Brandt, 2020). This research has several limitations to be resolved in future work. For instance, it is necessary to increase the sample size and representativeness, which will allow for the study of gender, age, and social class differences. Experimental procedures of non-repeated measures should also be tested. Finally, it is necessary to work with hate messages at different levels and categories in the future, also manipulating the type of senders and their ideological affiliations.

### Authors' Contributions

Idea, U.C., N.A.; Literature review (state of the art), N.A.; Methodology, U.C.; Data analysis, C.B.; Results, U.C.; Discussion and conclusions, U.C.; Writing (original draft), J.I.N.; Final revisions, J.I.N.; Project design and funding agency, N.A.

### Funding Agency

This research has been carried out as part of two projects: 1) MINECO PID2019-105613GB-C31 “Cartography of hate speech in Spain from the field of Communication: sports, bullfighting and politics”, financed by the Ministry of Economy and Competitiveness; and 2) MINECO RTI2018-097670-B-I00 “Sentiment and popularity of pro- and anti-vaccine messages in social networks: analysis of explicit and implicit responses using EEG, GSR, facial recognition and eye-tracking”, funded by the Ministry of Economy and Competitiveness (Spain).

### References

- Allport, G.W. (1954). *The nature of prejudice*. Addison-Wesley. <https://bit.ly/33EVBXR>
- Atienza-Cerezo, E., & Van-Dijk, T.A. (2010). Identidad social e ideología en libros de texto españoles de Ciencias Sociales. *Revista de Educación*, 353, 67-106. <https://bit.ly/3DOD263>
- Bermejo-Berros, J., & Soto-Sanfiel, M.T. (2015). Efectos psicofisiológicos longitudinales de desensibilización a la violencia en videojugadores. In I. M. de Salazar-Muñoz, & D. Alonso-Urbano (Eds.), *Videojuegos: Diseño y sociología* (pp. 93-109). Editorial ESNE. <https://bit.ly/3oYzYA2>
- Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, 51(1), 3-33. <https://doi.org/10.1111/pops.12670>
- Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, 3(1), 27-52. <https://doi.org/10.1002/ejsp.2420030103>
- Canto-Ortiz, J., & Toranzo, F. (2005). El sí mismo desde la teoría de la identidad social. *Escritos de Psicología-Psychological Writings*, 1(7), 59-70. <https://doi.org/10.24310/espiespsci.vi1.13399>
- Carnagey, N.L., Anderson, C.A., & Bushman, B.J. (2007). The effect of video game violence on physiological desensitization to real-life violence. *Journal of Experimental Social Psychology*, 43(3), 489-496. <https://doi.org/10.1016/j.jesp.2006.05.003>
- Cline, V.B., Croft, R.G., & Courier, S. (1973). Desensitization of children to television violence. *Journal of Personality and Social Psychology*, 27(3), 360-365. <https://doi.org/10.1037/h0034945>
- Cole, J.G., & McCroskey, J.C. (2003). The association of perceived communication apprehension, shyness, and verbal aggression with perceptions of source credibility and affect in organizational and interpersonal contexts. *Communication Quarterly*, 51(1), 101-110. <https://doi.org/10.1080/01463370309370143>
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317-332. <https://doi.org/10.1111/jcom.12084>
- Crawford, J.T., & Brandt, M.J. (2020). Ideological (A)symmetries in prejudice and intergroup bias. *Current Opinion in Behavioral Sciences*, 34, 40-45. <https://doi.org/10.1016/j.cobeha.2019.11.007>

- Cuesta, U., Canel, M.J., & Gurrionero, M. (2012). *Comunicación y terrorismo*. Tecnos. <https://bit.ly/3GJq7nA>
- Cuesta, U., Niño, J.I., & Martínez, L. (2021). *Viaje al fondo del neuromarketing*. Fragua. <https://bit.ly/3q4ZQcz>
- Cuesta, U., Niño, J.I., Martínez, L., & Paredes, B. (2020). The neurosciences of health communication: An fNIRS analysis of prefrontal cortex and porn consumption in young women for the development of preventive health programs. *Front Psychol*, 11, 1-16. <https://doi.org/10.3389/fpsyg.2020.02132>
- Delgado, R., & Stefancic, J. (1995). Ten arguments against hate-speech regulation: How valid symposium: Political correctness in the 1990s and beyond. *The Northern Kentucky Law Review*, 23, 475-490. <https://bit.ly/3IhZpi>
- Fasoli, F., Maass, A., & Carnaghi, A. (2014). Labelling and discrimination: Do homophobic epithets undermine fair distribution of resources? *British Journal of Social Psychology*, 54(2), 383-393. <https://doi.org/10.1111/bjso.12090>
- Fonfria, A., Poy, R., Segarra, P., López, R., Esteller, A., Ventura, C., Ribes, P., & Moltó, J. (2011). Variabilidad de la tasa cardíaca (HRV) y regulación emocional. *Fòrum de Recerca*, 16, 903-914. <https://bit.ly/3ur5tUq>
- Hovland, C.I., Janis, I.L., & Kelley, H.H. (1953). *Communication and persuasion; psychological studies of opinion change*. Yale University Press. <https://bit.ly/33u0SHv>
- Leader-Maynard, J., & Benesch, S. (2016). Dangerous speech and dangerous ideology: An integrated model for monitoring and prevention. *Genocide Studies and Prevention: An International Journal*, 9(3), 70-95. <https://doi.org/10.5038/1911-9933.9.3.1317>
- Leets, L. (2002). Experiencing hate speech: Perceptions and responses to anti-semitism and antigay speech. *Journal of Social Issues*, 58(2), 341-361. <https://doi.org/10.1111/1540-4560.00264>
- Marañón-Lazcano, F.D.J. (2015). *El spot como herramienta de persuasión política. Análisis del impacto de la publicidad política en la desafección política a través de la ruta central y periférica*. [Doctoral Dissertation, Universidad Autónoma de Nuevo León]. <https://bit.ly/3ywMdxP>
- Ministerio del Interior (Ed.) (2018). *Informe 2018 sobre la evolución de los delitos de odio en España*. <https://bit.ly/3ygSkzj>
- Ministerio del Interior (Ed.) (2019). *Informe 2019 sobre la evolución de los delitos de odio en España*. <https://bit.ly/3ygSkzj>
- Ministerio del Interior (Ed.) (2020). *Informe 2020 sobre la evolución de los delitos de odio en España*. <https://bit.ly/3ygSkzj>
- Nau, C. (2012). *The impact of gender and verbal aggressiveness on speaker and message perception in political speeches*. International Communication Association Conference. <https://bit.ly/3IRxnpj>
- Olaz-Capitán, A.J., & Ortiz-García, P. (2021). Polarización afectiva sobre las élites políticas. *Más Poder Local*, 45, 41-55. <https://bit.ly/3ikHelM>
- Paredes, B., Martínez, L., Cuesta, U., Briñol, P.A., Petty, R.E., & Moreno, L.M. (2021). The role of meta-cognitive certainty on pornography consumption. *Psicothema*, 33(3), 442-448. <https://doi.org/10.7334/psicothema2021.8>
- Paz-García, A.P., Brussino, S., & Alonso, D. (2020). Efectos del tratamiento periodístico sesgado ideológicamente en el procesamiento cognitivo de información política. *Opinión Pública*, 26(2), 351-376. <https://doi.org/10.1590/1807-01912020262351>
- Petty, R.E., Cacioppo, J.T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research*, 10(2), 135-146. <https://doi.org/10.1086/208954>
- Picazo-Sánchez, L., De-Frutos-Torres, B., & Gutiérrez-Martín, A. (2020). La COVID-19 en la prensa española. Encuadres de alarma y tranquilidad en las portadas de El País, El Mundo y La Vanguardia. *Revista de Comunicación y Salud*, 10(2), 355-384. [https://doi.org/10.35669/rcys.2020.10\(2\).355-384](https://doi.org/10.35669/rcys.2020.10(2).355-384)
- Saha, P., Mathew, B., Goyal, P., & Mukherjee, A. (2018). *Hate miners: Detecting hate speech against women*. <https://bit.ly/3q5kU2C>
- Smith, R. (2020). The three-process model of implicit and explicit emotion. *Neurosci Enduring Change*, 89(3), 25-55. <https://doi.org/10.1093/oso/9780190881511.003.0003>
- Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, 44(2), 136-146. <https://doi.org/10.1002/ab.21737>
- Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. Academic Press. <https://bit.ly/3p1TJXn>
- Tajfel, H., & Turner, J.C. (1979). An integrative theory of intergroup conflict. In *The social psychology of intergroup relations* (pp. 33-47). Brooks/Cole. <https://bit.ly/3m5BuhV>
- Tarnowski, P., Kołodziej, M., Majkowski, A., & Rak, R.J. (2018). Combined analysis of GSR and EEG signals for emotion recognition. *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, (pp. 137-141). <https://doi.org/10.1109/IIPhDW.2018.8388342>
- Waisbord, S. (2018). The elective affinity between post-truth communication and populist politics. *Communication Research and Practice*, 4(1), 17-34. <https://doi.org/10.1080/22041451.2018.1428928>
- Woodruff, S.L., & Cashman, J.F. (1993). Task, domain, and general efficacy: A reexamination of the self-efficacy scale. *Psychological Reports*, 72(2), 423-432. <https://doi.org/10.2466/pr0.1993.72.2.423>





SCHOOL OF  
REVIEWERS



# SCHOOL OF REVIEWERS

Scientific Communication Journals



[www.escueladerevisores.com](http://www.escueladerevisores.com)



# A systematic literature review of the representations of migration in Brazil and the United Kingdom

Una revisión sistemática de la literatura de las representaciones de la migración en Brasil y Reino Unido

**Id** Isabella Gonçalves. PhD Candidate, Department of Communication, Johannes Gutenberg University, Mainz (Germany) (idesousa@uni-mainz.de) (<https://orcid.org/0000-0002-6492-9236>)  
**Id** Dr. Yossi David. Professor, Department of Communication, Johannes Gutenberg University, Mainz (Germany) (ydauid@uni-mainz.de) (<https://orcid.org/0000-0001-8020-0317>)

## ABSTRACT

In recent decades, increased scholarly attention has been paid to the interactions between immigration, media coverage, framing, and the rise of populism. This paper draws on these interactions to systematically review peer-reviewed articles related to media representations of immigrants and refugees in Brazil and the United Kingdom (UK). The objective was to identify the tone used in such representations and the research methods applied in the articles. This is the first systematic literature review that compares studies on media and immigration including both Western, Educated, Industrialized, Rich, and Democratic (WEIRD) and non-WEIRD countries. This paper uses Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) to collect, map, and systematize 47 peer-reviewed articles published in the past two decades. The findings show that, in the studies we analyzed, the tone of representation of refugees in the media in both countries is negative. Mass media underrepresent the main actors and use stereotypes and discourses of otherization. This paper contributes to our understanding of the differences between studies, including those conducted in the UK and Brazil, and calls for more comparative studies that include countries from the global south and global north. It also demonstrates the standardization of frames and tone of representation on immigration in both countries, suggesting similar patterns across different countries.

## RESUMEN

En las últimas décadas, se ha prestado una mayor atención académica a las interacciones entre la inmigración, la cobertura de los medios de comunicación, los tipos de encuadres y el aumento del populismo. Este trabajo revisa sistemáticamente un grupo de artículos revisados por pares cuya temática son las representaciones mediáticas de los inmigrantes y los refugiados en Brasil y el Reino Unido con el fin de identificar el tono de dichas representaciones y los métodos de investigación utilizados. Esta es la primera revisión bibliográfica sistemática que compara estudios sobre los medios de comunicación y la inmigración que incluye una muestra de países del sur y del norte global. Este trabajo utiliza las directrices de PRISMA para recopilar, mapear y sistematizar 47 artículos revisados por pares y publicados en las últimas dos décadas. Encontramos tres tipos de tonos en los medios de comunicación analizados en los estudios de ambos países, ordenados aquí de mayor a menor frecuencia: negativo, neutro y positivo. Los medios de comunicación infrarrepresentan a los principales actores y utilizan estereotipos y discursos que trazan una frontera entre nosotros/as y los/as otros/as. Este trabajo contribuye a nuestra comprensión de las diferencias entre los estudios realizados en el Reino Unido y Brasil, y reclama más estudios comparativos que incluyan países del sur y del norte global. También demuestra la existencia de patrones de representación similares en ambos países.

## KEYWORDS | PALABRAS CLAVE

Immigrants, refugees, representation, systematic review, mass media studies, scientific production.  
 Inmigrantes, refugiados, representación, revisión, estudios de medios de comunicación, producción científica.



## 1. Introduction

Populistic rhetoric is characterized by discourses and frames dividing society into "us" and "them" (Araújo & Prior, 2020). Immigrants and refugees, as the ultimate other, are a frequent target of populist discourses, portrayed as threats to the economy, culture, and security (Levy et al., 2016; Wirz et al., 2018). Negative representation is associated with the increase of xenophobia and hate speech, weakening existing anti-discriminatory regulations and strengthening avoidance and discrimination towards immigrants and refugees (Bilewicz & Soral, 2020; Kroon et al., 2020). Media discourses may help shape and construct negative stereotypical frames that present immigrants and refugees as bad, threat or victim, in a process that divides in-groups from outgroups (Seate & Mastro, 2017; Matthes & Schmuck, 2017). Since the media influences the public agenda (McCombs & Shaw, 1972; McCombs & Valenzuela, 2007), it is essential to understand media representation and frames over time.

There are numerous studies focusing on media framing and representations of immigrants and refugees in Europe (Amores et al., 2019; Eberl et al., 2018; Lams, 2018). A robust literature review from 2018 examines the diversity of findings from those studies; however, it is noteworthy that none of these studies compares European and Latin American countries (Eberl et al., 2018). A vast majority of the studies have focused on European territories, with a minority also comparing European countries with the United States or African countries (Benson, 2013; Fengler et al., 2020). The prevalence of studies that focus on Western Educated, Industrialized, Rich and Democratic (WEIRD) societies was highlighted in communication and media studies (Chakravarty et al., 2018).

This systematic review seeks to contribute in filling this gap by focusing on studies about media and migration in Brazil and the United Kingdom (UK). Drawing on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines for systematizing literature, this paper aims to provide a better understanding of the differences and similarities in research related to media representations of immigrants and refugees in Brazil and the UK. It focuses on the following two questions: (RQ1) which research methods were used; and (RQ2) how the representation varied between Brazil and the UK in the findings of previous scholarly literature.

The study focuses on these two countries because of their recent interactions with the populist and far-right end of the political spectrum. Scholars have shown that immigration is an important factor in the success of right-wing political parties (Shehaj et al., 2021). In the UK, two events represent the political bond with populism: the UK Independence Party's ascension and the Brexit Referendum in 2016 (Hughes, 2019). In Brazil, the most prominent event is the election of President Jair Bolsonaro in 2018 (Lugo-Ocando, 2020; Novoselova, 2020). In both cases, politicians portrayed immigrants as enemies (Hughes, 2019; Silva, 2020).

Although populism rose in different countries worldwide, this study aims to compare one country from the global south and one country from the global north with different political and media systems with similar experiences on populism in the past years. The UK has a parliamentary form of democracy, while Brazil has a presidential system. Moreover, the UK belongs to the liberal model of media, while Brazil has characteristics from different models. Some aspects of Brazilian media are: private media enterprises' control of the press; the politicization of public broadcasting media; low level of newspapers circulation; and tradition of non-objective journalism style (Albuquerque, 2012; Hallin & Mancini, 2004). A comparison between these two countries may help us uncover the prevalent discourse and the similarities and differences between countries with different political and economic systems and traditions, and those that have experienced far-right and populist political movements in the 21st century. Moreover, this study can lead to more comparisons, including countries from the global south and north.

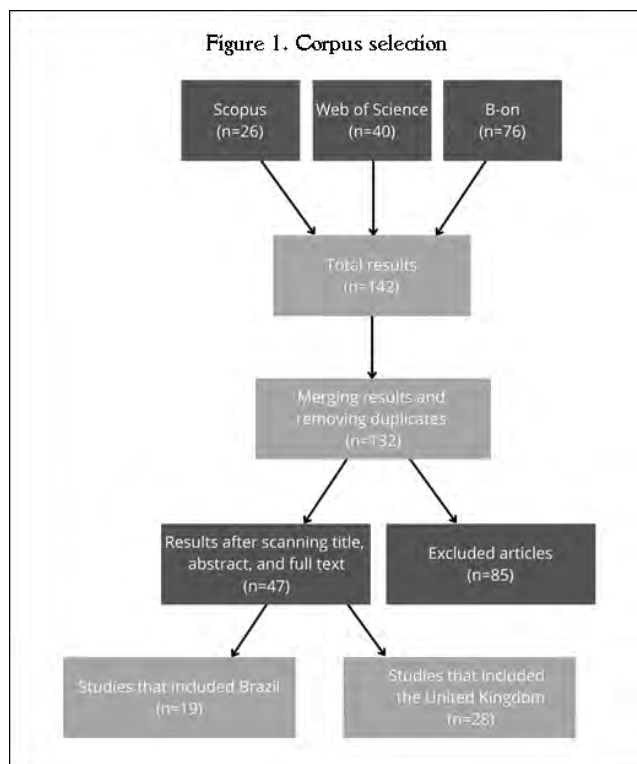
## 2. Materials and methods

Our study is based on a systematic review of 47 peer-reviewed articles published in indexed journals over the past two decades (2001-2020) in either English or Portuguese. The articles focus on references to immigration and/or refugees in both legacy and digital media. We followed PRISMA guidelines when collecting, mapping, and systematically reviewing the literature (Moher et al., 2015). The articles were collected from the Web of Science Core Collection, Scopus, and B-on databases.

## 2.1. Search and selection procedure

We systematically searched for articles from the past two decades that focus on the representation of immigrants and refugees in Brazil and the UK. First, we conducted a search of all databases using a Boolean search string that combined the terms: "content analysis"; "discourse analysis"; "media analysis"; "media"; "coverage"; "news"; "migrat\*"; "immigrat\*"; "migrant\*"; "refugee"; "Brazil", and "the United Kingdom"<sup>1</sup>. We performed the search in English for research related to the UK and in Portuguese and English for research related to Brazil. In the case of Brazil, we also included the B-on database, since only a very small number of Brazilian communication articles was found in the Web of Science Core Collection and Scopus databases (five and three articles, respectively). We therefore decided to add the Portuguese database B-on to our criteria, since this database searches Scopus, Web of Science, and other important academic indexers in Latin America, such as Latindex, Sumários.org, and DOAJ (Costa, 2015; Vasconcelos, 2004). We opted for including B-on to find results beyond the major western databases, hence avoiding the exclusion of important academic literature in Brazil and Latin America.

We used the following inclusion criteria (CR) when searching for papers: (CR1) peer-reviewed articles related to the UK and Brazil; (CR2) peer-reviewed articles related to discourses about immigration in mass media, political campaigns, social media, or cultural productions; (CR3) peer-reviewed articles written in English or Portuguese. We found 40 articles in Web of Science (35 on the UK and five on Brazil), 26 in Scopus (three on Brazil) and 76 on the B-on platform (all related to Brazil). From the total of 142 articles, 10 were excluded because they were duplicates. Of the remaining 132 articles, 85 more were excluded because they did not adhere to the abovementioned criteria. The final corpus of analysis thus included 47 peer-reviewed articles published in scientific journals, 19 related to Brazil, and 28 to the UK (Figure 1).



## 2.2. Coding procedure

When analyzing the literature, we identified the object of analysis; research question(s); method(s); and findings. Additionally, we used the SPSS software for quantitative analysis. We added variables, including methods (quantitative, qualitative, mixed); types of study (comparative or single-country), and



tone of representation in media analyzed in the studies (negative, neutral, and positive). We quantified the methods that were used by coding them as: 1 (quantitative); 2 (qualitative); or 3 (mixed). The types of study were coded as: 1 (comparative); or 2 (single-country). When analyzing the tone of representation, we coded this as: -1 (negative), 0 (neutral), and 1 (positive). Articles that did not specifically mention the tone of representation were excluded from this analysis.

### 3. Analysis and findings

#### 3.1. Methods and objects of analysis

We found some important cross-national differences when examining the use of mixed, quantitative, or qualitative methods in the two countries. Table 1 shows that quantitative methods were most prevalent in the UK (67.9%), followed by qualitative (17.9%), and mixed approaches (14.3%). Studies that included Brazil were mostly qualitative (94.7%), with only one quantitative example. The Fisher's exact test demonstrates significant differences ( $p < .001$ ) between Brazil and the UK in terms of these studies' use of mixed, quantitative, or qualitative methods.

<b>Dependent variable</b>	<b>Brazil</b>	<b>United Kingdom</b>	<b>Total</b>
Quantitative	1 (5.3%)	19 (67.9%)	20 (42.6%)
Qualitative	18 (94.7%)	5 (17.9%)	23 (48.9%)
Mixed	0 (0%)	4 (14.3%)	4 (8.5%)
Total	19 (100%)	28 (100%)	47 (100 %)

Note.  $\chi^2(2, N=47)=26.807$ ;  $p < 0.001$ .

The most frequently used method in UK studies was quantitative content analysis, which was applied in 39% of the reviewed articles (Amores et al., 2019; Aydemir & Vliegenthart, 2018; Hughes, 2019; Kaleda, 2014; Koopmans, 2004; Langer & Gruber, 2021; Masini, 2019; Masini et al., 2018; Shan-Jan, 2019; Walter, 2019). Other studies used a combination of methods to answer their research questions, such as semi-structured interviews, focus groups, social network analysis, and panel surveys (Balabanova, 2019; Froio & Ganesh, 2019; Hamlin, 2016; Scalvini, 2016; Wirz et al., 2018). Discourse analyses were carried out in 15% of the articles (Bates, 2017; Gibson & Booth, 2017; Lams, 2018; Pruitt, 2019). A minority of papers used other methods, such as an experimental approach to media effects, topic modeling, sentiment analysis, thematic analysis, case studies, and public opinion polls (Evans & Mellon, 2019; Harper & Hogue, 2019; Heidenreich et al., 2019; Risam, 2018; Ryan & Reicher, 2019; Schumann et al., 2020; Vautier, 2009).

The vast majority (94%) of the papers related to research from Brazil used qualitative methods (Aragão & Santil, 2018; Moreira et al., 2019; Silva, 2020; Zanforlin & Cogo, 2019); by contrast, quantitative methods were prevalent in the UK (accounting for 67% of all articles reviewed). Most articles that included Brazil (31%) combined different methods, such as interviews and questionnaires (Brignol & Costa, 2018b; Javorski & Brignol, 2017; Theodoro & Cogo, 2020). The second most frequently used method (15%) was discourse analysis (Mizga & Trovão, 2018; Moreira et al., 2019; Sacramento & Machado, 2015). Other methods included a case study approach, narrative analysis, experience reports, framing analyses, reception studies, and exploratory research. Most of the UK studies analyzed newspapers (75%). Some (19%) analyzed newspapers alongside other data sources, broadcast transcripts, social media, official immigration rates, longitudinal monitoring surveys, and in-depth interviews (Evans & Mellon, 2019; Hamlin, 2016; Langer & Gruber, 2021; Ryan & Reicher, 2019; Wirz et al., 2018). A few used other media sources, such as Twitter, photos, political speeches, or questionnaires (Amores et al., 2019; Froio & Ganesh, 2019; Schumann et al., 2020; Vautier, 2009). Similarly, most research on Brazil (35.3%) analyzed newspapers (Moreira et al., 2019; Ranincheski & Uebel, 2018; Sacramento & Machado, 2015). However, other media outlets were also investigated, and a significant portion of the studies (35.3%) examined a variety of media (Cogo, 2018; Silva, 2020; Zanforlin & Cogo, 2019). Some articles (11.8%) studied television broadcast material (Javorski & Brignol, 2017; Mizga & Trovão, 2018); others (11.8%) analyzed social media (Brignol & Costa, 2018b; Morais & Santos, 2017); and, lastly, a single study focused on literary books related to migration and refugees in Brazil (Tonus, 2018).

### 3.2. Comparative studies

Table 2 provides an overview of which types of study were most frequently seen in each country. In total, 57.1% of studies that included the UK were comparative, while in Brazil, there was only a single comparative study. In Brazil, most studies (94.7%) focused only on national media, in contrast to the sample from the UK (42.9%). The chi-square test yielded a significant difference ( $p < .001$ ) between Brazil and the UK when considering comparative versus non-comparative studies (Table 1).

Table 2. Comparative versus non-comparative studies			
Dependent variable	Brazil	United Kingdom	Total
Comparative	1 (5.3%)	16 (57.1%)	17 (36.2%)
Non-comparative	18 (94.7%)	12 (42.9%)	30 (63.8%)
Total	19 (100%)	28 (100%)	47 (100%)

Note.  $\chi^2(1, N=47)=13.196$ ;  $p < 0.001$ .

Of the 16 comparative studies including the UK, 12 (75%) focused on European countries (Aydemir & Vliegenthart, 2018; Mancini et al., 2019; Masini, 2019), and of those 12, only three (25%) compared countries from eastern and western Europe (Balabanova, 2019; Heidenreich et al., 2019; Lams, 2018). Four studies (25%) compared the UK with non-European countries, such as Canada, Hong Kong, Kenya, the United States, and Taiwan (Kaleda, 2014; Lawlor, 2015; Risam, 2018; Shan-Jan, 2019). The analysis shows that, even when a study made an attempt to adopt a more cross-national perspective, this primarily focused on comparisons between countries in the global north.

One study that did conduct a comparative, cross-national analysis of media framing of immigration in the UK, United States, Hong Kong, and Taiwan (Shan-Jan, 2019) showed higher levels of negative tone toward immigrants in the UK and United States than in Hong Kong and Taiwan. According to this study's findings, most media frames in the United States and the UK portray immigrants as threats to the economy or national security, while the media in Hong Kong and Taiwan adopt a cultural integration frame which presents the assimilation immigrants into their host culture. In addition, studies which compared the UK and Canada showed that negative representation and an emphasis on crime and security frames prevailed in both (Lawlor, 2015); however, Canada adopted a less negative tone by using an economic frame on immigrants' contributions to economic growth.

Comparative studies that included the UK and the United States found that reporting on refugees were characterized by suspicion, lack of sympathy, and uniformization (Risam, 2018; Kaleda, 2014). Further studies have found that most of the actors quoted directly or indirectly in legacy media reports on refugees were state authorities or actors in favor of or against immigrants, but not the immigrants themselves (Masini, 2019; Masini et al., 2018). Alternative media, on the other hand, usually gave immigrants a voice by quoting refugees or immigrants directly (Kaleda, 2014).

In contrast to the prevalence of comparative studies in the UK, only one article in the Brazilian corpus compared different countries, both in Latin America (Silva, 2020). Most research focused on the national level and analyzed important Brazilian newspapers, such as *O Globo* and *Folha de S. Paulo* (Morais & Santos, 2017; Moreira et al., 2019; Ranincheski & Uebel, 2018; Sacramento & Machado, 2015). Findings related to national media demonstrate a process of silencing immigrants and refugees, who are rarely quoted or referenced as sources of information and are not presented as the protagonists in media stories (Moreira et al., 2019). In addition, articles also diminish immigrants by portraying aspects of their home countries as negative, contributing to the process of othering and the separation between "us" and "them" (Ranincheski & Uebel, 2018). When newspapers show the terrible conditions and difficulties faced by immigrants, these immigrants are also presented as a threat to public health (Sacramento & Machado, 2015) or as victims (Morais & Santos, 2017). When comparing national and regional television news, it is clear that the theme of immigration has little visibility in Brazil and that international reporting trends are reproduced on the national level (Javorski & Brignol, 2017). Results from an analysis focusing on southern Brazil also show that the negative tone is observed locally, with news items focusing on regional problems and immigrants' difficulties, again using a victim frame (Almeida & Brandão, 2015).

### 3.3. Patterns in the representation of immigration

Table 3 shows that the tone of representation in media analyzed in the studies that included Brazil and the UK is mainly negative (50%), followed by neutral (32.5%), and then positive (17.5%). When comparing the research on both countries, no major differences were found. In both Brazil and the UK, the aggregate media representation from the data analyzed in the literature we reviewed was primarily negative (50%), followed by neutral (33.3% and 31.3% for Brazil and the UK, respectively), and then positive (16.7% and 18.8% for Brazil and the UK, respectively).

Table 3. Tone of representation			
Dependent variable	Brazil	United Kingdom	Total
Positive	3 (18.8%)	4 (16.7%)	7 (17.5%)
Neutral	5 (31.3%)	8 (33.3%)	13 (32.5%)
Negative	8 (50%)	12 (50%)	20 (50%)

Note. Due to the sample size, we could not perform the Chi-Square test. Results show that no significant differences were found.

A majority (78.5%) of studies including the UK focused on the general category of immigration and refugees (Amores et al., 2019; Lawlor, 2015; Mancini et al., 2019; Masini, 2019), while in Brazil, most studies (79%) focused on specific groups of immigrants, such as Senegalese, Haitians, or Venezuelans (Almeida & Brandão, 2015; Brignol & Costa, 2018a; Morais & Santos, 2017; Sacramento & Machado, 2015). The results reveal that representation differs depending on the nationality of the immigrants in question, with European immigrants receiving more positive representation in Brazil (Cogo, 2018; Dadalto, 2013) and immigrants to the UK who originate from Muslim countries being the worst represented (Froio & Ganesh, 2019; Scalvini, 2016). Studies in both countries lack intersectional perspectives, uniformizing the figure of the immigrant or refugee. For example, only one study about immigrants in Brazil references invisibility dynamics in the LGBTQI+ diaspora (Theodoro & Cogo, 2020).

As found in one study, conducted by Benson (2013), this systematic literature review revealed standardization and repetition of the threat, victims, and hero frames. The hero frame is usually related to neoliberal values, presenting stories that describe immigrant success stories, helping to shape the idea that outsiders have to fight for their achievements (Zanforlin & Cogo, 2019). Media also represent refugees and immigrants as victims, for instance, in articles that show overcrowded spaces, describe conflicts with authorities, or mention a lack of resources, creating the idea of a uniform mass of victims (Lima et al., 2016; Mizga & Trovão, 2018). Finally, another frequent issue is presenting these groups as a threat by associating them with words such as "invasion," stating that there is a necessity to control them, and mentioning state-sanctioned benefits for migrants (Cogo & Silva, 2015; Lima-Pimentel & Castro-Cotinguiba, 2014; Ranincheski & Uebel, 2018). The results also showed that the terms "refugees" and "immigrants" are generally used interchangeably (Brignol & Costa, 2018a; Pereira-Lima & Faria-Campos, 2015).

Of studies including the UK, only a minority (16.7%) of results showed a positive tone (Balabanova, 2019; Bates, 2017; Langer & Gruber, 2021; Ryan & Reicher, 2019). In some cases, positive representation was the result of special circumstances, for instance, the United Nations' Refugee Agency campaign that sought to increase positive sentiment related to refugees (Balabanova, 2019). Other examples of positive representation were in response to scandals related to hostile behavior against immigrants or newspapers portraying immigrants positively in opposition to human rights violations against them (Bates, 2017; Langer & Gruber, 2021; Ryan & Reicher, 2019). In rare cases (4.3%), studies in both countries referred to positive discourses that were linked to hero frames (Ryan & Reicher, 2019; Zanforlin & Cogo, 2019). Nevertheless, hero frames portray immigrants positively when immigrants have been accepted by and adapted to fit the culture of a more welcoming country (Shan-Jan, 2019). Positive representation is also found in alternative media, as some articles related to the Brazilian context show (Aragão & Santil, 2018; Brignol & Costa, 2018b; Cogo, 2018). In these media outlets, some prevalent themes are citizenship, political and social participation, integration, culture valorization, exposing racism, and helpful information for those groups.

In political discourses, liberal immigration programs are presented as a "fair" option founded on meritocratic methods, provided these programs dictate which immigrant groups should be accepted

(Gibson & Booth, 2017). Research shows connections between the growth of the United Kingdom Independence Party (UKIP), anti-migration and anti-European Union (EU) sentiment, the EU referendum in 2016, and Brexit (Evans & Mellon, 2019). UKIP regularly used populist language to appeal to "ordinary" people, diminishing elites and evoking the differences between British people and others (Hughes, 2019) in a bid to boost its success. Politically, Brazil faced a similar situation. Until the election of President Jair Bolsonaro, Brazilian politicians tended to use pro-immigrant and refugee discourses (Cogo, 2018; Dadalto, 2013); Bolsonaro, on the other hand, portrays immigrants as the enemy (Silva, 2020). Like UKIP, Bolsonaro gained visibility by adopting a populist discourse that separated "us" from "them," criticized the system and the establishment, condemned corruption, and adopted a nationalistic perspective.

#### 4. Discussion

This systematic literature review of peer-reviewed studies examining media representations of immigrants and refugees in Brazil and the UK revealed similarities and differences in how research was conducted in these two countries. Studies focusing on Brazil were predominantly local, while in the UK, they usually adopted a comparative approach that incorporated other countries. However, these comparative studies were mainly conducted in European countries and the global north, focusing on WEIRD countries. The results also show differences in the methods chosen for studies based on the country it took place in. In Brazil, a qualitative approach was more popular, while quantitative methods were more prevalent in the UK. One possible explanation for the difference in which methods were used could be the different research traditions and inequity in terms of access to research infrastructure. These differences also demonstrate how varied the research traditions are in each country. Brazil, for instance, is influenced by Anglophone, Latin-American and French communication research and focuses on critical approaches to media and communication. In addition, WEIRD countries in the global north have more significant resources to devote to research, while universities in Latin America, including Brazilian communication faculties, often lack access to news databases and statistical software (Skoric, 2014; Zelenkauskaitė & Bucy, 2016).

Different studies in this review also focused on different media platforms. Studies focusing on Brazil looked at a diverse range of media, including newspapers, television, social media, or a combination of two or more. In contrast, research focusing on the UK concentrated mainly on newspapers; the prevalence of newspaper articles in these studies is most likely due to the fact that they are easily accessible via digital archives, while collecting data from broadcast media is more difficult. The results also demonstrate that research in both Brazil and the UK tends to focus primarily on legacy media, with alternative communication platforms rarely being studied. The emergence of new voices and their possible participation in social media has gone almost unnoticed. This is most likely a result of the significant influence that legacy media has at the agenda across society (Langer & Gruber, 2021).

Although both countries have different media and political systems and significant cultural differences, the results show that discourses on immigration are predominantly negative. In line with previous systematic reviews showing the dominance of political elite sources in legacy media (Consterdine, 2018; Eberl et al., 2018), our findings demonstrate that immigrant groups are usually underrepresented and they are rarely included as sources in these discourses. Instead, other groups communicate for them, such as national authorities and pro- or anti-immigration groups or representatives. The studies revealed that immigrants and refugees are better represented in alternative media, which offers them a platform to express their opinions with less stereotypical framing in comparison to mainstream media (Aragão & Santil, 2018; Cogo, 2018; Kaleda, 2014).

Refugees are framed as vulnerable, suffering, helpless victims, and as a mass of people (Kaleda, 2014; Morais & Santos, 2017). In some cases, concepts are used interchangeably, with "immigrants" and "refugees" being used synonymously, or the categories overlap, generating and reinforcing stereotypes (Mizga & Trovão, 2018). Both immigrants and refugees are represented as threats to culture, the welfare state, and public health (Hamlin, 2016; Pruitt, 2019; Sacramento & Machado, 2015). Metaphors describing refugees and immigrants as an "invasion" that is "flooding" a country promote in the public imagination a sense that these groups are dangerous or threatening, increasing fear toward them. These



findings are in line with previous studies, in which immigrants and refugees are often framed as cultural, economic, or criminal threats (Consterdine, 2018; Eberl et al., 2018). The presence in these studies of the threat, victim, and hero frames mirrors findings by Benson (2013), which suggest a standardization of frames in both countries.

Furthermore, this type of framing also contributes to the process of otherization, in which a society is divided into "us" and "them." These stereotypical frames are particularly prevalent in populist discourses, designed to create groups of others, or strangers, who are seen as dangerous enemies of the nation. Negative stereotyping is also articulated with hate speech (Paz et al., 2020), and a past study showed that forms of hate speech are more common today in traditional media (Winiewski et al., 2017), which often reproduces insults directed to minority groups through others' declarations, such as the ones from populist leaders. However, even before the election of populist leaders, there existed frames dividing ingroups from outgroups and an association between immigration and invasion. This makes it clear that frames prevail across many years and that, even if they can be dynamic, they create common cognitions and produce societal bonds that are collective and resistant to change. Previous results demonstrate that people more exposed to hate speech are more likely to use hate speech and to support repressive measures directed to immigrants and refugees, such as closing frontiers. Moreover, the increase of hate speech is also associated with the desensitization of people towards this type of discourse (Winiewski et al., 2017).

Although negative representations are prevalent, the results also revealed some positive discourses in both Brazil and the UK. Positive representations were usually associated with exceptional and unique events, such as illegal or controversial deportations, issues of discrimination, or amnesty for immigrants; or with memorable dates, such as World Refugee Day (Balabanova, 2019; Bates, 2017; Langer & Gruber, 2021; Ryan & Reicher, 2019; Zanforlin & Cogo, 2019). Another frequent positive representation is the hero frame, which encapsulates a neoliberal notion of immigration in which refugees or immigrants are responsible for their own success, with the state bearing no responsibility for it.

## 5. Conclusions

The aim of this systematic literature review was to analyze research concerning media representations of immigrants and refugees in Brazil and the UK, given the absence of comparative studies related to immigration and media in these two countries. This review first sought to provide answers to the following questions: which research methods were used in these studies, and how the representation varied between Brazil and the UK in the findings of the scholarly literature. This article evaluates patterns that exist in studies including both countries and highlights the primary differences and similarities between the two by systematically reviewing the existing scholarly literature on media representations of immigration in both countries. This is the first comparative examination of its type, and it aims to pave the way for further empirical comparative studies that incorporate the global south and global north by comparing WEIRD and non-WEIRD countries.

Our study also revealed differences in terms of which methods were used in research focusing on Brazil versus the UK. For studies related to the UK, we would suggest using additional methods, including mixed-method approaches and comparative studies that incorporate the global south. Research focusing on Brazil may benefit from more quantitative approaches. In all cases, our results revealed a focus on legacy media and newspapers, to the exclusion of alternative communication platforms and other media outlets. Thus, future research should also analyze alternative media, compare legacy and alternative media, and focus on other media outlets. Alternative communication platforms may also offer an invaluable perspective into immigration coverage for those who might wish to craft an ideal ethical model for covering the topic.

We demonstrated that comparative studies, including those from the UK, tend to concentrate on WEIRD countries, while there is a lack of comparative studies including Brazil and other countries in the global south. We would like to emphasize that comparative studies incorporating both the global south and north are essential for a better understanding of the similarities and differences at play in relation to media and immigration. The limitations of this research lie on the fact that it concentrates on only two countries. Further research focusing on additional nations from the global north and global south can offer more insight into other similarities and differences. We would also recommend further research

comparing different countries across Asia, Africa and Latin America with countries from the global north. Strengthening and internationalizing research that is produced in the global south is valuable for a broader, decolonizing perspective, since scientific production is often westernized.

Finally, a systematic literature review has by its very nature several limitations; a keyword search, for example, can sometimes exclude essential research results. Further systematic literature reviews and meta-analyses using additional research criteria may be able to point to other important avenues of research. We searched for terms such as migration, immigration, and refugees. Additional studies can include keywords such as alien, illegal aliens, asylum seekers, international students, or expatriate.

This study contributes to our understanding of the representation of immigration in Brazil and the UK media analyzed in the studies. This paper reveals the similarities and differences between studies that examine media and migration in both the UK and Brazil. It highlights a standardization of framing, the prevalence of a negative tone in media analyzed in the studies, the silencing of migrant and refugee voices, and differences in applied research methods. Our research and its results cover countries from both the global north and the global south; it is evident that there is a repetition of frames related to immigration issues in both. Legacy media usually represents the topic negatively, focusing on aspects such as threats to the economy, security, or safety. Hence, the findings show that mass media generally contributes to negative stereotyping, which is shown to contribute to the increase of hate speech and violence towards immigrants and refugees. To improve the representation, mass media should emphasize immigrants' and refugees' contributions to increase positive associations and adopt a balanced view that highlights different angles.

### Authors' Contribution

Idea, I.G.; Literature review (state of the art), I.G.; Methodology, I.G.; Data analysis, I.G.; Results, I.G.; Discussion and conclusions, I.G.; Writing (original draft), I.G.; Y.D.; Final revisions, Y.D.; Project design and funding agency, D.A.A.D.

### Notes

<sup>1</sup>We used the following search string to identify relevant publications related to media representations of immigrants or refugees between January 2000 and November 2020: ((AB=(content analysis OR discourse analysis OR media analysis) AND AB=(media OR coverage OR news) AND AB=(media OR coverage OR news) AND AB=(migrant\* OR immigrant\* OR refugee\* OR immigrant\* OR migrant\*)) AND AB=(Brazil OR Brasil OR the United Kingdom OR the UK)).

### Funding Agency

This study was supported by the Deutscher Akademischer Austauschdienst (DAAD) Doctoral Research Grant number 57507871.

