

Psicothema



Volumen 37, no. 4

ISSN 0214–9915 • eISSN 1886-144X

Colegio Oficial de Psicología del Principado de Asturias



Editor-in-Chief: Laura E. Gómez. Univ. de Oviedo, Spain
Deputy Editor: Rebeca Cerezo. Univ. de Oviedo, Spain
Associate Editors: Susana Al-Halabi. Univ. de Oviedo, Spain
 Jorge L. Arias, Univ. de Oviedo, Spain
 Isabel Benítez, Univ. de Granada, Spain
 Jorge F. del Valle, Univ. de Oviedo, Spain
 Eduardo Fonseca, Univ. de La Rioja, Spain
 José Carlos Núñez, Univ. de Oviedo, Spain
 Javier Suárez-Álvarez. Univ. of Massachusetts Amherst, USA
 Paz Suárez-Coalla, Univ. de Oviedo, Spain
Honorary Editor: José Muñiz, Univ. Nebrija, Spain

Managing Editor: Gloria García-Fernández, Univ. de Oviedo, Spain
Editorial Office: Leticia García, COPPA, Spain
 M.ª Ángeles Gómez, COPPA, Spain

ADVISORY BOARD

Olivia Afonso. Oxford Brookes Univ., UK
 Leandro Almeida. Univ. do Minho, Portugal
 David Álvarez. Univ. de Oviedo, Spain
 Marta Álvarez Cañizo. Univ. de Valladolid, Spain
 Rui Alves. Univ. of Porto, Portugal
 Antonio M. Amor González. Univ. de Salamanca, Spain
 Constantino Arce. Univ. de Santiago, Spain
 Diego Ardura. UNED, Spain
 Natalia Arias. Univ. Nebrija, Spain
 Ignacia Arruabarrena. Univ. del País Vasco, Spain
 Roger Azevedo. Univ. of Central Florida, USA
 Giulia Balboni. Univ. de Perugia, Italy
 Elisardo Becoña. Univ. de Santiago, Spain
 Mónica Bernaldo de Quirós. Univ. Complutense de Madrid, Spain
 Ana Bernardo. Univ. de Oviedo, Spain
 Verner P. Bingman. Univ. Browling Green, USA
 Teresa Bobes Bascarán. Univ. de Oviedo, Spain
 Roser Bono. Univ. de Barcelona, Spain
 Amaia Bravo. Univ. de Oviedo, Spain
 José Luis Carballo. Univ. Miguel Hernández, Spain
 Thomas J. Carew. New York Univ., USA
 Juan L. Castejón. Univ. de Alicante, Spain
 José Pedro Espada. Univ. Miguel Hernández, Spain
 Martin Debbané. Univ. de Ginebra, Switzerland
 Paula Elosua. Univ. del País Vasco, Spain
 Joyce L. Epstein. Univ. John Hopkins, USA
 Rubén Fernández Alonso. Univ. de Oviedo, Spain
 Sergio Fernández Artamendi. Univ. de Sevilla, Spain
 Concepción Fernández. Univ. de Oviedo, Spain
 María Fernández Sánchez. Univ. de Salamanca, Spain

Pere J. Ferrando. Univ. Rovira i Virgili, Spain
 Victoria A. Ferrer. Univ. de las Islas Baleares, Spain
 Antonio León García Izquierdo. Univ. de Oviedo, Spain
 José M. García Montes. Univ. de Almería, Spain
 Alba González de la Roz. Univ. de Oviedo, Spain
 Francisco González-Lima. Univ. of Texas, USA
 Ana González Menéndez. Univ. de Oviedo, Spain
 Steve Graham. Arizona State Univ., USA
 Eve Griffin. National Suicide Research Fdn., Ireland
 Mattias Grünke. Univ. of Cologne, Germany
 Ana Hernández. Univ. de Valencia, Spain
 M.ª Dolores Hidalgo. Univ. de Murcia, Spain
 Stephen T. Higgins. Univ. Vermont, USA
 Cándido Inglés. Univ. Miguel Hernández, Spain
 Valentina Ladera. Univ. de Salamanca, Spain
 Christa Labouliere. Columbia University, USA
 Juan Alfonso Lara Torralbo. Univ. de Córdoba, Spain
 Alfonso Lara Torralbo. Univ. de Córdoba, Spain
 Susana Lázaro Visa. Univ. de Cantabria, Spain
 Pablo Livacic Rojas. Univ. de Santiago de Chile, Chile
 Mónica López. Univ. Groningen, The Netherlands
 José Antonio López Pina. Univ. de Murcia, Spain
 Urbano Lorenzo Seva. Univ. Rovira i Virgili, Spain
 Verónica Marina Guillén. Univ. de Cantabria, Spain
 Eduardo Martín Cabrera. Univ. de La Laguna, Spain
 Víctor Martínez Loreda. Univ. de Sevilla, Spain
 Ana Miranda. Univ. de Valencia, Spain
 Santiago Monleón Verdú. Univ. de Valencia, Spain
 Fabia Morales Vives. Univ. Rovira i Virgili, Spain
 Mª Lucía Morán. Univ. de Cantabria, Spain
 Patricia Navas Macho. Univ. de Salamanca, Spain