### References

- Albuquerque, A. (2012). On models and margins: Comparative media models viewed from a Brazilian perspective. In D. C. Hallin, & P. Mancini (Eds.), *Comparing media systems beyond the western world* (pp. 72-95). <https://doi.org/10.1017/CBO9781139005098.006>
- Almeida, C., & Brandão, B. (2015). Imigração, mídia e sociabilidade dos Haitianos [Immigration, media and the Haitian sociability]. *Revista Observatório, 1*, 62-79. <https://doi.org/10.20873/ufv.2447-4266.2015v1n3p62>
- Amores, J.J., Calderón, C.A., & Stanek, M. (2019). Visual frames of migrants and refugees in the main western European media. *Economics and Sociology, 12*(3), 147-161. <https://doi.org/10.14254/2071-789X.2019/12-3/10>
- Aragão, T., & Santil, V. (2018). Somos migrantes: O uso das redes sociais na produção midiática alternativa sobre a migração venezuelana em Roraima [We are Migrants: The use of social networks in alternative media production on Venezuelan immigration in Roraima]. *Aturá Revista Pan-Amazônica de Comunicação, 2*(1), 136-156. <https://bit.ly/3s2eYKx>
- Araújo, B., & Prior, H. (2020). Framing political populism: The role of media in framing the election of Jair Bolsonaro. *Journalism Practice, 15*(2), 226-242. <https://doi.org/10.1080/17512786.2019.1709881>
- Aydemir, N., & Vliegthart, R. (2018). Public discourse on minorities: How discursive opportunities shape representative patterns in the Netherlands and the UK. *Nationalities Papers, 46*(2), 237-251. <https://doi.org/10.1080/00905992.2017.1342077>
- Balabanova, E. (2019). Communicating cosmopolitanism during times of crisis: UNHCR and the World Refugee Day campaign in the UK and Bulgaria. *Journal of Human Rights Practice, 11*(3), 467-485. <https://doi.org/10.1093/jhuman/huz029>
- Bates, D. (2017). The "red door" controversy: Middlesbrough's asylum seekers and the discursive politics of racism. *Journal of Community and Applied Social Psychology, 27*(2), 126-136. <https://doi.org/10.1002/casp.2300>
- Benson, R. (2013). *Shaping immigration news: A French-American comparison*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139034326>

- Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The Dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, 41(1), 3-33. <https://doi.org/10.1111/pops.12670>
- Brignol, L.D., & Costa, N.D. (2018a). A saga e o sofrimento do outro senegalês: A construção do racismo em representações midiáticas da migração [Saga and suffering of the Senegalese other: building racism into media representations of migration]. *Chasqui*, 138, 135-151. <https://bit.ly/2Y7QC51>
- Brignol, L.D., & Costa, N.D. (2018b). Diáspora senegalesa e mediação tecnológica: Entre tempos e lugares na observação do Magal de Touba. *Contracampo*, 37(1), 9-29. <https://doi.org/10.22409/contracampo.v0i0.1079>
- Chakravarty, P., Kuo, R., Grubbs, V., & McIlwain, C. (2018). #CommunicationSoWhite. *Journal of Communication*, (2), 254-266.
- Cogo, D. (2018). O Haiti é Aqui: Mídia, imigração haitiana e racismo no Brasil [Haiti is here: Media, Haitian immigration and racism in Brazil]. *Chasqui*, 139, 427-448. <https://bit.ly/39Rs82F>
- Cogo, D., & Silva, T. (2015). Entre a “fuga” e a “invasão”: lteridade e cidadania da imigração haitiana na mídia brasileira [Between escape and invasion: otherness and citizenship of Haitian immigration in the Brazilian media]. *Revista Famecos*, 23(1), 21885. <https://doi.org/10.15448/1980-3729.2016.1.21885>
- Consterdine, E. (2018). *State-of-the-art report on public attitudes, political discourses and media coverage on the arrival of refugees*. CEASEVAL Research on the Common European Asylum System. <https://bit.ly/3oRAgIV>
- Costa, T. (2015). A b-on e a preservação digital de periódicos eletrônicos: O projeto Portico. In *Ligar. Transformar. Criar Valor. 12º Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*. Évora. <https://bit.ly/3AVVsOuT>
- Dadalto, M.C. (2013). Imigração e permanência do sonho [Immigration and staying in the dream]. *Matrizes*, 7(2), 249-263. <https://doi.org/10.11606/issn.1982-8160.v7i2p249-263>
- Eberl, J.M., Meltzer, C.E., Heidenreich, T., Herrero, B., Theorin, N., Lind, F., Berganza, R., Boomgaarden, H.G., Schemer, C., & Strömbäck, J. (2018). The European media discourse on immigration and its effects: A literature review. *Annals of the International Communication Association*, 42(3), 207-223. <https://doi.org/10.1080/23808985.2018.1497452>
- Evans, G., & Mellon, J. (2019). Immigration, Euroscepticism, and the rise and fall of UKIP. *Party Politics*, 25(1), 76-87. <https://doi.org/10.1177/1354068818816969>
- Fengler, S., Bastian, M., Brinkmann, J., Zappe, A.C., Tatah, V., Andindilile, M., Assefa, E., Chibita, M., Mbaine, A., Obonyo, L., Quashigah, T., Skleparis, D., Splendore, S., Tadesse, M., & Lengauer, M. (2020). Covering migration -in Africa and Europe: Results from a comparative analysis of 11 countries. *Journalism Practice*, (pp. 1-21). <https://doi.org/10.1080/17512786.2020.1792333>
- Froio, C., & Ganesh, B. (2019). The transnationalisation of far-right discourse on Twitter: Issues and actors that cross borders in Western European democracies. *European Societies*, 21, 513-539. <https://doi.org/10.1080/14616696.2018.1494295>
- Gibson, S., & Booth, R. (2018). An Australian-style points system”: Individualizing immigration in radical right discourse in the 2015 U.K. General election campaign. *Peace and Conflict: Journal of Peace Psychology*, 24(4), 389-397. <https://doi.org/10.1037/pac0000267>
- Hallin, D.C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511790867>
- Hamlin, R. (2016). Foreign criminals”, the human rights act, and the new constitutional politics of the United Kingdom. *Journal of Law and Courts*, 4(2), 437-461. <https://doi.org/10.1086/687409>
- Harper, C.A., & Hogue, T.E. (2019). The role of intuitive moral foundations in Britain’s vote on EU membership. *Journal of Community and Applied Social Psychology*, 29(2), 90-103. <https://doi.org/10.1002/casp.2386>
- Heidenreich, T., Lind, F., Eberl, J.M., & Boomgaarden, H.G. (2019). Media framing dynamics of the “European refugee crisis”: A comparative topic modelling approach. *Journal of Refugee Studies*, 32(1), 172-182. <https://doi.org/10.1093/jrs/fez025>
- Hughes, C. (2019). It’s the EU immigrants stupid! UKIP’s core-issue and populist rhetoric on the road to Brexit. *European Journal of Communication*, 34(3), 248-266. <https://doi.org/10.1177/0267323119830050>
- Javorski, E., & Brignol, L. (2017). Journalists and their role in selecting and building news on contemporary transnational migrations. *Brazilian Journalism Research*, 13(2), 28-45. <https://doi.org/10.25200/bjr.v13n2.2017.985>
- Kaleda, C. (2014). Media perceptions: Mainstream and grassroots media coverage of refugees in Kenya and the effects of global refugee policy. *Refugee Survey Quarterly*, 33(1), 94-111. <https://doi.org/10.1093/rsq/hdt025>
- Koopmans, R. (2004). Migrant mobilisation and political opportunities: Variation among German cities and a comparison with the United Kingdom and the Netherlands. *Journal of Ethnic and Migration Studies*, 30(3), 449-470. <https://doi.org/10.1080/13691830410001682034>
- Kroon, A.C., Trilling, D., Van-Der-Meer, T.G.L.A., & Jonkman, J.G.F. (2020). Clouded reality: News representations of culturally close and distant ethnic outgroups. *Communications*, 45(s1), 744-764. <https://doi.org/10.1515/commun-2019-2069>
- Lams, L. (2018). Discursive constructions of the summer 2015 refugee crisis: A comparative analysis of French, Dutch, Belgian francophone and British centre-of-right press narratives. *Journal of Applied Journalism and Media Studies*, 7(1), 103-127. [https://doi.org/10.1386/ajms.7.1.103\\_1](https://doi.org/10.1386/ajms.7.1.103_1)
- Langer, A.I., & Gruber, J.B. (2021). Political agenda setting in the hybrid media system: Why legacy media still matter a great deal. *The International Journal of Press/Politics*, 26(2), 313-340. <https://doi.org/10.1177/1940161220925023>
- Lawlor, A. (2015). Local and national accounts of immigration framing in a cross-national perspective. *Journal of Ethnic and Migration Studies*, 41(6), 918-941. <https://doi.org/10.1080/1369183X.2014.1001625>
- Levy, D., Aslan, B., & Bironzo, D. (2016). *UK press coverage of the EU referendum*. Reuters Institute for the Study of Journalism. <https://bit.ly/39P5Rm6>
- Lima, M., Ávila, O., & Silva, A. (2016). Aspectos da representação dos imigrantes haitianos no jornal Gazeta do Povo, de Curitiba/PR. *Intexto*, 35, 51-75. <https://doi.org/10.19132/1807-8583201635.51-75>

- Lima-Pimentel, M., & Castro-Cotinguiba, G. (2014). Wout, raketè, fwontyè, anpil mizè: Reflexões sobre os limites da alteridade em relação à imigração haitiana para o Brasil [Wout, raketè, fwontyè, anpil mizè: Reflections on the limits of otherness in relation to Haitian immigration to Brazil]. *Universitas: Relações Internacionais*, 12(1), 73-86. <https://doi.org/10.5102/uri.v12i1.2861>
- Lugo-Ocando, J. (2020). The 'changing' face of media discourses on poverty in the age of populism and anti-globalisation: The political appeal of anti-modernity and certainty in Brazil. *International Communication Gazette*, 82, 101-116. <https://doi.org/10.1177/1748048519880749>
- Mancini, P., Mazzoni, M., Barbieri, G., Damiani, M., & Gerli, M. (2021). What shapes the coverage of immigration. *Journalism*, 22(4), 845-866. <https://doi.org/10.1177/1464884919852183>
- Masini, A. (2019). Consonance and diversity of voices and viewpoints: A new paradigm to study actors' cumulative influence on viewpoints in immigration news. *International Journal of Communication*, 13, 674-692. <https://bit.ly/2YaUbyr>
- Masini, A., Van-Aelst, P., Zerback, T., Reinemann, C., Mancini, P., Mazzoni, M., Damiani, M., & Coen, S. (2018). Measuring and explaining the diversity of voices and viewpoints in the news: A comparative study on the determinants of content diversity of immigration news. *Journalism Studies*, 19(15), 2324-2343. <https://doi.org/10.1080/1461670X.2017.1343650>
- Matthes, J., & Schmuck, D. (2017). The effects of anti-immigrant right-wing populist ads on implicit and explicit attitudes: A moderated mediation model. *Communication Research*, 44(4), 556-581. <https://doi.org/10.1177/0093650215577859>
- Mccombs, M., & Valenzuela, S. (2007). The agenda-setting theory. *Cuadernos.Info*, 1(20), 44-50. <https://doi.org/10.7764/cdi.20.111>
- Mccombs, M.E., & Shaw, D.L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176-187. <https://doi.org/10.1086/267990>
- Mizga, L.S., & Trovão, S.S. (2018). Conflitos contemporâneos: A categoria refugiado no telejornalismo brasileiro [Contemporary conflict: The refugee category in Brazilian telejournalism]. *Revista Observatório*, 4(5), 512-535. <https://doi.org/10.20873/ufv.2447-4266.2018v4n5p512>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., & Group, P.P. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1-9. <https://doi.org/10.1186/2046-4053-4-1>
- Morais, V.M., & Santos, L.C. (2017). Mediações e cibercultura: Estudo de comentários em matéria jornalística sobre a imigração venezuelana em Roraima [Mediation and Cyberculture: Study of journalistic comments on Venezuelan immigration in Roraima]. *Aturá*, 1(2), 120-144. <https://bit.ly/3inBqli>
- Moreira, S., Sousa, M., & Fortuna, V. (2019). Imigrantes sem voz: A produção de silêncio na cobertura jornalística das manifestações anti-imigração [Voiceless immigrants: The lack of press coverage on anti-immigration demonstrations]. *Contracampo*, 38(2), 23-35. <https://doi.org/10.22409/contracampo.v38i2.28514>
- Moreira, S.V. (2005). Trends and new challenges in journalism research in Brazil. *Brazilian Journalism Research*, 1(2), 9-24. <https://doi.org/10.25200/bjr.v1n2.2005.51>
- Novoselova, O.V. (2020). Peculiarities of modern nationalist messages in online political communication: The analyses of Donald Trump's and Jair Bolsonaro's election campaigns. *Changing Societies and Personalities*, 4, 350-371. <https://doi.org/10.15826/csp.2020.4.3.105>
- Paz, M.A., Montero-Díaz, J., & Moreno-Delgado, A. (2020). Hate speech: A systematized review. *SAGE Open*, 10(4), 1-12. <https://doi.org/10.1177/2158244020973022>
- Pereira-Lima, R.C., & Faria-Campos, P.H. (2015). Campo e grupo: Aproximação conceitual entre Pierre Bourdieu e a teoria moscoviciana das representações sociais [Field and group: A conceptual approximation between Pierre Bourdieu and the social representation theory of Moscovici]. *Educação e Pesquisa*, 41(1), 63-77. <https://doi.org/10.1590/s1517-97022015011454>
- Pruitt, L.J. (2019). Closed due to 'flooding'? UK media representations of refugees and migrants in 2015-2016- creating a crisis of borders. *British Journal of Politics and International Relations*, 21(2), 383-402. <https://doi.org/10.1177/1369148119830592>
- Ranincheski, S., & Uebel, R.R.G. (2018). "Polícia Federal deporta 450 imigrantes ilegais venezuelanos de Roraima": Um estudo sobre remediation, positivismo e pós-positivismo no jornal O Globo. *Diálogo*, 37, 39-52. <https://doi.org/10.18316/dialogo.v0i37.3879>
- Risam, R. (2018). Now you see them: Self-representation and the refugee selfie. *Popular Communication*, 16(1), 58-71. <https://doi.org/10.1080/15405702.2017.1413191>
- Ryan, C., & Reicher, S. (2019). An analysis of the nature and use of promigrant representations in an antideportation campaign. *Political Psychology*, 40(3), 583-598. <https://doi.org/10.1111/pops.12526>
- Sacramento, I., & Machado, I.B. (2015). A imigração como risco para a saúde: Análise das representações do imigrante africano na cobertura da Folha de S. Paulo sobre o ebola. *Comunicação e Sociedade*, 28, 25-47. [https://doi.org/10.17231/comsoc.28\(2015\).2269](https://doi.org/10.17231/comsoc.28(2015).2269)
- Scalvini, M. (2016). A crisis of religious diversity: Debating integration in post-immigration Europe. *Discourse and Communication*, 10(6), 614-634. <https://doi.org/10.1177/1750481316674779>
- Schemer, C. (2012). The influence of news media on stereotypic attitudes toward immigrants in a political campaign. *Journal of Communication*, 62(5), 739-757. <https://doi.org/10.1111/j.1460-2466.2012.01672.x>
- Schumann, S., Osborne, D., Gill, P., Fell, B.F., & Hewstone, M. (2020). Radical news? Immigrants' television use, acculturation challenges, and support for terrorism. *Behavioral Sciences of Terrorism and Political Aggression*, 13(4), 1-17. <https://doi.org/10.1080/19434472.2020.1779782>
- Seate, A.A., & Mastro, D. (2017). Exposure to immigration in the news: The impact of group-level emotions on intergroup behavior. *Communication Research*, 44(6), 817-840. <https://doi.org/10.1177/0093650215570654>



- Shan-Jan, S.L. (2019). Framing immigration: A content analysis of newspapers in Hong Kong, Taiwan, the United Kingdom, and the United States. *Politics, Groups, and Identities*, 9(4), 1-25. <https://doi.org/10.1080/21565503.2019.1674162>
- Shehaj, A., Shin, A.J., & Inglehart, R. (2021). Immigration and right-wing populism: An origin story. *Party Politics*, 27(2), 282-293. <https://doi.org/10.1177/1354068819849888>
- Silva, A.L. (2020). Imigrações na América Latina em tempos de globalização e mídia [Immigration of populations in latin america in globalization times and the media]. *Tópicos Educacionais*, 26(1), 183-204. <https://doi.org/10.51359/2448-0215.2020.245667>
- Skoric, M.M. (2014). The implications of big data for developing and transitional economies: Extending the triple helix? *Scientometrics*, 99(1), 175-186. <https://doi.org/10.1007/s11192-013-1106-5>
- Theodoro, H., & Cogo, D. (2020). Sociocommunicational (in)visibilities in the/of the queer diaspora. *Palavra Clave*, 23(3), 1-29. <https://doi.org/10.5294/pacla.2020.23.3.2>
- Tonus, J.L. (2018). Migrantes e refugiados: À (a) espera de uma narrativa? *Letras de Hoje*, 53, 476-483. <https://doi.org/10.15448/1984-7726.2018.4.33009>
- Vasconcelos, D. (2004). Biblioteca do conhecimento online: um novo impulso para a sociedade da informação. *Cadernos BAD*, (1), 116-123. <https://doi.org/https://bit.ly/3ii2Csc>
- Vautier, E. (2009). Playing the "race card": White anxieties and the expression and repression of popular racisms in the 1997 UK election. *Patterns of Prejudice*, 43(2), 122-141. <https://doi.org/10.1080/00313220902793880>
- Walter, S. (2019). Better off without you? How the British media portrayed EU citizens in Brexit news. *International Journal of Press/Politics*, 24(2), 210-232. <https://doi.org/10.1177/1940161218821509>
- Winiewski, M., Hansen, K., Bilewicz, M., Soral, W., Widerska, A., & Bulska, D. (2017). *Contempt speech, hate speech. Report from research on verbal violence against minority groups*. Stefan Batory Foundation. <https://bit.ly/2ZX2RCS>
- Wirz, D.S., Wettstein, M., Schulz, A., Müller, P., Schemer, C., Ernst, N., Esser, F., & Wirth, W. (2018). The effects of right-wing populist communication on emotions and cognitions toward immigrants. *International Journal of Press/Politics*, 23(4), 496-516. <https://doi.org/10.1177/1940161218788956>
- Zanforlin, S., & Cogo, D. (2019). Mídia, mobilidade e cidadania no contexto do capitalismo global: Reflexões a partir da trajetória de um refugiado sírio [Media, mobility and citizenship in global capitalism: reflexions based on the trajectory of a Syrian refugee]. *Contemporânea*, 17, 7-28. <https://doi.org/10.9771/contemporanea.v17i1.25222>
- Zelenkauskaitė, A., & Bucy, E.P. (2016). A scholarly divide: Social media, big data, and unattainable scholarship. *First Monday*, 21(5). <https://doi.org/10.5210/fm.v21i5.6358>



# When negativity is the fuel. Bots and political polarization in the COVID-19 debate

Cuando la negatividad es el combustible. Bots y polarización política en el debate sobre el COVID-19

- Dr. José-Manuel Robles.** Professor, Department of Applied Sociology, Complutense University of Madrid (Spain) (jmrobles@ucm.es) (<https://orcid.org/0000-0003-1092-3864>)
- Juan-Antonio Guevara.** Postdoctoral Research Fellow, Department of Applied Sociology, Complutense University of Madrid (Spain) (juanguev@ucm.es) (<https://orcid.org/0000-0003-3946-3910>)
- Dr. Belén Casas-Mas.** Assistant Professor PhD, Sociology Department, Complutense University of Madrid (Spain) (bcasas@ucm.es) (<https://orcid.org/0000-0001-8329-0856>)
- Dr. Daniel Gómez.** Professor, Department of Statistics and Data Science, Complutense University of Madrid (Spain) (dagomez@estad.ucm.es) (<https://orcid.org/0000-0001-9548-5781>)

## ABSTRACT

The contexts of social and political polarization are generating new forms of communication that affect the digital public sphere. In these environments, different social and political actors contribute to extreme their positions, using bots to create spaces for social distancing where hate speech and incivility have a place, a phenomenon that worries scientists and experts. The main objective of this research is to analyze the role that these automated agents played in the debate on social networks about the Spanish Government's management of the global COVID-19 pandemic. For this, "Social Big Data Analysis" techniques were applied: machine learning algorithms to know the positioning of users; bot detection algorithms; "topic modeling" techniques to learn about the topics of the debate on the web, and sentiment analysis. We used a database comprised of Twitter messages published during the confinement, as a result of the Spanish state of alarm. The main conclusion is that the bots could have served to design a political propaganda campaign initiated by traditional actors with the aim of increasing tension in an environment of social emergency. It is argued that, although these agents are not the only actors that increase polarization, they do contribute to deepening the debate on certain key issues, increasing negativity.

## RESUMEN

Los contextos de polarización social y política están generando nuevas formas de comunicar que inciden en la esfera pública digital. En estos entornos, distintos actores sociales y políticos estarían contribuyendo a extremar sus posicionamientos, utilizando «bots» para crear espacios de distanciamiento social en los que tienen cabida el discurso del odio y la «incivilidad», un fenómeno que preocupa a científicos y expertos. El objetivo principal de esta investigación es analizar el rol que desempeñaron estos agentes automatizados en el debate en redes sociales sobre la gestión del Gobierno de España durante la pandemia global de COVID-19. Para ello, se han aplicado técnicas de «Social Big Data Analysis»: algoritmos de «machine learning» para conocer el posicionamiento de los usuarios; algoritmos de detección de «bots»; técnicas de «topic modeling» para conocer los temas del debate en la red, y análisis de sentimiento. Se ha utilizado una base de datos compuesta por mensajes de Twitter publicados durante el confinamiento iniciado a raíz del estado de alarma español. La principal conclusión es que los «bots» podrían haber servido para diseñar una campaña de propaganda política iniciada por actores tradicionales con el objetivo de aumentar la crispación en un ambiente de emergencia social. Se sostiene que, aunque dichos agentes no son los únicos actores que aumentan la polarización, sí coadyuvan a extremar el debate sobre determinados temas clave, incrementando la negatividad.

## KEYWORDS | PALABRAS CLAVE

COVID-19, political bots, political polarization, digital propaganda, public opinion, social networks analysis.  
COVID-19, bots políticos, polarización política, propaganda digital, opinión pública, análisis de redes sociales.



## 1. Background and introduction

The polarisation that occurs in social and political debates on social networks such as Twitter or Facebook has become an increasingly relevant phenomenon for the social sciences. This is not only because it can create a divide between the parties involved in public debate, but also because this divide occurs as a consequence of strategies such as “incivility” or “flaming”, which are based on hatred, the discrediting of one of the parties, name-calling, etc. In short, polarisation is not only relevant because of its consequences, but also because of the emergence of “modes of communication” that can generate a state of “failed communication.”

Generally speaking, political polarisation is associated with selective exposure to information. It is believed that this limitation facilitates the development of extreme values, attitudes, and political stances, or is at least based on the strengthening of the earlier stances of the persons involved in the debate (Prior, 2013). In this sense, digital social networks favour polarisation by allowing greater control over the type and the sources of information (Sunstein, 2001, 2018). In this context, polarisation experts have warned about how important different social and political players and stakeholders are in the development of polarisation processes. It has been shown that when extreme positions are adopted by political leaders, it causes a ripple effect that ends up influencing the position of their followers. When leaders slant or present biased information or interpretation of a media event, such bias makes it harder for their followers to understand representatives of opposing views (Boxell et al., 2017).

Our work aims to explore this area by focusing on the role of an agent that can be key in polarisation processes: the “bots”. Although the literature on the interference of this type of non-human agent in social, political, and electoral processes is extensive (Ferrara et al., 2016; Howard et al., 2018; Keller & Klinger, 2019), the study of their role in polarisation processes is less so. The question arises as to whether “bots” are contributing to the polarisation of political debate during times of social upheaval. In this sense, our main objective is to study to what extent and by means of what strategies these agents create or exaggerate the processes of political polarisation in debates that take place on digital social networks. To achieve this objective, we are using as a case study the public debate that took place on social networks regarding the Spanish government’s management of the COVID-19 global pandemic health crisis during the first few months.

### 1.1. Digital political polarisation

Of course, there is no universally accepted definition of polarisation. Abramowitz (2010) sees it as a process that gives greater consistency and strength to the attitudes and opinions of citizens and parties; and others such as Fiorina and Abrams (2008) state that it is the distancing of people’s views in a given political context. Similarly, experts are divided between those who argue for the centrality of an ideological polarisation and others who support the emergence of an emotional polarisation (Lelkes, 2016). However, there are polarisation processes that might be “to be expected” (Sartori, 2005); for example, in two-party contexts, especially when the electoral system is presidential. That is, when two candidates face each other in an electoral process, it is to be expected that their followers will be structured around two poles (generated by each of the candidates). The same can happen when a debate arises around an issue that has two clearly defined positions (e.g. for or against). Polarisation studies have recently focused their attention not so much on the study of the formation of poles, but on those aspects that lead towards the process of negative public debate. Numerous studies suggest that not being in contact with multiple and/or reliable information, intentionally negativised and/or incomplete information is what transforms polarisation into a negative process. Authors such as Prior (2013) state that combining selective exposure to information with negative polarisation results in the entrenching, or radicalisation of, the starting positions of the followers on both sides.

Experts have highlighted different mechanisms to explain this process. From a methodological and experimental individualistic point of view, Taber and Lodge (2006) initiated a field of analysis on the psychological and rational components of the selection of those pieces of information that best matched the individuals’ prior conceptions. In this sense, humans tended to filter information in such a way as to identify that which reinforced their previous conceptions or emotional dispositions as more relevant

or truthful. Sunstein (2001; 2018) highlights the significant role played by the internet and digital social networks in the polarisation process. From this point of view, the internet offers the possibility of selecting the sources of information to which people are exposed more precisely, as well as the people with whom they choose to debate. Finally, different authors show how major political players are the key agents in polarisation by creating a ripple effect of their potentially biased and/or negative messages (Allcott & Gentzow, 2017). We know that people's levels of polarisation are closely linked to the membership of certain social groups and the consumption of certain political information (Boxell et al., 2017). This circumstance may be a consequence of the polarisation they receive from the main political agents: parties, organisations, media, etc., which could lead to the contagion effect. The danger of this contagion effect is what Lelkes (2016) calls "emotional polarisation", a process in which citizens tend to radicalise their emotions and/or affections on various issues. They also become fixed on them, following the polarised discourses of political parties and public representatives. In a similar vein, Calvo and Aruguete (2020: 60-70) believe that emotional polarisation on the networks constitutes "a fiery defence of one's own beliefs in the face of the other's communicational objectives" and affirm that "hating the networks is an emotional, cognitive and political act".

Recent studies (Mueller & Saeltzer, 2020) also refer to this network contagion being caused by messages that evoke negative emotions and they suggest that negative emotional communication freely arises as a result of strategic campaigns (Martini et al., 2021). In this context, "incivility" emerges as a communicative strategy which, resulting from the generation of negative emotions through insult or social discrediting, attempts to exclude the adversary from public debate (Papacharissi, 2004). Our work is based on this latter theoretical context, in which polarisation is understood to be the process of extension of the attitudes of political leaders or, as in this case, of the digital tools that play central roles in the debate.

## 1.2. The role of "bots" in digital propaganda

Advanced AI and micro-targeting systems mobilise users through social networks. Specifically, in digital political communication, these propaganda tactics go beyond for-profit fake news and conspiracy theories because they involve the deliberate use of misinformation to influence attitudes on an issue or towards a candidate (Persily, 2017). Among these propaganda strategies are political "bots", which are social media accounts controlled in whole or in part by computer algorithms. They create automatic content to interact with users, often by impersonating or mimicking humans (Ferrara et al., 2016). The main purpose is to break the flow of debate in the networks by denigrating opponents or users who have opposing views (Yan et al., 2020). Some authors point out that the use of "bots" is not always linked to malicious purposes, highlighting as an example the usefulness of Twitter in informing the population about the risks of the COVID-19 pandemic, in disseminating accurate breaking news, and in urging citizens to stay at home (Al-Rawi & Shukla, 2020).

However, while the use of "bots" in pandemics is still an under-researched field of study, there is already empirical evidence that they have been used to promote conspiracy theories in the multimedia political sphere regarding the dissemination of controversial and polarised messages (Moffit et al., 2021). In contrast to social "bots" of an inclusive nature, research highlights political "bots" whose function is to disseminate messages containing negative emotions (Stella et al., 2018; Neyazi, 2019; Yan et al., 2020; Adlung et al., 2021), to spread fake news on a large scale, (Shao et al., 2018; Shu et al., 2020; Yan et al., 2020) and to draw on users' private information for partisan political purposes (Boshmaf et al., 2013; Persily, 2017; Yan et al., 2020).

Several authors warn that, in an increasingly polarised and troubled network environment, political "bots" increase the level of vulnerability of users because they are better able to segment them and target their propaganda (Stella et al., 2018; Yan et al., 2020). Price et al. (2019) also highlight the difficulties that the various "bot" detection tools face due to the constant developments in their ability to improve and modify behaviour. The rise of political tension in the public digital sphere is also linked to disinformation campaigns on social media such as "astroturfing": an activity initiated by political players on the internet. This is strategically manufactured in a top-down manner, mimicking bottom-up activity by autonomous individuals (Kovic et al. 2018:71). They consist of a set of robots coordinated by grassroots activists who



emulate ordinary citizens and act independently. They have the potential to influence electoral outcomes and political behaviour by positioning themselves for or against various causes (Howard, 2006; Walker, 2014; Keller et al., 2019). Also known as "Twitter bombs" (Pastor-Galindo et al., 2020) or "cybertroops" (Bradshaw & Howard, 2019), they operate by disseminating comments that are very similar to each other and consistent with the objective of the propaganda campaign (Keller et al., 2019). Often, the features that characterise this type of robotic messaging are the use of false information, uncivil language and hate messages against minority or opposing opinion groups. Moreover, attempts are made to harass and exclude these groups from the debate (Keller et al., 2019; Santana & Huerta-Cánepa, 2019).

Beyond the damage that this practice wreaks on the natural flow of conversations on social networks, the main problem is that it often leads to the process of strong polarisation. When these dynamics involve extreme positions that prevent dialogue, it is called centrifugal polarisation (Sartori, 2005) and can pose a threat to democracy (Morgan, 2018). Papacharissi (2004) refers to this type of polarisation as "incivility" and specifies that it involves the use of inappropriate, insulting, or demeaning language. Messages that invade the polarised online debate and infringe on personal freedoms or the freedoms of certain social groups (Rowe, 2015). Sobieraj and Berry (2011) refer to "outrage" as a type of political discourse that aims to provoke visceral audience responses such as fear or moral outrage through the use of exaggerations, sensationalism, lies, inaccurate information or partial truths that affect specific individuals, organisations or groups.

The literature points to the fact that hate speech in cyberspace can be fuelled by non-human agents such as social bots, leading to a global problem that was exacerbated by the current COVID-19 crisis (Uyheng & Carley, 2020). We understand that this use of harmful AI tools polarised and increased "incivility" in the political debate regarding the pandemic. The polarisation derived from this health crisis has been the subject of study on platforms such as Youtube (Serrano-Contreras et al. 2020; Luengo et al., 2021). We believe that the presence of "bots" in the debates that took place during the health crisis is an area that needs to be investigated further. Our hypothesis is that these agents are not the only key players in the polarisation process but that they do use polarised situations to heighten the debate by including a greater degree of negativity, particularly on certain key issues. We therefore believe it necessary to investigate the involvement of these agents and the effect they had on Twitter during the "State of Alarm."

## 2. Material and methods

Our database is comprised by tweets downloaded throughout the entire "State of Alarm" period of lockdown that was imposed in Spain. To achieve our objective, we applied "Social Big Data Analysis" techniques such as machine learning algorithms to find out the stance of users on the network towards the Spanish government, algorithms for detecting "bots," "topic modelling" techniques to ascertain the topics of debate on social networks, and sentiment analysis.

### 2.1. Data source and cleansing

The data was downloaded from the Twitter API via R-Studio with the "rtweet" library (Kearney, 2019). The data was downloaded according to a set of keywords composed of the names of the accounts of the main political parties in Spain and those of their political leaders, and they also included the words "state of alarm", "coronavirus" and "COVID". The database that was generated covers the period of 16 March 2020 to 29 June 2020. Data were downloaded in 5 different batches, during the first week of each of the phases of the "State of Alarm", to cover the whole period. By the end, 4,895,747 messages were collected.

A cleansing of the data was carried out to remove any downloaded messages that were not relevant to the aim of the study. For this purpose, the "machine learning" methodology was applied. Also, because of how lively the debate was on the networks, many algorithms trained as batches were downloaded. For this purpose, a simple random sample of 1,500 tweets per batch was generated for manual pre-coding by a trained expert. This coding consisted of labelling the messages as "belongs" or "does not belong" to the target of the study. During this process, machine learning algorithms were applied, with lineal Support Vector Machines (SVMs) showing the best performance, finding an average inter-batch accuracy of 0.8

and an average F-measure of 0.768. To carry out this task, text processing was applied to ensure the correct compatibility with machine learning algorithms. Firstly, the content was “tokenised”, separating any given tweet into all the words it included. Secondly, words whose content did not provide relevant “stopwords”, such as determiners, prepositions, etc., were eliminated. Finally, a tf-idf (“term frequency - inverse document frequency”) matrix was constructed as input for the machine learning algorithms, where each row represented a tweet, and the columns represented all the words that appeared in the body. Following the application of SVM-linear, a total of 1,208,631 messages corresponding to the study target, posted by 469,616 users, were highlighted.

## 2.2. Detection of “bots”

To detect and classify users as “bots” or “non-bots”, the algorithm proposed by Kearney (2019) called “tweetbotornot” in its “FAST-gradient boosted” version, incorporated in the R-Cran statistical package “tweetbotornot”, was applied. Also, to be as conservative as possible, only users in the highest quartile of probability of being “bots” were actually identified as “bots.” We believe it is vitally important to maintain this conservative stance because it is preferable to detect fewer “bots” than to include any real users in the “bot” category.

## 2.3. Polarisation measurement

To measure polarisation on social networks, the measurement means of Guevara et al was applied (2020), it is based on fuzzy logic, known as JDJ. The authors go by the premise that reality is not clearly defined one way or the other, but rather that there are different nuances in people’s attitudes. It is understood that while a given individual may be, for example, a supporter of one political party, this does not necessarily mean that they may not still agree with some of the proposals of other, different political parties. In this way, instead of looking in a clear-cut way at a person’s attitudinal score, the degree to which that person’s attitude belongs (or is close) to the poles of the attitudinal axis being measured is computed. In this way, the position of an individual towards the extremes of a variable is considered simultaneously. Thus, the risk of polarisation between an individual “i” and an individual “j” is understood as the joint consideration of the following scenarios:

- How close the individual “i” is to pole A and how close the individual “j” is to pole B.
- How close the individual “i” is to pole B, and how close the individual “j” is to pole A.

Thus, the total polarisation of a population set is the sum of all the possible comparisons between the individuals that compose it.

- Given a variable X.
- Each individual  $i \in N$ .
- $X_A, X_B$  are the poles of X and  $\mu_{X_A}, \mu_{X_B}$  the membership functions of an individual to the poles:  $\mu_{X_A}, \mu_{X_B}: N \rightarrow [0, 1]$  are functions, and for each  $i \in N$   $\mu_{X_A}^{(i)}$  and  $\mu_{X_B}^{(i)}$  are the membership functions of the individual “i” to both poles.

$$JDJ(X) = \sum_{i, j \in N, i \leq j} \varphi \left( \phi(\mu_{X_A}(i), \mu_{X_B}(j)), \phi(\mu_{X_B}(i), \mu_{X_A}(j)) \right)$$

Where,  $\Phi$  is an “overlapping” aggregation operator and  $\varphi$  is the grouping function. In this study, the product has been used as the “overlapping” operator and as the maximum grouping function. This measure presents its maximum value when 50% of the population has a maximum degree of membership to pole A and a null degree of membership to pole B, and the other 50% of the population has a maximum degree of membership to pole B and a null degree of membership to pole A. On the other hand, a null level of polarisation is found not only when 100% of the population has the same attitude level, but also when this value is situated around an extreme, this scenario being the one that presents the greatest distance from the maximum value of polarisation. Since the above equation gives as a result the sum of the polarisation risk for all possible combinations of pairs of individuals, the following calculation is made to facilitate its interpretation:

$$JDJ\_INDEX = \frac{JDJ}{N} * 2$$

Where  $N$  is the total number of individuals. In this way, the measurement shows its minimum value at 0 and its maximum value at 1. Guevara et al. (2020) provides a detailed comparison of this proposal with other measurements in the literature.

## 2.4. Stopping topics

To detect the discourse topics, present in the downloaded messages, the Latent Dirichlet Allocation (LDA) algorithm, present in the R package called "topicmodels" (Grün & Hornik, 2011), was used. This algorithm is based on creating distances between words according to their occurrence together. The algorithm has the particularity of indicating the number of topics a priori, so that for the correct determination of the number of topics, measures of semantic coherence by topics can be applied. It is also recommended that an expert scan the content to determine the appropriate number of topics.

## 2.5. Sentiment Analysis

Sentiment analysis dictionaries were used to detect the amount of negative or positive content present in the digital debate. The Afinn dictionary was used (Hansen et al., 2011), consisting of 2,477 words, scored from most negative to most positive, on a scale of five to five.

## 3. Analysis and results

### 3.1. Classification of messages as "for" or "against" the government

Firstly, machine learning algorithms were applied to encode a given message as "for" or "against" the government. Here, too, the support vector machines showed better results (Table 1). The results indicate satisfactory performance levels, allowing for the correct automatic classification of all the messages present in the database.

Table 1. Results of the SVM-linear classifier for message coding as "in favour" or "against" the Spanish government					
Precision measurements					
Batch	Accuracy	Sensitivity	Kappa	F-Score	AUC
1	0.8492	0.9854	0.4816	0.9122	0.6950
2	0.8960	0.9619	0.7761	0.9277	0.8780
3	0.8392	0.8488	0.6675	0.8439	0.8366
4	0.9133	0.9048	0.8225	0.9090	0.9121
5	0.8318	0.8600	0.6638	0.8456	0.8335

### 3.2. Detection of "bots"

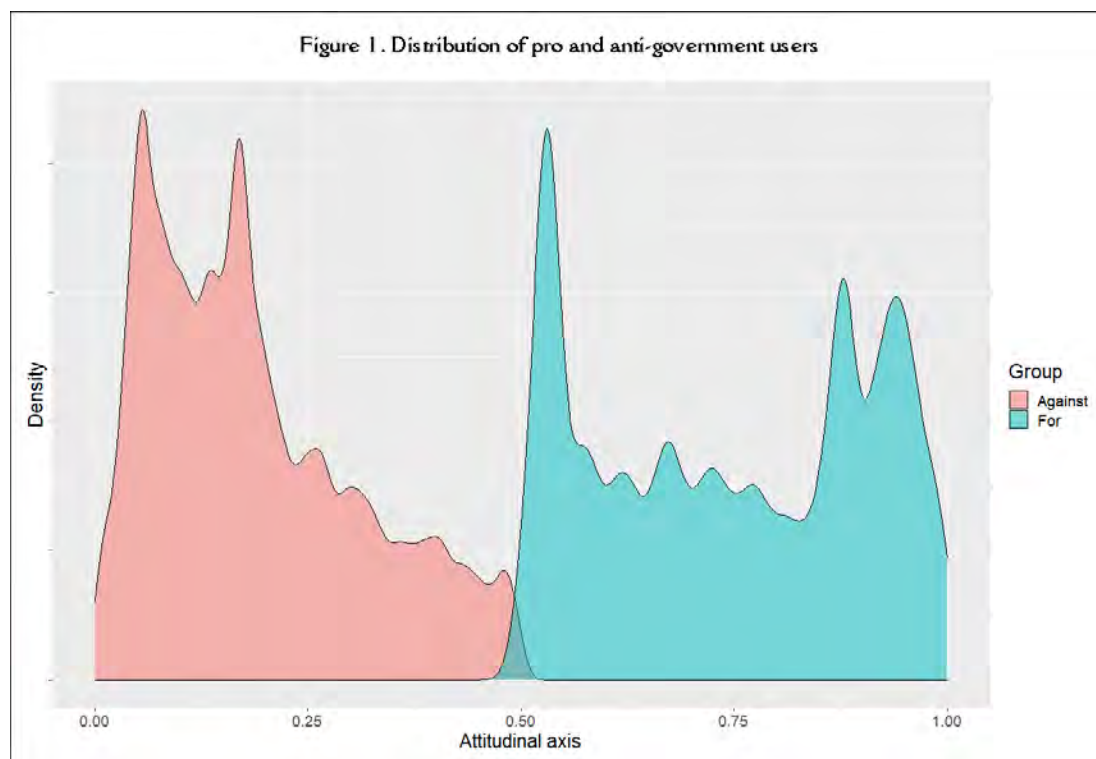
The bot detection algorithm was then applied to identify automated accounts as "bots". The criterion used corresponded to classifying as a "bot" those users whose probability of being a "bot" was located at the highest quartile of probability, which is  $> 0.975$ . When the algorithm was applied to the 469,616 users in the database, 69,033 accounts that could be defined as "bots" were detected. This amounts to a total of 15% of all accounts that were present in the digital debate. Similarly, the 69,033 bots posted 172,704 of the 1,208,631 messages in the filtered database, accounting for 14.28% of the total.

### 3.3. Polarisation measurement

It is important to remember that the calculation of polarisation uses the probabilities of being "for" or "against" the government that are offered by the machine learning algorithms as the degrees of membership of users to both poles, for or against. So, for the given individual "i," there are two values: 1) their probability of being in favour of the government and 2) their probability of being against the government. However, the "input" of the automatic classifiers are tweets, the objective being to calculate the polarisation of users. For each user, the average probability of being for and against the government was calculated for all their posted messages. Thus, for each user, the necessary two degrees of membership were obtained and could be used to calculate the polarisation measurement. On the other hand, due to the computational

costs of applying the measure to 469,616 users, which meant comparing all users with each other, a total of  $(469,616^2)/2 = 110,269,593,728$  JDJ calculations was necessary. The JDJ index was then calculated as the average of 1,500 JDJ iterations for a simple random sample of  $N=200$  users per iteration.

First, polarisation was measured for the overall sample ("non-bots" and "bots"), obtaining a level of  $(JDJ\_mean)_{1500} = 0.76$ ;  $sd = 0.027$ , resulting in a high level of polarisation  $JDJ\_mean \rightarrow [0,1]$ . Figure 1 shows the graphical representation of the distribution of users to be in favour of or against the government. A probability of 0.5 means to be against the government, while  $>0.5$  means to be in favour.



Polarisation levels were also calculated for 1) "non-bots" and 2) "bots". Thus, JDJ showed an average polarisation level for the 1,500 iterations of  $JDJ\_mean\_no\_bots_{1500} = 0.761$ ;  $sd = 0.026$ . For the "bots" group, it was  $JDJ\_mean\_bots_{1500} = 0.765$ ;  $sd = 0.026$ . Finally, the Mann-Whitney test for two independent variables was applied to determine whether the differences found in the averages were statistically significant. Thus, with a statistic  $U = 1021715$ , we found a significance level  $<0.000$ , finding a higher level of polarisation for "bot" users compared to "non-bot" users.

### 3.4. Detection of topics and polarisation

Firstly, the LDA topic detection algorithm was applied. Since the groups were established a priori, the algorithm was applied repeatedly by changing the parameter "number of groups" from one to ten to calculate the coherence for each number of clusters. It is important to remember that this coherence index is based on the semantic similarity between words. Therefore, after an expert had checked the algorithm's different suggestions, the decision was made to opt for the detection of three topics (coherence of 0.42) in the dialogue since, as can be seen in Figure 2, they are well defined (health, economy, and politics). On the other hand, it is worth mentioning that due to the computational costs of the LDA algorithm, it was possible to access all the messages published by "bots" ( $N = 172,704$ ), while a simple random sample of  $N = 200,000$  was applied to the 1,035,927 messages from the group of "non-bots". According to the formulas for calculating sample size (Fernández, 1996), with a population of  $N = 1,035,927$ , a confidence



[illegible]

**Figure 3. Polarisation values on messages**

Category	Real	Fict
Healthcare	0.814	0.820
Economics	0.839	0.835
Politics	0.845	0.821

<https://doi.org/10.3916/C71-2022-05> • Pages 59-71

politics) (Table 2). Given the significance levels found ( $p < 0.00$ ), it can be concluded that both user type and subject matter affect polarisation levels. Furthermore, the interaction effect is also significant, which leads to the conclusion that the levels of polarisation found in the "topic" variable are conditioned by whether or not one is a "bot."

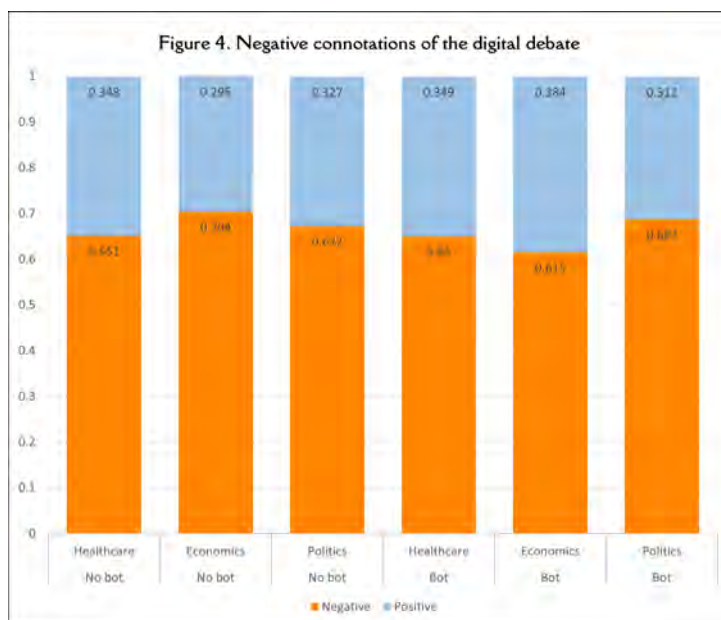
Table 2. Tests for inter-subject effects					
Dependent variable: Polarisation					
Origin	Type III sum of squares	df	Root mean square	F	Sig.
Corrected model	1,123 <sup>a</sup>	5	.225	327,641	.000
Intersection	6,194,468	1	6,194,468	9,032,376,674	.000
USER	.121	1	.121	176,075	.000
TOPIC	.668	2	.334	486,967	.000
USER * TOPIC	.335	2	.167	244,100	.000
Error	6,168	8,994	.001		
Total	6201,759	9,000			
Corrected total	7,292	8,999			

a. R squared = .154 (Adjusted R-squared=.154).

Finally, a new combination variable ("user" x "topic") with six levels was created to determine which of these scenarios had a higher level of polarisation. Thus, a one-factor ANOVA was applied, showing a statistic  $F_{(5)} = 327.641$ ,  $p < 0.000$ , where multiple comparisons were performed with Tukey's test, finding statistically significant differences between all levels except for the "no bot" - health and "no bot" - politics levels. It is safe to assume that the three highest levels of discourse polarisation were found in the political topic in the "bot" debate, followed by the economic topic in the "bot" debate and the economic topic in the "non-bot" debate (Figure 3).

### 3.5. Sentiment analysis and topics

Finally, the Afinn dictionary for sentiment analysis was applied to each of the detected topics. As can be seen in Figure 4, words with negative connotations predominate throughout the digital debate, finding a greater presence in messages that talk about the economy for the "non-bots" group (0.704%), followed by politics in the discourse of "bots" (0.687%) and politics in "non-bots" (0.672).



#### 4. Discussion and conclusions

In this paper we set out to analyse to what extent and with what strategies political "bots" participate in the public discussion process through digital social networks. We focused particularly, as a sample case, on the Twitter debate surrounding the Spanish government's management of the global COVID-19 pandemic. The theoretical background of this research was to advance our understanding of a complex and potentially damaging communication process. That is, political polarisation and the emergence of a "communication failure" scenario. In short, we are talking about situations in which communication between participants tends to become fixated on strong viewpoints and to ignore, if not attack, those who think differently ("incivility").

Many experts point to the presence of political polarisation in public debate processes, such as the one we are analysing here. However, the very definition of polarisation given in this article warns that this phenomenon does not depend solely on the existence of two or more opposing poles, but on a tendency towards isolation and a breakdown in communication between these different poles. A process derived from the so-called "echo chambers" on the web (Colleoni et al., 2014), in which information exchanges take place mainly between individuals with similar ideological preferences, especially when it comes to political issues (Barberá et al., 2015). Thus, we have shown (Figure 1) how the debate around governance was strongly polarised. In fact, our data suggests that the political "bots" identified in the analysis have a greater tendency to polarise public opinion than "non-bot" accounts. This finding is the necessary but not sufficient prelude to finding the type of polarisation that alerts us and encourages a rupture in communication.

The key factor (reason enough) for this rupture is, in our view, the identification of a strategy of distancing and entrenchment of the parties involved in the debate. Beyond the actions of political and media leaders and certain opinion leaders and thanks to sentiment analysis, in this study we have identified a strategy of negativisation of the debate that is more present among the "bots" than it is among the "non-bot" accounts. Our interpretation shows that this strategy seeks to skew opinion on the Spanish government's performance. It is biases of this kind that lead the way to polarisation in its most negative sense of rupture and distancing. An emotional polarisation that can have serious repercussions, particularly in times of political turmoil (Iyengar, 2019) such as that brought about by the onset of the pandemic.

This strategy becomes even clearer when we analyse the central topics of our case study. Here we observe how the "bots" tend to focus the debate on politics, rather than on economic or health issues (topics of debate that are potentially more subject to scientific and objectifiable criteria). This is a field in which it is easier to attack one or several figures rather than talk about general issues (as we have seen, the literature points to this strategy as a source of polarisation). In other words, it is an area in which it is more accessible to develop "ad hominem" strategies, represented in Figure 3 by the constant references to the Spanish Prime Minister, which are clearly more biased and focused on the flaws and the negative circumstances surrounding the person. In other words, it is an individual strategy centred on "incivility" as a form of communication management.

Political issues are also the most polarised in this debate. They also show the greatest difference between "bots" and "non-bots". The latter present a more polarised debate on the management of the pandemic in political terms, being, out of the three topics, the one most negativised by the "bots". In line with the concept of outrage outlined by Sobieraj and Berry (2011), the topics found in the "bots" reinforce the melodrama and improbable predictions of impending doom that are attributed to decisions made by the government, a discourse distinguished by the tactics used to provoke emotion, rather than to evoke emotion in the political arena. Therefore, the use of "bots" does not seem to be oriented towards informing society about the risks of the pandemic or promoting prevention dynamics (Al-Rawi & Shukla, 2020), but rather is mostly focused on mobilising public opinion against the government by negativising it. As noted by Howard (2006), Walker (2014) and Keller et al. (2019), "bots" have the potential to influence users' political positioning online because, according to our results, they emulate ordinary citizens concerned with purely health issues. They can be considered political rather than social "bots" because they spread messages with negative sentiments (Stella et al., 2018; Neyazi, 2019; Yan et al., 2020; Adlung et al., 2021), specifically towards the government. These robots could have been designed to launch a

political propaganda campaign of "astroturfing" initiated by traditional agents with the aim of increasing tension in a context of social emergency. Iyengar et al. (2019) warned that this is a type of strategy closely linked to the theory of the spiral of silence because, in a context of uncertainty and general frustration, it makes it difficult for users to express favourable opinions regarding any of the health measures taken by the government.

Our impression is that the polarisation-negativisation binomial is the ammunition chosen by these types of accounts to alienate and confront the parties involved in this public debate, as well as to create an environment of tension, lack of civility and attacks on those who think differently. In an already polarised context, whether it is due to the actions of other agents (political, social or media) or due to the situation of exceptionality and uncertainty itself, the ammunition used by the "bots" in this debate has consisted of making positions more extreme. This finding is one that can serve as a basis for future research and can be contrasted in various case studies with similar or different characteristics to the one carried out here.

### Authors' Contributions

Idea, JMR, JAG, BC-M; Literature review (state of the art), JMR, BC-M; Methodology, JAG, DG; Data analysis, JMR, JAG, BC-M, DG; Results, JMR, JAG, BC-M; Discussion and conclusions, JMR, BC-M; Writing (original draft), JMR, JAG, BC-M; Final revisions, JMR, JAG, BC-M; Project design and sponsorship, JMR.

### Funding Agency

Research Group for Data Science and Soft Computing for Social Analytics and Decision Aid. This research has been supported by national research projects funded by the Spanish Government, with reference to R&D&I, PID2019- 106254RB-I00 funding: MINECO (Period: 2020-2024) and PGC2018-096509B-I00.

### References

- Abramowitz, A.I. (2010). *The disappearing center*. Yale University Press. <https://bit.ly/3s7UlwC>
- Adlung, S., Lünenborg, M., & Raetzsch, C. (2021). Pitching gender in a racist tune: The affective publics of the # 120decibel campaign. *Media and Communication*, 9, 16-26. <https://doi.org/10.17645/mac.v9i2.3749>
- Al-Rawi, A., & Shukla, V. (2020). Bots as active news promoters: A digital analysis of COVID-19 tweets. *Information*, 11(10), 461-461. <https://doi.org/10.3390/info11100461>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-247. <https://doi.org/10.1257/jep.31.2.211>
- Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26, 1531-1542. <https://doi.org/10.1177/0956797615594620>
- Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2013). Design and analysis of a social botnet. *Computer Networks*, 57(2), 556-578. <https://doi.org/10.1016/j.comnet.2012.06.006>
- Boxell, L., Gentzkow, M., & Shapiro, J. (2017). *Is the internet causing political polarization? Evidence from demographics*. National Bureau of Economic Research. <https://doi.org/10.3386/w23258>
- Bradshaw, S., & Howard, P.N. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation*. Oxford Internet Institute. <https://acortar.link/puyazU>
- Calvo, E., & Aruguete, N. (2020). *Fake News, trolls y otros encantos. Cómo funcionan (para bien y para mal) las redes sociales*. Siglo XXI. <https://doi.org/10.22201/fcpys.24484911e.2020.29.76061>
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317-332. <https://doi.org/10.1111/jcom.12084>
- Fernández, P. (1996). Determinación del tamaño muestral. *Cad Aten Primaria*, 3, 1-6. <https://bit.ly/3DYcijz>
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96-104. <https://doi.org/10.1145/2818717>
- Fiorina, M.P., & Abrams, S.J. (2008). Political polarization in the American public. *Annual Review of Political Science*, 11, 563-588. <https://doi.org/10.1146/annurev.polisci.11.053106.153836>
- Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13). <https://doi.org/10.18637/jss.v040.i13>
- Guevara, J.A., Gómez, D., Robles, J.M., & Montero, J. (2020). Measuring polarization: A fuzzy set theoretical approach. In M. Lesot, S. Vieira, M. Reformat, J. Carvalho, A. Wilbik, & B. B.-M. R. Yager (Eds.), *Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 510-522). Springer. [https://doi.org/10.1007/978-3-030-50143-3\\_40](https://doi.org/10.1007/978-3-030-50143-3_40)
- Hansen, L.K., Arvidsson, A., Nielsen, F.A., Colleoni, E., & Etter, M. (2011). Good friends, bad news-affect and virality in Twitter. In J. J. Park, L. T. Yang, & C. Lee (Eds.), *Future information technology* (pp. 34-43). Springer. [https://doi.org/10.1007/978-3-642-22309-9\\_5](https://doi.org/10.1007/978-3-642-22309-9_5)



- Howard, P.N. (2006). *New media campaigns and the managed citizen*. Cambridge University Press.  
<https://doi.org/10.1080/10584600701641532>
- Howard, P.N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the U.S. 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, 15(2), 81-93. <https://doi.org/10.1080/19331681.2018.1448735>
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S.J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22(1), 129-146.  
<https://doi.org/10.1146/annurev-polisci-051117-073034>
- Kearney, M.W. (2019). Rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software*, (42), 4-4.  
<https://doi.org/10.21105/joss.01829>
- Keller, F.B., Schoch, D., Stier, S., & Yang, J.H. (2019). Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication*, 37(2), 256-280. <https://doi.org/10.1080/10584609.2019.1661888>
- Keller, T.R., & Klinger, U. (2019). Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, 36(1), 171-189. <https://doi.org/10.1080/10584609.2018.1526238>
- Kovic, M., Rauchfleisch, A., Sele, M., & Caspar, C. (2018). Digital astroturfing in politics: Definition, typology, and countermeasures. *Studies in Communication Sciences*, 18, 69-85. <https://doi.org/10.24434/j.scoms.2018.01.005>
- Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, 80(1), 392-410.  
<https://doi.org/10.1093/poq/nfw005>
- Luengo, O., García-Marín, J., & De-Blasio, E. (2021). COVID-19 on YouTube: Debates and polarisation in the digital sphere. [COVID-19 en YouTube: Debates y polarización en la esfera digital]. *Comunicar*, 69, 9-19.  
<https://doi.org/10.3916/C69-2021-01>
- Martini, F., Samula, P., Keller, T.R., & Klinger, U. (2021). Bot, or not? Comparing three methods for detecting social bots in five political discourses. *Big Data & Society*, 8(2). <https://doi.org/10.1177/20539517211033566>
- Moffitt, J.D., King, C., & Carley, K.M. (2021). Hunting conspiracy theories during the COVID-19 pandemic. *Social Media & Society*, 7(3). <https://doi.org/10.1177/20563051211043212>
- Morgan, S. (2018). Fake news, disinformation, manipulation and online tactics to undermine democracy. *Journal of Cyber Policy*, 3(1), 39-43. <https://doi.org/10.1080/23738871.2018.1462395>
- Mueller, S.D., & Saeltzer, M. (2020). Twitter made me do it! Twitter's tonal platform incentive and its effect on online campaigning. *Information, Communication & Society*, (pp. 1-26). <https://doi.org/10.1080/1369118X.2020.1850841>
- Neyazi, T.A. (2019). Digital propaganda. *India. Asian Journal of Communication*, 30(1), 39-57.  
<https://doi.org/10.1080/01292986.2019.1699938>
- Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New media & society*, 6(2), 259-283. <https://doi.org/10.1177/1461444804041444>
- Pastor-Galindo, J., Nespoli, P., Gómez-Mármol, F., & Martínez-Pérez, G. (2020). Spotting political social bots in Twitter: A use case of the 2019 Spanish general election. *IEEE Transactions on Network and Service Management*, 8, 10282-10304.  
<https://doi.org/10.1109/access.2020.2965257>
- Persily, N. (2017). The 2016 U.S. election: Can democracy survive the internet. *Journal of Democracy*, 28(2), 63-76.  
<https://doi.org/10.1353/jod.2017.0025>
- Price, K.R., Prisalu, J., & Nomin, S. (2019). Analysis of the impact of poisoned data within twitter classification models. *IFAC-PapersOnLine*, 52(19), 175-180. <https://doi.org/10.1016/j.ifacol.2019.12.170>
- Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16, 101-127.  
<https://doi.org/10.1146/annurev-polisci-100711-135242>
- Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, 18(2), 121-138. <https://doi.org/10.1080/1369118X.2014.940365>
- Santana, L.E., & Huerta-Cánepa, G. (2017). ¿Son bots? Automatización en redes sociales durante las elecciones presidenciales de Chile 2017. *Cuadernos.info*, 44, 61-77. <https://doi.org/10.7764/cdi.44.1629>
- Sartori, G. (2005). *Parties and party systems: A framework for analysis*. ECPR press.
- Serrano-Contreras, I.J., García-Marín, J., & Luengo, O.G. (2020). Measuring online political dialogue: Does polarization trigger more deliberation? *Media and Communication*, 8, 63-72. <https://doi.org/10.17645/mac.v8i4.3149>
- Shao, C., Ciampaglia, G.L., Varol, O., Flammini, A., Menczer, F., & Yang, K.C. (2018). The spread of low-credibility content by social bots. *Nature Communication*, 9(1), 1-10. <https://doi.org/10.1038/s41467-018-06930-7>
- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*, 8, 171-188.  
<http://doi.org/10.1089/big.2020.0062>
- Sobieraj, S., & Berry, J.M. (2011). From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication*, 28(1), 19-41. <https://doi.org/10.1080/10584609.2010.542360>
- Stella, M., Ferrara, E., & De-Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. In J. Kleinberg (Ed.), *Proceedings of the National Academy of Sciences*, volume 115 (pp. 12435-12440).  
<https://doi.org/10.1073/pnas.1803470115>
- Sunstein, C.R. (2001). *Designing democracy: What constitutions do?* Oxford University Press.
- Sunstein, C.R. (2018). *#Republic*. Princeton university press.
- Taber, C.S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, 50(3), 755-769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>

- Uyheng, J., & Carley, K.M. (2020). Bots and online hate during the COVID-19 pandemic: Case studies in the United States and the Philippines. *J Comput Soc Sc*, 3, 445-468. <https://doi.org/10.1007/s42001-020-00087-4>
- Walker, E.T. (2014). *Grassroots for hire: Public affairs consultants in American democracy*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139108829>
- Yan, H.Y., Yang, K., Menczer, F., & Shanahan, J. (2020). Asymmetrical perceptions of partisan political bots. *New Media & Society*, 23(10), 3016-3037. <https://doi.org/10.1177/1461444820942744>



# AlfaMed



**EuroAmerican Interuniversity Research Network**  
on Media Literacy for Citizenship

[www.redalfamed.org](http://www.redalfamed.org)





**MOOC** MIRÍADAX

# How to publish in high **impact** **journals?**

**Sign up now!**

Learn how to publish scientific articles in high impact international journals with this MOOC. This course will help you to understand the publishing process involved in the publication of a scientific manuscript in peer-reviewed journals, focusing on writing, references, promotion through academic social media, and other essential guidelines to publish successfully and rigorously.



**Certified by**



**Universidad  
de Huelva**



Course available in Spanish.

**[www.miriadax.net](http://www.miriadax.net)**







Comunicar 71

---

# Kaleidoscope

Researchs

Studies

Proposals



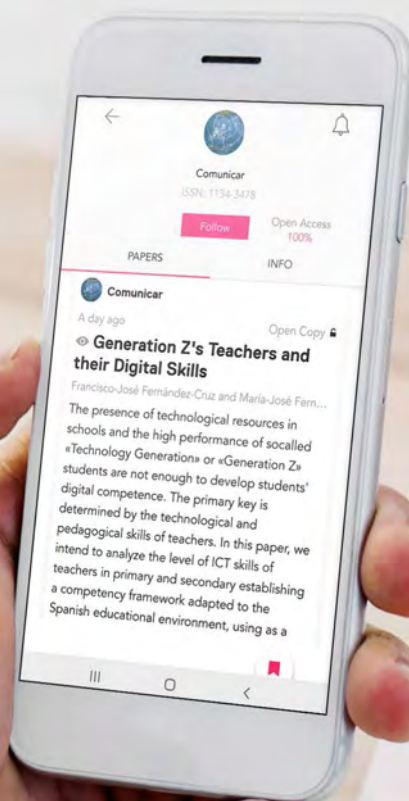


# Comunicar

Follow us in



## Researcher



GET IT ON  
Google Play

Download on the  
App Store

Available on  
AppGallery



# Twitter and human trafficking: Purposes, actors and topics in the Spanish-speaking scene

La trata de personas en Twitter: Finalidades, actores  
y temas en la escena hispanohablante

- Alba Sierra-Rodríguez. Researcher, University Institute for Research in Women's and Gender Studies, University of Granada (Spain) (albasierra@ugr.es) (<https://orcid.org/0000-0003-3016-7963>)
- Wenceslao Arroyo-Machado. FPU Predoctoral Fellow, Department of Information and Communication, University of Granada (Spain) (wences@ugr.es) (<https://orcid.org/0000-0001-9437-8757>)
- Dr. Domingo Barroso-Hurtado. Assistant Professor, Department of Education Studies, University of Extremadura, Cáceres (Spain) (dobarrosoh@unex.es) (<https://orcid.org/0000-0001-6482-9860>)

## ABSTRACT

Human trafficking is a phenomenon linked to several forms of exploitation, such as sexual exploitation, forced labour, forced marriage, begging, forced criminal activity, or organ removal. There are different debates about a potential overrepresentation of some of these exploitation purposes, such as the sexual one, in the discourses underlying international regulations, news or institutional campaigns. This may have consequences on the identification of some of the population affected by this phenomenon and the assistance provided to them. The aim of this paper is to analyse the temporal evolution of the purposes of exploitation, actors and topics most represented in the activity on human trafficking and exploitation on Twitter during the 2011-2020 period. The results show that sexual exploitation is the purpose most represented in Twitter activity and the fact that international organisations and the criminal prosecution of the crime of human trafficking have progressively become more relevant. Furthermore, the networks of topics suggest that sexual exploitation seems to be linked to the notion of prostitution. This may have consequences for the displacement of other purposes of exploitation or the approaches centred on the promotion of human rights to less relevant positions, as well as for the construction of certain images of victims of human trafficking.

## RESUMEN

La trata de personas es un fenómeno vinculado a diferentes finalidades de explotación, como la sexual, laboral, para el matrimonio, la mendicidad, la comisión de delitos o la extracción de órganos. Existen numerosos debates que remiten a una posible sobrerrepresentación de algunas de estas finalidades, como la sexual, en los discursos presentes en normativa internacional, prensa escrita o campañas institucionales. Esto puede derivar en posibles consecuencias sobre la detección de parte de la población afectada por este fenómeno y la atención que se le proporciona. El presente trabajo tiene como objetivo analizar la evolución temporal de las finalidades de explotación, actores y temas más representados en la actividad sobre trata de personas y explotación acontecida en Twitter durante el periodo 2011-2020. Los resultados arrojan que la explotación sexual es aquella con una mayor presencia en la actividad en esta red social, así como que los organismos internacionales y la persecución penal del delito de trata se tornan progresivamente más relevantes. Además, las redes de comunidades de temas analizadas sugieren que la explotación sexual parece estar vinculada a la idea de prostitución en los tuits publicados. Todo ello puede tener consecuencias sobre el desplazamiento de otras finalidades de explotación o de enfoques como los centrados en la promoción de Derechos Humanos a posiciones menos relevantes, así como sobre la construcción de determinadas imágenes de víctimas de este fenómeno.

## KEYWORDS | PALABRAS CLAVE

Trafficking, sexual exploitation, Twitter, social media, social media analysis, prostitution.  
Trata, explotación sexual, Twitter, redes sociales, análisis de redes sociales, prostitución.



## 1. Introduction and state of play

The United Nations Protocol to Prevent, Suppress and Punish Trafficking in Persons (2000), also known as the Palermo Protocol, established the most widely accepted definition of trafficking in persons, embracing the act of recruitment, transfer and receipt of persons through the use of coercive means for the purpose of their exploitation, whether sexual, labour, for forced removal of organs or other slavery-like practices such as forced marriages, begging, or the commission of crimes (Bernstein, 2018). Women and girls are the most severely affected by this phenomenon and sexual exploitation is the most widespread purpose (United Nations Office on Drugs and Crime — UNODC, 2018), although the numbers of cases of labour exploitation seem to have increased in recent years (UNODC, 2020). The lines established for the eradication of human trafficking are prosecution, prevention, and victim protection.

The Palermo Protocol has been criticised for its focus on women and children, influenced by the ‘white slave trade’ policies initiated in the early 1900s to control the sexuality and mobility of European, Anglo-Saxon and American women engaged in prostitution, which, regardless of their consent, were considered forms of slavery (Attwood, 2021; Doezema, 2010; Kempadoo, 2015). Currently, this influence has contributed towards the assimilation of all forms of prostitution with situations of exploitation, thereby leading to an overrepresentation of sexual exploitation (Andrijasevic & Anderson, 2009; González, 2019). Indeed, there are concerns that other exploitation purposes such as labour exploitation have been underrepresented due to the difficulties of their detection (Hebert, 2016; Rodríguez-López, 2020; Zhang, 2012). The Palermo Protocol has also been criticised for prioritising prosecution of the crime as a strategy, reinforcing a criminal approach that displaces others, such as the human rights approach (Edwards, 2007; Reina-Peñas, 2019).

Coverage of the phenomenon in the media seems to reinforce dominant discourses regarding human trafficking (Couto et al., 2012), promoting understanding of its status as a crime, and placing the focus on sexual exploitation (Austin & Farrell, 2017; Denton, 2010; Meneses-Falcón & Urío, 2021). This could contribute to people suffering from other types of exploitation receiving less visibility and attention in public policies (Albright & D’Adamo, 2017; Rodríguez-López, 2020). The issue of sexual exploitation predominates in press articles (Marchionni, 2012; Sanford & Weitzer, 2016) and also in other fields such as academic publications (Gerassi, 2015; Sweileh, 2018), which emphasise criminal prosecution to achieve its eradication.

Other studies analysing the press point to the frequent presence of violated female bodies that reinforce an image of trafficking victims linked to contexts of prostitution (Rodríguez-López, 2018; Sobel, 2014). Advertising campaigns to raise awareness of the phenomenon which are based on the hypervisibility of women and the violence they suffer also seem to reiterate this prototypical idea of the sex slave (Saiz-Echezarreta et al., 2018). This representation may have led to the criminalisation of some of the people subject to human trafficking, especially those present in migration contexts, as they do not fit this image of victims of this phenomenon (Clemente, 2017; Gregoriou & Ras, 2018; O’Connell-Davidson, 2006; Volken, 2006).

The analysis by Papadouka et al. (2016) of the interaction between journalists and readers also highlighted that news about human trafficking engenders debate regarding prostitution among readers. Adopting a similar approach, the analysis of a social media platform such as Twitter, allows us to explore the interaction between groups of actors and debates regarding different topics, although covering a period of up to 15 years given that Twitter was created in 2006. Twitter has proven useful in many studies, such as the analysis of topics in publications about climate change that may or may not be shared on Twitter or in the press (Haunschild et al., 2019), communities related to anti-vaccine movements (van-Schalkwyk et al., 2020), the relationship networks between educational influencers and the topics of their publications, (Marcelo & Marcelo, 2021) and the networks of actors in the debate regarding dual vocational training in Spain (Barroso-Hurtado et al., 2021).

However, there are hardly any studies focusing on the debate and activity relating to human trafficking on social media. Those that do exist have focused their efforts on understanding how trafficking networks use Twitter to recruit new victims (Burbano & Hernández-Álvarez, 2017; Granizo et al., 2020; Hernández-Álvarez & Granizo, 2021; Tundis et al., 2019). A different survey and interview research

suggests that increased use of social media such as Facebook is associated with increased knowledge about human trafficking (Ruiz-Herrera et al., 2018) or it points to the existence of digital anti-trafficking activism (Gong, 2015).

This article aims at identifying the exploitation purposes linked to trafficking that are most widely disseminated on Twitter, along with some of the main actors and topics present in the activity on this social network. This analysis is considered relevant because the representations and imaginaries contained in debates and discussions on human trafficking can have consequences for the population that experiences it; because of the scarcity of publications addressing this issue and activity on Twitter from the context of social media analysis; and because this social network allows us to explore the activity and interactions of numerous actors in different fields, and of different types and scales. Despite the steady increase in the number of people suffering labour exploitation, the expected result is that Twitter activity on human trafficking and exploitation focuses primarily on sexual exploitation, given the various references to the prominence of this type of exploitation in academic articles, the press, and publications by international agencies. For all these reasons, the following objectives are proposed:

- Objective 1: To describe the level of representation of the different exploitation purposes associated with human trafficking and their evolution over time in Twitter activity.
- Objective 2: To identify the total number of actors and their activity, as well as some of the main actors who tweet and are mentioned in tweets for each of the different exploitation purposes linked to human trafficking, considering their evolution over time.
- Objective 3: To detect the topics present in tweets on human trafficking, considering their evolution over time.

## 2. Materials and methods

### 2.1. Data

Different combinations of terms were established to develop a search strategy divided into six blocks to retrieve all the tweets in Spanish that highlight the main purposes of exploitation linked to human trafficking. These combinations are composed of different root terms that refer to trafficking in persons per se and are complemented by specific words for each type of exploitation: a) sexual; b) labour; c) for begging; d) for forced marriage; e) for the commission of crimes; and f) for forced removal of organs.

Accordingly, each tweet was labelled considering the type of exploitation to which it referred, and some of them could appear in more than one. It is assumed that by naming one or more of the word combinations included in Table 1MC (supplementary materials), a tweet is introducing one or more types of exploitation into the activity and public debate on this phenomenon. In addition, other data were extracted from the tweet, such as the actor who posted it and the mentions made.

The terms associated with each of the exploitation purposes were selected according to different inclusion criteria. The first criterion was the relevance of the terms, defined as their potential to locate a large number of tweets for a particular purpose and their ability to represent some of the most important concepts associated with that purpose. Secondly, the relevance and unambiguity of the terms were assessed, defined as the degree of specificity of the term for the type of exploitation it addresses and the number of potentially misleading results its use could lead to, either because it does not refer to the exploitation purpose in question or because it includes results referring to other purposes.

The download of tweets was carried out on 8 June 2021 using Graphext's Tractor<sup>1</sup> tool, setting 31 December 2020 as the deadline for the publication of tweets and retrieving tweets, retweets and replies. However, tweets dated before 2011 were removed from the analysis due to the low activity prior to that year (two crime tweets, 84 labour tweets, seven marriage tweets, three begging tweets, 18 organ tweets and 764 sex tweets). The final set of tweets consists of 101,051 tweets, of which 90,328 are unique.

### 2.2. Method

Firstly, a descriptive analysis of the tweets was carried out using R. Descriptive statistics were applied and the activity was explored in general and for each type of exploitation in particular, at the level of tweets and actors and differentiating between active actors (those who tweet about trafficking) and mentioned



actors. Semantic networks were also generated with Graphext based on the text of the tweets. In these networks, each node represents a tweet and the connections between pairs of tweets are given by their semantic relationships. By detecting communities using the Louvain algorithm (Blondel et al., 2008), the main topics of interest were identified and the main communities were tagged after manually reviewing the tweets and the most relevant terms in each of them. Two researchers were involved in this labelling task, reaching a consensus on the solution in the event of discrepancies. In both analyses, the activity was considered for the full period (2011-2020) and also for two different periods (2011-2015 and 2016-2020).

### 3. Analysis and results

#### 3.1. Evolution of general activity for each exploitation type

Of the 101,051 tweets analysed (90,328 unique tweets), the most frequent exploitation purpose linked to human trafficking is sexual exploitation (Table 1). Not only does it account for the largest volume of tweets (66,615), it also has the highest number of unique tweets, making it the purpose most frequently addressed exclusively (60,301 tweets; 90.52%). It is followed by labour exploitation (13,493 tweets; 69.2%), while the rest are much further behind. The type of exploitation with the lowest number of tweets is 'for commission of crimes' (242 tweets; 54.88%).

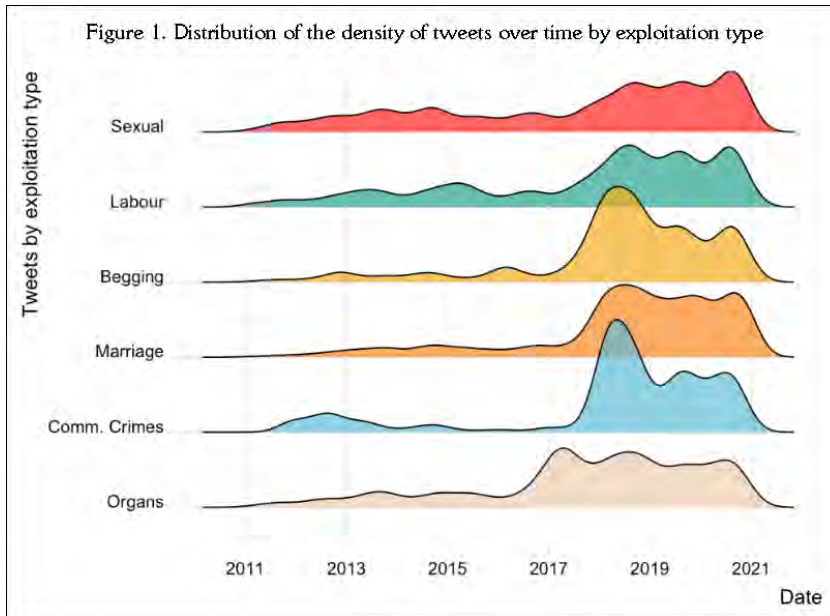
Exploitation type	Tweets	Unique tweets (% unique)
Sexual	66,615	60,301 (90.52%)
Labour	19,498	13,493 (69.2%)
Begging	4,506	2,582 (57.3%)
Forced marriage	4,016	2,458 (61.21%)
Commission of crimes	441	242 (54.88%)
Organ removal	5,975	3,794 (63.5%)
Total	101,051	90,328

Note: The number of tweets and the sum of the 'unique' column do not match because tweets that were repeated in more than one exploitation type were only counted once.

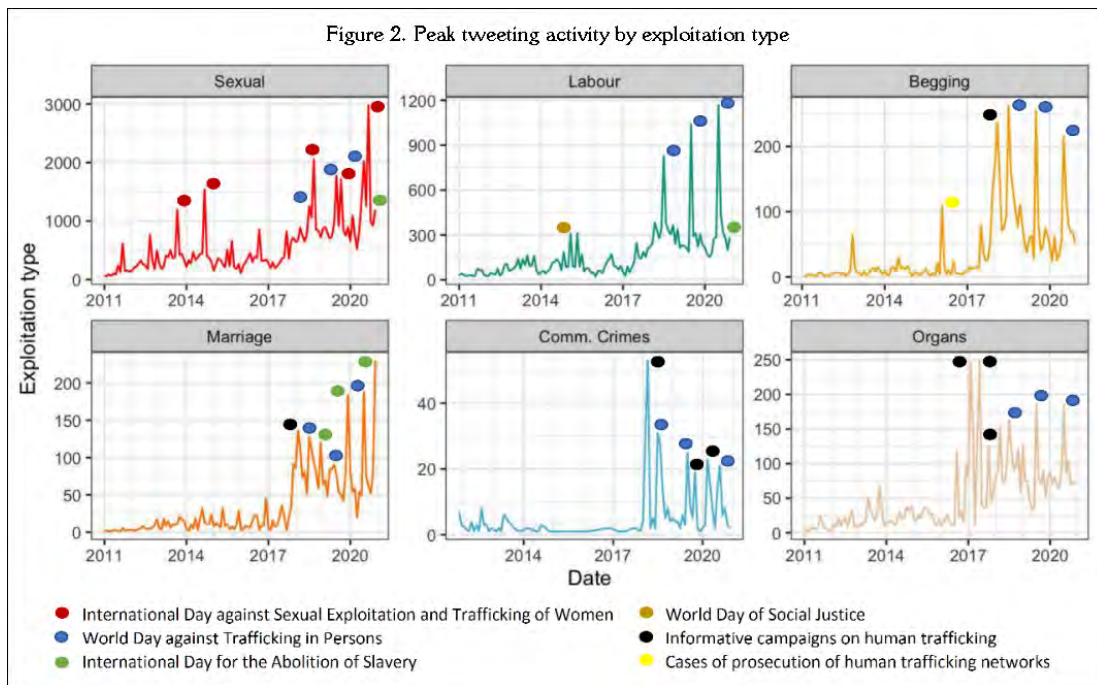
If we take sexual exploitation (the purpose most frequently present among the tweets) as a reference to analyse the overlap between tweets that talk about two exploitation types, it converges with 25.54% of the tweets that address labour exploitation, 31.89% of 'begging', 32.59% of 'forced marriage', 31.97% of 'commission of crimes' and 23.66% of 'forced removal of organs'. Thus, sexual exploitation is not only the exploitation type most frequently represented exclusively, it also comprises roughly a quarter of the tweets about other exploitation types, sharing some of the attention they receive.

The evolution of the level of presence of the different exploitation types in tweets over time has a series of common and divergent features. Firstly, an analysis of the distribution of tweets based on a density graph (Figure 1) shows that activity on Twitter began timidly in 2010 for most exploitation types. Some events in 2010 that could be related to the beginning of the Twitter debate activity include: a) the fact that Twitter use became popular at that time; b) consolidation of international guidelines and national legislation, especially relating to the codification of the crime of human trafficking (General Congress of the United Mexican States, 2012; United Nations, 2000).

The exponential increase in activity is concentrated between 2017 and 2018 in the case of sexual and labour exploitation, begging, forced marriage and commission of crimes and during 2016 and 2017 in the case of forced removal of organs. Most countries that have ratified the Palermo Protocol (2000) had criminalised both human trafficking as a stand-alone offence and any kind of purpose of human trafficking by 2016 (Sweileh, 2018). In the case of exploitation for forced organ removal, the growth in visibility from 2017 onwards could be related to the dissemination of news based on UNICEF reports pointing to the high percentage of people exploited for this purpose in Mexico (Arena Pública, 2017). In fact, Mexico's National Human Rights Commission (CNDH) implemented a strategy in 2019 that aims to collect statistical data at the state level on people affected by this phenomenon, among others, with the involvement of the State Attorney-General's Offices (FGRs) (CNDH, 2019), which are also very active in promoting prevention campaigns that have an impact on this greater visibility.



Following this exponential growth during 2017 and 2018, the cumulative activity on exploitation for begging, for the commission of crimes, and to a lesser extent, for forced removal of organs, declined. Exploitation for marriage and especially labour and sexual exploitation did not experience such a notable decrease, with the activity even increasing for the latter two.



It is worth noting that sexual and labour exploitation have the highest volume of tweets, so the variability of their growth in the previous graphs is not as dependent on such a small number of tweets as might be the case for other types of exploitation, suggesting consolidation of the debate in general terms. Some of these fluctuations in activity may be related to the presence of certain peaks of activity on Twitter (Figure 2), identified based on days or periods with a higher number of tweets, although the events to which

these peaks relate vary according to the exploitation type and do not represent the total amount of activity for each exploitation type. A large proportion of the tweet production peaks is related to the celebration of different international days sponsored by various international actors and organisations. Among these, World Day against Trafficking in Persons is particularly noteworthy, it has been organised since 2013 by the United Nations. This event seems to stand out, in particular since 2018, although there are other prior activities related to this Day, and it also seems to be related to peaks in all exploitation types. The International Day for the Abolition of Slavery, held since 1985 and also sponsored by the United Nations, is more prominent in the case of exploitation for marriage, and in 2020 for sexual and labour exploitation. In contrast, the International Day against Sexual Exploitation and Trafficking of Women, sponsored by the World Conference of the Coalition Against Trafficking in Persons and the Women's Conference, seems to be linked to the main peaks of sexual exploitation, and, although to a lesser extent, also to publications on other types of exploitation such as labour exploitation.

Also, noteworthy are the media impact of the dismantling of a network of human traffickers who engage in forced begging and, even more so, the implementation of informative campaigns on human trafficking carried out for days or weeks by the FGRs of different Mexican regions. These usually refer to several exploitation types and, on fewer occasions, to just one, as in 2017 with the campaigns relating to exploitation for forced removal of organs destined for trafficking. Other events that also seem to lead to increased activity in the debate are the celebration of the World Day against Child Slavery and International Women's Day; the increase in agricultural activity, given that a high percentage of trafficked persons end up working in this sector (UNODC, 2018; 2020); the implementation of campaigns in Spain — #trabajoforzoso— in 2019 and 2020; and the migration crises relating to the migrant caravans in Mexico (05/11/2018 and 18/01/2019), where migration seems to be linked to trafficking and organised crime and where institutions such as the Mexican regional FGRs seem to highlight exploitation for organ trafficking (second caravan).

### 3.2. Analysis of the actors participating in publications and their activity

The number of active actors—those who tweet—who are mentioned and the average dissemination of their activity in terms of tweets, retweets, favourites, and replies also varies depending on the exploitation type and the period: 2011-2015 or 2016-2020 (Table 2).