Javier Ortuño-Sierra. Univ. de La Rioja, Spain
 José Luis Padilla. Univ. de Granada, Spain
 Mercedes Paino. Univ. de Oviedo, Spain
 Alicia Pérez de Albéniz. Univ. de La Rioja, Spain
 Mª Carmen Pérez Fuentes. Univ. de Almería, Spain
 Gina Quirarte. UNAM, Mexico
 Celestino Rodríguez. Univ. de Oviedo, Spain
 Susana Rodríguez. Univ. da Coruña, Spain
 Javier Rodríguez Ferreiro. Univ. de Barcelona, Spain
 Mª Fe Rodríguez Muñoz. UNED, Spain
 Sonia Romero. Univ. Aut. de Madrid, Spain
 Pedro Rosário. Univ. do Minho, Portugal
 Mª Consuelo Sáiz Manzanares. Univ. de Burgos, Spain
 Louis A. Sass. Univ. Rutgers, USA
 David Scanlon. Boston College, USA
 Roberto Secades-Villa. Univ. de Oviedo, Spain
 Albert Sesé. Univ. de las Islas Baleares, Spain
 Giorgios Sideridis. Harvard Medical School, USA
 Steve Sirecci. Univ. of Massachusetts, USA
 Linda C. Sobell. Univ. Nova Southeastern, USA
 Miguel Ángel Sorrel. Univ. Autónoma de Madrid, Spain
 Mark Torrance. Nottingham Trent Univ., UK
 Luis Valero Aguayo. Univ. de Málaga, Spain
 Antonio Valle. Univ. de A Coruña, Spain
 Wouter Vanderplasschen. Ghent Univ., Belgium
 Antonio Verdejo-García. Monash Univ., AUS
 Eva Vicente. Univ. de Zaragoza, Spain
 Andreu Vigil. Univ. Rovira i Virgili, Spain
 Jianzhong Xu. Mississippi State Univ., USA
 Jin H. Yoon. Univ. of Texas, USA
 Izabela Zych. Univ. de Córdoba, Spain

Psicothema is indexed by Social Sciences Citation Index (WOS), Scopus, Google Scholar, SciELO, Dialnet, EBSCO Essentials Academic Search Premier, PubMed, PsycINFO, IBECs, Redinet, Psycodoc, Pubpsych, Fuente Académica Plus, IBZ Online, Periodicals Index Online, MEDLINE, EMBASE, ERIH PLUS, Latindex, MIAR, CARHUS Plus+ 2018, Rebiun, DOAJ, & Crossref.

D.L. AS 3779-1989 ISSN: 0214 - 9915 CODEN PSOTEG



This paper meets the requirements of ISO 9706:1994, Information & documentation – Paper for documents – Requirements for permanence, effective with Volume 7, Issue 2, 1995.

Publisher address:

► Colegio Oficial de Psicología del Principado de Asturias
 Ildefonso Sánchez del Río, 4 - 1º B
 33001 Oviedo (Spain)
 Tel.: +34 985 28 57 78 • Fax: +34 985 28 13 74
 E-mail: psicothema@cop.es • <http://www.psicothema.com>

PUBLICATION GUIDELINES

Psicothema publishes empirical work in English which is done with methodological rigor and which contributes to the progress of any field of scientific psychology. As an exception, the Editorial Board may accept publication of work in Spanish if the content justifies such a decision. Theoretical work may also be accepted, if requested by the Editorial Board, with preference given to articles that engage with critical research issues or which discuss controversial approaches.

Submission of articles

- Articles should be submitted via the journal's web page: www.psicothema.com (Authors section – submission of articles): <http://www.psicothema.es/submit>
- Submissions must comply with the rules for preparation and publication of articles, as well as the ethical standards specified below.
- Studies must be unpublished. Articles which have been fully or partially published elsewhere will not be accepted, nor will articles that are in the process of publication or which have been submitted to other journals for review. It will be assumed that all those who appear as authors have agreed to do so, and all those cited for personal correspondence have consented.
- The activities described in the published articles will comply with generally accepted ethical standards and criteria, both in terms of work with human beings and animal experimentation, as well as all aspects of professional and publishing ethics.
- The original work may be submitted in Spanish initially and receipt will be acknowledged immediately. If so, and if it is accepted, the authors will be responsible for translating it into English for publication.
- Authors may only submit one article for consideration by Psicothema per year.**
- Names and surnames should be entered on the platform in the form they will be cited (a single surname, two separate surnames, hyphenated surnames, etc.). The affiliation of all authors must be indicated. **A maximum of two affiliations per author may be indicated. Affiliations must follow the format "entity or university (country, in English)".** Do not include information about research groups or departments. Only one person may appear as corresponding author, who will be responsible for ensuring that the author names, order, and affiliations are correct.
- Authors should suggest three people who they believe would be suitable reviewers for the article, clearly indicating their institutional affiliation and email address. Authors may also indicate people who, for whatever reason, they do not wish to be involved in the review process for their work. Please bear in mind the recommendations from the Committee on Publication Ethics (COPE) when suggesting the three reviewers https://publicationethics.org/files/Ethical_guidelines_for_peer_reviewers_0.pdf
- Manuscripts are screened by the Editorial Board to assess relevance and interest for the journal and whether it follows the rules. Articles must faithfully conform to the editorial rules and fall within the editorial scope of the journal. It is a necessary, though not sufficient, condition that articles must comply with the rules for publication. Articles which do not follow Psicothema's rules will be rejected. In general, within around 10 days the Editorial Board will communicate a decision of interest to begin the review process.
- Psicothema is only able to publish about 10% of the manuscripts it receives, which is why we apply a very rigorous screening and selection system. Many submissions are considered **non-priorities** by the Editorial Board without being sent for review.
- If an article passes the Editorial Board screening, it will be sent to a minimum of two reviewers to evaluate its scientific quality. The journal has a **policy of "double blind" reviews**, meaning that both authors and reviewers are anonymous during the review process. To that end, manuscripts must