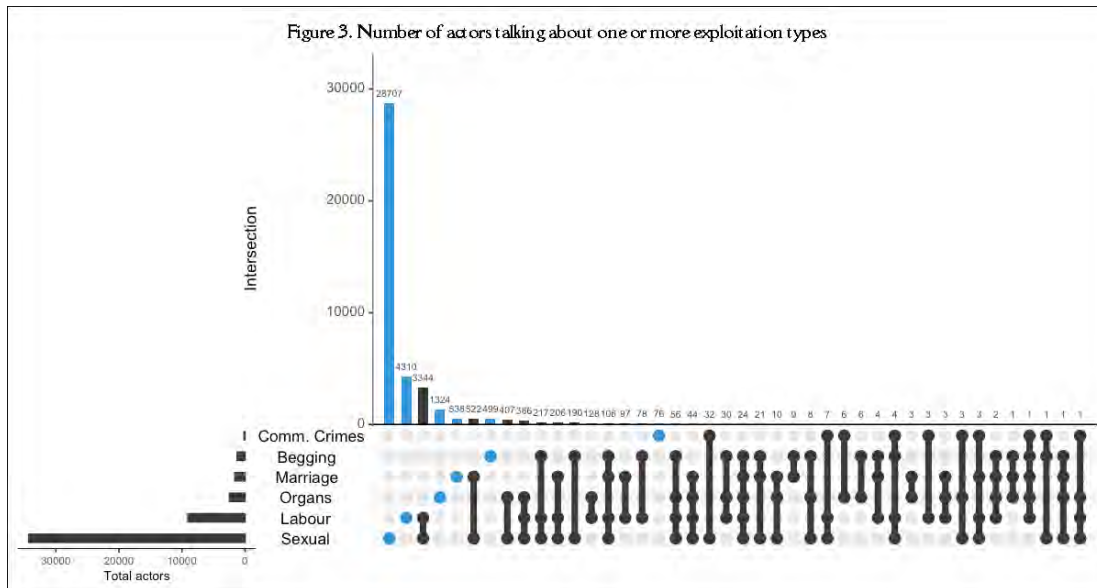
Table 2. Descriptions of the actors participating in tweets and dissemination of their activity, differentiated by exploitation type and period									
Exploitation	Actors and mentions				Activity				
	Active actors	Actors mentioned	Mentions	Avg mentions	Tweets	Avg tweets	Average retweets	Avg favourites	Avg replies
2011-2015									
Sexual	12,107	3,427	6,920	0.35	19,796	1.64	1.50	0.50	0.17
Labour	3,109	830	1,323	0.27	4,922	1.58	1.57	0.49	0.14
Begging	330	114	153	0.32	484	1.47	1.53	0.40	0.12
Forced Marriage	399	107	143	0.27	529	1.33	1.54	0.34	0.11
Comm. Offences	54	13	21	0.33	64	1.19	2.50	0.13	0.23
Organs	789	272	368	0.32	1,143	1.45	1.28	0.39	0.15
2016-2020									
Sexual	23,870	14,098	31,455	0.67	46,819	1.96	5.34	8.89	0.63
Labour	6,796	3,242	6,040	0.41	14,576	2.14	6.58	11.37	0.46
Begging	1,372	575	1,338	0.33	4,022	2.93	3.69	4.25	0.37
Marriage	1,675	671	1,036	0.3	3,487	2.08	3.94	6.34	0.33
Comm. Offences	196	144	164	0.44	377	1.92	4.37	6.74	0.29
Organs	2,116	1,725	2,381	0.49	4,832	2.28	4.57	7.14	0.36

Although sexual and labour exploitation have the highest volume of active actors during both periods, the number of actors tweeting increased in 2016-2020 for all exploitation types. Most actors are mentioned in relation to sexual exploitation, especially in the period from 2016-2020. This suggests a greater inquiry

to other actors in the tweets posted, recognising them as relevant in the debate. Other exploitation types such as the forced removal of organs also have a higher average number of mentions than the rest.

On the other hand, there is also an increase in the dissemination indicators referring to the number and average number of tweets and the average number of retweets, favourites and replies for all exploitation types during the period from 2016-2020. However, while the average number of favourites is higher in the case of labour exploitation than sexual exploitation, there is a greater increase in the number of replies in the case of sexual exploitation. This might suggest that while actors are directly aligned with other tweets on labour exploitation without introducing their convergent or divergent opinion, activity on sexual exploitation seems to elicit greater participation and expression of actors' own opinions on third-party posts.

The emphasis on sexual and labour exploitation is also reflected in the number of actors talking about one or several exploitation types (Figure 3), considering that at least 20% of the tweets made by an actor must be about an exploitation type to be assigned to that group. The highest volume of actors corresponds to those who talk only about sexual exploitation (28,707), followed by labour exploitation (4,310), those who talk about both sexual and labour exploitation (3,344) and those who talk about forced organ removal (1,324 actors).



The actors posting the highest volume of tweets for each exploitation type vary for the periods between 2011-2015 and 2016-2020 (Table 2MC). During the period from 2011-2015, the activity of actors such as social organisations, academics and activists stands out, as is the case for sexual and labour exploitation. On the other hand, especially in the case of exploitation for forced organ removal but also exploitation for forced marriage, actors related to the criminal prosecution of human trafficking stand out, such as the police, the Ministry of the Interior and an international network to counter organised crime. During the period from 2016 to 2020, which had a substantial increase in activity, actors related to the criminal prosecution of human trafficking generally gained relevance. The presence of FGRs, Public Ministries, Ministries of the Interior, their advisors and police forces (all from Latin American countries) stands out. Although less frequently represented, the media and some social organisations are also present. In the case of sexual exploitation, for example, the most active actor was a social organisation.

The most frequently mentioned actors also vary according to the exploitation type and period (Table 3MC). During the period from 2011-2015, some of the most frequently mentioned actors are the media and audiovisual products (e.g. films), anti-trafficking campaigns, initiatives or institutional bodies, politicians and some social organisations. International organisations such as UNODC and UNICEF also appear, albeit discreetly. It is precisely during the period between 2016-2020 that these international organisations become particularly relevant. For almost all exploitation types, the UNODC division in Mexico, the

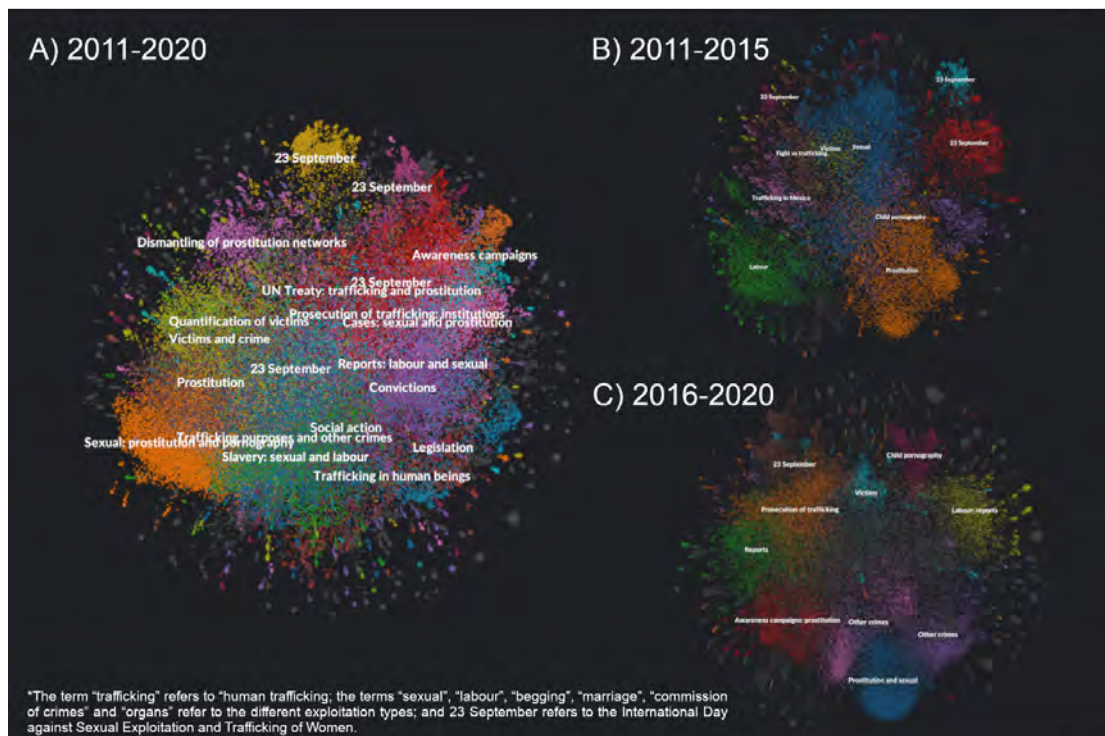


UNODC itself as an umbrella organisation, and the United Nations stand out. Other bodies involved in the criminal prosecution of human trafficking such as the Guardia Civil and the FGRs are also noteworthy. In addition, personalities such as Pope Francis, López Obrador (President of Mexico), the documentary filmmaker Mabel Lozano, and institutions such as the CNDH stand out.

### 3.3. Identifying topic communities in the tweet network

The main communities of topics identified in the general network of tweets on trafficking and exploitation purposes during the period from 2011-2020 and the sub-periods of 2011-2015 and 2016-2020 (Figure 4) show that sexual exploitation is the purpose with the highest representation among the topics addressed.

Figure 4. Semantic networks of tweets about trafficking in: A) 2011-2020, B) 2011-2015 and C) 2016-2020



Note. The term 'trafficking' refers to 'human trafficking'; the terms 'sexual', 'labour', 'begging', 'marriage', 'commission of crimes' and 'organs' refer to the different exploitation types; and 23 September refers to the International Day against Sexual Exploitation and Trafficking of Women.

The network for the entire period (2011-2020) includes a total of 90,328 nodes, each representing one tweet. 112 communities were detected, of which the 20 main ones were analysed (77.74% of the nodes are included in them). This network shows that the community with the highest volume of tweets (8% of the total) compares human trafficking with other crimes, making references to South, Central and North American countries. The communities relating to sexual exploitation, prostitution and pornography (7%), and prostitution (6%) are contiguous within the network. The International Day against Sexual Exploitation and Trafficking of Women is also well represented (7%, 3%, 3% and 2%). In addition, there are different communities on the criminal prosecution of the crime of human trafficking (4%, 3% and 2%); cases and convictions for sexual exploitation and prostitution (4% and 3%); social action against the crime of trafficking (3% and 2%), and legislation to combat this problem (1% and 1%).

An analysis of the evolution of the network between the periods 2011-2015 (25,515 nodes and 23 communities) and 2016-2020 (64,813 nodes and 84 communities), with 10 communities identified in each (94.99% and 62.6% of nodes, respectively), shows that most of them deal with sexual exploitation and prostitution. During 2011-2015 they represent the two largest communities (25% and 22%, respectively),



although during 2016-2020, when the network fragments, these topics are presented transversally within the rest of the communities (Table 4MC). Prostitution, which seems to gain relevance, has an exclusive community (10%). Also noteworthy is the increase in references to criminal prosecution of the crime of human trafficking, different cases and their reporting, representing three of the 10 communities from 2016-2020 (21% of the nodes). During 2011-2015 there was also a large community on labour exploitation (16%), which decreased from 2016-2020 (6%), although two communities appear on other exploitation purposes and other crimes (6% and 6%).

The topic networks of the tweets were also analysed (2011-2020) for the two exploitation purposes most frequently present in the analysis: sexual (Figure 1MC) and labour (Figure 2MC). As it is the main purpose, the top 20 communities in the sexual exploitation network were identified. As can be seen, it is very similar to the general network of tweets where all exploitation types are included. On the other hand, and as in this general network, most of the communities in the network of tweets on labour exploitation, with 10 communities identified, also include references to sexual exploitation (Table 4MC).

The network on sexual exploitation focuses on topics such as prostitution and sexual exploitation (communities with 11% and 7% of tweets), together with references to pornography and women; other organised crime offences (7%) where the word 'prostitution' stands out, forming a community located in the centre of the network and close to the two previous ones, as well as that on sexual exploitation as a crime, especially child sexual exploitation (6%); criminal prosecution of sexual exploitation and human trafficking, reports of offences, convictions, and dismantling of networks (5%, 4%, 4%, 3% and 2%), some of which are located in the centre of the network; and the International Day against Sexual Exploitation and Trafficking of Women (5%, 3% and 3%).

In the network on labour exploitation, the majority community on the criminal prosecution of the crime of human trafficking and the institutions and agencies involved stands out (9% of tweets). Also prominent are other communities on labour exploitation which highlight its forced nature or slavery (6% and 8%); the victims of this exploitation (the previous community with 8% and another with 8%); different purposes of exploitation (7% and 5%); human trafficking in America (4%); and, finally, the quantification of people experiencing this phenomenon (4%), forming a larger community than in the general tweet network, although it is impossible to clarify whether it refers to victims of human trafficking, labour exploitation or both.

#### 4. Discussion and conclusions

The process established for data collection and analysis allowed us to explore a set of 101,051 tweets (90,328 unique tweets) that were recovered and classified according to six different exploitation purposes linked to human trafficking: sexual, labour, begging, forced marriage, commission of crimes and forced removal of organs. Through this process, we have analysed the evolution of the presence of these different types of exploitation in Twitter activity over time; the volume of actors posting about each of these exploitation types; and the topics present in the general network of tweets and in the specific networks on two of the most widespread types of exploitation, sexual exploitation and labour exploitation. Thus, the research conducted used substantially more data and is far removed from other more mainstream approaches to studies of human trafficking and Twitter, such as victim recruitment (Burbano & Hernández-Álvarez, 2017; Granizo et al., 2020; Hernández-Álvarez & Granizo, 2021; Tundis et al., 2019).

However, this process of data collection and analysis has a number of limitations: not all the tweets on each exploitation type were retrieved because the process for inclusion and exclusion of terms meant that not all possible words associated with each exploitation type were used; the category 'other exploitation purposes' (e.g. armed conflict) was not created in order to avoid making the presentation of data more complex; and the difficulties encountered to differentiate the geographical scope of the actors did not allow specific analyses by country or region. Despite these limitations, the research carried out yielded unique results that allow us to characterise Twitter activity on human trafficking and exploitation in Spanish, with certain outstanding aspects and trends that are related to the objectives of this study: the prominence of sexual exploitation over other types of exploitation in this social network's activity; the importance of international actors and organisations in the capitalisation of attention; the current focus on the criminal

prosecution of the crime and the reporting of cases of human trafficking; and the link between sexual exploitation and prostitution in the networks of topics in the tweets.

The high percentage of tweets that highlight human trafficking for sexual exploitation either exclusively or overlapping with other purposes, the high volume of active or mentioned actors for this exploitation type, the high number of replies to other actors' tweets, the prominence of sexual exploitation in the topics of the tweet network and the incidence of the International Day against Sexual Exploitation and Trafficking of Women in the highest peaks of activity and in that network are all examples of the overrepresentation of this purpose in the overall visibility of the phenomenon of human trafficking and the exploitation with which it is associated. In this sense, the analysis conducted yields similar conclusions to analyses of the representation of human trafficking in other areas of communication such as the press (Marchionni, 2012) and academic publications (Sweileh, 2018).

This high presence and the level of overlap and influence in Twitter network topics on other types of exploitation also points to the fact that the overrepresentation of sexual exploitation hinders the visibility of other exploitation purposes, as other media studies have pointed out (Ruiz-Herrera et al., 2018). This is despite the fact that the gap between those experiencing sexual and labour exploitation appears to have narrowed considerably according to the latest UNODC reports (2018; 2020), with figures of 59% sexual and 34% labour in 2018 and 50% and 38% respectively in 2020. In the future, it might be possible to discern whether this could be related to the fact that the Twitter debate could be carried out in terms and dimensions that are far removed from the real volume of the different exploitation purposes; whether the characteristics of this social debate could ultimately be related to the difficulties in identifying victims; whether the volume of people identified as affected has changed due to the construction of the statistics themselves; or whether the social debate will change in the face of these new data.

Precisely, the increasing prominence of actors related to the criminal prosecution of the crime of trafficking, whether international organisations such as the UNODC or other actors such as the Mexican regional FGRs, the impact on networks of the campaigns deployed by some of these actors, the increase in the number of communities of topics in the networks of tweets related to the criminal prosecution of the crime and the development of criminal measures against the crime of trafficking in different countries (e.g. General Congress of the United Mexican States, 2012) give cause for reflection regarding the impact of the interpretation of the phenomenon of human trafficking and exploitation mainly as a criminal issue. The results do not show a prominent presence of other alternative approaches, suggesting a displacement of approaches such as human rights or broader macro-structural approaches that relate human trafficking to migration policy, for example (Mendel & Sharapov, 2016; O'Connell-Davidson, 2006).

In addition to the hypervisibility of the sexual exploitation purpose, the activity on Twitter and the interpretation of the phenomenon of human trafficking as a crime, there is also a recurring connection between sexual exploitation and prostitution as reflected in the main communities of the tweet networks, whereby it is sometimes linked to notions such as slavery. This is also reflected in the prominence of the International Day against Sexual Exploitation and Trafficking of Women, and the increased exposure of this exploitation purpose during the Days for the Abolition of Slavery. The relationship between prostitution and sexual exploitation and its link to slavery reinforces the criticism of the current concept of human trafficking as being influenced by policies to combat white slavery which assimilated the phenomena of prostitution and slavery, conjuring an image of the trafficked person as a 'white sex slave' that occupies centre stage in representations related to the phenomenon (Doezema, 2010; Kempadoo, 2015). This is related to what happens in other areas, such as human trafficking awareness campaigns, where the hypervisibility of women's violated bodies seems to promote the image of a trafficking victim as a sex slave (Saiz-Echezarreta et al., 2018).

Furthermore, the results and the construction of this image invite reflection on the risk of overlooking or criminalising part of the population that does not fit this dominant image, especially those located in migratory contexts, once again displacing other approaches such as the human rights approach (Andrijasevic & Anderson, 2009; Gregorio & Ras, 2018; O'Connell-Davidson, 2006; Wolken, 2006). In fact, the appearance on Twitter networks of certain topics relating to migratory contexts suggests a need for further research on the relationship between human trafficking, migration, humanitarian crises and the

consequences that the emphasis on the criminal prosecution of the crime of human trafficking in these contexts may have.

The dominant presence of sexual exploitation, heightening of its visibility as a crime, its link to prostitution in Twitter activity (and, to a lesser extent, to migratory contexts) as well as new data from the UNODC (2020) all pose an urgent need to rethink and diversify approaches for the interpretation and visibility of human trafficking for other exploitation purposes. This is necessary for the design and deployment of effective eradication measures and to meet the needs of a greater proportion of the population suffering human trafficking. The implications of this study for education purposes may also include diversification and promotion of alternative approaches to human trafficking in the training of professionals and decision-makers in the socio-educational sector and also in awareness-raising campaigns for the general public, including through social media such as Twitter.

More research is needed on human trafficking and the dissemination of certain discourses and issues on social media such as Twitter where a wide range of actors is involved. The results of this initial descriptive study, which is useful for mapping actors and issues in areas that have been subject to limited exploration in social media (Hodge et al., 2020), may be used as a basis for further analysis of the specific content or ideas disseminated by some of the actors identified as most relevant in this research, the propagation of the tweets posted and reflection on other terms present in the Palermo Protocol, such as 'victim-aggressor'.

### Authors' Contributions

Idea: A.S.R., D.B.H.; Literature review: A.S.R., D.B.H.; Methodology: W.A.M., A.S.R.; Data analysis: A.S.R.; W.A.M., D.B.H.; Results: A.S.R., W.A.M., D.B.H.; Discussion and conclusions: A.S.R., D.B.H.; Writing (original draft), A.S.R., W.A.M., D.B.H.; Final revisions, D.B.H.; Project design and sponsorship, A.S.R., W.A.M., D.B.H.

### Notes

<sup>1</sup> <https://www.graphext.com/tractor>

### Funding Agency

This study received financial support under the project 'Influencia - Científicos socialmente influyentes: un modelo para medir la transferencia del conocimiento en la sociedad digital' (PID2019-109127RB-I00 / AEI / 10.13039/501100011033), funded by the Spanish Ministry of Science and Innovation; the Department of Teaching Studies of the University of Granada; and the research group 'Educación Transformadora para una Sociedad Global y Digital' (EduTransforma-T) of the University of Extremadura (SEJ054). Wenceslao Arroyo-Machado is the holder of an FPU grant (FPU18/05835) from the Spanish Ministry of Universities.

### References

- Albright, E., & Adamo, K. (2017). The media and human trafficking: A discussion and critique of the dominant narrative. In M. Chisolm-Straker, & H. Stoklosa (Eds.), *Human trafficking is a public health issue: A paradigm expansion in the United States* (pp. 363-378). Springer. [https://doi.org/10.1007/978-3-319-47824-1\\_21](https://doi.org/10.1007/978-3-319-47824-1_21)
- Andrijasevic, R., & Anderson, B. (2009). Anti-trafficking campaigns: Decent? Honest? Truthful? *Feminist Review*, 92(1), 151-155. <https://doi.org/10.1057/fr.2009.1>
- Arena Pública (Ed.) (2017). *México, 'foco rojo' en tráfico de órganos, niños migrantes principal blanco: Unicef*. Arena Pública. <https://bit.ly/3mYe9zM>
- Attwood, R. (2021). A very un-English predicament: 'The White Slave Traffic' and the construction of national identity in the suffragist and socialist movements' coverage of the 1912 Criminal Law Amendment Bill. *National Identities*, (pp. 1-30). <https://doi.org/10.1080/14608944.2021.1895096>
- Austin, R., & Farrell, A. (2017). Human trafficking and the media in the United States. *Oxford Research Encyclopedia of Criminology*. <https://doi.org/10.1093/acrefore/9780190264079.013.290>
- Barroso-Hurtado, D., Arroyo-Machado, W., & Torres-Salinas, D. (2021). Formación Profesional Dual: Evolución de red de actores en Twitter. *Educación XXI*, 1, 209-230. <https://doi.org/10.5944/educXXI.28136>
- Bernstein, E. (2018). *Brokered subjects sex, trafficking, and the politics of freedom*. The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226573809.001.0001>
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), 10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Burbano, D., & Hernandez-Alvarez, M. (2017). Identifying human trafficking patterns online. In *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ETCM.2017.8247461>
- Clemente, M. (2017). Human trafficking in Portugal: An ethnography of research and data. *Studi Emigrazione*, 208, 663-687. <https://bit.ly/3plyKrO>

- Comisión Nacional de los Derechos Humanos (Ed.) (2019). *Diagnóstico sobre la situación de la trata de personas en México 2019. Resumen ejecutivo*. Quinta Visitaduría General/Programa contra la Trata de Personas. <https://doi.org/10.5377/cuadernojurypol.v5i13.11134>
- Congreso General de los Estados Unidos Mexicanos (Ed.) (2012). *Ley General para prevenir, sancionar y erradicar los delitos en materia de trata de personas y para la protección y asistencia a las víctimas de estos delitos*. Primera Sección: Diario Oficial de la Federación, de 14 de junio de 2012, Primera Sección, 2-32. <https://bit.ly/3kJW1XC>
- Couto, D., Machado, C., Martins, C., & Abrunhosa, R.A. (2012). A construção mediática do tráfico de seres humanos na imprensa escrita portuguesa. *Análise Psicológica*, 30, 231-246. <https://doi.org/10.14417/ap.551>
- Denton, E. (2010). International news coverage of human trafficking arrests and prosecutions: A content analysis. *Women & Criminal Justice*, 20(1-2), 10-16. <https://doi.org/10.1080/08974451003641321>
- Doezema, J. (2010). *Sex slaves and discourse masters: The construction of trafficking*. Bloomsbury Publishing. <https://doi.org/10.5040/9781350222502>
- Edwards, A. (2007). Traffic in human beings: At the intersection of criminal justice, human rights, asylum/migration and labor. *Denver Journal of International Law and Policy*, 36, 9-53. <https://bit.ly/31FrPqV>
- Gerassi, L. (2015). A heated debate: Theoretical perspectives of sexual exploitation and sex work. *Journal of sociology and social welfare*, 42(4), 79-100. <https://bit.ly/3EFb0Lv>
- Gong, R. (2015). Indignation, inspiration, and interaction on the Internet: Emotion work online in the anti-human trafficking movement. *Journal of Technology in Human Services*, 33(1), 87-103. <https://doi.org/10.1080/15228835.2014.998988>
- González, E. (2019). Apropiación del término 'puta' desde el activismo por los derechos en el trabajo sexual. *Gazeta de Antropología*, 35(1). <https://bit.ly/3dBIUEY>
- Granizo, S., Valdivieso, A., Barona, L., & Hernández-Álvarez, M. (2020). Detection of possible illicit messages using natural language processing and computer vision on Twitter and linked website. *IEEE Access*, 8, 44534-44546. <https://doi.org/10.1109/ACCESS.2020.2976530>
- Gregoriou, C., & Ras, I.A. (2018). Representations of transnational human trafficking: A critical review: Present-day news media, true crime, and fiction. In *Representations of transnational human trafficking* (pp. 1-24). Palgrave MacMillan. [https://doi.org/10.1007/978-3-319-78214-0\\_1](https://doi.org/10.1007/978-3-319-78214-0_1)
- Haunschild, R., Leydesdorff, L., Bornmann, L., Hellsten, I., & Marx, W. (2019). Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags. *Journal of Informetrics*, 13, 695-707. <https://doi.org/10.1016/j.joi.2019.03.008>
- Hebert, L.A. (2016). Always victimizers, never victims: Engaging men and boys in human trafficking scholarship. *Journal of Human Trafficking*, 2(4), 281-296. <https://doi.org/10.1080/23322705.2016.1144165>
- Hernández-Álvarez, M., & Granizo, S.L. (2021). Detection of human trafficking ads in twitter using natural language processing and image processing. In *Advances in artificial intelligence, software and systems engineering. Proceedings of the AHFE 2020 Virtual Conferences on Software and Systems Engineering, and Artificial Intelligence and Social-Computing* (pp. 77-83). Springer. [https://doi.org/10.1007/978-3-030-51328-3\\_12](https://doi.org/10.1007/978-3-030-51328-3_12)
- Hodge, E., Childs, J., & Au, W. (2020). Power, brokers, and agendas: New directions for the use of social network analysis in education policy. *Education Policy Analysis Archives*, 28, 1-17. <https://doi.org/10.14507/epaa.28.5874>
- Kempadoo, K. (2015). The modern-day white (wo)man's burden: Trends in anti-trafficking and anti-slavery campaigns. *Journal of Human Trafficking*, 1(1). <https://doi.org/10.1080/23322705.2015.1006120>
- Marcelo, C., & Marcelo, P. (2021). Educational influencers on Twitter. Analysis of hashtags and relationship structure. [Influencers educativos en Twitter. Análisis de hashtags y estructura relacional]. *Comunicar*, 68, 73-83. <https://doi.org/10.3916/C68-2021-06>
- Marchionni, D.M. (2012). International human trafficking: An agenda-building analysis of the US and British press. *International Communication Gazette*, 74, 145-158. <https://doi.org/10.1177/1748048511432600>
- Mendel, J., & Sharapov, K. (2016). Human trafficking and online networks: Policy, analysis, and ignorance. *Antipode*, 48(3), 665-684. <https://doi.org/10.1111/anti.12213>
- Meneses-Falcón, C., & Urío, S. (2021). La trata con fines de explotación sexual en España: ¿Se ajustan las estimaciones a la realidad. *Revista Española de Investigaciones Sociológicas*, 174, 89-108. <https://doi.org/10.5477/cis/reis.174.89>
- Naciones Unidas (Ed.) (2000). *Protocolo de Naciones Unidas, para prevenir, reprimir y sancionar la trata de personas, especialmente mujeres y niños*. Naciones Unidas. <https://bit.ly/31KS3sf>
- O'Connell-Davidson, J. (2006). Will the real sex slave please stand up? *Feminist Review*, 83(1), 4-22. <https://doi.org/10.1057/palgrave.fr.9400278>
- Papadoukaa, M.E., Evangelopoulos, N., & Ignatow, G. (2016). Agenda setting and active audiences in online coverage of human trafficking. *Information, Communication & Society*, 19, 655-672. <https://doi.org/10.1080/1369118X.2016.1139615>
- Reina-Peñas, M. (2019). Análisis del fenómeno de la trata de personas desde la perspectiva de los derechos humanos. *Icade. Revista de la Facultad de Derecho*, 107. <https://doi.org/10.14422/icade.i107.y2019.006>
- Rodríguez-López, S. (2018). De)constructing stereotypes: Media representations, social perceptions, and legal responses to human trafficking. *Journal of Human Trafficking*, 4(1), 61-72. <https://doi.org/10.1080/23322705.2018.1423447>
- Rodríguez-López, S. (2020). The invisibility of labour trafficking in Spain. A critical analysis of cases and policies. *Revista Española de Investigación Criminológica*, 18(2), 1-25. <https://doi.org/10.46381/reic.v18i2.329>
- Ruiz-Herrera, A., Ruiz-Guevara, S., & López-Cantero, E. (2018). El papel de los medios de comunicación masiva en la comprensión del fenómeno de la trata de personas. *Revista Criminología*, 60(2), 25-39. <https://bit.ly/33jyA2F>
- Saiz-Echezarreta, V., Alvarado, M.C., & Gómez-Lorenzini, P. (2018). Advocacy of trafficking campaigns: A controversy story. [Incidencia política de las campañas contra la trata: Un relato controvertido]. *Comunicar*, 55, 29-38.

- <https://doi.org/10.3916/C55-2018-03>
- Sanford, R., Martínez, D.E., & Weitzer, R. (2016). Framing human trafficking: A content analysis of recent US newspaper articles. *Journal of Human trafficking*, 2(2), 139-155. <https://doi.org/10.1080/23322705.2015.1107341>
- Sobel, M.R. (2014). Chronicling a crisis: Media framing of human trafficking in India, Thailand, and the USA. *Asian Journal of Communication*, 24(4), 315-332. <https://doi.org/10.1080/01292986.2014.903425>
- Sweileh, W.M. (2018). Research trends on human trafficking: A bibliometric analysis using Scopus database. *Globalization and Health*, 14(106), 1-12. <https://doi.org/10.1186/s12992-018-0427-9>
- Tundis, A., Jain, A., Bhatia, G., & Muhlhauser, M. (2019). Similarity analysis of criminals on social networks: An example on Twitter. In *2019 28th International Conference on Computer Communication and Networks (ICCCN)* (pp. 1-9). <https://doi.org/10.1109/icccn.2019.8847028>
- United Nations Office on Drugs and Crime (Ed.) (2018). *Global Report on Trafficking in Persons 2018*. United Nations. <https://bit.ly/2YfbOG0>
- United Nations Office on Drugs and Crime (Ed.) (2020). *Global report on trafficking in persons 2020*. United Nations. <https://bit.ly/3DHTkx>
- Van-Schalkwyk, F., Dudek, J., & Costas, R. (2020). Communities of shared interests and cognitive bridges: The case of the anti-vaccination movement on Twitter. *Scientometrics*, 125, 1499-1516. <https://doi.org/10.1007/s11192-020-03551-0>
- Volken, C.L. (2006). Feminist legal theory and human trafficking in the United States: Towards a new framework. *U. MD. L.J. Race, Religion, Gender & Class*, 6(2), 407-438. <https://bit.ly/3rVW3597>
- Zhang, S.X. (2012). Measuring labor trafficking: A research note. *Crime, Law and Social Change*, 58, 469-482. <https://doi.org/10.1007/s10611-012-9393-y>



# Comunicar

www.comunicarjournal.com





# Special Education Teacher's professional development through digital storytelling

## Desarrollo profesional de maestros de educación especial a través de la narración digital

- Id** Ozgur Yasar-Akyar. Research Assistant, Dep. of Physical Education and Sports Teaching, Hacettepe University, Ankara (Turkey) (ozguryasar@hacettepe.edu.tr) (<https://orcid.org/0000-0002-9658-8392>)
- Id** Cinthia Rosa-Feliz. Research Assis., Dep. of Science, Tech., and Innov., Federico Henríquez y Carvajal Uni., S. Domingo (Dominican Rep.) (cinthia.delarosa@ufhec.edu.do) (<https://orcid.org/0000-0002-9696-4132>)
- Id** Dr. Solomon Sunday-Oyelere. Assis. Professor, Dep. of Comp. Science, Electrical and Space Engineering, Luleå Uni. of Technology (Sweden) (solomon.oyelere@ltu.se) (<https://orcid.org/0000-0001-9895-6796>)
- Id** Dr. Darwin Muñoz. Professor, Dep. of Science, Tech., and Innovation, Federico Henríquez y Carvajal Uni., S. Domingo (Dominican Republic) (dmunoz@ufhec.edu.do) (<https://orcid.org/0000-0003-4691-2614>)
- Id** Dr. Gıyasettin Demirhan. Professor, Dep. of Physical Education and Sports Teaching, Hacettepe University, Ankara (Turkey) (demirhan@hacettepe.edu.tr) (<https://orcid.org/0000-0002-5370-2036>)

### ABSTRACT

This research presents the results of an exploration of special education teachers' understanding of how their participation in workshop-based digital storytelling (DST) would enhance their professional development concerning inclusive education. This study evaluates the usability of the Smart Ecosystem for Learning and Inclusion (SELI) platform for supporting teachers during the workshop-based digital storytelling process. We used a convergent parallel mixed-method research design approach with 47 secondary school teachers working with disabled people in the Dominican Republic. The results of this study indicated that the SELI smart learning platform had shown good usability in supporting teachers during the workshop-based digital storytelling pedagogical process. Besides, two themes emerge regarding how workshop-based digital storytelling can contribute to teacher professional development for promoting inclusive education. The resulting themes are expressing, listening, and learning through digital storytelling; and driving change with digital storytelling to create more inclusive environments. Teachers who participated in the interviews were optimistic about DST implementation. They expressed that the workshop worked for multiple ways of expression, listening from and connecting with other stories, and learning through DST. Teachers could reflect their idea about using DST in terms of its potential impact on inclusion in the classrooms for driving change, building meaningful learning, and influential practice when used in the classroom.

### RESUMEN

Este artículo presenta los resultados de un estudio exploratorio de la forma en que los maestros de educación especial comprenden que las narrativas digitales, basadas en talleres, mejorarían el desarrollo profesional en relación con la educación inclusiva. El estudio evalúa la usabilidad de la plataforma Smart Ecosystem for Learning and Inclusion (SELI) para apoyar a los docentes durante el proceso de narración digital basada en talleres. Utilizamos un enfoque de diseño de investigación de método mixto paralelo convergente con 47 profesores de secundaria que trabajan con personas discapacitadas en la República Dominicana. Los resultados de este estudio indicaron que la plataforma de aprendizaje inteligente SELI había demostrado una buena usabilidad para apoyar a los docentes durante el proceso pedagógico de la narración digital basada en talleres. Además, surgen dos temas sobre cómo la narración digital basada en talleres puede contribuir al desarrollo profesional de los docentes para promover la educación inclusiva. Los temas resultantes son la expresión, la escucha y el aprendizaje a través de la narración digital; e impulsar el cambio con la narración digital para crear entornos más inclusivos. Los maestros fueron optimistas respecto a la implementación de narrativas digitales. Los maestros reflexionaron en términos del potencial impacto de las narrativas digitales en inclusión en el aula, como promotoras del cambio, construyendo aprendizaje significativo y promoviendo una práctica influyente.

### KEYWORDS | PALABRAS CLAVE

Professional development, teacher training, ICT, inclusive education, digital storytelling, special education. Desarrollo profesional, formación del profesorado, TIC, educación inclusiva, Narración digital, educación especial.



## 1. Introduction

There has been a long discussion about inclusive education regarding the need to change the assumptions, systems, and procedures in schooling with new ways of thinking and working to support all learners rather than "most and some" (Florian, 2007). A challenging question is how teachers might best be prepared to work better with an increasingly diverse student population. Therefore, continuing professional development opportunities for teachers with a focus on inclusive education is essential. Moreover, innovative training opportunities for mid-career teachers through continuing professional development (CPD) may play a vital role in promoting pro-inclusion changes in education systems (Bařková & Closs, 2013). Technology-enhanced learning can cover the need for such innovative ways of promoting inclusive education.

Using ICTs for disadvantaged groups to achieve technology-enhanced learning in inclusive education has been an interest of scholars. For example, researchers suggest using ICT-based learning technologies for students with disabilities to support and enable them to access learning (Hersh, 2017; Hersh & Mouroutsou, 2019; Sánchez-Serrano et al., 2020). Moreover, the research highlighted the need for teacher training on ICT use to support students with disabilities (Fernández-Batanero et al., 2019). This study examines the use of digital storytelling (DST) within a smart learning platform for teacher professional development in inclusive education.

In addition, this study reports the outcome of one aspect of the recently finalized international research project called SELI (Smart Ecosystem for Learning and Inclusion). The SELI project aimed to identify the challenges in using ICTs as tools for learning and inclusion and initiating a broad stakeholder dialogue and consultation to screen potential educationally, technical, and business solutions for the challenges (Imek et al., 2021). In this scope, one of the main aspects of the project is the implementation of workshop-based DST for enhancing teacher training through integrating innovative pedagogical practices of using ICTs for inclusive education (Akyar et al., 2020; Tomczyk et al., 2020).

Hwang (2014) highlighted that knowing more about learners' learning performances and perceptions is valuable research for developing more effective smart learning environments. Moreover, a recent study reported a need to investigate the views and perceptions of teachers on smart learning technologies (Li & Wong, 2021). Based on this previous research, the present study advances the field by exploring how special education teachers' participation in workshop-based DST through the innovative learning platform supports professional development in inclusive education in the Dominican Republic. We acknowledge the need to know more about the technologies used to support teachers' continuing professional development based on teachers' actual experiences in the context of inclusive education. In line with this need, we adopted a mixed-methods approach to understanding teachers' perceptions better. In particular, the study focused on the following research questions (RQs):

- RQ 1: To what extent do teachers find the SELI smart learning platform usable?
- RQ 2: How can workshop-based DST be used as an educational strategy for teacher professional development in inclusive education?

Although DST for teacher education in the context of inclusive education can provide a creative showcase for teachers, the literature is scarce regarding delivering continuing professional development of teachers to enhance teachers' use of ICT for inclusive education. Mainly there is a lack of research on understanding the impact of DST to support continuing professional development of teachers based on the actual experiences of teachers in the context of inclusive education. This article aims to address the gap in the literature by exploring special education teachers' understanding of how their participation in workshop-based DST through the SELI smart learning platform supports professional development in the context of inclusive education.

## 2. Materials and methods

### 2.1. Research design

In this study, we used a convergent parallel mixed-method design. According to Creswell and Clark (2011), a convergent parallel mixed-method design entails that the researcher concurrently conducts the quantitative and qualitative elements in the same phase of the research process. We have chosen an

embedded mixed-method design to understand the results of our intervention with teachers using the SELI smart learning platform by incorporating perspectives of participant views within the context of inclusive education. After the workshop, we surveyed for quantitative analysis for usability of the platform and a case study as qualitative research to understand how workshop-based DST can contribute to teacher professional development as an educational strategy for inclusive education.

## 2.2. Participants

A total of 47 special education teachers who volunteered to attend the Dominican Republic workshop between 22-24 January 2020 comprised the participants in the present study. Participants were secondary school teachers working with students with hearing disabilities in special education schools. These teachers consisted of 38 females and nine males in their mid-career (5 to 15 years' experience). In addition, within the qualitative part of the research, three female and one male volunteer participated in the study. Then we had face-to-face semi-structured interviews with a total of four special education teachers.

## 2.3. Data collection tools

We collected quantitative and qualitative data after the workshop through two data collection tools. As a quantitative data collection tool, we administered the USE questionnaire developed by Lund (2001) with the author's permission to understand the general overview usability of the SELI smart learning platform. The USE questionnaire comprises four dimensions; including usefulness, ease of use, ease of learning, and satisfaction, on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) with one (N/A) option. Native Spanish speakers among the authors translated questionnaires based on scale adaptation suggestions (Dantas et al., 2017). The USE questionnaire is a widely used, valid, and reliable tool for assessing self-perceived usability in different cultural settings. Studies that examined psychometric properties of the USE questionnaire reported that it is a valid and reliable instrument (Gao et al., 2018). In addition, we looked at Cronbach's alpha values to test reliability in order to ensure that the questionnaire provided stable and consistent results by using SPSS 21.0. The instrument is reliable as the Cronbach's alpha's cutoff value is at least 0.7 (Landauer, 1997). According to Table 1, all construct items (usefulness, ease of use, ease of learning, and satisfaction) had a value greater than 0.7. Thus, Cronbach's alpha values show good values for all constructs of the USE questionnaire.

Table 1. USE questionnaire reliability examination		
Variables	Cronbach's Alpha	N of Items
Usefulness	.795	8
Ease of Use	.886	11
Ease of Learning	.940	4
Satisfaction	.830	7

The second data collection tool was qualitative and based on the DST process. The research team created it to explore teachers' experiences in the DST workshop process as one of the tools offered by the SELI smart learning platform. The questions aimed at insights of participants based on their experiences in workshop-based DST, including feelings about one's personal story, the process of creation and sharing, DST role for personal/professional change, watching others' stories.

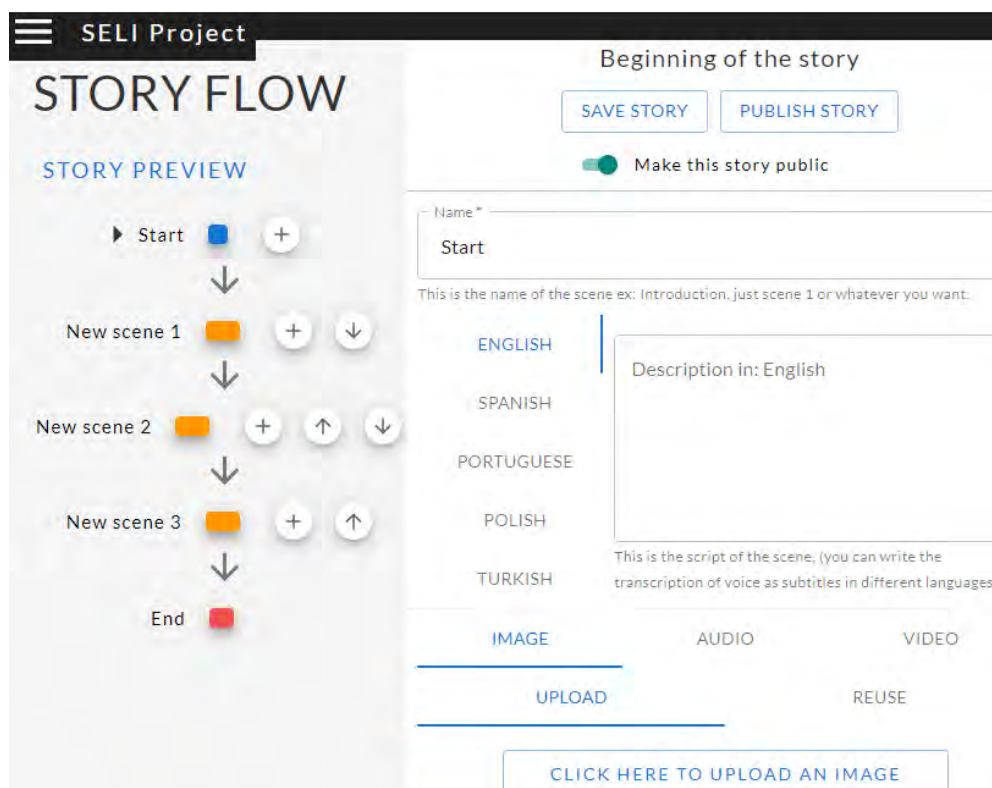
## 2.4. Procedure

We worked with the teachers to understand their experience of creating digital stories through the SELI smart learning platform. A lead facilitator who is a certified DST facilitator joined the workshop online and shared the background regarding workshop-based practices adopted by Lambert and Hassler (2018). Later, we created small groups among teachers, including a co-facilitator in each group. All the groups involved a maximum of 8 teacher participants. During this workshop, teachers first listened to the online facilitator's story projected in the workshop room. Then teachers told their stories face to face in the story circle. After that, teachers continued digital processes through a DST tool in the SELI smart learning platform. The SELI smart learning platform allowed functionalities of using DST as well as accessibility module. Details of the open-source application are reachable in the open-source project platform. Co-facilitators focused on listening to the participants. Co-facilitators lead six phases of the workshop within the



groups. Story circle was the first phase that aimed at establishing trust by telling inclusion-related personal stories in small groups. There were participants with hearing disabilities, however, and a sign language interpreter could express personal stories to them. The second phase was text writing as a dialogic stage where participants wrote down their text and discussed similarities and differences among their accounts. The third phase of image search allowed participants to use their mobile phones for collecting personal images from WhatsApp or other applications. Some others preferred to search the Internet. The co-facilitators considered the ICT skills of participants and provided help when needed. The fourth phase was for recording the audio throughout the SELI smart learning platform. The fifth phase allowed users to merge image and audio through a storyboard on the SELI smart learning platform. Figure 1 is a screenshot of the storyboard used in the SELI smart learning platform.

Figure 1. View of storyboard in the SELI smart learning platform



The sixth phase was publishing the created stories and sharing final versions of the story, which allowed further discussions before the data collection.

## 2.5. Data collection and analysis

In this study, we analyzed both quantitative and qualitative data after data collection. Firstly, the research team distributed the printed version of the USE questionnaire (Lund, 2001) to teachers at the end of the workshop process for quantitative data collection. Based on the suggestion of Nielsen (1994), we calculated the mean value of each variable to describe the outcome of the usability measurement. The qualitative collection involved participatory observation and semi-structured interviews because of the nature of the qualitative analysis. We carried out an inductive thematic analysis to code the data without fitting it into a pre-existing coding frame explained by Braun and Clarke (2006). We then gathered data in an online document.

As facilitators of the workshop, we collected observational data regarding the implementation of the process and participants' attitudes. At the end of the workshop, we conducted semi-structured interviews.



Open-ended questions of the interview were oriented toward the process of the DST workshop (Lambert & Hessler, 2018). The university ethical commission provided ethical approval, and participants gave informed consent throughout the SELI smart learning platform to participate in the study. We made constant comparisons of data within the research team through online collaborative documents and video conference discussions. Codes with example quotes from data analysis are provided in the Appendix.

### 3. Results

This paper focuses primarily on special education teachers' understanding of how their participation in workshop-based DST through the SELI smart learning platform supports professional development in the context of inclusive education. Besides, the study explores the usability of the SELI smart learning platform and teachers' perceptions of using workshop-based DST for inclusive education. We present our findings according to two research questions based on quantitative and qualitative results.

#### 3.1. Usability of SELI learning platform

The first research question focuses on the usability of the SELI smart learning platform. We identified the usability of the SELI smart learning platform used in the workshop based on the data analysis. The mean scores for usefulness, ease of use, ease of learning, and satisfaction in the current study are 6.61, 6.07, 5.90, 6.40 on a seven-point Likert scale (Table 2). The scores converted from the mean for usefulness, ease of use, ease of learning, and satisfaction are 94.43, 86.71, 84.29, and 91.43, respectively.

<b>Variables</b>	<b>Mean Score</b>	<b>0-100 Score</b>
Usefulness	6.61	94.43
Ease of Use	6.07	86.71
Ease of Learning	5.90	84.29
Satisfaction	6.40	91.43
Average Score	6.25	89.60

Based on these results, the SELI smart learning platform has shown good usability. The high results of usefulness, ease of use, ease of learning, and satisfaction prove that the users accept the SELI smart learning platform.

#### 3.2. Workshop-based DST as a way of enjoyment for driving change for inclusive education

The second research question focuses on the use of workshop-based DST for inclusive education. Data analysis resulted in overarching themes as (a) enjoyment of expressing, listening, and learning through DST and (b) "Driving change" with DST to being "able to create more inclusive environments."

##### 3.2.1. Enjoyment of expressing, listening, and learning through DST

In analyzing the responses from interviews with teachers regarding the workshop-based DST process, findings revealed the contribution of DST when used for the professional development of teachers in the context of inclusive education. All the teachers were optimistic about the DST implementation. More specifically, teachers highlighted three aspects regarding how DST workshops worked for (a) Multiple ways of expression, (b) Listening from and connecting with other stories, and (c) Learning through DST.

Notably, most of the teachers shared their appreciation about the DST process they were able "to express ideas" (P1), "never participated in a workshop like this before" (P2), "to see how to relate what I like and don't like" (P3), to "learn new things and refine them" (P4) when reflecting about the process of workshop-based DST. For instance, P1, who is a mid-career female teacher, explained how DST allows expressing ideas and feelings in multiple ways: "Sometimes it is not easy to understand or express ideas, but in that way, it is possible to understand how to deliver the message you want to achieve. It was interesting because you can express ideas, not necessarily textually. My story was completely conceptual, but working with images and audio enables people to understand the feelings in the story. I liked that other people felt identified with it, accepted the message. ...I wouldn't change the story as it was extremely open to get the message across and get people to identify with the story." (P1)

Not only sharing stories but also listening to and connecting with other stories was also highlighted by teachers. A male teacher stated: "I identified with the story of the little black bean, where the person

felt discriminated because of the color of their skin, and over time you learn that it wasn't a disability but just things of life." (P2) Besides personal stories, work-related stories allowed us to build connections for teachers by listening to and learning from colleagues. A female teacher stated: "I identified myself with the story of inclusion, as this is what I experience every day in my center." (P4) These personal and professional stories made P3 connect with stories and construct her understanding of inclusion: "I like to see how to relate what I like and don't like and to understand the stories of inclusion. I identified with my CONADIS colleagues' stories. I could see how the work environment is part of our daily knowledge. As everyone tells and listens to a story, a person can see the disparity that exists in society." (P3)

Teachers could also learn something related to inclusive education throughout stories by identifying themselves with others' stories as a surprisingly educational experience "how something simple leads you to find great learning." (P1) Teachers participating in the DST workshop connected to each other through diverse stories, which opened various understandings in a comfortable atmosphere. Engagement in telling and listening to stories served to negotiate different points and diverse definitions of inclusion: "Well, it was interesting to hear and see others' perspectives and see how the word inclusion is defined in different ways, which allowed me to add a plus to my previous knowledge. I felt comfortable, and I found it easy." (P3)

Moreover, conversation through stories enhanced professional engagement for mutually extending knowledge as "It helped me to learn new things and refine them. I was able to see the development very well, where good tools were used, and we had excellent facilitators." (P4) Besides storytelling benefits, the digital aspect of DST is rarely mentioned by teachers as P2 shared when discussing the DST workshop process, "Yes, I'd never participated in a workshop like this before. the new technologies for education nourished me."

### 3.2.2. Driving change with DST to create more inclusive environments

As a result of participation in the workshop-based DST, teachers could reflect on their idea about using DST in terms of its potential impact on "inclusion in the classrooms," "driving change," "building meaningful learning," and finding it as "influential" when used in the classroom. Two aspects are highlighted as (a) DST for an inclusive learning process and (b) Increasing awareness through stories. Most of the teachers highlighted DST as providing an inclusive learning process. Participating in DST embodied interest in increasing their professional competencies by becoming aware of knowledge deficits. Trust built during the story circle allowed teachers a space where they could share obligations without feeling deficient, "I learned that it is easy for me to tell a story to promote a learning process for the children" (P3). Ultimately, the DST workshop facilitated a safe environment in which participants got inspired to improve their practices. When asked about the meaning of her story, P4 connected it with topics that must be considered to ensure the quality of teaching and learning. "My story is related to what we do, what we want to do, and to be able to differentiate what is integration with inclusion, to guarantee a quality of teaching and learning. I learned I must handle ICTs better to have them as accessible tools." (P4)

The value of DST for the learning process is emphasized by teachers not only focusing on its digital dimension but also creating an interactive, social, collaborative environment. DST workshop practice provided a space for constructing collective knowledge and professional growth that teachers thought could be transferred to the classroom. The teachers' wish to transfer what they learned in the workshop to the classroom setting was clear when P2 spoke about her commitment to contributing to education: "I have learned that it is necessary to give more from ourselves to continue helping and contributing to education so that the classes are interactive and generate an environment of collective knowledge construction. I understand that (DST) was very good for inclusion in the classrooms, as it allows the classes to be more dynamic and interactive for students with different disabilities. I learned about new technologies that can be applied in the classroom to innovate lessons." (P2)

It is possible to associate the inclusiveness of DST with the diversity of formats, with content provided as both audio and visual. The strength of multimodal communication and the ability to spread content through digital stories made teachers feel empowered to innovate. P1 shared the motivation to drive change with DST: "it (DST) is an audiovisual content, many people with disabilities can learn from this

kind of tool to achieve meaningful learning.... I can now drive change since now I know a new tool, which I can work not only in person, but I can make content to be observed by various people in the world... (I learned) that we have many tools available to transform traditional education processes and introduce them in our campuses where students can integrate and positively see the learning process.” (P1)

Teachers found the workshop-based DST process inclusive. In addition, digital stories as a product are considered helpful in improving inclusion. Most of the teacher’s saw the potential of stories to increase awareness about inclusion. Indeed, there was an unmistakable sense of inclusion in all the stories as the DST workshop built around the context of inclusion. For example, P1 highlighted the need for accepting each other and getting united in diversity when asked about the meaning of her story, “the (meaning of my story) is that we are all special in one way or another, so we must all accept and understand each other to break down barriers for diversity, no matter what special circumstances or conditions exist, it is all a matter of the mind” (P1).

Teachers considered personal examples in stories related to exclusion cases to increase awareness about inclusion. When asked whether she considers changing the story and the meaning of her story, P2 stated: “I wouldn’t change it (my story) since I suffered bullying against my personality many times... and I would like more people to know about my story and become aware... so that people don’t try to bully disabled people. My story reflects how one person’s lack of knowledge about a disability can denigrate another person. It (DST) allows us to use our example as a strategy for learning through DST, and the testimony will serve as a model for others. I enjoyed giving a message from my perspective of how it is the most direct inclusion to help students with disabilities.” (P2)

Similarly, P4 was also happy about her story as it is, since the report highlights the uniqueness of every person and increasing awareness about this to achieve inclusion: “It is very influential because it shows that there are no barriers for students to feel included. ... I would not change it because I like the one I use, and I do not see any problem with it. (My story) ... makes it clear that each person is different, and if we are aware of this, we will create more inclusive environments. ...I must be open to change, and acknowledge that the process and handling of ICTs are important, to use them with people who have disabilities, and I must handle ICTs better.” (P4)

However, keeping the story as it is being not always the case because there are multiple stories one might tell in the context of inclusion, “Yes, I would change it because there are many stories to tell and much global information” (P3). Diversity of stories can also be considered an essential characteristic that allows participants to increase awareness about inclusive education. While these teachers realized the DST workshop process and product as a story, significant commitment shaped their motivation to “drive change” with DST to be “able to create more inclusive environments.”

## 4. Discussion and conclusions

### 4.1. Discussion

Acknowledging short-term educational workshops is an essential part of continuing professional development. Our study explored special education teachers’ perceptions about implementing workshop-based DST as an educational strategy through the SELI project. The study provides significant findings regarding the usability of the SELI smart learning platform and an in-depth understanding of the workshop in the context of inclusive education. Teachers in this study evaluated the usability of the SELI smart learning platform well in all dimensions of the USE Questionnaire. Although this workshop included only limited use of the SELI smart learning platform, involving DST in a short period that lacks extended user experience, these results are significant as the first usability results of the SELI smart learning platform. Some other studies that have assessed the usability of educational systems reported similar good usability results (Faria et al., 2016; Filippidis & Tsoukalas, 2009; Lattie et al., 2017; Hariyanto et al., 2020).

Teachers valued workshop-based DST to allow enjoyment for expressing, listening, and learning. This finding is similar to the result of the research conducted by Park (2019), which aimed to identify experiences of graduate students in a program for Teaching English to Speakers of Other Languages (TESOL) when completing DST. DST facilitated the participants’ dialogic hybrid learning, which Park (2019) defines as learning via traditional and technologically assisted methods with internal and external dialogic interactions.

Participants explained dialogic interaction as "I like to see how to relate what I like and don't like and to understand the stories of inclusion. ...I identified with the stories of my colleagues" (P3) and "how something simple leads you to find great learning" (P1). Similarly, the results of another short teacher professional development study offer optimistic insights regarding the effectiveness of dialogic teaching and learning practice (Rapanta et al., 2021). Therefore, dialogue towards DST can foster an inclusive dialogic stance among teachers.

Participants enjoyed sharing their stories, and they felt motivated and empowered when listening to the stories of others. This finding is similar to the result of another study that aimed to examine the effect of the DST process on the digital literacy skills of pre-service teachers and examine the creation process in detail (Çetin, 2021). Çetin (2021) reported that pre-service teachers found digital story creation makes a course fun and attractive as it increases participation, motivation, interaction and provides better learning and permanence. Differently from previous studies that had an emphasis on the improvement of digital literacy skills (Ranieri & Bruni, 2018; Çetin, 2021) through DST, in this study, teachers perceived DST as a way of using "new technologies for education" (P2) to "have in hand and thus be able to adapt it to the context of the students" (P8). The findings of the present study are in line with the studies reported use of DST in inclusive methodologies as an instrument of constructive and experiential learning based on the diversity of each student (Kouvara et al., 2019) and a way of starting up different learning channels (Albano & Iacono, 2019).

Moreover, in this study, we implemented DST with teachers who work with students with hearing disabilities. Therefore, inclusive education was the main topic of the DST workshop, which aimed to provide learning opportunities through teachers' stories. Similarly, DST is also applied in a different context of students as a strategy for children's and adolescents' self-expression, particularly immersed in unequal and ethnically diverse educational contexts (Valdivia, 2017). Although the Covid 19 circumstances revealed inequalities in access to education caused by access to technology and online delivery, where teaching approaches may not always address the student voice with an appreciation of their culture (Istemi, 2021), DST can be a solution for teachers learning to make sense of inclusive education. Apart from this technical or technological aspect, implementing workshop-based DST facilitates the value of listening, as it provides multiple ways of expression, which allows connecting with other stories in a specific topic such as inclusion. Therefore, teachers valued DST beyond improving digital skills to learn from each other in the context of inclusion. As facilitators of the workshop, we experienced using DST flexibly by implementing workshops with two participant teachers with hearing disabilities. During the implementation of the DST workshop, participants had a collaboration in terms of interpreting sign language. Facilitators in small groups joined the conversation in an equal participation principle. This dialogue allowed a collaborative, co-creative learning environment between participants and facilitators. Creating such learning environments and co-creative habits gains importance for both students and educators (Tomczyk et al., 2019). In addition, digital story creation was not limited to only one-way communication and focused on sharing experiences that provided multimodal communication, including at least one of text, audio, and visual elements.

Teachers expressed open curiosity about each others' stories by "identifying" themselves with the stories of inclusion related to their work contexts. Sharing stories allowed collaboration with and learning from peers throughout sharing stories of inclusion which provided an endless opportunity for teachers to align technology with meaningful context for learning which can be called a meaningful technology-integrated approach (Sadik, 2008). Therefore, the DST workshop provided an inclusive example for teachers to use in their setting. For example, P1 highlighted "DST as an audiovisual content which many people with disabilities can learn to build meaningful learning. I can now drive change since now I know a new tool."

A previous study reported that reforms have failed to significantly improve access to regular schools and classes for students with a disability (De-Bruin, 2019). However, the teachers in the present study look for new ways to change traditional practices. Therefore, workshop-based DST can provide a promising approach for improving inclusive education. Several other studies highlighted DST as being a source for social change in different areas such as women's participation (Imek, 2012; Hlalele & Brexa, 2015), teaching to Children with hearing disabilities (Flórez-Aristizábal et al., 2019). Like these studies, teachers in this study considered DST to drive change in the context of teacher education as it promotes "inclusion



in the classrooms,” “building,” meaningful learning” when used in the classroom. The participants in the study did not have any experience in using DST; therefore, a step-by-step DST workshop process was implemented. Participants joined a story circle where they listened to facilitators’ stories and had dialogic interactions about their inclusive education experiences. Unlike studies which focus on acquiring digital competencies by the teacher or the student through DST implementation, workshop-based DST adopted in this study may contribute to the development of teachers’ emotional intelligence as it “was interesting to hear and see the different points of the others”, thereby establishing enhanced professional development in inclusive education. For teachers to be effective co-agents, they need “the capacity to act purposefully and constructively to direct their professional growth and contribute to the growth of their students and colleagues” (Calvert, 2016). To achieve this, teachers need continuing professional development in designing learning environments that support student agency. The present findings show that workshop-based DST can be an engaging, inclusive approach to achieving meaningful learning outcomes to contribute professional development of teachers for promoting inclusive education.

#### 4.2. Conclusion and suggestions

This study provides the first usability results of the SELI smart learning platform. Further studies which focus on the usability of the SELI Learning platform might give detailed information for validation of results. To achieve sustainability in inclusive education, Vanderpuye et al. (2020) call for intensive staff training and continuing professional development to sufficiently prepare teachers for inclusive education. Our example of using DST to enhance teachers’ professional development demonstrates the benefits of sharing stories for reflecting and understanding inclusion through diverse stories from the Dominican Republic. We experienced a case of inclusion through the inclusion of two participant teachers with hearing disabilities in the DST workshop. This experience provided learning from diverse stories and provided a practical example of implementation for participants and facilitators.

When facilitating DST workshops, there are restrictions and risks as facilitators need to be confident with the process and technologies to help solve technical problems and support participants through ethical issues. Much can be solved initially through good preparation and setting specific ground rules and by actively promoting dialogic mechanisms such as story circles and digital processes like sound and image merging. The SELI smart platform fosters the inclusion of disabled people through DST by facilitating participation in ICT-enhanced learning. In addition, ICT use for inclusive education should be considered beyond disability research and focus on quality education for all students with diverse backgrounds.

Workshop-based DST goes beyond the dominant discourses and initiatives of digital inclusion, focusing mainly on the use and acquisition of digital devices and digital competence development. Nevertheless, workshop-based DST was one of the strategies of the SELI project to allow different levels of communication among teachers, technology experts, and teacher educators to use the potential of digital technologies to trigger a transformation in education. We used workshop-based DST as an educational strategy to understand and use digital technologies in teachers’ professional development with an active, respectful, and critical perspective. Teachers can experience inclusive ways of using ICT-based technologies to transfer inclusive competencies to their classroom settings. This practice allowed listening differences among various levels of stakeholders for promoting inclusive education.

#### Authors’ Contribution

Idea, O.Y.A., C. R, S.S.O., D.M., G.D.; Literature review (state of the art), O.Y.A., C.R, S.S.O.; Methodology, O.Y.A., C.R, S.S.O., D.M., G.D.; Data analysis, O.Y.A., C. R, D.M.; Results, O.Y.A., C. R, S.S.O, G.D.; Discussion and conclusions, O.Y.A., C.R, S.S.O., D.M., G.D.; Writing (original draft), O.Y.A., C. R, S.S.O.; Final revisions, O.Y.A., D.M., G.D.; Project design and funding agency, O.Y.A., C.R, S.S.O., D.M., G.D.

#### Funding Agency

This work was supported by the ERANET-LAC project which has received funding from the European Union’s Seventh Framework Programme. The project entitled Smart Ecosystem for Learning and Inclusion (SELI)– with number ERANet17/ICT-0076 was also supported by the Scientific and Technological Research Council of Turkey (118K447) with the project number and Ministerio de Educacion Superior, Ciencia y Tecnologia, MESCyT.

## References

- Akyar, O.Y., Demirhan, G., Oyelere, S.S., Flores, M., & Jauregui, V.C. (2020). Digital storytelling in teacher education for inclusion. In A. Rocha, H. Adeli, L. Reis, S. Costanzo, I. Orovic, & F. Moreira (Eds.), *Trends and Innovations in Information Systems and Technologies. World Conference on Information Systems and Technologies* (pp. 367-376). Springer. [https://doi.org/10.1007/978-3-030-45697-9\\_36](https://doi.org/10.1007/978-3-030-45697-9_36)
- Albano, G., & Iacono, U.D. (2019). Designing digital storytelling for mathematics special education: An experience in support teacher education. *The Mathematics Enthusiast*, 16(1), 263-288. <https://bit.ly/399uYj3>
- Bacakova, M., & Closs, A. (2013). Continuing professional development (CPD) as a means to reducing barriers to inclusive education: Research study of the education of refugee children in the Czech Republic. *European Journal of Special Needs Education*, 28(2), 203-216. <https://doi.org/10.1080/08856257.2013.778108>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Calvert, L. (2016). *Moving from compliance to agency: What teachers need to make professional learning work*. Learning Forward and NCTAF. <https://bit.ly/2VDvqTR>
- Çetin, E. (2021). Digital storytelling in teacher education and its effect on the digital literacy of pre-service teachers. *Thinking Skills and Creativity*, 39, 100760. <https://doi.org/10.1016/j.tsc.2020.100760>
- Creswell, J.W., & Clark, V.L.P. (2011). *Designing and conducting mixed methods research*. SAGE. <https://bit.ly/3ymZPom>
- Dantas, C., Jegundo, A.L., Quintas, J., Martins, A.I., Queirós, A., & Rocha, N.P. (2017). European portuguese validation of usefulness, satisfaction and ease of use questionnaire (USE). In A. Rocha, A. Correia, H. Adeli, L. Reis, & S. Costanzo (Eds.), *Recent advances and ease of use questionnaire (USE)* (pp. 561-570). World Conference on Information Systems and Technologies. Springer. [https://doi.org/10.1007/978-3-319-56538-5\\_57](https://doi.org/10.1007/978-3-319-56538-5_57)
- De-Bruin, K. (2019). The impact of inclusive education reforms on students with disability: an international comparison. *International Journal of Inclusive Education*, 23(7-8), 811-826.
- Faria, T.V.M., Pavanelli, M., & Bernardes, J.L. (2016). Evaluating the usability using USE questionnaire: Mindboard system use case. In P. Zaphiris, & A. Ioannou (Eds.), *Learning and Collaboration Technologies* (pp. 518-527). Springer. [https://doi.org/10.1007/978-3-319-39483-1\\_47](https://doi.org/10.1007/978-3-319-39483-1_47)
- Fernández-Batanero, J.M., Cabero, J., & López, E. (2019). Knowledge and degree of training of primary education teachers in relation to ICT taught to students with disabilities. *British Journal of Educational Technology*, 50(4), 1961-1978. <https://doi.org/10.1111/bjet.12675>
- Filippidis, S.K., & Tsoukalas, I.A. (2009). On the use of adaptive instructional images based on the sequential-global dimension of the Felder-Silverman learning style theory. *Interactive Learning Environments*, 17(2), 135-150. <https://doi.org/10.1080/10494820701869524>
- Flórez-Aristizábal, L., Cano, S., Collazos, C.A., Benavides, F., Moreira, F., & Fardoun, H.M. (2019). Digital transformation to support literacy teaching to deaf children: From storytelling to digital interactive storytelling. *Telematics and Informatics*, 38, 87-99. <https://doi.org/10.1016/j.tele.2018.09.002>
- Florian, L. (2007). Reimagining special education. In L. Florian (Ed.), *The SAGE handbook of special education* (pp. 1-4). SAGE. <https://doi.org/10.4135/9781848607989.n2>
- Gao, M., Kortum, P., & Oswald, F. (2018). Psychometric evaluation of the use (usefulness, satisfaction, and ease of use) questionnaire for reliability and validity. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62, 1414-1418. <https://doi.org/10.1177/1541931218621322>
- Hariyanto, D., Triyono, M.B., & Köhler, T. (2020). Usability evaluation of personalized adaptive e-learning system using USE questionnaire. *Knowledge Management & E-Learning: An International Journal*, 12(1), 85-105. <https://doi.org/10.34105/j.kmel.2020.12.005>
- Hersh, M. (2017). Classification framework for ICT-based learning technologies for disabled people. *British Journal of Educational Technology*, 48(3), 768-788. <https://doi.org/10.1111/bjet.12461>
- Hersh, M., & Mouroutsou, S. (2019). Learning technology and disability-Overcoming barriers to inclusion: Evidence from a multicountry study. *British Journal of Educational Technology*, 50(6), 3329-3344. <https://doi.org/10.1111/bjet.12737>
- Hlalele, D., & Brexa, J. (2015). Challenging the narrative of gender socialisation: Digital storytelling as an engaged methodology for the empowerment of girls and young women. *Agenda*, 29(3), 79-88. <https://doi.org/10.1080/10130950.2015.1073439>
- Hwang, G.J. (2014). Definition, framework and research issues of smart learning environments-a context-aware ubiquitous learning perspective. *Smart Learning Environments*, 1(4), 1-14. <https://doi.org/10.1186/s40561-014-0004-5>
- Istemic, A. (2021). Shifting to digital during COVID-19: Are teachers empowered to give voice to students? *Educational Technology Research and Development*, 69, 43-46. <https://doi.org/10.1080/0022027980280601>
- Kouvara, T.K., Karasoula, S.A., Karachristos, C.V., Stavropoulos, E.C., & Verykios, V.S. (2019). Technology and school unit improvement: Researching, reconsidering and reconstructing the school context through a multi-thematic digital storytelling project. *Social Sciences*, 8(2), 49-49. <https://doi.org/10.3390/socsci8020049>
- Lambert, J., & Hessler, B. (2018). *Digital storytelling: Capturing lives, creating community*. Routledge. <https://doi.org/10.4324/9781351266369>
- Landauer, T.K. (1997). Behavioral research methods in human-computer interaction. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 203-227). Elsevier. <https://doi.org/10.1016/B978-044481862-1.50075-3>

- Lattie, E.G., Duffecy, J.L., Mohr, D.C., & Kashima, K. (2017). Development and evaluation of an online mental health program for medical students. *Academic Psychiatry*, (5), 642-645. <https://doi.org/10.1007/s40596-017-0726-0>
- Li, K.C., & Wong, B.T.M. (2021). Review of smart learning: Patterns and trends in research and practice. *Australasian Journal of Educational Technology*, 37(2), 189-204. <https://doi.org/10.14742/ajet.6617>
- Lund, A.M. (2001). Measuring usability with the use questionnaire12. *Usability Interface*, 8(2), 3-6. <https://bit.ly/3GL5Jmi>
- Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. In *Proc. ACM CHI'94 Conf.* (pp. 152-158). <https://doi.org/10.1145/259963.260333>
- Park, H.R. (2019). ESOL pre-service teachers' experiences and learning in completing a reflection paper and digital storytelling. *Australasian Journal of Educational Technology*, 35(4). <https://doi.org/10.14742/ajet.4117>
- Ranieri, M., & Bruni, I. (2018). Digital and media literacy in teacher education: Preparing undergraduate teachers through an academic program on digital storytelling. In *Handbook of Research on Media Literacy in Higher Education Environments* (pp. 90-111). IGI Global. <https://doi.org/10.4018/978-1-5225-4059-5.ch006>
- Rapanta, C., Garcia-Mila, M., Remesal, A., & Gonçalves, C. (2021). The challenge of inclusive dialogic teaching in public secondary school. [El reto de la enseñanza dialógica inclusiva en la escuela pública secundaria]. *Comunicar*, 66, 9-20. <https://doi.org/10.3916/C66-2021-02>
- Sadik, A. (2008). Digital storytelling: A meaningful technology-integrated approach for engaged student learning. *Educational Technology Research and Development*, 56(4), 487-506. <https://doi.org/10.1007/s11423-008-9091-8>
- Sánchez-Serrano, J.L.S., Jaén-Martínez, A., Montenegro-Rueda, M., & Fernández-Cerero, J. (2009). Impact of the information and communication technologies on students with disabilities. A systematic review. *Sustainability*, 12(20). <https://doi.org/10.3390/su12208603>
- Simsek, B. (2012). Enhancing women's participation in Turkey through digital storytelling. *Journal of Cultural Science*, 5(2), 28-46. <https://doi.org/10.5334/csci.45>
- Simsek, B., Akyar, O.Y., Oyelere, S.S., & Demirhan, G. (2021). *Reflections on inclusion and ICT in the context of smart ecosystem for learning and inclusion project*. Hacettepe University. <https://bit.ly/3DV3KtD>
- Tomczyk, L., Jáuregui, V.C., Amato, C., Muñoz, D., Arteaga, M., Oyelere, S.S., Akyar, O.Y., & Porta, M. (2020). Are teachers techno-optimists or techno-pessimists? A pilot comparative among teachers in Bolivia, Brazil, the Dominican Republic, Ecuador, Finland, Poland, Turkey, and Uruguay. *Education and Information Technologies*, 26, 2715-2741. <https://doi.org/10.1007/s10639-020-10380-4>
- Tomczyk, L., Oyelere, S.S., Puentes, A., Sanchez-Castillo, G., Muñoz, D., Simsek, B., Akyar, O.Y., & Demirhan, G. (2019). Flipped learning, digital storytelling as the new solutions in adult education and school pedagogy. *Adult Education 2018-Transformation in the Era of Digitization and Artificial Intelligence*, (pp. 69-83). <https://bit.ly/3DN0tMW>
- Valdivia, A. (2017). What was out of the frame? A dialogic look at youth media production in a cultural diversity and educational context in Chile. *Learning, Media and Technology*, 42, 112-125. <https://doi.org/10.1080/17439884.2016.1160926>
- Vanderpuye, I., Obosu, G.K., & Nishimuko, M. (2020). Sustainability of inclusive education in Ghana: Teachers' attitude, perception of resources needed and perception of possible impact on pupils. *International Journal of Inclusive Education*, 24(14), 1527-1539. <https://doi.org/10.1080/13603116.2018.1544299>

# Editors' Club

[revistacomunicar.com/wp/club-de-editores](http://revistacomunicar.com/wp/club-de-editores)



Comunicar  
Editing resources for  
scientific journals





# Detection of traits in students with suicidal tendencies on Internet applying Web Mining

## Detección de rasgos en estudiantes con tendencia suicida en Internet aplicando Minería Web

- Id** Dr. Iván Castillo-Zúñiga. Professor, Systems and Computing Department, National Technological Institute of Mexico, Llano Aguascalientes Campus (Mexico) (ivan.cz@llano.tecnm.mx) (https://orcid.org/0000-0001-8017-5908)
- Id** Dr. Francisco-Javier Luna-Rosas. Professor, Systems and Computing Department, National Technological Institute of Mexico, Aguascalientes Campus (Mexico) (fcoluna2000@yahoo.com.mx) (https://orcid.org/0000-0001-6821-4046)
- Id** Dr. Jaime-Iván López-Veyna. Professor, Systems and Computing Department, National Technological Institute of Mexico, Zacatecas Campus (Mexico) (ivanlopezveyena@zacatecas.tecnm.mx) (https://orcid.org/0000-0002-9225-3202)

### ABSTRACT

This article presents an Internet data analysis model based on Web Mining with the aim to find knowledge about large amounts of data in cyberspace. To test the proposed method, suicide web pages were analyzed as a study case to identify and detect traits in students with suicidal tendencies. The procedure considers a Web Scraper to locate and download information from the Internet, as well as Natural Language Processing techniques to retrieve the words. To explore the information, a dataset based on Dynamic Tables and Semantic Ontologies was constructed, specifying the predictive variables in young people with suicidal inclination. Finally, to evaluate the efficiency of the model, Machine Learning and Deep Learning algorithms were used. It should be noticed that the procedures for the construction of the dataset (using Genetic Algorithms) and obtaining the knowledge (using Parallel Computing and Acceleration with GPU) were optimized. The results reveal an accuracy of 96.28% on the detection of characteristics in adolescents with suicidal tendencies, reaching the best result through a Recurrent Neural Network with 98% accuracy. It is inferred that the model is viable to establish bases on mechanisms of action and prevention of suicidal behaviors, which can be implemented in educational institutions or different social actors.

### RESUMEN

Este artículo presenta un modelo de análisis de datos en Internet basado en Minería Web con el objetivo de encontrar conocimiento sobre grandes cantidades de datos en el ciberespacio. A fin de probar el método propuesto, se analizaron páginas web sobre el suicidio como caso de estudio con la intención de identificar y detectar rasgos en estudiantes con tendencias suicidas. El procedimiento considera un Web Scraper para localizar y descargar información de Internet, así como técnicas de Procesamiento de Lenguaje Natural para la recuperación de los vocablos. Con el propósito de explorar la información, se construyó un conjunto de datos basado en Tablas Dinámicas y Ontologías Semánticas, especificando las variables predictivas en jóvenes con inclinación suicida. Por último, para evaluar la eficiencia del modelo se utilizaron algoritmos de Aprendizaje de Máquina y Aprendizaje Profundo. Cabe destacar que se optimizaron los procedimientos para la construcción del dataset (utilizando Algoritmos Genéticos) y obtención del conocimiento empleando Cómputo Paralelo y Aceleración con Unidades de Procesamientos de Gráfico (GPU). Los resultados revelan una precisión del 96,28% sobre la detección de las características en adolescentes con tendencia suicida, alcanzando el mejor resultado a través de una Red Neuronal Recurrente con un 98% de precisión. De donde se infiere que el modelo es viable para establecer bases sobre mecanismos de actuación y prevención de comportamientos suicidas, que pueden ser implementados en instituciones educativas o distintos actores de la sociedad.

### KEYWORDS | PALABRAS CLAVE

Suicidal behavior, cybersuicide, web mining, machine learning, deep learning, recurrent neural networks. Conducta suicida, ciber-suicidio, minería web, aprendizaje de máquina, aprendizaje profundo, redes neuronales recurrentes.



## 1. Introduction

The benefits that new technologies have brought to our lives cannot be denied, however, they have also generated problems that were previously unknown, one of them is known as cybersuicide. According to López-Martínez (2020), cybersuicide is a phenomenon that refers to the influence of the information that circulates on the Internet and incites a person to commit a suicidal act. Suicide is the second leading cause of death in the world population aged 10 to 24, which represents 100,000 dead adolescents per year. Suicide is a complex behavior that is constructed over time and which depends on multiple biological, family, social, and educational factors, among others.

The internet and social media have rapidly and substantially transformed the way adolescents and people communicate and access suicide-related information. This large number of websites and volumes of information is commonly referred to as Big Data. Molina and Restrepo (2018), mention that users not only search web pages that provide information related to methods and ways to commit suicide, but also seek help, support, and guidance in the face of the suffering they experience derived from suicidal thoughts, sadness, loneliness, and anxiety. These reasons require early detection and immediate response to this public health problem, where information is essential for its analysis and a useful means to educate and prevent suicidal behavior. Educational institutions play a very important role in promoting healthy lifestyles and providing early support to at-risk youth.

From different approaches, many efforts have been made to study the large volumes of information that are generated on the Internet by applying automatic techniques to analyze and predict events that affect society, such as suicide. With this perspective, there are programs, developments, algorithms, and evolving processes that continue to be studied by researchers. According to Nalini and Sheela (2014), understanding the relationship between analytical skills and the characteristics of a criminal event can help researchers use these techniques more efficiently to identify trends and patterns, address different issues, and even predict crime. For these reasons, this study proposes a new approach to Web Mining processes, integrating Machine Learning and Deep Learning techniques using GPU-accelerated and Parallel Computing to process large volumes of data in order to identify and detect traits in adolescents with suicidal tendencies.

Research such as those of Bonami et al. (2020), Denia (2020), Kim and Chung (2019), Anggraini et al. (2018), and Roy et al. (2017), have aimed to treat large volumes of data, analyzing information with Artificial Intelligence techniques and Natural Language Processing (NLP), which are summarized in Table 1.

Related Jobs	Big Data	Artificial Intelligence					Research Focus
	Data Source	Data Mining	NLP	Computation			
				Sequential	Parallel	GPU Acceleration	
Bonami, Piazentini and Dala-Possa (2020)	Sensor	Yes	No	Yes	No	No	Educational
Denia (2020)	Social Network	No	Yes	Yes	No	No	Social impact of scientific speeches on Twitter
Kim and Chung (2019)	Websites	Yes	Yes	Yes	No	No	Health
Anggraini, Sucipto and Indriati (2018)	Social Network	Yes	Yes	Yes	No	No	Cyberbullying
Roy et al. (2017)	Websites	Yes	Yes	Yes	No	No	Cybercrime Intrusion detection
Current investigation	Websites	Yes	Yes	Yes	Yes	Yes	Suicide

Some of these research projects present knowledge-seeking methods aimed at data mining (Bonami et al., 2020; Kim & Chung, 2019; Anggraini et al., 2018; Roy et al., 2017), other studies have focused on information retrieval using Natural Language Processing techniques (Denia, 2020; Kim & Chung, 2019; Anggraini et al., 2018). One of the complexities of this research is that the information is unstructured and located in large amounts of data dispersed on websites with different security and communication protocols, in addition, the texts have spelling errors and abbreviations. To begin the process of detecting the characteristics of students with suicidal tendencies, it is necessary to transform the information into structured data, continuing with the analysis using Machine Learning techniques to obtain patterns that become knowledge and added value.

This proposal is based on the combination of Big Data and Artificial Intelligence techniques with a genetic strategy using Parallel Computing. Also, we adapted some processes of the Semantic Web with procedures based on Semantic Ontologies, Vocabularies, and Dynamic Tables, integrating NLP methods for the processing of data, such as cleaning, separation of information from the computational code and removing meaningless words, tokenization (word separation), synonyms, stemming (word root) and term frequency, in order to generate valuable and value-added information, obtained from the analysis of a large number of pages with suicidal content downloaded from the web. It should be noted that, with these processes, we build a Semantic Ontology that classifies information through concepts validated with Machine Learning and Deep Learning techniques, applying GPU-accelerated Parallel Computing.

The main purpose of our research is to explore the benefits that can be obtained from the information that circulates on the web, seeking to discover patterns, unknown correlations, additional information that can be very useful to support the decision-making process in problem-solving. The study objectives contemplate the detection of traits in students with suicidal tendencies, as well as, carrying out a technical practice of recovering web pages, pre-processing, analyzing and classifying data. Finally, we create a categorization of the linguistic corpus of suicide. The main contribution of the article focuses on a study model on data related to cyberbullying problems on the Internet, specifically cybersuicide. Within the framework of the study, the limit is specifically the detection of cybersuicide words on the Internet through semantic ontologies using predictive analysis on a computer. It should be noted that the descriptive analysis of the words is not considered within the scope of the research.

### 1.1. Cybersuicide and its impact on students

López-Martínez (2020) mentions that suicide is a multifactorial phenomenon, and that, at present, with the emergence of the new Information and Communication Technologies (ICTs), it constitutes a new scenario and with it, a new problem for the prevention of suicidal behavior. In this context, a new concept is created, cybersuicide, which refers to the action of taking one's own life, motivated by the influence of prosuicidal pages, forums, and chat rooms on the Internet, among other variables. Within the same framework, Olivares (2019) points out that cybersuicide refers to the influence of information circulating on the internet, as well as the incitement that exists in those media to exercise it. In addition to the situation, Moreno & Blanco (2012) indicate that it is a phenomenon that spreads rapidly throughout the planet, increasing suicide cases year after year.

Durkheim (2008) indicates that suicide is any case of death that results directly or indirectly, from any act, positive or negative, performed by the victim itself, knowing that the victim should produce this result. From another perspective, Berengueras (2018) mentions that suicide is when individuals violate the laws of their own nature by making the decision to end their own life, in a deeply expressive and essential act of being understood and heard in their previous manifestations. For his part, Marchiori (2015) establishes that the most frequent instruments to commit suicide are firearms, knives, ropes, wires, fabrics for suffocation, medicaments, drugs, jumping from bridges, buildings, to train tracks or passing cars, asphyxia by immersion, poisons, fuels (gas, coal, kerosene, naphtha), among others.

Research such as Arevalos (2020), SeGob (2021), Luna and Dávila (2018), Sánchez-García et al. (2018), Blanco (2019), Healy (2019), and the World Health Organization (WHO) (2019) reveal that suicide is a global problem, which can occur at any age. However, it is preventable through timely interventions, where the education sector plays an important role, since it is reported that in 2016 it was the second leading cause of death for young people between the ages of 15 and 19. In Argentina, testimonies from young people enrolled in peripheral urban junior high schools related to suicidal behavior were obtained, in which the following features stand out: belittling, humiliation, bullying, neglect, obesity, family abuse and lack of support, parental divorce, rape, abuse, complexes with their body or image, dating problems, drugs, alcohol, social indifference and visualization of few opportunities in the future (Arevalos, 2020).

In Mexico, the Secretary of the Interior reports the impact of the Covid-19 pandemic on girls and boys concerning suicide, wherein 2021 a record figure of 1,150 suicides were registered, with an increase of 37% in children between 10 and 14 and of 12% in adolescent women between 15 and 19 (SeGob,

2021), along the same lines it is identified that aggression, family violence, educational delay, alcohol or tobacco consumption are risk factors associated with the suicide attempt in adolescents, especially in younger women (Luna & Dávila, 2018). In Spain, it is revealed that 7.7% of adolescents between 12 and 19 years old showed difficulties in emotional adjustment and greater suicidal ideation derived from behaviors related to bullying, and tobacco and cannabis use (Sánchez-García et al., 2018). Along the same line, the records of the National Institute of Statistics reported 3,679 deaths by suicide in 2017 with 74% for men and 26% for women, with an average of 10 suicides per day (Blanco, 2019). In the United States, suicide claimed the lives of 5,016 men and 1,225 women between the ages of 15 and 24 in 2017, with a youth suicide rate of 14.6 per 100,000 inhabitants (Healy, 2019). For its part, the WHO (2019) reveals that in the world about 800,000 people take their own lives and many more try to do it, where 79% of suicides took place in low, and middle-income countries. Ingestion of pesticides, hanging, and firearms are the most common suicide methods.