not contain information that would allow the authors to be identified. Most reviewers report back within the agreed three week period. The review process, from receiving an article to the decision to modify it or reject it, usually takes around two months.

- If, after receiving the reviewers' reports, the Editorial Board decides that the article needs "modifications" to be published, the authors should send the modifications in the requested format together with a point-by-point response to all the comments made by the reviewers and the Editorial Board. Failure to respond in the required format within the set timescale will lead to the article being rejected and removed from the management platform, with no possibility of re-submission.
- The Editorial Board is responsible for the final decision to accept the article for publication or not. The editors usually make their decisions as quickly as possible once they have received all the necessary reports.
- After an article has been accepted, and before publication, the authors must sign a copyright agreement. Printing rights and rights of reproduction in any format or medium belong to Psicothema, who will not reject any reasonable request from authors for permission to reproduce their contributions.
- It is the authors' responsibility to obtain relevant permissions to reproduce copyright-protected material. They are also responsible for disclosing possible conflicts of interest, declaring sources of funding and their participation in the research, and providing access, where necessary, to databases, procedure manuals, scores, and other experimental material that may be relevant. These aspects must be declared in the articles, as described below.

For any questions or clarifications, the journal can be contacted via the email address psicothema@cop.es

Manuscript preparation

- File format:** Articles must be sent in DOC or DOCX format. Microsoft Word documents must not be locked or password-protected, they should not have comments in the margins or information that might reveal the authors' identities. The file should be anonymised in "file properties" so that author information does not appear.
- Length:** The maximum length for articles is **6,000 words** (including the title, abstracts, key words, in-text references, acknowledgements, figures, and tables). The 6,000 word limit **does not include the list of references**. If authors wish to provide supplementary material, the article should include a unique, persistent web link (see point 18 about supplementary material).
- Format:** The articles must be in Microsoft Word format, using **12-point Times New Roman**, in a single column with 3 cm margins, paragraphs left-aligned and double spaced (except for tables and figures which may use single spacing). Page numbers must be included in the lower right corner. Limit sections and subsections to three levels of headings and follow the recommendations in the APA 7th edition about "Sentence case" in the list of references. Psicothema does not allow footnotes, annexes, or appendices. Any such content should be incorporated appropriately into the text (see point 18 about supplementary material).
- Language:** Although articles may be submitted and reviewed in Spanish, accepted articles are usually published in English. Once articles are accepted, the authors must provide an English translation of the reviewed article, within the indicated timeframe, for publication. Psicothema accepts American and British English, but not a mix of the two. Any text in English must be of appropriate professional quality, which will be reviewed by a professional native-speaking translator. Following that review, Psicothema may suggest changes, or if necessary, request a new translation or revision of the translation, the costs of which will be borne by the article's authors.
- Title page:** The first page of the article contains the article title in English and in Spanish, the running title (in English), the total number of words

in the article (not counting references) and a **declaration of authorship, originality and the fact that the work is previously unpublished**. This declaration is obligatory as one of the measures the journal takes to avoid plagiarism. The submitted text must be anonymized, avoiding use of the authors names or anonymizing other possible references that may identify them. Follow the APA 7th edition rules for capitalization of titles and subtitles (i.e., “Title case”). Use upper case for the first letter of all nouns, verbs, adjectives, adverbs, pronouns, and any word longer than three letters.

6. Title: The title should be short, descriptive, clear, accurate, and easy to read. It should engage the reader’s interest and name variables or topics addressed. Ensure that the main key phrase of the topic is in the article title and avoid superfluous words. Remember that searches normally use key phrases rather than individual words (for example, “mental health in people with disability” not just “health”). Try to include the topic at the start of the title. If the title is “creative”, add a more descriptive subtitle after a colon. A descriptive title will help the article to be found in databases. The Editorial Board reserves the right to change titles and abstracts of articles accepted for publication in order to follow the above rules and enhance the article’s impact and dissemination.

7. Abstracts and key words: the second page of the article contains the abstracts (in Spanish and English) and 3-5 key words or terms. Abstracts must be no more than 200 words and **structured** in four sections: Background, Method, Results, and Conclusions. The abstract should be a single paragraph with these titles in bold, followed by colons and upper case. The key words cover essential elements of the paper such as the research topic, population, method, or application of the results. Avoid general terms and empty words (pronouns, adverbs etc.), or redundant words such as analysis, description, research, etc. Nouns are preferred. Pay particular attention to selection of key words as they are used to index the article.