### 1.2. Suicidal behavior in adolescents

Beaven-Ciagara et al. (2018), describe a study on the risk factors associated with suicidal behavior in young people aged 13 to 18 in the community of Guaymas, Sonora, Mexico. The results indicate that the psychosocial factor that is being presented is dysfunctional families, causing depression and low self-esteem, where the model used relates suicidal behavior with this finding in 37%, occurring more frequently in females. The study was carried out on 120 middle and high school students (41% males and 59% females), and its statistical analysis was done using SPSS software version 21.0. In turn, Mosquera (2016), presents a non-systematic review of the literature on child suicidal behavior, revealing that among the most prominent risk factors are: being male, having previous suicide attempts, social exclusion, emotional conflict. In addition, high comorbidity is observed with depressive disorders, bipolar disorder, and schizophrenia. Among the most effective treatments are dialectical-behavioral and cognitive-behavioral therapy. On the other hand, Carballo-Belloso and Gómez-Peñalver (2017) found a strong causal association between individual vulnerability factors and stressors, such as bullying experiences in childhood, and the subsequent development of thoughts and/or self-inflicted-injury behaviors, highlighting the importance of an adequate detection of this potentially modifiable risk factor.

### 1.3. Suicide analysis with Artificial Intelligence

Table 2 shows a comparison of related works for the analysis and prediction of suicide with Machine Learning techniques, in which Ramírez-López et al. (2021) seek to predict possible cases of suicide in the city of Aguascalientes, Mexico, using an SQL Server database of the 911 emergency service that records suicides. In their tests, they implement a geospatial analysis with Weka's EBK (Empirical Bayesian Kriging) method to locate the places where suicides have taken place. The results reveal a 99.22% prediction with a Bayesian classifier in MatLab identifying graphically the probability areas where suicide occurs and where it does not. Gen-Min et al. (2020) research how to predict suicidal ideas in army personnel since they have greater psychological stress and are at a higher risk of suicide attempts compared to the general population. The analysis uses Machine Learning techniques that include Logistic Regression, Decision Trees, Random Forests, Regression Trees, Vector Support Machines, and Multilayer Perceptron, considering five psychopathological domains (BSRS-5), anxiety, depression, hostility, interpersonal sensitivity, and insomnia. Their results exceed 98% accuracy in the classification using a questionnaire-based dataset of 3546 people. For their part, Pérez-Martínez et al. (2020) describe a Twitter analysis in which school harassment, rape, and suicide-related to the Netflix series "13 Reasons Why" are discussed in several countries, retrieving 154,470 tweets for exploration.

The results reveal that 51% of tweets were about suicide, 24% about bullying, and 23% about rape, where the United States and Spain were the countries with the greatest participation. The hashtags #13ReasonsWhy and #PorTreceRazones (in Spanish) were used to recover tweets, eliminate duplicate tweets, and form 3 samples with words in Spanish, English, French, Portuguese and Italian. The projections and calculations were made with statistical analysis of the SPSS software. Chiroma et al. (2018), identify suicide text on Twitter with Decision Trees, Bayes, Random Forests, and Vector Support



Machines, obtaining an accuracy between 34.6% and 77.8%, achieving better results with the first. On the other hand, Du et al. (2018) extract psychiatric stressors from suicide-related Twitter data using a Deep Learning approach and transfer learning strategy, where an accuracy of 78% is obtained with convolution Neural Networks and 67.94% with Recurrent Neural Networks. Finally, Hermosillo-De la Torre et al. (2015) show the relationship of depressive symptoms, hopelessness, and psychological resources on suicide attempts in a sample of 96 adolescents in Aguascalientes, Mexico. Using SPSS they applied descriptive statistics for the analysis of proportions and estimation of population parameters, and nonparametric statistics for the comparisons of study groups. Subsequently, they implemented Spearman's Rho statistics to find out how variables were associated, and linear regression to observe the relationships between them. The results show that the development of the capacity to properly manage sadness is one of the factors to be considered as a suicide prevention measure to be encouraged and developed in adolescents.

Related studies	Data source	Data Analysis Techniques						Software Used
		Statistics	Machine Learning	Deep Learning	Data Pre-processing			
					Web Scraper	NLP	Semantic Ontology	
Ramírez-López et al. (2021)	SQL Server	No	Yes	No	No	Yes	No	Weka y MatLab
Gen-Min et al. (2020)	Questionnaire	Yes	Yes	No	No	No	No	Non-specific
Perez-Martínez et al. (2020)	Twitter	Yes	No	No	No	No	No	SPSS
Chiroma et al. (2018)	Twitter	No	Yes	No	No	Yes	No	Non-specific
Du et al. (2018)	Twitter	No	Yes	Yes	No	Yes	No	Non-specific
Hermosillo-De la Torre et al. (2015)	Psychological resource scales	Yes	No	No	No	No	No	SPSS
Present research	Websites	Yes	Yes	Yes	Yes	Yes	Yes	Java and Python

## 2. Materials and methods

### 2.1. Methodology to help to detect traits in students with a suicidal tendency on websites

The methodology applied in this study begins by obtaining data from cyberspace and ends with the detection of the traits of students with suicidal tendencies. The procedure consisted of three stages that integrated Big Data Analytics, Natural Language Processing, Semantic Web, and Artificial Intelligence, which are described below.

#### 2.1.1. Location and download of suicide web pages (stage 1)

For locating and downloading websites with suicidal content, an open-source crawler was developed with the JSOUP libraries using the Java programming language, followed by scraping techniques to get the name of the website, the internet address (URL), and to make a copy of the file on the computer hard drive. The information is stored in databases to avoid duplication, along with control fields that allow the pre-processing of documents.

#### 2.1.2. Construction of datasets for tests (stage 2)

One of the most crucial and complex parts for the detection of traits on students with suicidal tendencies, is the construction of the dataset since it is the main part where the classification and prediction tests are carried out using Artificial Intelligence techniques. To represent different sets of suicidal characteristics in adolescents, we used an approach based on semantic ontologies that allow us to associate concepts through object-oriented techniques (classes-objects-attributes), facilitating the grouping of different suicide conditions such as signs of suicide, ways to carry it out, types of suicide, risk factors, prevention, influences, and synonyms. Where the class symbolizes the main theme as suicide, the objects are the subtopics as signs of suicide, and the attributes are the characteristics, such as euphoria, distress, sleep, farewell, and isolation, among others.

The dataset is the result of the transformation of unstructured data to structured data, this part begins with the elimination of the computational code of the text, followed by the separation of words (tokenization), suppressing the words without meaning (stop word), such as prepositions, pronouns, articles, adverbs, conjunctions, and some verbs. Finally, the terms that will be part of the linguistic corpus of suicide were stored in a database. To indicate the importance of each suicide trait and increase the analytical accuracy, the techniques of Term Frequency (TF) and Document Inverse Frequency (IDF) were used, where the term was replaced with the root word, obtained with the Porter algorithm (2006) through the technique of lemmatization (stemming), where for example we considered for the search of the term suicide, the words suicide, suicides, suicidal and commit suicide, generating a more accurate result.

Subsequently, dynamic tables were constructed in MySQL using the characteristics defined in the Semantic Ontology as metadata linking them to the linguistic corpus to generate the dataset. It is important to mention that the dataset construction process was optimized with Parallel Computing through a Genetic Strategy to equally distribute the transformation of the web pages establishing a cluster with the processor cores simulating a chromosome and its genes. The process of the Genetic Algorithm evolves until it reaches the optimum distribution of web pages and meets the completion criteria with an adaptation function based on the mean. The evolution of the population is based on elitism, selection by tournament, point crossing, and mutation with random replacement. Finally, the target binary variable is established, with the name "correct" and the values "yes/no" that determine the response to be generated by the Artificial Intelligence algorithm from the predictor variables defined in the Semantic Ontology. To carry out this experiment, the variable to predict is determined with the value "yes", when the frequency of groups (signs of suicide, influences, types of suicide, synonyms, ways of carrying it out, prevention, and risk factors) is greater than zero, and "no" otherwise.

### 2.1.3. Artificial Intelligence applied to the detection of traits of students with a suicidal tendency (stage 3)

To evaluate the dataset, Artificial Intelligence techniques were selected: Random Forests, Neural Network, Decision Tree, and Logistic Regression, because they present greater similarities with related works (Gen-Min et al., 2020, Chiroma et al. 2018). In the construction of the algorithms, Python programming language was used, employing Machine Learning techniques with Sequential Computation based on the Sklearn API, Pandas, and NumPy.

Within the same testing framework, we considered, to optimize the above strategies using GPU-Accelerated Parallel Computing, applying Apple's scalable technology that allows the development of custom models through the Turicreate API with the object SFrame, which can mutate and scale to Big Data, considering the algorithms Random Forests, Logistic Regression, and Decision Tree. Finally, Deep Learning tests were performed using Recurrent Neural Networks (RNN), where the greater the number of layers and neurons, the greater the depth of the network and its learning capacity. The proposed RNN model integrates a 4-layer sequence (input layer, recurrent hidden layer with Long Short-Term Memory (LSTM), hidden layer, and output layer) that are densely connected, where all the neurons in one layer are connected to all the neurons in the next layer.

Among the optimization algorithms, Adam's algorithm, the SGD Stochastic Gradient Descent Method, and the RMSProp incremental learning technique were considered. Adam's algorithm combines the advantages of the AdaGrad and RMSProp algorithms, which calculates the learning rate of adaptive parameters based on the mean value of the first moment and makes full use of the mean value of the second moment of the gradient based on the non-centered variance. The SGD method maintains a single learning rate to update all weights throughout the training. Finally, the RMSProp technique considers a different training factor for each dimension, where the scaling of the training factor is performed by dividing it by the mean of the exponential decline of the square of the gradient.

The RNN algorithm proposed is based on the synchronous data parallelism of Keras and TensorFlow, where each layer is represented by a tensor with global information (called a global lot) and is divided into sublots according to the number of GPUs (called local lots), where the gradient calculations are performed

considering the loss of the model. Subsequently, the updates originated in the local gradients are merged with the rest of the replicates, thus remaining synchronized with the process.

## 2.2. Materials

The operating systems implemented in the different tests include MacOS Big Sur, Linux Ubuntu 20.04, and Windows 10. The software tools integrate Java programming language, 8.0 edition, and the MySQL database manager, 5.7 edition, used for the localization and download of web pages (stage 1). Also, for the construction of the dataset through the Genetic Algorithm, NLP techniques, vocabulary construction, and the design of semantic ontologies based on objects (stage 2). For the Artificial Intelligence processes, Python programming language, 3.8 edition, was used, with the Sklearn, Pandas, Numpy, Turicreate, Keras, and TensorFlow libraries, with Parallel Computation and GPU acceleration (stage 3). Finally, the computer equipment used in this study was a MacBook Pro with 2.4 GHz Intel Core i9 processor, 16 GB DDR4 memory, 500 GB Flash hard drive, Intel UHD Graphics 630 GPU, and Radeon Pro 560X GPU, connected to a 100MB internet service for locating and downloading web pages.

## 3. Tests and results

### 3.1. Testing procedure

The test procedure developed allows the analysis of small and large datasets, where thousands to millions of records can be processed according to the procedure described below:

Stage 1:

- A set of web pages related to suicide is located and a copy is made on the hard disk for analysis.

Stage 2:

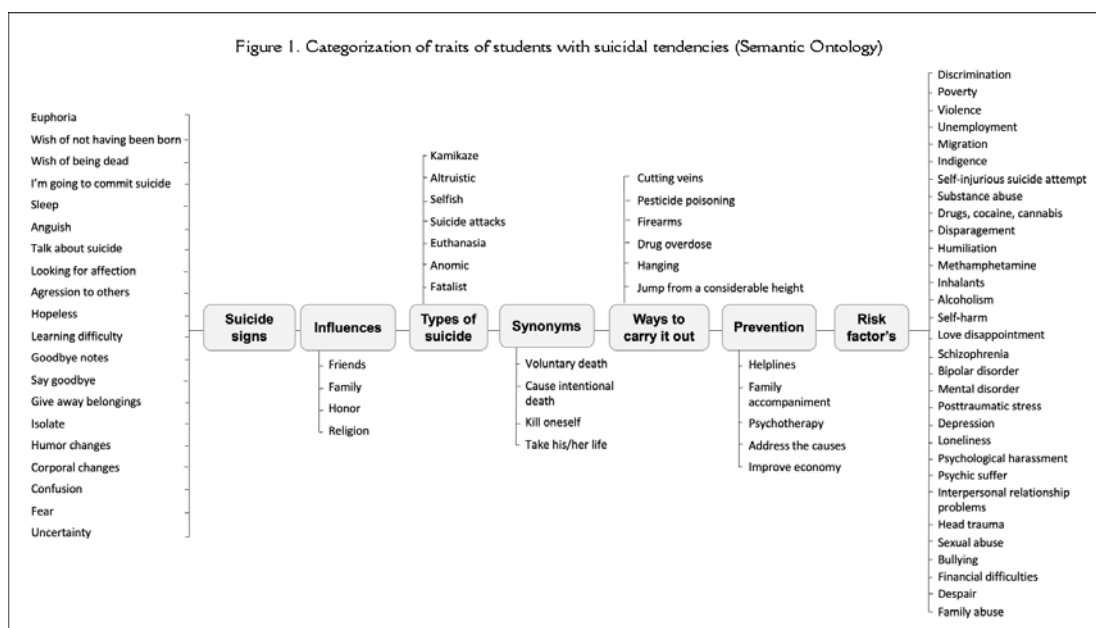
- A Semantic Ontology is established on the domain related to traits of students with a suicidal tendency extracting a set of words from books and processed using the classes-objects-attributes technique.
- A set of meaningless words is selected based on a Google SEO standard (Landaeta, 2014).
- A linguistic corpus of suicide is generated using NLP and semantic web techniques, optimized with Parallel Computing through Genetic Algorithms to balance the load.
- The cybersuicide dataset is constructed for the tests linking Semantic Ontology with the linguistic corpus, and the binary target variable is established.

Stage 3:

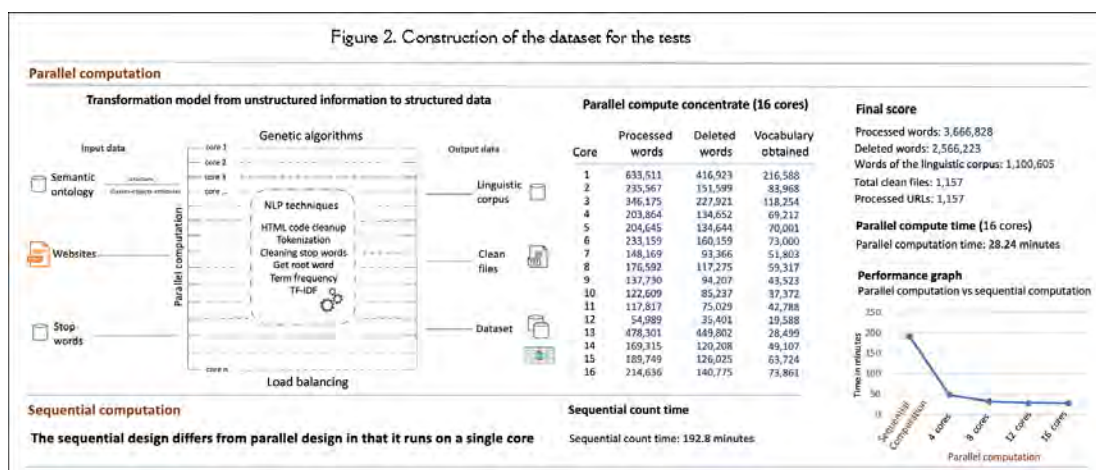
- For the construction of programs with Sequential Computation, the Sklearn library was used, specifically with the algorithms: Random Forests (RandomForestClassifier), Neural Network (MLPClassifier), Decision Trees (DecisionTreeClassifier) and Logistic Regression (LogisticRegression).
- For the optimization of the programs with Parallel Computation and acceleration with GPU, the algorithms based on the Turicreate Library were used, specifically: Random Forests (tc.random\_forest\_classifier), Decision Trees (tc.decision\_tree\_classifier) and Logistic Regression (tc.logistic\_classifier).
- Finally, for the construction of Recurrent Neural Networks with Deep Learning, an algorithm that connects several layers was considered. The Input Layer (inputLayer) receiving 97 suicidal traits, the recurrent occult layer with Short-Term Memory (RecurrentLayer LSTM) defined with 97 neurons, which consider possible combinations of suicidal traits, where the parameters obtained are the result of linear and non-linear transformations by calculating the current prediction considering the previous result of weights and biases, before using the activation function (using what has been learned), which allows having a short-term memory, the Hidden Layer dense (HiddenLayer) with 48 neurons, reducing the calculations of the transformations and filtering the optimal results to interpret the output of the hidden LSTM layer, finally the Output Layer (outputLayer\_Sigmoid) of one neuron compatible with binary predictions, which returns the final prediction.

### 3.2. Analysis and results

Regarding suicide-related web pages for testing purposes, 1,157 websites were located using the crawler and were downloaded to the hard drive. Concerning Semantic Ontology, the result is presented in Figure 1, where the traits of students with a suicidal tendency are described, organized by groups that reflect risk factors: ways of prevention, ways of carrying out suicide, similarities of suicide, types of suicide, factors that influence carrying suicide out, and signs indicating suicidal tendency. It should be emphasized that the information is based on the books: Suicide (Durkheim, 2008), The thought of suicide in adolescence (Villardón-Gallego, 2013), Suicide criminological approach (Marchiori, 2015), When nothing makes sense: reflections on suicide from logotherapy (Rocamora, 2017), Suicide the unbearable need to be another (Berengueras, 2018), The footprint of hopelessness: prevention strategies and coping with suicide (Urrea, 2019), and Suicide: a comprehensive and integrative look (García-Peña, 2020).



The results of the dataset construction for the tests are shown in Figure 2, illustrating the model of transformation from unstructured information to structured data, the Parallel Computing concentrate (taking as an example the 16 processor cores), specifying the core used, words processed, words removed, and vocabulary obtained.



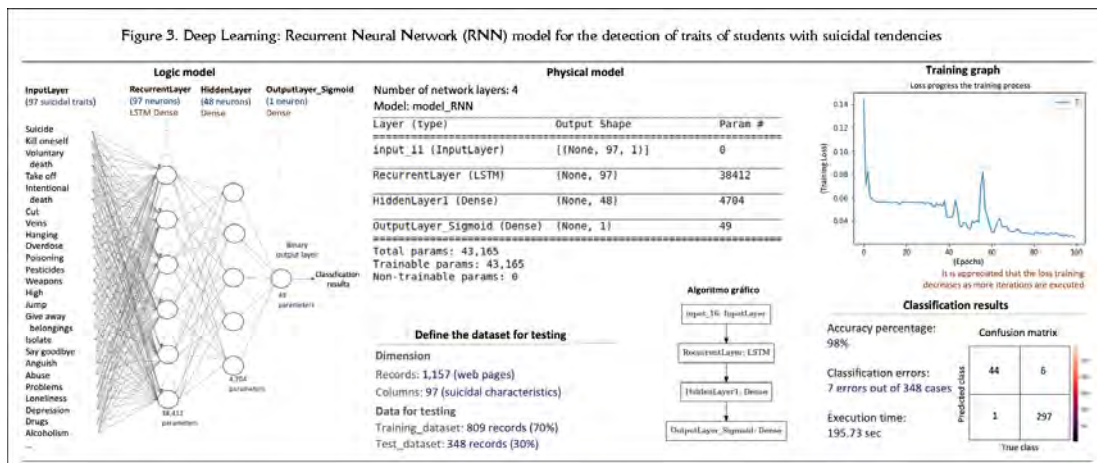


Likewise, the overall results of the process are presented, where 3,666,828 words were extracted, from which 2,566,223 empty words were eliminated, obtaining a linguistic corpus of 1,100,605 words of cybersuicide and 1,157 clean files with information extracted from websites. Finally, the time results of the Sequential Computation are included.

It should be noted that in the execution of the tests a better performance is obtained with the Parallel Computation (16 cores) on the Sequential Computation, obtaining an optimization of 682%, with a time of 28 minutes 24 seconds, against 192 minutes 8 seconds, obtaining the best time response when using a 16-core chromosome, as shown in the performance graph. Similarly, it can be seen, that the computation time begins to stabilize at 12 cores. On the other hand, the results of trait classification in students with suicidal tendencies are shown in Table 3. These were organized using Machine Learning techniques (Sequential Computation and Parallel Computation) and Deep Learning. Specifying the algorithm used, the values of the confusion matrix (true positives, true negatives, false positives, and false negatives), the percentage of precision and execution time.

Learning Types	Learning Techniques with Python	Algorithm	Learning Assessment				Precision (%)	Execution time (seconds)
			Confusion matrix					
			True Positive	True Negative	False Positive	False Negative		
Machine Learning	Sequential Computation (Sklearn)	Random Forests	293	44	6	5	96.84%	0.58 seg
		Neural Network	294	43	6	5	96.84%	2.09 seg
		Decision Tree	290	46	8	4	96.55%	0.03 seg
		Logistic Regression	293	46	6	3	97.41%	0.06 seg
	Parallel Computation (Turicreate) CPU + GPU AMD	Random Forests	289	52	14	4	94.98%	0.02 seg
		Logistic Regression	281	47	10	9	94.52%	0.07 seg
		Decision Tree	288	43	10	7	95.11%	0.01 seg
Deep Learning	Parallel Computation (TensorFlow y Keras) CPU	Recurrent Neural Network	297	44	6	1	98.00%	195.73 seg

For Machine Learning with Sequential Computation (Sklearn library), the algorithm that obtained the best classification result was Logistic Regression with 97.41% accuracy and a processing time of 0.06 seconds. The Random Forests, Neural Network, and Decision Tree showed stability above 96.55% very close to the first. The best time was obtained with the Decision Tree algorithm with 0.03 seconds.



Regarding Machine Learning with GPU accelerated Parallel Computing (Apple's Turicreate Library), the algorithm that obtained the best classification result was a Decision Tree with 95.11% accuracy and a response time of 0.01 seconds. Random Forest and Logistic Regression algorithms obtained a stable

percentage close to 95%. According to the tests performed, Turicreate API algorithms do not operate on Windows Operating Systems, in addition, no support for Neural Networks was found on the official website of the library.

Figure 3 presents the solutions of the RNN model for the detection of traits of students with suicidal tendency. It shows the logical and physical model with the outputs generated (layers), the graphical algorithm, the preparation of the dataset for the tests, and the training and classification results.

The results of the RNN reveal a 98% prediction and 195.73 seconds of computation, obtaining the best classification result compared to the rest of the algorithms. The procedure examines the dataset that represents the 1,157 web pages and the 97 characteristics of Semantic Ontology, employing 70% for training and 30% for testing. It should be noted that to optimize the model, the method based on Adam's Learning Rhythm was implemented since it obtained better performance in the tests than the SGD and RMSProp methods. The accuracy of the model can be seen in the training graph; in which it is observed that the loss decreases as more iterations of the algorithm are performed. In the same line, the confusion matrix reveals the accuracy of the classification by reporting 7 errors and 341 hits. It is worth mentioning that 43,165 solutions were generated in the different layers of the RNN model to obtain the optimal solution.

#### 4. Discussion

First, it is important to emphasize that the combination of different techniques and procedures has a greater technological and methodological scope, which can be seen in the excellent results. Putting Big Data Analytics techniques in practice into the information recovery and transformation processes, as well as Machine Learning for the discovery of knowledge, allows speeding up the analysis and classification of web pages for the user with an acceptable response time. One of the most important findings in the study is that the downloaded sample of 1,157 URLs, represents the web pages about suicide in Spanish because the web crawler searches showed repeated URLs mostly after 1,000.

As for computation time related to Artificial Intelligence processes, the algorithm that showed the best performance was Decision Tree using Parallel Accelerated Computation with GPU, and in contrast, the longest time was obtained by the RNN algorithm, associated with the same nature of recurrence of the algorithm. On the other hand, in the transformation of unstructured data to structured data, it was possible to optimize time by 682% by applying Parallel Computation compared to Sequential Computation.

Regarding the objective of the research (to detect traits or characteristics in students with a suicidal tendency in websites), the proposed methodology and architecture reached 98% accuracy in the classification, identifying patterns related to suicide signs, risk factors, ways to carry it out, people who influence the suicidal tendencies, types of suicide and forms of prevention. This allows the establishment of foundations for mechanisms of action and prevention of suicidal behaviors, which can be implemented by different social entities, including educational institutions, government agencies, and pro-suicide associations.

In relation to other authors, such as Gen-Min et al. (2020), Chiroma et al. (2018), we agree that the Machine Learning techniques employed (Neural Network, Decision Tree, Logistic Regression, and Random Forests) are highly effective in the classification and prediction of suicide. Regarding Deep Learning with RNN, the present research obtained better classification results than Du et al. (2018), with 98% and 67.94% accuracy respectively, derived from applying different techniques for the construction of the dataset.

Finally, functionality tests were carried out with the Artificial Intelligence algorithms in different operating systems, finding that the Linux Ubuntu 20.04 Operating System is ideal for working with the different hardware and software technologies.

#### 5. Conclusions

Cyberspace becomes a shared ecosystem of information on websites, social networks, and people who comment on suicidal experiences based on anonymity, in which they interact and express a specific opinion or information without the need to expose their identity. Suicidal behavior in students highlights the

cracks in our contemporary society and confronts academic communities because they generate frustration, impotence, guilt for not having done what was needed at the time. This results in questioning the current educational system, which sometimes can be severe. It is necessary to understand education as a critical process in society since it is the starting point of socialization outside the family nucleus and in which citizenship is incubated. Therefore, it concerns a privileged place to train new generations in human values, which enable the support and cultural recreation of a specific society and can prevent problems such as instances of suicidal events.

It should be noted that the model proposed in this study may represent an interesting contribution to the analysis of data in cyberspace related to suicidal tendencies in students and adolescents. Thus, as the model is applied to new websites, it acts as an expert system, to identify patterns related to suicide signs, ways to carry it out, risk factors, ways to prevent it, and influences (terms defined in the semantic ontology). Thus, it is possible to establish grounds for the development of protocols (pro-suicide) in educational institutions, which enable the prevention of suicidal behavior through timely information and sensitization of both students and parents.

Among the results, high percentages are reported in the detection of traits in students with a suicidal tendency on websites, associated with the different dataset construction and Machine Learning techniques used. The results show an improvement in the transformation time from unstructured data to structured data using Parallel Computing Techniques with Genetic Algorithms, obtaining a 682% time saving compared to Sequential Computing. Likewise, an average accuracy of 96.28% is obtained, reaching an optimal value of 98% accuracy with the RNN algorithm. Therefore, it is inferred that a Recurrent Neural Network is a robust architecture for dealing with text analysis, where the output of the previous state is the feedback to preserve the memory of the network over time or sequence of words. Similarly, it is concluded that the proposed methodology and architecture are suitable for identifying and classifying suicidal signs in students with information from the web.

In relation with the results of the study, the exploration of descriptive analysis techniques on the suicidal traits obtained in the present study were considered to establish the degree of association between the groups and their variables. As future work, it is proposed to explore the suicidal behavior in adolescents based on short texts (tweets) from the social network Twitter to help in the detection of suicidal incidents, and to establish a hybrid data analysis model that combines sentiment analysis techniques, Semantic Ontologies, and Natural Language Processing, integrating Convolutional Neural Networks with Deep Learning for text classification and pattern search.

### Authors' Contribution

Idea, I.C.Z.; Literature review (state of the art), I.C.Z.; Methodology, I.C.Z., F.J.L.R.; Data analysis, I.C.Z., F.J.L.R.; Results, I.C.Z., J.I.L.V.; Discussion and conclusions, I.C.Z., J.I.L.V.; Writing (original draft), I.C.Z.; Final revisions, I.C.Z., F.J.L.R., J.I.L.V.; Project Design and sponsorships, I.C.Z.

### Funding Agency

National Technological Institute of Mexico, Llano Aguascalientes Campus.

### References

- Angraini, I.Y., Sucipto, S., & Indriati, R. (2018). Cyberbullying detection modelling at Twitter social networking. *Jurnal Informatika*, 6(2), 113-118. <https://doi.org/10.30595/juita.v6i2.3350>
- Arealos, D.H. (2020). El sentido de la vida y las prácticas ligadas al suicidio. Testimonio de jóvenes escolarizados. *Revista Latinoamericana de Estudios sobre cuerpos, emociones y sociedad. RELACES*, 32, 52-63. <https://bit.ly/3pXSVC8>
- Beaven-Ciapara, N.I., Campa-Álvarez, R.A., Valenzuela, B.A., & Guillen-Lúgigo, M. (2018). Inclusión educativa: Factores psicosociales asociados a conducta suicida en adolescentes. *Prisma Social*, 23, 185-207. <https://bit.ly/3GRIO9X>
- Berengueras, M. (2018). *Suicidio la insoportable necesidad de ser otro*. Universidad Autónoma del Estado de Morelos. <https://bit.ly/3F7a8iVV>
- Blanco, C. (2019). El suicidio en España, respuesta institucional y social. *Revista de Ciencias Sociales*, 33(46), 79-106. <https://doi.org/10.26489/rvs.v33i46.5>
- Bonami, B., Piazzentini, L., & Dala-Possa, A. (2020). Education, Big Data and Artificial Intelligence: Mixed methods in digital platforms. [Educación, Big Data e Inteligencia Artificial: Metodologías mixtas en plataformas digitales]. *Comunicar*, 65, 43-52. <https://doi.org/10.3916/C65-2020-04>

- Carballo-Belloso, J.J., & Gómez-Peñalver, J. (2017). Relación entre bullying, autolesiones, ideación suicida e intentos auotolíticos en niños y adolescentes. *Revista de estudios de Juventud*, 115, 207-218. <https://doi.org/10.3916/C65-2020-04>
- Chiroma, F., Liu, H., & Cocca, M. (2018). Text Classification For Suicide Related Tweets. In *International conference on Machine Learning and Cybernetics (ICMLC)* (pp. 587-592). <https://doi.org/10.1109/ICMLC.2018.8527039>
- Denia, E. (2020). The impact of science communication on Twitter: The case of Neil deGrasse Tyson. [El impacto del discurso científico en Twitter: El caso de Neil deGrasse Tyson]. *Comunicar*, 65, 21-30. <https://doi.org/10.3916/C65-2020-02>
- Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., & Xu, H. (2018). Extracting psychiatric stressors for suicide from social media using deep learning. *BMC Medical Informatics & Decision Making*, 18(43), 77-87. <https://doi.org/10.1186/s12911-018-0632-8>
- Durkheim, E. (2008). *El suicidio*. Grupo Editorial Éxodo. <https://bit.ly/3p6C8h7>
- García-Peña, J.J. (2020). *El suicidio: Una mirada integral e integradora*. Universidad Católica Luis Amigó. <https://doi.org/10.21501/9789588943619>
- Gen-Min, L., Szu-Nian, Y., & Yueh-Ming, T. (2020). Machine Learning based suicide ideation prediction for military personnel. *IEEE Journal of Biomedical and Health Informatics*, 24(7), 1907-1916. <https://doi.org/10.1109/JBHI.2020.2988393>
- Healy, M. (2019). *Alcanzan máximo histórico los índices de suicidio de adolescentes y adultos jóvenes en EE.UU.* Los Angeles Times. <https://lat.ms/3IYhebO>
- Hermosillo-De-La-Torre, A.E., Vacío-Muro, M.A., Méndez-Sánchez, C., Palacios-Salas, P., & Sahagún-Padilla, A. (2015). Sintomatología depresiva, desesperanza y recursos psicológicos: una relación con tentativa de suicidio en una muestra de adolescentes mexicanos. *Acta Universitaria*, 25(2), 52-56. <https://doi.org/10.15174/au.2015.900>
- Kim, J., & Chung, K. (2019). Associative feature information extraction using text mining from health big data. *Wireless Pers Commun*, 105, 691-707. <https://doi.org/10.1007/s11277-018-5722-5>
- Landaeta, G. (2014). *Lista de stop words o palabras vacías en español*. SEO para Google. <https://bit.ly/3p2ysg3>
- López-Martínez, L.F. (2020). Suicidio, adolescencia, redes sociales e Internet. *Norte de SaludMental*, 17, 25-36. <https://bit.ly/3sg25g2>
- Luna, M., & Dávila, A. (2018). Adolescentes en riesgo: factores asociados con el intento de suicidio en México. *Revista Gerencia y Política de Salud*, 17(34), 1-14. <https://doi.org/10.11144/Javeriana.rgsp17-34.arfa>
- Marchiori, H. (2015). *El suicidio enfoque criminológico*. Editorial Porrúa. <https://bit.ly/3sfMYDu>
- Molina, M.J., & Restrepo, D. (2018). Internet y comportamiento suicida en adolescentes: ¿Cuál es la conexión? *Revista Pediatría*, 51, 30-39. <https://doi.org/10.14295/pediatr.v51i2.109>
- Moreno, P., & Blanco, C. (2012). Suicidio e Internet. Medidas preventivas y de actuación. *Revista Psiquiatria.com*, 16(18). <https://bit.ly/3E110Lb>
- Mosquera, L. (2016). Conducta suicida en la infancia: Una revisión crítica. *Revista de Psicología Clínica con Niños y Adolescentes*, 3, 9-18. <https://bit.ly/3p5lkG8>
- Nalini, K., & Sheela, L. (2014). A survey on Datamining in Cyber Bullying. *International Journal on Recent and Innovation Trends in Computing and Communication*, 2(7). <https://bit.ly/3q7GHqt>
- Olivares, S. (2019). Uso de Internet y conductas suicidas en adolescentes de 14 a 18 años en México. *Visión Criminológica-Criminalística*, (pp. 6-21). <https://bit.ly/3p7uKSF>
- Organización Mundial de la Salud (Ed.) (2019). *Suicidio. Información obtenida el 6 de abril de 2021 en la dirección de Internet*. OMS. <https://bit.ly/3q8TkBG>
- Pérez-Martínez, V.M., Aparicio-Vinacia, B., & Rodríguez-González, M.D. (2020). Acoso escolar, violación y suicidio en Twitter: Segunda temporada de «Por trece razones». *Vivat Academia*, 153, 137-168. <https://doi.org/10.15178/va.2020.153.137-168>
- Porter, M.F. (2006). An Algorithm for suffix stripping. *Program: Electronic Library and Information Systems*, 40, 211-218. <https://doi.org/10.1108/00330330610681286>
- Ramírez-López, C.M., Montes, M., Ochoa-Zezzatti, A., Ponce-Gallegos, J.C., & Guzmán-Mendoza, J.E. (2021). Identification of possible suicide cases using a Bayesian Classifier with the database the Emergency Service 911 of Aguascalientes. *International Journal of Combinatorial Optimization Problems and Informatics*, 12(1), 43-57. <https://bit.ly/32gamVvt>
- Rocamora, A. (2017). *Cuando nada tiene sentido: Reflexiones sobre el suicidio desde la logoterapia*. Editorial Desclée de Brouwer. <https://bit.ly/3F9TxLa>
- Roy, S.S., Mallik, A., Gulati, R., Obaidat, M.S., & Krishna, P.V. (2017). A deep learning based artificial neural network approach for intrusion detection. In D. Giri, R. Mohapatra, H. Begehr, & M. Obaidat (Eds.), *Mathematics and Computing. ICMC 2017. Communications in Computer and Information Science* (pp. 44-53). Springer. [https://doi.org/10.1007/978-981-10-4642-1\\_5](https://doi.org/10.1007/978-981-10-4642-1_5)
- Sánchez-García, M.A., Pérez-De-Albéniz, A., Paño, M., & Fonseca, P. (2018). Ajuste emocional y comportamental en una muestra de adolescentes españoles. *Actas españolas de psiquiatria*, 46, 205-216. <https://bit.ly/3yAOv8a>
- SeGob (Ed.) (2021). *Impacto de la pandemia en niñas y niños*. Secretaría de Gobernación de México. <https://bit.ly/3J1bsGg>
- Urra, J. (2019). *La huella de la desesperanza: Estrategias de prevención y afrontamiento del suicidio*. Ediciones Morata. <https://bit.ly/3OH1QiQ>
- Villardón-Gallego, L. (2013). *El pensamiento de suicidio en la adolescencia*. Publicaciones de la Universidad de Deusto. <https://bit.ly/33FtwFV>



# Booktokers: Generating and sharing book content on TikTok

Booktokers: Generar y compartir contenidos sobre libros a través de TikTok

- Nataly Guíñez-Cabrera. Lecturer, Business Management Department, University of Bío-Bío, Chillán (Chile) (nguinez@ubiobio.cl) (<https://orcid.org/0000-0002-6109-8457>)
- Katherine Mansilla-Obando. Lecturer, School of Business and Economics, University Finis Terrae, Santiago (Chile) (kmansillao@uft.edu) (<https://orcid.org/0000-0003-4273-4259>)

## ABSTRACT

Reading is essential for learning in education. However, the digital revolution has transformed the ways to encourage reading and learning about books. People who generate and share content about books on social networks are the so-called booktokers, influencers on TikTok who have become an emerging phenomenon in this connected society. However, there is scarce research that studies them in depth. Therefore, it is crucial to explore booktokers to understand them better and to assess why they accept and use technologies to generate and share content about books on TikTok. This paper intends to explore the acceptance and use of TikTok to generate and share content about books on this social network as a booktoker. For this, a qualitative methodology of semi-structured interviews with thirteen Latin American booktokers was used. The analysis was developed through a thematic analysis based on the UTAUT2 theory to explain the findings. The results indicated the categories: performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, habit, price value, and a new identified category related to the generation of community and networks. This study constitutes one of the pioneering works in the exploration of the booktoker phenomenon.

## RESUMEN

La lectura es fundamental para el aprendizaje en la educación. Si bien, la revolución digital ha transformado la manera de incentivar la lectura y conocer sobre libros. Las personas que generan y comparten contenidos sobre libros en las redes sociales son los denominados booktokers, influencers de la red social TikTok, quienes se han convertido en un fenómeno emergente en esta sociedad conectada, pero aún con poca investigación que los estudie en profundidad. Por lo tanto, es crucial explorar para una mayor comprensión a los booktokers y conocer por qué aceptan y usan tecnologías para generar y compartir contenidos sobre libros en la red social TikTok. Se pretende explorar la aceptación y uso de TikTok para generar y compartir contenido sobre libros en esta red social siendo un booktoker. Se utiliza una metodología cualitativa de entrevistas semiestructuradas a trece booktokers latinoamericanos. El análisis se desarrolló por medio de un análisis temático basándonos en la teoría UTAUT2 para explicar los hallazgos. Los resultados indicaron las categorías: la expectativa de desempeño, la expectativa de esfuerzo, la influencia social, las condiciones facilitadoras, la motivación hedónica, el hábito, el precio y una nueva categoría identificada que es generar comunidad y redes. La contribución de este estudio es ser uno de los pioneros en la exploración del fenómeno de los booktokers.

## KEYWORDS | PALABRAS CLAVE

Virtual communities, education, influencer, reading, books, social media.  
Comunidades virtuales, educación, influencer, lectura, libros, redes sociales.





## 1. Introduction

Reading enables learning and communication; it helps us understand and learn about the world. In this sense, recreational reading of books is consistently associated with literacy benefits (Jerrim & Moss, 2019), reading comprehension (Torppa et al., 2020), spelling (Mol & Bus, 2011), and other areas such as mathematics (Sullivan & Brown, 2013). However, statistics indicate that reading motivation among adolescents is declining (Darmawan, 2020). According to UNESCO (2017), there is a learning crisis among children and adolescents in terms of minimum reading literacy, compromising Sustainable Development Goal number 4 (SDG 4), which seeks to achieve quality education. While the statistics are troubling, social media use among teenagers is on the rise. This poses a great opportunity to encourage reading through social media platforms among teenagers. Currently, TikTok is one of young people's favorite social networks, exceeding 1 billion monthly active users in the world (La Tercera, 2021); and by the year 2025, an estimated 1,327 million users (Fernández, 2021). This platform enables users to easily create, edit and share short videos (Hudders et al., 2021), akin to what occurs among Spanish and British teenage tiktokers (Suárez-Álvarez & García-Jiménez, 2021).

In social networks we find the social media influencer (SMI), who is: "First and foremost a content generator: someone who has a status of expertise in a specific area, who has cultivated a considerable number of captive followers, who has marketing value for brands by regularly producing valuable content delivered through social networks" (Lou & Yuan, 2019: 59). A specific type of SMI is one that generates and shares content about books on social media. While there are similar phenomena such as booktubers who share content about books on YouTube (Albrecht, 2017; Tomasena, 2019; Vizcaíno-Verdú et al., 2019), we did not find a specific definition of SMIs who share content about books in literature, therefore, in this article a definition is proposed: "A book SMI is one who has a status and expertise in books, with a considerable number of followers on social media platforms, and who has a marketing value for publishers and other related brands, regularly producing and sharing valuable content about books on social networking sites".

On the other hand, books and reading can be promoted on TikTok (Merga, 2021), here we find a type of book SMI, which is the booktoker. Booktokers generate and share content about books on TikTok. They create a community of readers where recommendations and discussions of books, writings, characters, and fictional places are allowed, known as booktok. This community can stimulate adolescents' interest in books and participation in reading (Merga, 2021).

There are studies focusing on the ways in which TikTok encourages learning (e.g., Escamilla-Fajardo et al., 2021; Literat, 2021). However, the importance of booktokers has received limited academic attention. According to our literature review (conducted as of December 13, 2021) there are only a few scientific publications, including Merga (2021), Jerasa and Boffone (2021) obtained from the Web of Science and the conference proceedings by Dezuanni (2021) who explore the booktok community phenomenon. Merga (2021) explores the presence of hashtags, authors, books, and book topics promoted in 116 TikTok videos analyzed. Meanwhile, Jerasa and Boffone (2021) refer to booktok as a useful literacy practice in English Language Arts classrooms among youth.

Furthermore, Dezuanni (2021) analyzes young Australians, indicating that adolescents can purposefully learn about books from microcelebrities on digital platforms. Given the above, more research is needed on SMI in other areas of study, such as books, which is an area that has not been explored in depth (Yé et al., 2021). In addition, studies are needed in emerging regions such as Latin America, as the study of SMI is mainly concentrated in developed countries (Vrontis et al., 2021). Therefore, the current literature review shows that previous SMI research has not yet incorporated TikTok, and it is important to explore it in detail (Hudders et al., 2021; Vrontis et al., 2021).

While further exploration of this phenomenon is needed to gain a deeper understanding of booktokers, it is necessary to understand their motivators on TikTok for educational and learning purposes. For this reason, we posed the following research question: why do booktokers accept and use TikTok to generate and share book content? To answer this exploratory question, a qualitative method was implemented with semi-structured interviews with 13 booktokers in Latin America.

### 1.1. Theoretical framework

There are several models that focus on studying technology acceptance and use. One of the most recognized is Venkatesh et al.'s (2012) Unified Theory of Acceptance and Use of Technology 2 (UTAUT2), an extension of Venkatesh et al.'s (2003) UTAUT. UTAUT is a theory that combines several models and includes four different central determinants that influence the behavior of technology acceptance and use. These are: performance expectancy, which in this study refers to booktokers' beliefs about the benefit of TikTok for sharing content about books; effort expectancy, understood as the degree of ease associated with the booktoker's perceived use; social influence, which refers to the booktoker's perception of how significant others (family, friends, followers, etc.) think he or she should use TikTok to share content about books; and facilitating conditions, such as the booktoker's perception of the resources and support available for using TikTok to share content about books (Venkatesh et al., 2003).

In contrast, the UTAUT2 (Venkatesh et al., 2012) incorporates three new determinants: hedonic motivation or enjoyment, understood as the booktoker's perception of pleasure or enjoyment derived from using TikTok to share content about books; price value, which refers to the booktokers' cognitive trade-off between the perceived benefits and cost of using TikTok to share content about books; and habit, which refers to the extent to which the booktoker tends to perform repeated behaviors on TikTok to share content about books.

In the context of educommunication, as a theoretical lens for UTAUT, it is perceived as a suitable framework to study the predisposition to use technology for learning. Awotunde et al. (2019) analyze university students' acceptance and willingness to use social networking sites for learning purposes, finding a direct influence of the model variables on learning. Furthermore, Lawson-Body et al. (2018) evidence that the acceptance and use of e-book technologies among students improve their learning. Thus, in this study, UTAUT2 is adapted to a social media context where SMIs share content about books.

## 2. Methodology

The acceptance and use of TikTok has not been explored for booktokers. This study adopted a qualitative approach to understand this new phenomenon (Miles et al., 2019). Therefore, to answer the research question, we considered data obtained through semi-structured interviews from the perspective of the booktokers themselves. The interviews were conducted with 13 booktokers in Latin America who use TikTok to generate and share content about books with their followers. The interview design was structured in three sections with standardized open-ended questions (personal background, general questions about book content generation, and specific questions about TikTok acceptance and use). This allowed the researchers to set guidelines for the data collection stage (Miles et al., 2019).

The researchers conducted two pilot interviews to verify that the participants understood the questions posed. The data collection instrument was then refined and applied to the Latin American booktokers in the sample. The procedure to contact the study participants involved identifying and creating a database of booktokers in Latin America. The creation of this database was carried out using the hashtags #booktok and #booktoker on TikTok, which yielded a total of 67 booktokers. Subsequently, we contacted each booktoker via instant messaging on the platform. The message indicated the objective of the research and requested collaboration and participation in the study. The booktokers who agreed to participate in the study were sent an informed consent indicating identity protection and authorization for the audio recording of the interview. The interview was conducted by telephone, which has been established as a sound data collection method (Cohen et al., 2017; Gaitán & Piñuel, 1998).

The interviews were conducted between the months of July and August 2021, the average duration was approximately 24 minutes; they were audio-recorded and transcribed verbatim, the latter being a crucial step for data analysis (Cohen et al., 2017). New booktoker participants were incorporated until there was sufficient information for the study; the data saturation point was reached with participant number 13, as no additional data were obtained (Cohen et al., 2017; Miles et al., 2019). Table 1 presents the participant data. To analyze the data, the six stages of thematic analysis (TA) proposed by Braun and Clarke (2006) were carried out to identify, interpret and define categories (themes). TA is a flexible method that provides detailed descriptions of phenomena in education (Xu & Zammit, 2020).

Table 1. Participant data						
Interviewee	Gender	Age	Country	July – August 2021		First TikTok
				Followers	Following	
E1	Female	25	Chile	1,800	71	06-jun-21
E2	Female	22	Chile	5,129	239	30-nov-20
E3	Female	18	Argentina	187.9 K	754	05-may-20
E4	Female	18	Ecuador	18.3 K	1,070	20-dec-19
E5	Male	18	Uruguay	4,407	151	18-jan-21
E6	Female	22	Peru	58.2 K	281	13-apr-20
E7	Female	25	Chile	13.4 K	58	03-jul-21
E8	Female	19	Argentina	112.9 K	126	10-nov-20
E9	Female	21	Mexico	3,695	176	8-dec-2018
E10	Female	23	Mexico	8,833	135	31-jan-21
E11	Male	23	Chile	17.4 K	38	02-jan-21
E12	Female	31	Chile	1,750	300	10-jul-20
E13	Male	24	Chile	44.7 K	121	21-jan-21

### 3. Results

Results are presented interpreting the determinants of UTAUT2 (Venkatesh et al., 2012), from the booktokers' point of view. The findings showed eight categories that influenced the acceptance and use of TikTok to generate and share content about books. The categories include performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, habit, and price value. These are related to the UTAUT2 theory. Additionally, there is the category related to the generation of community and networks, which is new and original to the UTAUT2 theoretical framework in the context of booktokers and TikTok. The eight categories and subcategories are described below using the booktoker responses.

#### 3.1. Performance expectancy

The participants perceived TikTok as a useful social networking site, which allowed them to generate and share content about books. As subcategories, they indicate that the most highlighted aspect of the platform mentioned by the interviewees is the algorithm. They also indicate that content generation is dynamic and fast. In addition, they highlight the achievement of attracting sponsors; and finally, they consider that their relationship with followers and the feedback they receive is rewarding.

##### 3.1.1. The TikTok algorithm

The participants perceived the TikTok algorithm as very useful for generating and sharing content about books. E3: "The truth is that TikTok has a good thing, you can become a hit, a video can go viral and that video can give you 100-200 followers, I was lucky enough that one of my videos reached a million views and I got about 3,000 followers and it was like wow, that's what TikTok has, people interact and that interaction has a result, so it is impressive how the TikTok algorithm works in that regard".

##### 3.1.2. Content creation

The generation of content on TikTok is highly valued by booktokers. E9: "I like it because it delimits the time you have, you have to make quick, digestible content, because you have a limited amount of time, so you have to make dynamic content, now that they have increased it to three minutes you can still make a video, you have more time to talk about a specific book".

##### 3.1.3. Sponsors

Booktokers indicated that it is positive to have sponsors and that the resources they receive enrich their videos and their TikTok profiles. E13: "Almost everyone comments positively, I had never been involved in social networks and everyone is surprised that I can do this kind of thing and that I can achieve so much, that publishers contact you, send you books, that they want to participate in activities with you".

#### 3.1.4. Relationship with followers

The interviewees indicated that the relationship with the followers is rewarding. E8: "Another positive thing is that every time someone writes to me, they tell me that they bought a book because I recommended it and they end up loving it and it becomes that person's favorite book, that also fulfills me a lot".

### 3.2. Effort expectancy

Most of the interviewees indicated that TikTok was easy to use to generate and share content about books. They emphasized that at the beginning it is hard to understand the application because of its wide variety of tools, but that habit of use makes it easier. In other words, the practice and collaboration that exists on this platform affects the acceptance and use of TikTok to generate and share content about books.

#### 3.2.1. Application's ease of use

Booktokers indicated that using TikTok is easy for generating and sharing content about books. E5: "I think it is a very complete application, it allows you to do everything there, that is, you can edit the videos there, it is quite easy I think, it was hard for me to understand some things, but because it has many features, it has many tools, but I find that it is quite interactive [...] And as I said, you can edit everything there, so it is quite easy".

#### 3.2.2. The practice

The booktokers related effort to practice. E7: "But it took me about one or two months to learn how to use the application well, because it is a new application, where I didn't understand why, if I liked a video, then another one would appear, etcetera; but yes, at the beginning it was complicated, but then you get into the rhythm, but of course, with some practice, because practice is what makes you good at something".

#### 3.2.3. Collaboration

Booktokers related the level of effort needed to use TikTok to collaboration. E12: "It was actually super easy, it is not very difficult to understand the platform, but also, within the platform itself, there are people who show you the tricks behind it, how to record certain types of videos, how to use certain transitions, so you can get ideas from there".

### 3.3. Social influence

Participants indicated that social influence is key to understanding the acceptance and use of TikTok to generate and share content about books. Given this, they identified that support from others such as family, followers, and friends is relevant. Likewise, influence from others and the effect of the COVID-19 pandemic.

#### 3.3.1. Support from others

Booktokers indicated that for them the influence of others, such as family, followers and friends is relevant to both initiating and continuing to share book content on TikTok. E1: "I have a younger sister who is 17 and she told me to download TikTok, I watched several videos on YouTube, such as some showing how to edit a video in TikTok, because I didn't know how to add the lyrics, how to forward it, I had no idea, it took me about two or three months to learn how to use the app to upload videos".

#### 3.3.2. Influence from others

Booktokers indicated that seeing other people talking about books prompted them to share knowledge about books on TikTok as well. E4: "I remember I saw a girl talking about books, and I was like, I still want to talk about books, I want to talk about that passion, without feeling like 'oh no, they want to listen to me!'".

### 3.3.3. The pandemic

Participants indicated that the COVID-19 pandemic influenced their use of TikTok to generate and share content about books. E6: "It was 2020, we were in pandemic, but it was almost the end of the year and during that whole year I had read a lot, and I just went to follow books and that kind of stuff on TikTok, there I saw what other girls who shared and listened to audios where I also came up with ideas and so I started creating videos, some of which had quite a few likes".

### 3.4. Facilitating conditions

Booktokers indicated that technological and literary resources (books) were relevant to generate and share content about books on TikTok with their followers. E2: «The truth is that I think you can make content with a phone, internet, a phone or a camera, I often share content when I don't even have the physical book, for example, I have the possibility of having a Kindle, but before, when I didn't have it I used the phone or printed the cover of the book and put it on another one and it was the book I had read, the thing was to show it more than having the book itself, and yes I think you don't need so much, to make content. Before I didn't have a ring of light, nor a tripod and, for example, I had those selfie tips and I tied them with shoelaces to a table leg so I could have a tripod and take pictures or record whatever, so I think that of course, over time I have been able to upgrade the devices I have, but with a cell phone, you can do many things, I think that a ring of light, having the book or having many books is not necessary to create content [...] with just a little you can do a lot".

### 3.5. Hedonic motivation

Booktokers indicated that for them, their acceptance and use of TikTok is related to enjoying what they do by sharing content about books. Specifically, they stated that they enjoy and like sharing their passion for books on TikTok.

#### 3.5.1. Enjoyment

Booktokers indicated that they enjoy making and sharing videos on TikTok. E10: "I have a lot of fun, in a way that you can't imagine, I stand in front of my window and start performing a trend, because I have a lot of fun performing trends, it also forces me to think creatively and that is very much appreciated, because I enjoy it a lot, if I didn't like it I would have quit a long time ago, but the truth is that it is very entertaining".

#### 3.5.2. Sharing a passion for books

Interviewees indicated that sharing their passion for reading with others motivates them to use and embrace TikTok. E4: "I think it's important that, if you are afraid to start a channel, that is, create booktok, create booktagram, leave that fear behind and start because it's a super nice adventure to talk about books, to share that passion, to meet new people makes you really happy".

### 3.6. Habit

Participants indicated that the routine of recording videos about books and feeling addicted to TikTok has an influence on their use and acceptance.

#### 3.6.1. Recording routine

Booktokers associated their video recording routine on TikTok with continued use of the platform, which would drive the use and acceptance of TikTok. E8: "Maybe I'm used to it, I don't know how to explain it, it's like when you get up in the morning and have breakfast, you don't think about it, you just do it because you're hungry, this is something similar, something happens to you and you record, in other words, it's sort of automatic".

#### 3.6.2. Sense of addiction

Respondents indicated that they experience a feeling of addiction when using TikTok to share content about books. E3: "I can't live without TikTok [...], in fact, there was a time when TikTok went down



and the follower count showed 000, I almost died, seriously, and then I realized that it was happening to everyone else”.

### 3.7. Price value

Participants did not comment on the fact that TikTok is free to use; rather, they refer to the time investment they make to prepare their content on books. E11: “I record when I feel like it, but I organize myself with the timing, because the videos I make generally recommend a book and I have to prepare a small script, and that takes some time, and recording involves several takes to get it right, so I give myself at least the day before I upload the video to prepare it”.

### 3.8. Generate community and networks

Finally, booktokers indicated that the community arises from common tastes, bonding, networking opportunities and the ability to influence others’ reading. This fosters the use and acceptance of TikTok to share content about books.

#### 3.8.1. Common tastes

Booktokers considered reading as a means of relating to others. E2: “It was just the fact that I liked to read, while I had a friend who also read, we didn’t read the same things, so I felt like, who can I discuss this with? I need someone to talk to, someone to tell how I’m feeling about the book, these are my favorite books, and things like that, I was already following other accounts, I was consuming a lot of BookTok”.

#### 3.8.2. Bonding

Participants indicated that generating and sharing content about books on TikTok has allowed them to form friendships. E4: “I think what I liked the most was meeting people from several Latin American countries, at the American and Canadian level, because we got to share that notion of books and I never thought in my life that I would have a friend who lives in Mexico, a friend who lives in Argentina, a friend who lives in Chile, so the notion of sharing it with someone and knowing that you are not the only one who has that passion for books was very nice”.

#### 3.8.3. Opportunities through the network

Booktokers identified that networking allows them to leverage their work. E7: “Two very beautiful things happened to me if I look at them from the booktokers side, one is that, through TikTok, people knew that I wrote a book because I was telling them about it [...] A girl met me and she introduced me to her publisher, that was like a very good thing: to find the publisher who trusted me to publish my book, so for me that was very important. And the second part is that I was invited to a school talk to share my experience writing a book, uploading it to the Wattpad platform and talking about that, I think that’s what I liked the most about the booktok on TikTok. But that is what being a booktoker is all about: being able to achieve your dreams when you persist, that was what I liked the most, being able to talk about my book, to talk to younger generations and get my book published”.

#### 3.8.4. Influencing others

The interviewees indicated that in their community they can influence others through the book-related content they generate. E13: “There was a lady who sent me a message saying that her 12-year-old daughter had started reading books that I like, books that I had recommended through my social networks, thanks to my videos and that she was very happy, because they had tried to instill reading in her and had not succeeded and somehow social networks convinced her, that day was the girl’s birthday and she told me if I could send a greeting to the girl, I really liked that, I like the closeness with people and having an impact on the girl’s life, hopefully it will be positive”.

## 4. Discussion

The performance expectancy category is identified by booktokers, who value TikTok, for the multiple benefits it gives them in generating and sharing book-related content. This is consistent with studies

conducted in a learning context by Mensah and Onyancha (2021) and Williams et al. (2021) who find a significant influence of performance expectancy on the acceptance of social networks for academic libraries. Likewise, Gunasinghe et al. (2019) found that performance expectancy among academics is considered relevant in predicting technology acceptance. Within this category one of the benefits is the TikTok algorithm, which booktokers indicated helps them grow by having more followers and interaction. Likewise, Su et al. (2020) in their study of TikTok stated that the algorithm promotes viral content.

On the other hand, booktokers also valued content generation because of the multiple tools that this platform provides in its role of generating attractive content, as indicated by Vizcaíno-Verdú and Tirocchi (2021). Another important output is the satisfaction generated among followers so that they buy books that a booktoker suggested. This is also evident in the case of booktubers (SMI of books on Youtube) (Albrecht, 2017). Another benefit of social networks are the profits generated. In the case of booktubers they engage in commercial collaborations with brands (with advertising contracts), with publishers, at book fairs or with self-published writers, which are similar for booktokers, these collaborations may or may not be monetary, such as invitations to events, premieres, meetings with writers, or free book shipments in exchange for reviews (Tomasena, 2019). This study also indicates that, in exchange for these collaborations, booktubers publicly thank publishers on their social media profiles (e.g., Instagram, Twitter and Facebook). Also, they indicate that they broadcast unboxing videos and hold contests that involve book raffles among their followers (Tomasena, 2019). This can also be evidenced on TikTok, according to the participants' responses. In the effort expectancy category, similar to Mensah and Onyancha (2021), Williams et al. (2021) found that this determinant influences the intention to use social networks in academic libraries; similarly, Gunasinghe et al. (2019) in an academic context. This is verified in the findings provided by the booktokers in this study, where they value the ease with which TikTok can generate and enable them to share content about books, indicating that it is even easier than other social networks. This is evidenced in the study by Su et al. (2020), where they indicate that TikTok fulfills a content-generating function thanks to numerous easy-to-use templates and unique visual effects.

Within the social influence category, aspects such as support and influence from others and the COVID-19 pandemic emerge as subcategories that explain this determinant. This is similar to the study by Williams et al. (2021) who analyzed the perceived acceptance of social networking sites for academic libraries in students, finding that social media acceptance is positively affected by social influence. From the field of educommunication, the results are in line with Awotunde et al. (2019) who point out that peer social influence among students is relevant in the acceptance of social networking sites for learning purposes. In relation to the facilitating conditions category, booktokers indicated that technological and literary resources are necessary to generate and share content about books on TikTok. Other studies such as Mensah and Onyancha (2021) in Ghana indicate that facilitating conditions have a significant influence on the intention and subsequent use of social networking sites in academic libraries; and Gunasinghe et al. (2019) found that facilitating conditions are important in predicting technology acceptance in a study conducted with academics for online educational purposes.

In the hedonic motivation category, Gunasinghe et al. (2019) found that this domain is important in predicting technology acceptance in an online academic context. This is also validated in the present study by booktokers who indicate that enjoyment and sharing a passion for books are subcategories that explain the motivation to use and accept TikTok. This supports findings by Iodice and Papapicco (2021) which suggest that people who use TikTok use it to satisfy their need for fun and creativity. In addition, Omar and Dequan (2020) indicated that TikTok users participate to relax and for entertainment. Likewise, athletes who generate videos on TikTok are characterized as funny and entertaining (Su et al., 2020).

The habit category in booktokers relates to having a recording routine and feelings of addiction to TikTok. This is similar to the study by Gunasinghe et al. (2019) who found that habitus is important in predicting technology use and acceptance in an online academic context. Tomasena (2019) indicated in his study that booktubers constantly review their page to evaluate their performance in order to adapt their content to their followers' reactions. The price value category among booktokers does not relate to the price of the digital platform, because it is free, as indicated by Zhang (2021) where TikTok users can log in for free, review and make videos, trim songs, use filters and virtual props, and use special effects on

their TikTok content. So, although booktokers do not comment on the free use of this social network, they refer to the time investment to create content about books. Thus, the time and effort ratio reflect the cost they perceive on TikTok. This is in line with what is stated in UTAUT2 (Venkatesh et al., 2012), where value is defined as the trade-off between perceived value, time, and effort invested.

Finally, there is the new category related to community and network building which contributes to the UTAUT2 model, which explains common tastes, the creation of ties, the opportunities that arise in the network and the ability to influence others. These results support Parratt-Fernández et al. (2021) who indicated that booktubers started talking about books because they did not find people or spaces to discuss and give their opinion to others with the same literary interests, where the motivation is initiated to share readings and meet people with the same interests. Likewise, Sorensen and Mara (2014) confirm that booktubers are part of a networked knowledge community. Tomasena (2019) indicates that booktubers work to improve their editing and filming to build a community, interacting with their followers. Also, other social networks, such as Instagram, help produce reader interactions that build communities (Thomas, 2021). This is also the case with TikTok (Merga, 2021). Moreover, TikTok has not only become an entertainment community, but also an online education platform (Zhang, 2021), which can be seen in the present study where booktokers promote and encourage their followers to read.

## 5. Conclusion

This study highlights a new phenomenon on social media. Specifically, a new type of SMI that generates and shares book-related content on TikTok known as a booktoker, who attracts followers interested in reading and influences their motivation to read. Since this study is exploratory and done from the booktoker's perspective, the use of the UTAUT2 enables an expanded understanding of the acceptance and use of TikTok to generate and share content about books on this platform. Findings indicate that the determinants that explain the acceptance and use of TikTok by booktokers are performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, habit, price value and a new category identified for this phenomenon which relates to community and network building. Therefore, this study contributes theoretically to SMI literature, understanding the acceptance and use of this platform by book-related content creators on TikTok known as booktokers. To our knowledge, there is limited research focused on book-related SMIs on TikTok, which opens the door for further research on this phenomenon that contributes to promoting reading among adolescents. Also, it contributes by studying a geographical area such as Latin America, which has limited research on SMIs (e.g., Gonzalez-Carrion & Aguaded, 2019; Guíñez-Cabrera et al., 2020; Lisdero & Duperré, 2021). This study also contributes to the area of education and SMIs (e.g., Izquierdo-Iranzo & Gallardo-Echenique, 2020; Marcelo & Marcelo, 2021). The practical implications are twofold. Firstly, this study is important for education, as a deeper understanding of this new booktoker phenomenon can encourage educators, administrators of educational institutions, and decision-makers in government institutions to work together with booktokers to promote and encourage reading habits among adolescents. This would make it possible to strengthen SDG 4 in order to achieve quality education. Secondly, the implications of this research concern publishers and marketing specialists by recognizing booktokers as allies to promote book marketing.

Finally, this study has limitations and opportunities for future research. First, the study sample focused only on Latin American booktokers, using non-probability sampling. Therefore, the results of this research may lack generalizability, which should be considered when interpreting the results. It is encouraged to replicate this study in other contexts and to conduct research using a quantitative methodology in a more representative sample, to achieve greater validation and generalization of the findings. Second, this study was limited to the exploration of other factors that could be influential, such as the specific characteristics of each booktoker. Therefore, future researchers could consider these factors to achieve a more complete understanding of the analyzed phenomenon. Third, this study is limited to TikTok. It invites further exploration and research on other social networks that provide opportunities to generate and share content about books that encourage reading. Finally, this study highlights opportunities for additional research that analyzes this phenomenon from the perspective of followers, such as considering their reading habits and interest in reading.

## Authors' Contribution

Idea, N.K., K.M.; Literature review (state of the art), N.K., K.M.; Metodology, N.K., K.M.; Data analysis N.K., K.M.; Results, N.K., K.M.; Discussion and conclusions, N.K., K.M.; Writing (original draft), N.K., K.M.; Final revisions, N.K., K.M.; Project design and funding agency, N.K., K.M.

## Funding Agency

Business Management Department, Faculty of Business Administration at the University of Bío-Bío, Chillán, Chile.

## References

- Albrecht, K. (2017). *Positioning BookTube in the publishing world: An examination of online book reviewing through the field theory*. [Master Dissertation, Leiden University]. <https://bit.ly/3nmNME3>
- Awotunde, J.B., Ogundokun, R.O., Ayo, F.E., Ajamu, G.J., Adeniyi, E.A., & Ogundokun, E.O. (2019). Social media acceptance and use among university students for learning purpose using UTAUT model. In L. Borzemski, J. Swiatek, & Z. Wilimowska (Eds.), *Information systems architecture and technology: Proceedings of 40th Anniversary International Conference on Information Systems Architecture and Technology* (pp. 91-102). Springer. [https://doi.org/10.1007/978-3-030-30440-9\\_10](https://doi.org/10.1007/978-3-030-30440-9_10)
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education*. Routledge. <https://doi.org/10.4324/9781315456539>
- Darmawan, I.G.N. (2020). The changes in attitudes of 15-year-old Australian students towards reading, mathematics and science and their impact on student performance. *Australian Journal of Education*, 64(3), 304-327. <https://doi.org/10.1177/0004944120947873>
- Dezuanni, M.L. (2021). *TikTok's peer pedagogies-learning about books through #booktok videos*. AoIR Selected Papers of Internet Research. <https://doi.org/10.5210/spir.v2021i0.11901>
- Escamilla-Fajardo, P., Alguacil, M., & López-Carril, S. (2021). Incorporating TikTok in higher education: Pedagogical perspectives from a corporal expression sport sciences course. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 28, 100302-100302. <https://doi.org/10.1016/j.jhlste.2021.100302>
- Fernández, R. (2021). *TikTok-datos estadísticos*. Statista. <https://bit.ly/3EchXDE>
- Gaitán, J.A., & Piñuel, J.L. (1998). *Técnicas de investigación en comunicación social*. Síntesis. <https://bit.ly/3GVVRTx9>
- Gonzalez-Carrion, E., & Aguaded, I. (2019). Los instagrammers más influyentes de Ecuador. *Universitas*, 31, 159-174. <https://doi.org/10.17163/uni.n31.2019.08>
- Guiñez-Cabrera, N., Mansilla-Obando, K., & Jeldes-Delgado, F. (2020). La transparencia publicitaria en los influencers de las redes sociales. *Retos, Revista de Ciencias de la Administración y Economía*, 10, 265-281. <https://doi.org/10.17163/ret.n20.2020.05>
- Gunasinghe, A., Abd-Hamid, J., Khatibi, A., & Ferdous-Azam, S.M. (2019). The adequacy of UTAUT-3 in interpreting academicians' adoption to e-Learning in higher education environments. *Interactive Technology and Smart Education*, 17(1), 86-106. <https://doi.org/10.1108/ITSE-05-2019-0020>
- Hudders, L., De-Jans, S., & De-Veirman, M. (2021). The commercialization of social media stars: a literature review and conceptual framework on the strategic use of social media influencers. *International Journal of Advertising*, 40(3), 327-375. <https://doi.org/10.1080/02650487.2020.1836925>
- Iodice, R., & Papapicco, C. (2021). To be a TikTok in COVID-19 era: An experience of social influence. *Online Journal of Communication and Media Technologies*, 11(1), e202103. <https://doi.org/10.30935/ojcm/9615>
- Izquierdo-Iranzo, P., & Gallardo-Echenique, E.E. (2020). Studygrammers: Learning influencers. [Estudigramers: Influencers del aprendizaje]. *Comunicar*, 62, 115-125. <https://doi.org/10.3916/C62-2020-10>
- Jerasa, S., & Boffone, T. (2021). BookTok 101: TikTok, digital literacies, and out-of-school reading practices. *Journal of Adolescent & Adult Literacy*.
- Jerrim, J., & Moss, G. (2019). The link between fiction and teenagers' reading skills: International evidence from the OECD PISA study. *British Educational Research Journal*, 45(1), 181-200.
- Lawson-Body, A., Willoughby, L., Lawson-Body, L., & Tamandja, E.M. (2018). Students' acceptance of E-books: An application of UTAUT. *Journal of Computer Information Systems*, 60(3), 256-267. <https://doi.org/10.1080/08874417.2018.1463577>
- Lisdero, P.M., & Duperré, J.L. (2021). Work, training and social sensibilities: Analysis from the experiences of fit-fluencers. *SN Social Sciences*, 1(8), 1-18. <https://doi.org/10.1007/s43545-021-00208-4>
- Literat, I. (2021). Teachers Act Like We're Robots": TikTok as a window into youth experiences of online learning during COVID-19. *AERA Open*, 7(1), 1-15. <https://doi.org/10.1177/2332858421995537>
- Lou, C., & Yuan, S. (2019). Influencer marketing: How message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, 19(1), 58-73.
- Marcelo, C., & Marcelo, P. (2021). Educational influencers on Twitter. Analysis of hashtags and relationship structure. [Influencers educativos en Twitter: Análisis de hashtags y estructura relacional]. *Comunicar*, 68, 73-83. <https://doi.org/10.3916/C68-2021-06>
- Mensah, M., & Onyancha, O.B. (2021). *Ghanaian academic libraries' use of social media: A structural equation modelling approach*. Information Development. <https://doi.org/10.1177/0266666921999429>

- Merga, M.K. (2021). How can Booktok on TikTok inform readers' advisory services for young people? *Library & Information Science Research*, 43(2), 101091-101091.
- Miles, M., Huberman, M., & Saldana, J. (2019). <https://bit.ly/31ZJ1bn>
- Mol, S.E., & Bus, A.G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, 137(2), 267-296.
- Omar, B., & Dequan, W. (2020). Watch, share or create: The influence of personality traits and user motivation on TikTok mobile video usage. *International Association of Online Engineering*, 14(4), 121-137.
- Parratt-Fernández, S., Mera-Fernández, M., & Sánchez, J. (2021). Nuevos prescriptores literarios: características sociodemográficas y autopercepciones del booktuber en España. *Ocnos*, 20, 56-67. [https://doi.org/10.18239/ocnos\\_2021.20.2.2454](https://doi.org/10.18239/ocnos_2021.20.2.2454)
- J. Sierra-Sánchez, & A. Barrientos-Báez (Eds.). *Cosmovisión de la comunicación en redes sociales en la era postdigital*. <https://bit.ly/30q1rRG>
- Sorensen, K., & Mara, A. (2014). Booktubers as a networked knowledge community. In M. Limbu, & B. Gurung (Eds.), *Emerging pedagogies in the networked knowledge society: Practices integrating social media and globalization* (pp. 87-99). IGI Global. <https://doi.org/10.4018/978-1-4666-4757-2.ch004>
- Su, Y., Baker, B.J., Doyle, J.P., & Yan, M. (2020). Fan engagement in 15 seconds: Athletes' relationship marketing during a pandemic via TikTok. *International Journal of Sport Communication*, 13(3), 436-446. <https://doi.org/10.1123/ijsc.2020-0238>
- Suárez-Álvarez, R., & García-Jiménez, A. (2021). TikTok, una aproximación de los tiktokers adolescentes españoles y británicos. In J. Sierra-Sánchez, & A. Barrientos-Báez (Eds.), *Cosmovisión de la comunicación en redes sociales en la era postdigital* (pp. 1037-1050). McGraw-Hill. <https://bit.ly/30q1rRG>
- Sullivan, A., & Brown, M. (2013). *Social inequalities in cognitive scores at age 16: The role of reading*. Centre for Longitudinal Studies. <https://bit.ly/3z3kiNH>
- L. Tercera (Ed.). (2021). *TikTok alcanza los 1000 millones de usuarios*. <https://bit.ly/31YF9Xo>
- Thomas, B. (2021). The# bookstagram: Distributed reading in the social media age. *Language Sciences*, 84. <https://doi.org/10.1016/j.langsci.2021.101358>
- Tomasena, J.M. (2019). Negotiating collaborations: BookTubers, the publishing industry, and YouTube's ecosystem. *Social Media+ Society*, 5(4). <https://doi.org/10.1177/2056305119894004>
- Torppa, M., Niemi, P., Vasalampi, K., Lerkkanen, M.K., Tolvanen, A., & Poikkeus, A.M. (2020). Leisure reading (but not any kind) and reading comprehension support each other-A longitudinal study across grades 1 and 9. *Child development*, 91(3), 876-900. <https://doi.org/10.1111/cdev.13241>
- UNESCO (Ed.) (2017). *Cifra en 617 millones a los niños y adolescentes sin conocimientos mínimos en lectura y matemáticas*. Naciones Unidas. <https://bit.ly/3laPxxE>
- Venkatesh, V., Morris, M.G., Davis, G.B., & Davis, F.D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 27(3), 425-478. <https://doi.org/10.2307/30036540>
- Venkatesh, V., Thong, J.Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS quarterly*, 36(1), 157-178. <https://doi.org/10.2307/41410412>
- Vizcaíno-Verdú, A., Contreras-Pulido, P., & Guzmán-Franco, M.D. (2019). Reading and informal learning trends on YouTube: The booktuber. [Lectura y aprendizaje informal en YouTube: El booktuber]. *Comunicar*, 59, 95-104. <https://doi.org/10.3916/C59-2019-09>
- Vizcaíno-Verdú, A., & Tirocchi, S. (2021). Tiktokers y objetivación sexual de género en retos musicales. *Media Education*, 12(1), 7-16. <https://doi.org/10.36253/me-9674>
- Vrontis, D., Makrides, A., Christofi, M., & Thrassou, A. (2021). Social media influencer marketing: a systematic review, integrative framework and future research agenda. *International Journal of Consumer Studies*, 45(4), 617-644. <https://doi.org/10.1111/ijcs.12647>
- Williams, M.L., Saunderson, I.P., & Dhoest, A. (2021). Students' perceptions of the adoption and use of social media in academic libraries: A UTAUT Study. *Communicatio*, 47(1), 76-94. <https://doi.org/10.1080/02500167.2021.1876123>
- Xu, W., & Zammit, K. (2020). Applying thematic analysis to education: A hybrid approach to interpreting data in practitioner research. *International Journal of Qualitative Methods*, 19, 1-9. <https://doi.org/10.1177/1609406920918810>
- Ye, G., Hudders, L., De-Jans, S., & De-Veirman, M. (2021). The value of influencer marketing for business: A bibliometric analysis and managerial implications. *Journal of Advertising*, 50(2), 160-178. <https://doi.org/10.1080/00913367.2020.1857888>
- Zhang, Z. (2021). Infrastructuralization of Tik Tok: Transformation, power relationships, and platformization of video entertainment in China. *Media, Culture & Society*, 43(2), 219-236. <https://doi.org/10.1177/0163443720939452>



## *Alfamed* media education curriculum for teachers

Ignacio Aguaded, Daniela Jaramillo-Dent & Águeda Delgado-Ponce (coords.)



Octaedro  Editorial

Updated guide on media and information literacy for educators, in which participated 22 researchers from 12 countries of America and Europe.



# The relationship of Twitter with teacher credibility and motivation in university students

Relación de la red social Twitter con la credibilidad docente y la motivación del alumnado universitario

- Dr. Facundo Froment. Researcher, Department of Social Psychology, University of Seville (Spain) (fpfroment@us.es) (<https://orcid.org/0000-0002-2337-3032>)
- Dr. Alfonso-Javier García-González. Professor, Department of Social Psychology, University of Seville (Spain) (alfonsoj@us.es) (<https://orcid.org/0000-0002-0839-162X>)
- Dr. Julio Cabero. Full Professor, Department of Instruction and School Organization, University of Seville (Spain) (cabero@us.es) (<https://orcid.org/0000-0002-1133-6031>)

## ABSTRACT

This paper aimed to analyse the perceptions of university students in relation to the credibility of university instructors according to the tweets posted on their Twitter profiles and the academic motivation that these can generate. Thus, students' perceptions of teacher credibility are affected by what instructors post on their social media profiles. The participants in the study were 166 students from the Faculty of Education Sciences at the University of Seville, carried out using a quasi-experimental methodology for which three Twitter profiles for a university professor were created with professional, social and mixed content. For the analysis of the results, normality and homoscedasticity tests were carried out on the sample to decide which statistical tests to use. The most salient results indicated that students perceived the professional profile as more competent, more caring, and more trustworthy, followed by the mixed profile and, lastly, by the social profile. Positive correlations were also found between students who perceived the university professor as reflected in the professional profile as credible in their academic motivation. The findings and their practical implications for instructors are discussed in terms of how they can incorporate Twitter into their teaching, considering the needs of their students to enhance their learning.

## RESUMEN

El presente trabajo pretendió analizar las percepciones del alumnado universitario en relación con la credibilidad del profesorado universitario según los tweets publicados en sus perfiles de Twitter y con la motivación académica que estos pueden generar. Así, las percepciones de los estudiantes sobre la credibilidad de los docentes se ven afectadas por lo que estos publican en sus perfiles de redes sociales. Los participantes del estudio fueron 166 estudiantes de la Facultad de Ciencias de la Educación de la Universidad de Sevilla, llevándose a cabo mediante metodología cuasiexperimental para lo que se crearon tres cuentas de Twitter de una profesora universitaria con contenido profesional, social y mixto. Para el análisis de resultados se ejecutaron pruebas de normalidad y homoscedasticidad de la muestra para decidir las pruebas estadísticas a emplear. Los resultados más destacados indicaron que los estudiantes percibieron como más competentes, con mejor voluntad y más confiables los perfiles profesionales, seguidos de los perfiles mixtos y, en último lugar, de los perfiles sociales. Asimismo, se encontraron correlaciones positivas entre los estudiantes que percibieron a la profesora universitaria del perfil profesional como creíble y su motivación académica. Se discutieron los hallazgos logrados y sus implicaciones prácticas para los docentes en cuanto que pueden incorporar Twitter en su docencia, considerando las necesidades de sus estudiantes para potenciar su aprendizaje.

## KEYWORDS | PALABRAS CLAVE

Credibility, social networks, Twitter, motivation, perceptions, communication.  
Credibilidad, redes sociales, Twitter, motivación, percepciones, comunicación.



## 1. Introduction and state of the art

In recent years, social networks have become an integral part of our daily lives (iabSpain, 2020). These platforms have not only changed the way we communicate and socialise, but also our ability to produce and share information with the rest of the world (García-Martín & García-Sánchez, 2015; Ricoy & Feliz, 2016). Although not originally created for educational purposes, social networks have gradually evolved into a new channel for communication between instructors and students across Spanish universities (Fondevila et al., 2015). As highlighted by Gettman and Cortijo (2015), social networks have become important communication tools for professors and college students, providing both parties with new opportunities for contact and interaction, as well as for professional development (Greenhow et al., 2020), educational innovation (Matosas-López et al., 2021; Rodríguez-Gallego et al., 2019), and improved academic performance (Rodelo & Lizárraga, 2018). In short, social networking sites enable instructors and students to remain in constant contact, moving beyond the conventional classroom and creating new teaching and learning environments (Ean & Lee, 2016). That said, their use will largely depend on students' perceptions of their pedagogical efficacy (Deaves et al., 2019), which has been demonstrated during the current COVID-19 pandemic (Greenhow et al., 2021).

One of the most widely used social media sites in higher education is Twitter (Carpenter et al., 2020), having become the main communication tool university instructors use to interact with their students (Higueras-Rodríguez et al., 2020; Santoveña-Casal & Bernal-Bravo, 2019). In particular, Twitter has emerged as an effective teaching platform in the higher education domain, already proving to be a suitable tool for instructors when sharing information with students (Desselle, 2017). Both parties can benefit from the advantages that Twitter offers in the education sector, most notably its potential to improve teacher-student relationships (Sunday, 2021). From this perspective, the use of Twitter creates a more closely knit classroom environment, thus enriching the teacher-student dynamic (Hull & Dodd, 2017). However, what professors post and share on their Twitter profiles affects students' perceptions of teacher credibility (McArthur & Bostedo-Conway, 2012).

Teacher credibility is defined as whether students perceive their instructor as believable or not (McCroskey, 1992) and encompasses three dimensions: competence, caring, and trustworthiness (McCroskey & Teven, 1999). Teacher competence refers to the student's perception of the instructor's knowledge and mastery of the subject they teach; caring is the degree to which students perceive instructors as showing interest in their wellbeing; and trustworthiness refers to students' perceptions of teacher reliability and kindness (Teven & McCroskey, 1997). In this regard, Johnson (2011) found that students who viewed an instructor's Twitter profile featuring social content rated them higher in credibility than those students who viewed the Twitter profile of an instructor posting academic content. In contrast, DeGroot et al. (2015) reported that an instructor's Twitter profile with professional content was deemed more credible than a profile with social or mixed content, also highlighting that a profile with mixed content was perceived as more credible than a profile with social tweets. Lastly, Clark-Gordon and Goodboy (2020) found that an instructor who tweeted professional information was perceived as higher in competence, whereas an instructor who tweeted personal information was perceived as more caring.

Teacher credibility also represents a key variable that impacts upon the teaching-learning process (Finn et al., 2009). From this perspective, teacher credibility affects the student's academic motivation (Martin et al., 1997), considered one of the most significant factors in student learning (Zheng, 2021). According to Brophy (2004), student motivation is defined as both a trait and a state. Motivation as a trait is a general and lasting predisposition towards learning, that is, it refers to an overall level of motivation across all learning situations, whereas motivation as a state refers to a specific learning situation and approach towards a particular class, task or content which is situation-dependent and ever-changing. Therefore, student motivation is, to an extent, influenced by environmental and contextual variables such as teacher behaviour (Jiang et al., 2021). As such, it is important to analyse students' perceptions of teacher behaviour in relation to academic motivation (Chan et al., 2021). Thus, to increase student academic motivation, instructors must be perceived as credible professionals by their students (Frymier & Thompson, 1992).

In terms of the influence that instructors' social media profiles have on the teaching-learning process, various research studies report that what instructors post on their Facebook profiles affects the student's

learning (Aubry, 2013; Imlawi & Gregg, 2014; Mazer et al., 2007; Saylag, 2013). Although most of the research to date on Twitter use in education has focused on higher education institutions (Greenhow et al., 2020), no published studies have been found which examine the effect of instructors' Twitter profiles on students' perceived credibility in relation to their academic motivation within the Spanish university context. Two research questions have emerged from the proposed theoretical framework:

- What type of content posted on Twitter by university instructors is likely to result in them being perceived by students as more credible?
- Is there a correlation between students' perceived teacher credibility according to the tweets posted by university instructors on their Twitter profiles and students' academic motivation?

## 2. Material and methods

### 2.1. Objectives

The overall objective of this study was to identify students' perceptions of teacher credibility based on the tweets that university instructors posted on their Twitter profiles. This overall objective can be broken down into two specific objectives:

- To establish which of the university instructors' Twitter accounts – professional, social, or mixed – are more credible based on the students' perceptions of the tweets posted on these profiles.
- To analyse the correlation between students' perceived teacher credibility based on tweets posted by instructors on their Twitter profiles and student academic motivation.