8. Article: The article introduction begins on the third page. The introductory section should not include the article title, or the subtitle “Introduction”, or subsections. Following that, the “Method” section should contain the following subsections “Participants”, “Instruments”, “Procedure”, and “Data Analysis”, and no others, in no other order, and with no other titles. Where appropriate, in the procedure section it is obligatory to provide information about ethical aspects of the study, the ethics committee that approved the study and the reference code (anonymized during the review process). For research with children, express mention must be made about obtaining informed consent. Pay particular attention to the APA rules about the presentation of statistical and mathematical results in the text, as well as tables and figures. At the end, there should be a single “Discussion” section which should include both discussion along with limitations and conclusions of the study. The discussion section should not have any subsections.

9. Declaration of author contributions: Where there is more than one author, there must be a declaration of responsibilities at the end of the article, before the references, specifying what contribution each of the authors made. To specify each author’s contribution, use the criteria established by the CRediT taxonomy (Contributor Roles Taxonomy; <https://credit.niso.org>). Please use the full name of each author as it appears in the manuscript to declare their contributions, followed by the CRediT roles performed. Follow this example: **John White:** Conceptualization, Methodology, Software. **Nuria García-Fernández:** Data curation, Writing - Original draft. **Lucinda Jackson:** Visualization, Investigation. **Laura Gayo:** Supervision, Software, Validation. **Michael Gutiérrez:** Writing - Review and Editing.

If a group of authors made equal contributions, please also use the CRediT taxonomy to specify their contributions: **John White:** Conceptualization, Writing – Original draft, Writing - Review and Editing. **Lucinda Jackson:** Conceptualization, Writing – Original draft, Writing review and Editing.

Psicothema does not permit the use of other formulas to indicate equal contributions, such as ‘contributed equally to this work’, co-first authors, co-last authors, or co-senior authors.

10. Corresponding author: Psicothema allows only **one corresponding author**, who will take primary responsibility for communication with the journal during the manuscript submission, peer review, and publication process, as well as for ensuring providing correct details of authorship

(including the names of co-authors, addresses and affiliations), ethics, acknowledgements, sources of funding, conflict of interests, and declarations. The corresponding author is responsible for having ensured that all authors have agreed to be so listed, and have approved the manuscript submission to the journal. After publication, the corresponding author is the point of contact for queries about the published paper. It is their responsibility to inform all co-authors of any matters arising in relation to the published paper and to ensure such matters are dealt with promptly.

11. Acknowledgements: any acknowledgements should be included at the end of the text, before the references, in a separate section titled “Acknowledgements”.

12. Sources of Funding: Priority will be given to work supported by competitive national and international projects. A section titled “Funding” must be included following the “Acknowledgements” section (if one is included) and before the list of references. The “Funding” section must clearly specify the funding body with the assigned code in brackets. It must also be clearly indicated whether the source of funding had any kind of participation in the study. If there was no participation, include the following sentence, “The source of funding did not participate in the design of the study, the data collection, analysis, or interpretation, the writing of the article, or in the decision to submit it for publication”. If no funding was received, add the following, “This study did not receive any specific assistance from the public sector, the commercial sector, or non-profit organizations”.

13. Conflict of interests: Authors must report any economic or personal relationship with other people or organizations that may inappropriately influence their work. If there are none, following the funding section, in a section titled “Conflict of Interest”, authors should state: “The author(s) declare(s) that there are no conflicts of interest”.

14. Declaration of availability of data: The authors should state, in a section titled “Data Availability Statement”, whether the research data associated with the article is available and where or under what conditions it may be accessed. They may also include links (where appropriate) to the dataset.

15. Reference style: Articles must be written following the guidelines in the 7th edition of the Publication Manual of the American Psychological Association. Articles that do not comply with these rules will be rejected. Some of the requirements are summarized below.

Bibliographical references in the text should include the author’s surname and year of publication (in brackets, separated by a comma). If the author’s name forms part of the narrative, it should be followed by the year in brackets. If there are more than two authors, only the first author’s surname is given, followed by “et al.” and the year; if there is confusion, add subsequent authors until the work is clearly identified. In every case, the references in the bibliography must be complete (up to 20 authors). When citing different articles in the same brackets, order them alphabetically. To cite more than one study from the same author or authors from the same year, add the letters a, b, c, as necessary, repeating the year (e.g., 2021a, 2021b).

The list of references at the end of the article must be alphabetical and comply with the following rules:

a) Books: Author (surname, comma, initials of first name(s) and a full stop); if there are various authors, separate them with a comma; before the final author use a comma and “&”; year (in brackets) and full stop. The full title in italics and full stop; finally, the publisher. For example:

Lezak, M., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). Oxford University Press.

b) Chapters of books with various authors, reports from conferences or similar: Author(s); year; title of the work being cited, followed by “In”, the director(s), editor(s), or compiler(s) and in brackets Ed., adding an s if plural; the title of the book in italics and in brackets the page numbers of the cited chapter; the publisher. For example:

de Wit, H., & Mitchell, S. H. (2009). Drug effects on delay discounting. In G. J. Madden & W. K. Bickel (Eds.), *Impulsivity: The behavioral and neurological science of discounting* (pp. 213-241). American Psychological Association.

c) **Journal articles:** Author(s); year; article title; full name of the journal in italics; volume number in italics; issue number in brackets with no space between it and the volume number; first and last page number. The doi should be included in URL format. For example:

Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>

For documents that do not have a doi, it is no longer necessary to use “Retrieved from”, instead give the URL directly. For example:

Walker, A. (2019, November 14). *Germany avoids recession but growth remains weak*. BBC News. <https://www.bbc.com/news/business-50419127>

d) Pay particular attention to the rules in the 7th edition of the APA manual for citing work presented in **conferences, doctoral theses, and software**, as well as the rules for the **use of acronyms in text** and in the references section.

e) When the original version of the cited work (book, chapter, or article) is **not in English**, cite the original title and give the English translation in square brackets (with no separation from the original, without using italics).

For further information and other cases, consult the 7th edition of the APA publication manual or the following page: <https://apastyle.apa.org/style-grammar-guidelines/references/examples>

16. Figures and tables should be included at the end of the manuscript, one per page. They should also follow the APA 7th edition guidelines, be appropriately numbered and cited in the text, indicating approximately where they should be placed. They must have a short, descriptive title that helps understand the content, and follow the APA recommendations about title case, with no full stop. They should be 7 or 14 cm wide and have clear, legible lettering and symbols. Avoid wasted space and make best use of the space available. Figures must be submitted in editable formats, consistent with the format of the rest of the article. If that is not possible, they must have a minimum resolution of 300ppp.

17. Pre-registration of studies and plans of analysis: as a general rule, Psicothema recommends pre-registering submitted studies. If authors have pre-registered studies or plans of analysis, links to that pre-registration should be provided in the article.

18. Supplementary material. Psicothema recommends sharing the data that has been used in the research and supplementary material in institutional or thematic open-access repositories, federated in the European Open Science Cloud (EOSC). Provide a web link if access is to be provided to databases or any other supplementary material, using unique, persistent identifiers.

19. We encourage authors to consult the following standard guidelines when preparing their manuscripts (although due to the multidisciplinary nature of the journal, this is not obligatory):

Case Reports - **CARE**

Diagnostic accuracy - **STARD**

Observational studies - **STROBE** (von Elm et al., 2008), **MQCOM** (Chacón et al., 2019) o **GREOM** (Portell et al., 2015)

Randomized controlled trial - **CONSORT** and **SPIRIT** (Hopewell et al., 2022)

Systematic reviews, meta-analyses – **PRISMA** (Page et al., 2020)

Test adaptation - **International Test Commission Guidelines** (Hernández et al., 2020)

Test development - **Ten steps for test development** (Muñiz & Fonseca, 2019)

Publication of articles

1. Publication rates: Psicothema is an “open access” journal. All of the articles will always be free to those who want to read or download them. In order to provide this open access, Psicothema charges a publication

fee which the authors or their funders must pay. The price depends on the length of the manuscript. In general, the average price per article is between €180 and €210, based on a mean of 6-7 pages per article, at €30 per laid-out page.

2. Print Proofs: Once an article has been accepted for publication, the contact person will receive an email with the print proofs in PDF format to check and correct spelling-typographical errors. Only minimal corrections can be made to the content of the article once it has been accepted. **Substantial modifications and changes will not be accepted** other than correcting printing or translation errors, possible errors detected during the review process, or incorporating suggestions made by the Editorial Board. No changes will be accepted in this phase to authorship, addition of new affiliations, or details such as including research groups or departments. Galley proofs should be checked carefully, following the instructions provided with them, to confirm that they match the accepted original. Corrected proofs should be returned within the requested timeframe (48-72 hours). Corrections must be made in the PDF file itself, no other means of correction will be accepted. It is vital to check that names, surnames, ORCID codes, and affiliations are all correct in this stage. The corresponding author is responsible for gaining approval from all co-authors for the corrected print proofs. If the proof article is not reviewed within the timeframe or manner specified, that version of the article will be published and subsequent changes or corrections will not be possible.

3. Published version: Once the edition of Psicothema containing the article is published, the author will receive a copy of their article in PDF format. The final version typeset by Psicothema will be available online via DOI. We strongly recommend sharing the final version published by Psicothema on social networks, (Facebook, Twitter, LinkedIn...), university and public repositories (Mendeley, Cosis...), scientific social networks (ResearchGate, Academia.edu, Kudos ...), personal and institutional websites, blogs, Google Scholar, ORCID, Web of Science ResearcherID, ScopusID...

Ethical standards

Psicothema is committed to the scientific community to ensure the ethical and quality standards of published articles. Its references are the “Core practices” defined by the **Committee on Publication Ethics (COPE)** for journal editors, the **American Psychological Association (APA)** Code of Conduct, and the Code of Ethics for Psychology from the **Spanish General Council of Psychology**.

Use of inclusive, non-sexist language. At Psicothema, we are firmly committed to equality and respect for all, recognizing and appreciating diversity. For this reason, authors should ensure that they use bias-free language, avoid stereotypes, and engage with inclusive, non-sexist language, albeit prioritizing grammatical correctness, economy of language, and accuracy, given the limitations of space. Pay particular attention to the presentation of data, so that participants’ characteristics are described and analysed properly, without presenting information that is irrelevant to testing hypotheses, achieving objectives, or presenting results of the study. Avoid condescending, obsolete, or inappropriate language, as well as the use of labels related to stereotypes. We recommend reporting where potential gender differences are found in the results.