### 2.2. Research hypotheses

For the first specific objective, three hypotheses were formulated:

- H1. The university instructor's professional Twitter profile will be perceived as more credible by students than the university instructor's social Twitter profile.
- H2. The university instructor's mixed Twitter profile will be perceived as more credible by students than the university instructor's social Twitter profile.
- H3. The university instructor's professional Twitter profile will be perceived as more credible by students than the university instructor's mixed Twitter profile.

For the second specific objective, three hypotheses were developed:

- H4. There will be no correlation between students' perceived teacher credibility upon viewing the university instructor's social profile and their academic motivation.
- H5. There will be a positive correlation between students' perceived instructor credibility upon viewing the university instructor's professional profile and their academic motivation.
- H6. There will be a positive correlation between students' perceived instructor credibility upon viewing the university instructor's mixed profile and their academic motivation.

### 2.3. Research design

Three Twitter profile types were created for the purpose of this study: a profile with social content, another featuring professional content, and a third with mixed content. The social profile shows tweets relating to the instructor's personal life, including family, friends, and non-academic activities; the professional profile contains tweets about the instructor's teaching and research work; and the mixed profile is a combination of the instructor's professional and social tweets. Twenty-five (25) tweets were initially drawn up for the professional profile, and 25 tweets were developed for the social profile. Following evaluation by nine experts, the final count was 17 tweets for the social profile; 22 tweets for the professional profile; and 19 tweets for the mixed profile. Subsequently, a group of students ( $N=65$ ) assessed (a) the level of social content across all profiles; and (b) the level of professional content, reporting statistically significant differences among them ( $\chi^2=283.799$ ,  $p<0.01$ ).

Examples of social profile tweets include "There's no better way to start the day than with a morning gym session" and "This evening dinner with friends and then onto karaoke"; and examples of professional profile tweets are "From 8 to 10 August, Indiana University (United States) will host the International Conference on Educational Technology" and "In today's class on Basic Psychological Processes, I explained

Classical Conditioning. In this video you can watch the experiment conducted by Pavlov, pioneer of Classical Conditioning: <https://www.youtube.com/watch?v=kuAVOQixBI8>".

To minimise the number of study variables, common Spanish names and surnames were used for the university instructors' Twitter accounts: Laura García for the social profile; Marta García for the professional profile; and Cristina García for the mixed profile. The profile picture was a beach sunset for all accounts, given that a photo of each instructor could affect the assessments of credibility made by the students (Sutherland et al., 2017). Participants were randomly assigned to the three experimental conditions: Twitter accounts of a university instructor containing professional, social, and mixed content (one account per type).

## 2.4. Research variables

### 2.4.1. Independent variable

- Central traits of the instructor reflected in their respective Twitter account at three defined levels: professional traits, social traits, and mixed traits.

### 2.4.2. Dependent variables

- Teacher credibility. This can be defined as the student's perception of whether the instructor is credible or not and encompasses three dimensions: competence, caring, and trustworthiness (McCroskey & Teven, 1999). Teacher credibility is the primary variable in the student's perception of the instructor, impacting heavily on teaching-learning processes (Froment et al., 2020).
- Academic motivation. According to Brophy (2004), motivation exhibited by students can be classified as both a trait and a state: motivation as a trait is a general and lasting predisposition towards learning, whereas motivation as a state is the student's attitude towards a class, task, or teacher in particular, and is affected by the student's perceptions of the instructor (Lin et al., 2017).

## 2.5. Participants

Purposive non-probability sampling (on the basis on accessibility) was used. Specifically, instructors teaching on the aforementioned degree programmes and willing to take part in the research were recruited, thus enabling access to students. This experimental study included 166 participants from the Faculty of Education Sciences at the University of Seville aged between 18 and 24 years ( $M=20.75$ ,  $SD=2.47$ ), namely students enrolled in the first, second and third years of the following undergraduate programmes: Primary Education, Early Childhood Education, and Education Studies. Participant distribution by degree was 29 in Primary Education (17.5%); 94 in Early Childhood Education (56.6%); and 43 in Education Studies (25.9%). The number of students by academic year was 49 for Year 1 (29.5%); 50 for Year 2 (30.1%); and 67 for Year 3 (40.4%). The distribution of participants by sex was 146 females (88%) and 20 males (12%).

Among the 166 participants, 144 had their own Twitter profile (68.7%), whereas 52 did not have a Twitter account (31.3%). Regarding frequency of use among participants with a Twitter account, 24.7% confirmed they log into their Twitter profile every day; 19.9% hardly ever log into their Twitter account; 12% never go into it; and 12% only occasionally go into their Twitter profile. Participants reported using Twitter 2.98 hours a week on average ( $ST=3.16$ ) and usually post an average of 2.60 tweets a week ( $SD=2.86$ ). 51.2% of participants reported never having searched for a university instructor on Twitter, whereas 17.5% confirmed having done so. Furthermore, 60.2% reported not following any university instructors on Twitter, whereas 8.4% did so. Specifically, 11 participants (6.6%) followed one; two participants (1.2%) followed two; and one participant (0.6%) followed four instructors.

As for students' opinions about university instructors using Twitter, 87.3% of participants considered it appropriate for university instructors to have their own Twitter profiles, whereas 12.7% deemed it inappropriate. However, 54.2% thought it inappropriate for university instructors to use Twitter as a means of communicating with students, whereas 45.8% found it appropriate.



## 2.6. Instruments

Socio-demographic data questionnaire. The research team developed and implemented a socio-demographic data questionnaire which captures information about sex, age, studies undertaken by students, and academic year. This survey also includes questions relating to the frequency and use of Twitter by participants in their daily lives, as well as questions that elicit students' opinions about how university instructors use Twitter.

Source Credibility Measure (Escala de Credibilidad; Froment et al., 2019). This tool includes 18 bipolar adjectives, specifically six per dimension: "competence", "caring", and "trustworthiness". The respondent is asked to rate their perception of the instructor on a scale ranging from one to seven, considering that the closer the number of adjectives, the more certain the teaching evaluation. The measure was submitted to reliability analysis, yielding a Cronbach's alpha of .98 for the global scale.

Motivated Strategies for Learning Questionnaire (Cuestionario de Estrategias de Aprendizaje y Motivación; Martínez & Galán, 2000), selecting the academic motivation scale only. This scale comprises 31 items, with responses ranging from one ("That doesn't describe me at all") to seven ("That describes me perfectly"). The questionnaire was subjected to internal consistency analysis, yielding a Cronbach's alpha of .97 for the global scale.

## 2.7. Procedure

Participants were first given an informed consent form to complete, which detailed the nature and aim of the research; participant anonymity and data confidentiality were also ensured. The students were handed the socio-demographic data questionnaire; an instructions worksheet outlining the steps to follow to access the university instructors' Twitter accounts; the Source Credibility Measure; and the academic motivation scale. Participants were randomly assigned to each experimental situation: N=55 to the instructor's social profile; N=55 to the instructor's professional profile; and N=56 to the instructor's mixed profile.

Next, participants were asked to complete the socio-demographic data questionnaire. They were then instructed to sign into their own Twitter accounts using a computer or mobile phone and search for the instructor described on the s worksheet. They were given four minutes to read all the tweets on the corresponding profiles; the instructions sheet included a username and password which participants without an account could use to access Twitter. Lastly, once they had read the posted tweets, they were asked to first complete the Source Credibility Measure followed by the academic motivation scale.

The principal investigator was joined in the classroom by an additional investigator, who had the task to monitor and make sure that each participant viewed a specific profile, and that no information was being exchanged between participants. Once the study had ended, the students were thanked for taking part and asked to keep their participation in the experiment confidential.

## 2.8. Data analysis

Normality and homoscedasticity tests were carried out on the sample to decide which statistical tests to use. Non-normal distribution was identified, as was non-homogeneity of variance for the main study variables (Kolmogorov-Smirnov and Shapiro-Wilk tests,  $<.05$ ). The decision was taken to perform non-parametric tests as part of the statistical analyses for this research.

For the first specific objective, namely H1, H2 and H3, non-parametric Kruskal-Wallis tests were performed to compare independent groups, followed by the respective post hoc pairwise Mann-Whitney U tests. Additionally, for those comparisons yielding statistically significant differences, an r effect size was calculated as per the threshold values established by Cohen (1988): .10 to .30 as small; .30 to .50 as medium; .50. to .70 as large; and greater than .70 as very large. Lastly, Spearman's rank-order correlation was performed to address the second specific objective, namely H4, H5 and H6.

### 3. Analysis and results

#### 3.1. Differences between students' perceived teacher credibility by viewed Twitter profile

Table 1 outlines the descriptive results of students' perceived teacher credibility for each Twitter profile viewed. As shown below, students perceived the professional profile as most credible, followed by the mixed profile and, lastly, the social profile.

Table 1. Descriptive results of teacher credibility for each Twitter profile					
Twitter profiles	N	Minimum	Maximum	Mean	SD
Social profile (Laura García)	55	18	68	51.78	9.75
Professional profile (Marta García)	55	81	126	103.25	11.19
Mixed profile (Cristina García)	56	37	126	83.59	22.92

Table 2 shows the descriptive results of all three perceived teacher credibility dimensions according to each viewed Twitter profile. Students perceived the professional profile as more competent, more caring, and more trustworthy, followed by the mixed profile and, lastly, the social profile.

Table 2. Descriptive results of the teacher credibility dimensions for each Twitter profile												
Twitter profiles	Competence				Caring				Trustworthiness			
	Min.	Max.	M	SD	Min.	Max.	M	SD	Min.	Max.	M	SD
Social (Laura García)	6	24	17.27	4.39	6	22	14.81	4.19	6	24	19.69	3.63
Professional (Marta García)	25	42	36.01	4.34	22	42	33.03	4.24	26	42	34.19	4.05
Mixed (Cristina García)	6	42	28.55	9.26	12	42	26.27	7.43	13	42	28.75	8.05

Statistically significant differences were found for students' perceptions of teacher credibility according to the viewed Twitter profile; this was the case for global credibility and across all three dimensions ( $X^2$  Kruskal–Wallis=202.45,  $p<.01$ ) (Table 3).

Table 3. Comparisons between Twitter profiles for teacher credibility			
	Chi-square	df	p
Credibility	202.45	5	.000**
Competence	193.42	5	.000**
Caring	206.57	5	.000**
Trustworthiness	184.56	5	.000**

Note. \*\* $p<.01$ .

Table 4 shows the descriptive and inferential results for the comparisons between perceived teacher credibility by students who viewed the instructors' Twitter profiles. Students perceived the instructor's professional profile as more credible than that of the instructor's social profile (Mann–Whitney U test=.000,  $p<.01$ ). Similarly, students perceived the instructor with the professional profile as more competent (Mann–Whitney U test=.000,  $p<.01$ ); more caring (Mann–Whitney U test=1.500,  $p<.01$ ); and more trustworthy (Mann–Whitney U test=.000,  $p<.01$ ) than the instructor with the social profile. Regarding the effect size calculation, the differences in teacher credibility perceived by students who viewed the professional profile and social profile, respectively, were very large ( $r=.86$ ). The effect size differences in teacher competence, caring, and trustworthiness perceived by students who viewed the professional profile and social profile, respectively, were also very large ( $r=.86$ , for competence;  $r=.86$ , for caring; and  $r=.86$ , for trustworthiness). Thus, H1 is accepted. Students perceived the instructor's mixed profile as more credible than that of the instructor's social profile (Mann–Whitney U test=367.000,  $p<.01$ ). Similarly, students perceived the instructor with the mixed profile as more competent (Mann–Whitney U test=438.000,  $p<.01$ ); more caring (Mann–Whitney U test=297.000,  $p<.01$ ); and more trustworthy (Mann–Whitney U test=536.000,  $p<.01$ ) than the instructor with the social profile. Regarding the effect size calculation, the differences in teacher credibility perceived by students who viewed the mixed profile and social profile, respectively, were large ( $r=.64$ ). The effect size differences in teacher competence, caring, and trustworthiness perceived by students who viewed the mixed profile and social profile, respectively, were also large ( $r=.60$ , for competence;  $r=.69$ , for caring; and  $r=.55$ , for trustworthiness). Thus, H2 is accepted.

Students perceived the instructor's professional profile as more credible than that of the instructor's

mixed profile (Mann–Whitney U test=714.500,  $p<.01$ ). Similarly, students perceived the instructor with the professional profile as more competent (Mann–Whitney U test=760.000,  $p<.01$ ); more caring (Mann–Whitney U test=626.000,  $p<.01$ ); and more trustworthy (Mann–Whitney U test=904.500,  $p<.01$ ) than the instructor with the mixed profile. Regarding the effect size calculation, the differences in teacher credibility perceived by students who viewed the professional profile and mixed profile, respectively, were medium ( $r=.45$ ). The effect size differences in teacher competence, caring, and trustworthiness perceived by students who viewed the professional profile and mixed profile, respectively, were also medium ( $r=.42$ , for competence;  $r=.50$ , for caring; and  $r=.34$ , for trustworthiness). Thus, H3 is accepted.

Table 4. Comparisons between university instructor profiles for teacher credibility									
	Social profile		Professional profile	Social profile		Mixed profile	Professional profile		Mixed profile
	<i>U</i>	<i>p</i>	<i>r</i>	<i>U</i>	<i>p</i>	<i>r</i>	<i>U</i>	<i>p</i>	<i>r</i>
Credibility	.000	.000**	.86	367.000	.000**	.64	714.500	.000**	.45
Competence	.000	.000**	.86	438.000	.000**	.60	760.000	.000**	.42
Caring	1.500	.000**	.86	297.000	.000**	.69	626.000	.000**	.50
Trustworthiness	.000	.000**	.86	536.000	.000**	.55	904.500	.000**	.34

Note. \*\* $p<.01$ .

### 3.2. Correlation between teacher credibility and academic motivation

Table 5 shows the relationship between teacher credibility and students' academic motivation according to the viewed Twitter profile. Teacher credibility ( $r=.06$ ,  $p>.05$ ), competence ( $r=.10$ ,  $p>.05$ ), caring ( $r=.15$ ,  $p>.05$ ), and trustworthiness ( $r=.15$ ,  $p>.05$ ) perceived by students who viewed the instructor's social profile does not correlate with academic motivation. Thus, H4 is accepted. Teacher credibility ( $r=.58$ ,  $p<.01$ ), competence ( $r=.54$ ,  $p<.01$ ), caring ( $r=.49$ ,  $p<.01$ ), and trustworthiness ( $r=.51$ ,  $p<.01$ ) perceived by students who viewed the instructor's professional profile correlates positively with academic motivation. Thus, H5 is accepted. Teacher credibility ( $r=.59$ ,  $p<.01$ ), competence ( $r=.65$ ,  $p<.01$ ), caring ( $r=.36$ ,  $p<.01$ ), and trustworthiness ( $r=.53$ ,  $p<.01$ ) perceived by students who viewed the instructor's mixed profile correlates positively with academic motivation. Thus, H6 is accepted.

Table 5. Correlation between teacher credibility and students' academic motivation				
Twitter profiles	Credibility Motivation	Competence Motivation	Caring Motivation	Trustworthiness Motivation
Social profile (Laura García)	.06	.10	.15	.15
Professional profile (Marta García)	.58**	.54**	.49**	.51**
Mixed profile (Cristina García)	.59**	.65**	.36**	.53**

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

## 4. Discussion and conclusions

The main aim of this study was to identify students' perceptions of teacher credibility based on the tweets posted by university instructors on their Twitter profiles. Regarding the first specific objective, the most salient results indicated that students perceived the professional profile as more credible, followed by the mixed and social profiles. This coincides with DeGroot et al. (2015), who found the professional profile to be perceived as the most credible, followed by the mixed and social profiles. However, our results contradict Johnson (2011), who found the social profile to be rated higher in perceived credibility than the professional profile. Interestingly, our findings both coincide with and contradict Clark-Gordon and Goodboy's (2020) study: on the one hand, professional profiles were also rated highest in perceived competence, yet on the other hand, these authors found that social profiles were not the highest rated for perceived caring.

Two main explanations can be drawn here. First, students are generally of the belief that university instructors' use of social media should be strictly professional, focusing primarily on academic-related matters (Hershkovitz & Forkosh-Baruch, 2017). Therefore, students tend to judge professional profiles more positively over other profile types, especially where we see instructors using social networking

accounts in ways which meet students' expectations of use. Second, and from the perspective that university instructors should use social networks in a professional manner, students might make a more negative evaluation of profiles that contain personal and irrelevant details, or which have no academic grounding; in other words, when students perceive information disclosed by teachers as lacking relevance, they view them in a negative light (Cayanus & Martin, 2004; Kromka & Goodboy, 2021). As Hosek and Presley (2018) argue, if the personal information shared does not address educational or academic matters, it will likely disorient students, and they will perceive these disclosures as being worthless and meaningless, thus affecting students' perceived teacher credibility.

Regarding the second specific objective, positive correlations were found between students' perceived teacher credibility – specifically, for those who viewed the professional and mixed profiles – and students' academic motivation. These findings coincide with earlier studies which indicate that students' perceived teacher credibility correlates positively with their academic motivation (Froment et al., 2021a; 2021b; Kulkarni et al., 2018; Pogue & AhYun, 2006). Similarly, our findings support previous research reporting how the information posted and shared on social networks by university instructors impacts upon student motivation (Aubry, 2013; Mazer et al., 2007; Saylag, 2013). Students generally believe that their motivation is, to some extent, determined by their perceptions of how the instructor conducts themselves (Amiryousefi & Geld, 2021), meaning that student motivation represents an important link between teacher behaviour and student learning (Liu, 2021).

However, no correlations were found between perceived teacher credibility by students who viewed the university instructor's social profile and academic motivation. As Beatty and Behnke (1980) claim, this may be since people are inclined to dismiss sources lacking credibility, meaning that students may have rejected information given by teachers which, in turn, results in little to no impact on their academic motivation. As pointed out by Hovland et al. (1954), for a message to be persuasive and exert an influence upon individuals, the source must be perceived as credible.

In terms of further research avenues, future studies would do well to establish which of the communication options offered by Twitter is the most determining factor in students' perceptions of teacher credibility. Besides the tweets posted by instructors, aspects including the number of tweets, who they follow on Twitter, the trends they tweet about, and how many followers they have could affect students' perceptions of teacher credibility. Thus, another suggestion would be to conduct a corpus-based discourse analysis of tweets using text mining tools and techniques to identify patterns and correlations derived from the wording of said tweets, given that this may impact on students' perceived credibility. Furthermore, students' own perceptions of Twitter may influence how they perceive teacher credibility. Hence, future research could benefit from examining students' perceptions of Twitter and its relationship to their perceptions of teacher credibility. In view of this, future studies should assess whether students' perceptions of teacher credibility based on instructors' Twitter profiles are likely to be affected by other variables such as the instructor's position or role within the university, or by their race or country of origin. Thus, this calls for studies in which students assess teacher credibility through social networks, to ascertain its impact on students' teaching evaluations. In this respect, Elhay and HersHKovitz (2019) argue that a teacher's behaviour outside of the classroom affects how students appraise classroom-based teaching practice. Lastly, and according to the impact that social networks have on the teacher–student relationship as well as on the student's learning (Camas et al., 2021; Sivakumar, 2020), there is a need to train future university instructors on how to best use social networking platforms as a digital educational tool. As suggested by Van-Den-Beemt et al. (2020), teachers should explore the educational benefits that social networks afford, focusing on how they impact the teaching–learning process. This highlights the need for research studies that analyse the relationship between instructors' Twitter profiles and other variables at play in the teaching–learning process, for example, student involvement, interest, and engagement. In short, social networks enrich pedagogical practice in education, making it crucially important to identify how they influence student learning (Fuentes-Cancell et al., 2021).

This research does, however, have some limitations, mainly relating to the sample's characteristics. The sample comprised mostly women; more male participation would have enabled comparisons between both sexes to determine whether sex is a contributing factor in students' perceptions of teacher credibility.

Moreover, only first, second and third-year undergraduate students took part in the research; involving students from across all academic years would have allowed for additional analyses to likely strengthen the results obtained. Lastly, participating students came from a limited number of education degrees. Inviting participants from across all Education Sciences degree programmes would have resulted in a more heterogeneous sample and would have enriched the achieved outcomes. Thus, a future research avenue would be to involve more instructors and students in this study.

Despite these limitations, the study findings have, on the one hand, broadened our understanding of the relationship between university instructors' use of social networks and students' perceptions of teacher credibility and, on the other hand, contributed to the advance of knowledge in the field by establishing a relationship between teacher credibility and academic motivation. Furthermore, the findings of this study raise important practical implications. For university instructors to be perceived as credible individuals by their students, they must post and share professional content on their Twitter profiles, that is, content relating to their teaching and research endeavours. As such, Twitter represents another platform via which students make judgements about teacher behaviours. As Qiu et al. (2012) point out, people perceive, form impressions, and make judgements about an individual's behaviour based on the tweets posted on their Twitter profiles. Thus, university instructors need to ensure that their Twitter profiles are as professional as possible (DiVerniero & Hosek, 2011). Furthermore, students have clear expectations about teacher behaviour (Frymier & Weser, 2001), and if students expect their instructors to engage with social networking platforms professionally, and these expectations are being met or exceeded through the instructor's behaviour, then these expectations are likely to break down positively, thus giving rise to more favourable evaluations of teacher credibility.

In summary, how university instructors use Twitter impacts on students' perceptions of teacher credibility, meaning that teachers need to exercise caution when it comes to the type of information they post on Twitter, given that credibility can suffer over the course of one's professional career (Helvie-Mason, 2011). Instructors need to be more aware of the importance attached to their social network use, and of the implications that said use has for students. Specifically, their approach to using social networking platforms can affect their credibility as education professionals and, consequently, the teaching-learning process. As highlighted by Zachos et al. (2018), social network use poses a challenge to the teacher community for promoting interactive environments that enhance student learning. What is more, instructors who envisage using Twitter in their teaching must consider the needs of their students to improve learning (Chapman & Marich, 2021). Therefore, it is important that instructors manage the type of content they share on their Twitter profiles to be perceived as credible teachers and to positively impact teaching-learning processes (Myers & Martin, 2018).

### Authors' Contribution

Idea, A.J.G.G., J.C.; Literature review (state of the art), F.F.; Methodology, F.F., A.J.G.G., J.C.; Data analysis, F.F.; Results, F.F., A.J.G.G.; Discussion and conclusions, F.F., A.J.G.G., J.C.; Drafting (original draft), F.F.; Final revisions, A.J.G.G., J.C.; Project Design and sponsorship, A.J.G.G.

### Funding Agency

This research was funded and supported as part of an Andalusian Autonomous Government ERDF R&D&I project (INV-2-2020-T-056) to promote intergenerational relationships via social networks within a European Union cofinancing framework, which links the reality of the young university population with older members involved in university programmes aimed at older groups.

### References

- Amiryousefi, M., & Geld, R. (2021). The role of redressing teachers' instructional feedback interventions in EFL learners' motivation and achievement in distance education. *Innovation in Language Learning and Teaching*, 15(1), 13-25. <https://doi.org/10.1080/17501229.2019.1654482>
- Aubry, J. (2013). Facebook-induced motivation shifts in a French online course. *TechTrends*, 57, 81-87. <https://doi.org/10.1007/s11528-013-0705-6>
- Beatty, M.J., & Behnke, R.R. (1980). Teacher credibility as a function of verbal content and paralinguistic cues. *Communication Quarterly*, 28(1), 55-59. <https://doi.org/10.1080/01463378009369358>
- Brophy, J. (2004). *Motivating students to learn*. Routledge. <https://doi.org/10.4324/9781410610218>
- Camas, L., Valero, A., & Vendrell, M. (2021). The teacher-student relationship in the use of social network sites for educational



- purposes: A systematic review. *Journal of New Approaches in Educational Research*, 10(1), 137-156. <https://doi.org/10.7821/naer.2021.1.591>
- Carpenter, J., Tani, T., Morrison, S., & Keane, J. (2020). Exploring the landscape of educator professional activity on Twitter: An analysis of 16 education-related Twitter hashtags. *Professional Development in Education*, (pp. 1-22). <https://doi.org/10.1080/19415257.2020.1752287>
- Cayanus, J.L., & Martin, M.M. (2004). An instructor self-disclosure scale. *Communication Research Reports*, 21(3), 252-263. <https://doi.org/10.1080/08824090409359987>
- Chan, S., Maneewan, S., & Koul, R. (2021). An examination of the relationship between the perceived instructional behaviours of teacher educators and pre-service teachers' learning motivation and teaching self-efficacy. *Educational Review*, (pp. 1-23). <https://doi.org/10.1080/00131911.2021.1916440>
- Chapman, A.L., & Marich, H. (2021). Using Twitter for civic education in K-12 classrooms. *TechTrends*, 65, 51-61. <https://doi.org/10.1007/s11528-020-00542-z>
- Clark-Gordon, C.V., & Goodboy, A.K. (2020). Instructor self-disclosure and third-party generated warrants: Student perceptions of professor social media use. *Western Journal of Communication*, 84(1), 79-97. <https://doi.org/10.1080/10570314.2019.1649453>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge. <https://doi.org/10.4324/9780203771587>
- Deaves, A., Grant, E., Trainor, K., & Jarvis, K. (2019). Students' perceptions of the educational value of Twitter: a mixed-methods investigation. *Research in Learning Technology*, 27, 1-15. <https://doi.org/10.25304/rlt.v27.2139>
- Degroot, J.M., Young, V.J., & Vanslette, S.H. (2015). Twitter Use and its Effects on Student Perception of Instructor Credibility. *Communication Education*, 64(4), 419-437. <https://doi.org/10.1080/03634523.2015.1014386>
- Desselle, S.P. (2017). The use of Twitter to facilitate engagement and reflection in a constructionist learning environment. *Currents in Pharmacy Teaching and Learning*, 9, 185-194. <https://doi.org/10.1016/j.cptl.2016.11.016>
- Diverniero, R.A., & Hosek, A.M. (2011). Students' Perceptions and Communicative Management of Instructors' Online Self-Disclosure. *Communication Quarterly*, 59(4), 428-449. <https://doi.org/10.1080/01463373.2011.597275>
- Ean, L.C., & Lee, T.P. (2016). Educational Use of Facebook by Undergraduate Students in Malaysia Higher Education: A Case Study of a Private University. *Social Media and Technology*, 41(1), 1-8. <https://doi.org/10.20897/lectito.201641>
- Elhay, A.A., & Hershkovitz, A. (2019). Teachers' perceptions of out-of-class communication, teacher-student relationship, and classroom environment. *Education and Information Technologies*, 24, 385-406. <https://doi.org/10.1007/s10639-018-9782-7>
- Finn, A.N., Schrodt, P., Witt, P.L., Elledge, N., Jernberg, K.A., & Larson, L.M. (2009). A meta-analytical review of teacher credibility and its associations with teacher behaviors and student outcomes. *Communication Education*, 58(4), 516-537. <https://doi.org/10.1080/03634520903131154>
- Fondevila, J.F., Mir, P., Crespo, J.L., Santana, E., Rom, J., & Puiggròs, E. (2015). La introducción de Facebook en el aula universitaria en España: la percepción del estudiante. *RELATEC*, 14(3), 63-73. <https://doi.org/10.17398/1695-288X.14.3.63>
- Froment, F., Bohórquez, M.R., & García, A.J. (2020). Credibilidad docente: Una revisión de la literatura. *Teoría de la Educación. Revista Interuniversitaria*, 32(1), 1-32. <https://doi.org/https://doi.org/10.14201/teri.20313>
- Froment, F., Bohórquez, M.R., & García, A.J. (2021). El impacto de la credibilidad docente y la motivación del estudiante en la evaluación de la docencia. *Revista Española de Pedagogía*, 79(280), 47-68. <https://doi.org/10.22550/REP79-3-2021-03>
- Froment, F., García, A.J., Bohórquez, M.R., & Checa-Esquivá, I. (2021). Adaptación y validación en español de la Escala de Motivación Estado en estudiantes universitarios. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 58(1), 117-126. <https://doi.org/10.21865/RIDEP58.1.10>
- Froment, F., García, A.J., Bohórquez, M.R., & García-Jiménez, E. (2019). Adaptación y validación en español de la escala de credibilidad en profesores universitarios. *Revista Iberoamericana de Diagnóstico y Evaluación*, 51(2), 61-76. <https://doi.org/10.21865/ridep51.2.05>
- Frymier, A.B., & Thompson, C.A. (1992). Perceived teacher affinity-seeking in relation to perceived teacher credibility. *Communication Education*, 41(4), 388-399. <https://doi.org/10.1080/03634529209378900>
- Frymier, A.B., & Weser, B. (2001). The role of student predispositions on student expectations for instructor communication behavior. *Communication Education*, 50(4), 314-326. <https://doi.org/10.1080/03634520109379258>
- Fuentes-Cancell, D.R., Estrada-Molina, O., & Delgado-Yanes, N. (2021). Las redes sociales digitales: una valoración socioeducativa. Revisión sistemática. *Revista Fuentes*, 23(1), 41-52. <https://doi.org/10.12795/revistafuentes.2021.v23.i1.11947>
- García-Martín, J., & García-Sánchez, J.N. (2015). Use of Facebook, Tuenti, Twitter and Myspace among young Spanish people. *Behaviour & Information Technology*, 34(7), 685-703. <https://doi.org/10.1080/0144929X.2014.993428>
- Gettman, H.J., & Cortijo, V. (2015). Leave Me and My Facebook Alone! Understanding College Students' Relationship with Facebook and its Use for Academic Purposes. *International Journal for the Scholarship of Teaching and Learning*, 9(1), 1-16. <https://doi.org/10.20429/ijstl.2015.090108>
- Greenhow, C., Galvin, S., Brandon, D., & Askari, E. (2020). A decade of research on K-12 teaching and teacher learning with social media: Insights on the state of the field. *Teachers College Record*, 122(6), 1-72. <https://doi.org/10.1177/016146812012200602>
- Greenhow, C., Staudt, K.B., & Galvin, S. (2021). Inquiring tweets want to know: #Edchat supports for #RemoteTeaching during COVID-19. *British Journal of Educational Technology*, 52(4), 1434-1454. <https://doi.org/10.1111/bjet.13097>
- Helvie-Mason, L. (2011). Facebook, "friending" and faculty-student communication. In C. Wankel (Ed.), *Teaching art and science with the new social media, cutting-edge technologies in higher education* (pp. 61-87). Emerald Group Publishing Limited. <https://doi.org/10.1108/s2044-9968>

- Hershkovitz, A., & Forkosh-Baruch, A. (2017). Teacher-student relationship and Facebook mediated communication: Student perceptions. [La relación profesor-alumno y la comunicación en Facebook: Percepciones de los alumnos]. *Comunicar*, 53, 91-101. <https://doi.org/10.3916/c53-2017-09>
- Higuera-Rodríguez, L., Medina-García, M., & Pegalajar-Palomino, M.C. (2020). Analysis of concepts of active and trainee teachers. *Education Sciences*, 10(8), 1-16. <https://doi.org/10.3390/educsci10080200>
- Hosek, A.M., & Presley, R. (2018). College Student perceptions of the (in) appropriateness and functions of teacher disclosure. *College Teaching*, 66(2), 63-72. <https://doi.org/10.1080/87567555.2017.1385587>
- Hovland, C.I., Janis, I.L., & Kelley, H.H. (1954). Communication and persuasion: Psychological studies of opinion change. *American Sociological Review*, 19(3), 355-357. <https://doi.org/10.2307/2087772>
- Hull, K., & Dodd, J.E. (2017). Faculty use of Twitter in higher education teaching. *Journal of Applied Research in Higher Education*, 9(1), 91-104. <https://doi.org/10.1108/JARHE-05-2015-0038>
- iabSpain (ed.) (2020). *Estudio redes sociales 2020*. iabSpain. <https://bit.ly/3vLF82R>
- Imlawi, J., & Gregg, D. (2014). Engagement in online social networks: The impact of self-disclosure and humor. *International Journal of Human-Computer Interaction*, 30(2), 106-125. <https://doi.org/10.1080/10447318.2013.839901>
- Jiang, Y., Lee, C.K.J., Wan, Z.H., & Chen, J. (2021). Stricter teacher, more motivated students? Comparing the associations between teacher behaviors and motivational beliefs of Western and East Asian learners. *Frontiers in Psychology*, 11, 1-10. <https://doi.org/10.3389/fpsyg.2020.564327>
- Johnson, K.A. (2011). The effect of Twitter posts on students' perceptions of instructor credibility. *Learning, Media and Technology*, 36(1), 21-38. <https://doi.org/10.1080/17439884.2010.534798>
- Kromka, S.M., & Goodboy, A.K. (2021). The effects of relevant instructor self-disclosure on student affect and cognitive learning: a live lecture experiment. *Communication Education*, 70(3), 266-287. <https://doi.org/10.1080/03634523.2021.1900583>
- Kulkarni, S., Afshan, N., & Motwani, J. (2018). The impact of faculty member's communication behaviors on student satisfaction: The role of cognitive and affective learning and student's motivation. *International Journal of Productivity and Quality Management*, 25(4), 444-458. <https://doi.org/10.1504/ijpqm.2018.096087>
- Lin, Y., Durbin, J.M., & Rancer, A.S. (2017). Perceived instructor argumentativeness, verbal aggressiveness, and classroom communication climate in relation to student state motivation and math anxiety. *Communication Education*, 66(3), 330-349. <https://doi.org/10.1080/03634523.2016.1245427>
- Liu, W. (2021). Does teacher immediacy affect students? A systematic review of the association between teacher verbal and non-verbal immediacy and student motivation. *Frontiers in Psychology*, 12, 1-13. <https://doi.org/10.3389/fpsyg.2021.713978>
- Martin, M.M., Chesebro, J.L., & Mottet, T.P. (1997). Students' perceptions of instructors' socio-communicative style and the influence on instructor credibility and situational motivation. *Communication Research Reports*, 14(4), 431-440. <https://doi.org/10.1080/08824099709388686>
- Martínez, J.R., & Galán, F. (2000). Estrategias de aprendizaje, motivación y rendimiento académico en alumnos universitarios. *Revista Española de Orientación y Psicopedagogía*, 11(19), 35-50. <https://doi.org/10.5944/reop.vol.11.num.19.2000.11323>
- Matosas-López, L., Luzardo-Briceño, M., Aguilar-Jiménez, A.S., & Jaimes-Carrillo, L. (2021). Relaciones entre redes sociales y recursos digitales de instrucción en la universidad: Comparativa España - Colombia. *Pixel-Bit*, 60, 77-93. <https://doi.org/10.12795/pixelbit.77522>
- Mazer, J.P., Murphy, R.E., & Simonds, C.J. (2007). I'll see you on "Facebook": The effects of computer-mediated teacher self-disclosure on student motivation, affective learning, and classroom climate. *Communication Education*, 56(1), 1-17. <https://doi.org/10.1080/03634520601009710>
- Mcarthur, J.A., & Bostedo-Conway, K. (2012). Exploring the relationship between student-instructor interaction on Twitter and student perceptions of teacher behaviors. *International Journal of Teaching and Learning in Higher Education*, 24(3), 286-292. <https://bit.ly/2UopLhu>
- McCroskey, J.C. (1992). *An introduction to communication in the classroom*. Burgess International. <https://bit.ly/31DZBNc>
- McCroskey, J.C., & Teven, J.J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communication Monographs*, 66(1), 90-103. <https://doi.org/10.1080/03637759909376464>
- Myers, S.A., & Martin, M.M. (2018). Instructor credibility. In M. L. Houser, & A. M. Hosek (Eds.), *The handbook of instructional communication: Rhetorical and relational perspectives* (pp. 38-50). Routledge. <https://bit.ly/396JMgm>
- Pogue, L., & Ahyun, K. (2006). The effect of teacher nonverbal immediacy and credibility on student motivation and affective learning. *Communication Education*, 55(3), 331-344. <https://doi.org/10.1080/03634520600748623>
- Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*, 46(6), 710-718. <https://doi.org/10.1016/j.jrp.2012.08.008>
- Ricoy, M.C., & Feliz, T. (2016). Twitter as a learning community in higher education. *Educational Technology & Society*, 19(1), 237-248. <https://bit.ly/31qNfUI>
- Rodelo, J.A., & Lizárraga, J. (2018). Correlación del uso de redes sociales con el rendimiento académico de estudiantes de nivel superior. *Revista de Investigación en Tecnologías de la Información*, 6, 62-67. <https://bit.ly/3eV61KV>
- Rodríguez-Gallego, M.R., López-Martínez, A., & Navarro-Montaño, M.J. (2019). Longitudinal study on social networks as a didactic method. *Magis*, 12(24), 85-104. <https://doi.org/10.11144/Javeriana.m12-24.lsn>
- Santoveña-Casal, S., & Bernal-Bravo, C. (2019). Exploring the influence of the teacher: Social participation on Twitter and academic perception. [Explorando la influencia del docente: Participación social en Twitter y percepción académica]. *Comunicar*, 27, 75-84. <https://doi.org/10.3916/c58-2019-07>
- Saylag, R. (2013). Facebook as a tool in fostering EFL teachers' establishment of interpersonal relations with students through self-disclosure. *Procedia - Social and Behavioral Sciences*, 82, 680-685. <https://doi.org/10.1016/j.sbspro.2013.06.329>

- Sivakumar, R. (2020). Effects of social media on academic performance of the students. *The Online Journal of Distance Education and e-Learning*, 8, 90-97. <https://bit.ly/3eUXcBR>
- Sunday, A.A. (2021). Use of Twitter technology in educational learning. *The Turkish Online Journal of Educational Technology*, 20(2), 189-196. <https://bit.ly/3EqChBC>
- Sutherland, C.A., Young, A.W., & Rhodes, G. (2017). Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties. *British Journal of Psychology*, 108(2), 397-415. <https://doi.org/10.1111/bjop.12206>
- Teven, J.J., & McCroskey, J.C. (1997). The relationship of perceived teacher caring with student learning and teacher evaluation. *Communication Education*, 46(1), 1-9. <https://doi.org/10.1080/03634529709379069>
- Van-Den-Beemt, A., Thurlings, M., & Willems, M. (2020). Towards an understanding of social media use in the classroom: A literature review. *Technology, Pedagogy and Education*, 29(1), 35-55. <https://doi.org/10.1080/1475939X.2019.1695657>
- Zachos, G., Paraskevopoulou-Kollia, E.A., & Anagnostopoulos, I. (2018). Social media use in higher education: A review. *Education Sciences*, 8(4), 194-105. <https://doi.org/10.3390/educsci8040194>
- Zheng, J. (2021). A functional review of research on clarity, immediacy, and credibility of teachers and their impacts on motivation and engagement of students. *Frontiers in Psychology*, 12, 1-10. <https://doi.org/10.3389/fpsyg.2021.712419>



# Comunicar

Next Issues



## **Comunicar 72 (2022-3):**

The disinformation society: The impact of fake news on the public sphere

### **Thematic Editors**

Dr. Guillermo López-García, University of Valencia (Spain)

Dr. Gianpietro Mazzoleni, University of Milan (Italy)

Dr. Eva Campos-Domínguez, University of Valladolid (Spain)

**Initial call:** 01-06-2021

**Last call:** 30-12-2021



## **Comunicar 73 (2022-4):**

Future Education: Prospective for sustainability and social justice

### **Thematic Editors**

Dr. Antoni Santisteban-Fernández, Autonomous Uni. of Barcelona (Spain)

Dr. Edda Sant Obiols, Manchester Metropolitan University (United Kingdom)

Dr. Gustavo A. González-Valencia, Autonomous Uni. of Barcelona (Spain)

**Initial call:** 01-09-2021

**Last call:** 28-02-2022



## **Comunicar 74 (2023-1):**

Education for digital citizenship: Algorithms, automation and communication

### **Thematic Editors**

Dr. Berta García-Orosa, University of Santiago de Compostela (Spain)

Dr. João Canavilhas, University of Beira Interior-Covilha (Portugal)

Dr. Suso Benítez-Baleato, Harvard University (United States)

**Initial call:** 01-11-2021

**Last call:** 30-05-2022



## **Comunicar 75 (2023-2):**

Youth, gender identity and power in digital platforms

### **Thematic Editors**

Dr. Adriana Gewerc, University of Santiago de Compostela (Spain)

Dr. Inés Dusel, Mexican National System of Researchers (SNI) (Mexico)

Dr. David Lee-Carlson, Arizona State University (United States)

**Initial call:** 01-04-2022

**Last call:** 30-09-2022



Quarterly scientific journal, bilingual in Spanish and English in all its articles, and abstracts in Portuguese, Chinese and Russian. With 29 years of edition and 1900 research articles published. Presence in 805 international databases, journal evaluation platforms, selective directories, specialized portals, newspaper catalogues... Rigorous and transparent blind system of evaluation of manuscripts audited in RECYT; International Scientific Board and public network of scientific reviewers of 1,020 researchers from 54 countries around the world.

Professional management of manuscripts through the OJS 3 Platform with ethical commitments of transparency and timeliness, anti-plagiarism systems (CrossCheck), alternative metrics (PlumX, Dimensions)... High level of visibility with multiple search systems, DOIs, ORCID, dynamic pdfs, epub and XML, with connection to document managers such as Mendeley, RefWorks, EndNote and scientific social networks such as academia.edu, ResearchGate.

Specialized in educommunication: communication and education, ICT, audiences, new languages...; monographs specialized in current issues. Double format: printed and online; digitally, accessible in full text, free of charge, for the entire scientific community and researchers around the world. Coeditions printed in Spanish and English for the whole world. Published by Comunicar, a non-profit professional association, veteran in Spain (33 years) in education and communication, which collaborates with many international centres and universities.

In active indexations in 2021/22, Comunicar is top worldwide: 2nd in the world in SCOPUS and 7th in the world in JCR (top 1% and 3% in the world; percentile 99% and 97%). In JCR-JIF it is Q1 in Education, in Communication and in Cultural Studies (1st in Spanish). It is 1st in FECYT Metrics; 1st in DIALNET METRICS. In GOOGLE SCHOLAR METRICS is the 2nd journal indexed in Spanish in all areas; 2nd in REDIB (out of 1,199 journals).



Sponsored by:



$u^b$

UNIVERSITAT  
BERN



Published by:



**Grupo Comunicar**

[www.comunicarjournal.com](http://www.comunicarjournal.com)

[info@grupocomunicar.com](mailto:info@grupocomunicar.com)

ISSN: 1134-3478 / e-ISSN: 1988-3293