Responsible authorship. Psicothema promotes transparency via the declaration of authors’ contributions. All signatories must have made substantial contributions in each of the following aspects: (1) conception and design of the study, or data acquisition, or analysis and interpretation of data, (2) drafting the article or critical review of the intellectual content, and (3) final approval of the submitted version. The list and order of authors should be carefully reviewed before the initial submission of the article. Any addition, removal, or re-ordering must be done before the article is accepted, with the approval of the Psicothema Editorial Board and the consent of all named authors. A form for this is available on request.

Open science. To facilitate the reproducibility of research and reuse of data, code, types of software, models, algorithms, protocols, methods, and any other useful material related to the project should be shared.

We recommend that authors publish the original study data in public open-access repositories online, such as FigShare (<http://figshare.com>), Mendeley Data (<https://data.mendeley.com/>), Zenodo (<http://zenodo.org/>), DataHub (<http://datahub.io>) and DANS (<http://www.dans.knaw.nl/>). Where data or supplementary material is shared, a corresponding reference should be included in the manuscript and the list of references, using unique, persistent identifiers.

Funding sources. In the acknowledgements section, authors should include data on the organizations that provided economic funding for the study or preparation of the article, and briefly describe the role any funding body played in designing the study, data collection, analysis, and interpretation, writing the article, or the decision to submit it for publication. If there was no participation from the funding body, this should be indicated as suggested in the “Preparation of Articles” section. The author responsible for submitting the article should include this metadata at the time of submission in the corresponding section.

San Francisco declaration on research assessment (DORA). As part of its commitment to open knowledge, Psicothema follows this initiative because it shares the need to address the quality assessment of scientific articles (not only the journals in which they are published), to consider the value and impact of all research outputs (including data and software), and to consider the societal impact of research from a broader perspective (including qualitative indicators, such as the influence on scientific policies and practices, together with a responsible use of quantitative indicators). To this end, it is committed to remove restrictions on the number of references that can be included in the bibliography, not counting them as part of the maximum number of words, to encourage responsible authorship practices and to provide information about the specific contributions of each author (CRediT), to mandate the citation of primary literature in favor of reviews in order to give credit to the group(s) who first reported a finding, and to make available a variety of journal-based metrics and article-level metrics (PlumX).

Good publishing practice in gender equality. Psicothema is committed to gender policies that lead to real equality between men and women in society through various actions: (1) pursuing equal proportions of women and men in the editorial team, as well as in those who review the articles; (2) recommending the use of inclusive language in scientific articles; (3) recommending that articles report whether the original study data considered sex or gender in order to identify possible differences; and (4) including the full names of the authors of published articles. To that end, authors must include their full names (not just first initials) in the metadata, which will appear in the published articles.

Authors' rights

Acknowledgement of receipt. Receipt of the article will be immediately communicated to the authors by email.

Screening. Articles will be reviewed by the Editor-in-Chief, Executive Editor, Managing Editor, and the Associate Editors. The editorial team may directly reject studies if, in their opinion, they do not follow the journal's publication rules, do not meet the minimum requirements, or do not fit the journal's objectives or priorities.

Review. Once past the Editorial Board screening, the articles will be reviewed by external reviewers and by the Associate Editor responsible for managing the article. The Associate Editor and the Managing Editor will consider the external reviewers' reports and will make the final decision on publication.

Reasoned reply. Except in cases of articles considered to be “non-priority” in the initial screening phase, authors will be given a reasoned response about the Editorial Board's final decision when that involves rejection (i.e., articles rejected after the peer review phase).

Confidentiality. Authorship of articles received will be kept anonymous and the evaluation process will be confidential, we commit to not disseminating the article more than necessary for the evaluation process and until the article is accepted for publication.

Use of data. The members of the Editorial Board will not use the results of unpublished work without the express consent of the authors.

Declaration of privacy. The names and email addresses provided to Psicothema will only be used for purposes established in the journal, they will not be given to third parties or used for commercial purposes.

Complaints and claims. Efforts will be made to respond to and resolve complaints and claims quickly and constructively. Complaints or claims should be sent by email to psicothema@cop.es, clearly and accurately specifying the nature of the complaint, the contact details of the person making it, and sufficient data to demonstrate any possible violation of the journal's declaration of ethics. Complaints about published content must be made as soon as possible after publication, and after having first contacted the corresponding authors to try and find a direct resolution. Psicothema may be contacted where it is not appropriate to contact the authors, if the authors do not respond, or if they do not resolve the issue. If possible, documentation must be included as evidence of the situation. Psicothema will acknowledge receipt of the complaint by email, and may request additional information or documentation for clarification. Depending on the nature or complexity of the issue, if the content is reviewed and sufficiently documented, the Editorial Board will study the case and make any decision in accordance with the directives of the Committee of Publishing Ethics (COPE). The Editor-in-chief will make the final decision and a response will be sent by email. Other people and institutions will be consulted as necessary, including university authorities or subject-matter experts, and legal advice may be sought if the complaint has legal implications. Complainants will have to expressly request that a complaint be treated confidentially and the Editorial Board will do so as far as appropriate and in line with our management processes. It is possible that complainants will not receive any information about the state of any investigation until a final decision is reached, and it is important to bear in mind that investigations may take some time. Complaints that are outside Psicothema scope or that are presented in an offensive, threatening, or defamatory manner will be dismissed. Personal criticism or comments are not acceptable. Communication will be terminated if it is not cordial and respectful, or if there is persistent vague or unfounded complaint. Psicothema reserves the right to take appropriate legal measures if a complainant insists on a complaint that is unfounded, false or malicious.

Authors' responsibilities

Editorial rules. Authors should read and accept the editorial rules and journal's instructions before sending a manuscript. While the article is undergoing the evaluation process at Psicothema, it must not be in any evaluation process at other journals.

Ethical rules. Authors must comply with the ethical standards specified in the Psicothema rules for authors.

License for public communication. The authors cede to Psicothema the public communication rights of their article for free dissemination through the internet, portals and electronic devices, through its free provision to users for online consultation, printing, download and archive, guaranteeing free, open access to the publication.

Publication licence. The authors accept the Psicothema copyright policy and cede it the right of publication. Psicothema publishes its articles under CC-BY-NC-ND license.

Reviewers' responsibilities

Editorial rules. Reviewers must read and accept the journal's editorial rules and instructions before reviewing an article. They must also follow the COPE ethical directives for reviewers.

Professional responsibility. Reviewers must only accept articles for review for which they have sufficient knowledge to perform a proper review.

Conflict of interests. Reviewers will constructively and impartially review articles for which they consider themselves qualified, abstaining from reviewing articles in which there might be a conflict of interest.

Confidentiality. Reviewers will respect the confidentiality of the review process and will not use information obtained during the peer review process for personal gain or to others' advantage, or to discredit or disadvantage others. They will not involve other people in the review process without the authorization of Psicothema.

Suspicion of a breach of ethics. Reviewers will inform the Editorial Board if they detect poor practice, fraud, plagiarism, or self-plagiarism, as well as any other irregularity related to research or publication ethics.

Deadline for reviews. Reviewers will commit to meeting the review timeframes set by Psicothema, informing the Editorial Board if they need additional time or are unable to send a report after having accepted a review request.

Preparation of the report. The format of Psicothema's review report is open, but reviewers must use a short scoring rubric. Reviewers must be objective and constructive in their reviews, offering feedback that will help authors improve their articles. Reviewers must make fair, impartial, constructive assessments of the article's strengths and weaknesses, and avoid disparaging personal comments or baseless accusations. They must not suggest that authors add references to the reviewer's own work (or that of colleagues) just to increase the number of citations or raise the visibility of their work or the work of associated; suggestions must only be based on valid academic reasons.

The journal's responsibilities

The Editorial Board is not responsible for the ideas or opinions expressed by the authors in the journal articles or the reviewers in their reports. The opinions and facts expressed in the articles are solely and exclusively the authors' responsibility and do not represent the journal's opinions or scientific policies. The editorial organization is not responsible in any case for the credibility or authenticity of the articles.

Psicothema will strive to avoid scientific fraud, which includes fabrication, falsification, or omission of data; plagiarism; duplicate publications; and authorial conflicts. Particular attention in plagiarism is paid to avoiding passing others' work off as one's own, co-opting others' ideas without recognition, giving incorrect information about the source of a reference, and paraphrasing a source without mentioning it. Detection of **fraud or plagiarism** will lead to the rejection of the submitted or published article.

The Psicothema Editorial Board undertakes to ensure that everyone involved (authors, reviewers, editors, and journal management) comply with the expected ethical standards in every phase of the publishing process, from reception to publication of an article, basing this on the recommendations from **The Committee on Publication Ethics (COPE)** to resolve possible conflicts.

The readers' rights

Readers have the right to read all articles published in Psicothema for free immediately after their publication.

Updating published articles

Psicothema is committed to correcting important scientific errors or ethical issues in published articles. In order to be transparent about any change, the following criteria and procedures have been established for updating our published articles.

Minor errors. Minor errors that do not affect the readability or meaning, such as spelling, grammar, or layout mistakes are not sufficient for and do not justify an update, regardless of the source of the error.

Metadata errors. Requests to correct errors in an article's metadata (for example, title, author name, abstract) must be made during the galley correction process. Once an article has been published, corrections can only be made if the Editorial Board believes the request to be reasonable and important. Once approved, the article will be updated and republished on the Psicothema website, with notification to relevant databases.

Author name and affiliation. Authors must make any desired changes to author names, surnames and affiliations during the galley correction process. Once the article has been published, no changes will be made without valid, convincing reasons, especially if the ORCID code has been supplied correctly. Changes to names after publication will only be in exceptional cases where the authors adopt a new name (such as for marriage or after gender transition) and want it updated. In such cases the Committee on Publication Ethics (COPE) recommendations will be followed.

Corrections. Requests may be submitted to correct errors that affect scientific interpretation. Once a request is approved, the article will be updated and re-published on the Psicothema website, together with a notice of correction. This notice will be a separate publication with a link to the updated article in the most recent edition of the journal, in order to notify readers that there has been a significant change to the article and that the revised version is available on the website. Relevant databases will then be notified about the update.

Retractions. If an article needs to be retracted from the research literature due to inadvertent errors during the review process, serious ethical violations, fabrication of data, plagiarism, or other reasons that threaten the integrity of the publication, Psicothema will follow the recommendations from the Committee on Publication Ethics (COPE) for retractions. In this case, the original publication will be amended with a "RETRACTED" mark but will remain available on the Psicothema website for future reference. Retractions will be published with the same authorship and affiliation as the retracted article, so that the notice and the original retracted article may be properly found in indexing databases. The retraction notice will be published in the most current edition of the journal. Partial retractions may be published in cases where results are partly incorrect. An article will only be removed completely from the Psicothema website and indexing databases in very exceptional circumstances, where leaving it online would constitute an illegal act or could cause significant harm.

Expression of concern. Psicothema may publish such a concern if the investigation about supposed bad conduct in research is inconclusive, complex, or very prolonged. In this case, a Psicothema editor may choose this option, detailing their concerns point by point, and any actions ongoing.

Comments and responses. Psicothema will only exceptionally publish comments on or responses to articles published in the journal, when the comments (i.e., short letters to the editors from readers who wish to publicly question a specific article) affect the editorial content of an article published in the journal, contain evidence of a claim, and the result of an investigation by the editorial team does not lead to rejecting the criticism, or correction or retraction of the article. In these exceptional cases, once the comment has been approved for peer review, the editorial team will contact the authors of the article in question and invite them to reply to the comment. The reply will allow the authors to publicly respond to the concerns raised. If the authors do not provide a reply within the set timescale or decide not to reply, the comment will be published along with a note explaining the absence of a response. Both comments and replies will be reviewed to ensure that the comment addresses significant aspects of the original article without becoming essentially a new article, the response directly addresses the concerns without evasion, and the tone of both is appropriate for a scientific journal. Only one round of comments and responses will be facilitated about any single article. Nonetheless, Psicothema recommends that readers direct their comments directly to the authors involved and use alternative forums for additional public discussion.

Archive and Digital Preservation Policies

Conditions for self-archiving preprints. Preprint versions of articles (the version initially submitted to the journal) may be shared at any time anywhere. Sharing an article (without review) on a preprint server, for example, is not considered prior publication. Because of that, before final publication, we recommend that authors self-archive the preprint

version on personal or institutional websites, scientific social networks, repositories, reference managers... Once published, if the preprint remains available in a repository, it must be specified that, "This is an electronic version of an article published on Psicothema (year). The final version is available on the official web page", also highlighting the full reference to the published article, including its DOI.

Psicothema's preservation policy. Psicothema publications are available on the journal website and in international open-access repositories online such as SciELO and Dialnet. Psicothema focuses on disseminating content and making it accessible through indexing services. Online access is free, while the printed version is subscription access. In addition, Psicothema allows self-archiving of preprint, postprint, and editorial version (immediately after publication). Exploitation rights and Psicothema's self-archiving permissions may be found at <https://dulginea.opensciencespain.org/ficha1034>.

Conditions for self-archiving postprint articles. Authors are encouraged to share the final version of their accepted articles (the authors' version including changes suggested by reviewers and editors) on social networks and repositories until the editorial version is published in an edition of the journal. It must be expressly indicated that this is an article "in press" in the Psicothema journal. Once the editorial version is published, if the postprint is still available in a repository, it must be specified that, "This is an electronic version of an article published on Psicothema (year). The final version is available on the official web page", also highlighting the full reference to the published article, including its DOI.

Archive. The journal undertakes to provide XML metadata or in other specific formats, immediately after its publication and within three

months in order to promote its dissemination in databases. Psicothema uses various national and international repositories: Clarivate Analytics, Scopus, Google Scholar, S2ciELO, Dialnet, EBSCO Essentials Academic Search Premier, PubMed, PsycINFO, IBECs, Redinet, Psycodoc, Pubpsych, Fuente Académica Plus, IBZ Online, Periodicals Index Online, MEDLINE, EMBASE, ERIH PLUS, Latindex, MIAR, CARHUS Plus+ 2018, Rebiun, DOAJ, & Crossref.

Digital preservation. In order to preserve permanent access to digital objects hosted on its own servers, Psicothema makes backups, monitors the technological environment to foresee possible migrations of obsolete formats or software, preserves digital metadata, uses DOI Digital Object Identifier) and ORCID. All articles published in Psicothema are also hosted and available in the institutional repository of the Universidad de Oviedo (REUNIDO). The articles published on Psicothema's website are available in easily reproducible format (PDF).

Anti-Plagiarism Policy

In compliance with our code of ethics and in order to guarantee the originality of the manuscripts submitted for evaluation, Psicothema applies anti-plagiarism software to all manuscripts that meet the minimum criteria of the preliminary review and are to be subjected to the review process.

The submission will be rejected in case of detecting practices of plagiarism or scientific fraud, either during the preliminary review by the editorial committee or once the peer review has started.

Articles

- Likert scales: A practical guide to design, construction and use
Pere J. Ferrando, Fabia Morales-Vives,
José M. Casas, & José Muñiz 1-15
- Selective review of empirically supported psychological treatments for pediatric health problems
Isabel Cuéllar-Flores, Izaskun Basterra, Silvia Cámara, Borja Esteso-Orduña,
Deborah Fuentes-Casany, Laura González-Riesco, Amaia Izquierdo,
María del Pilar Mundo-Cid, Estela Salcedo,
Raquel Seijas-Gómez, & María Serrano-Villar..... 16-28
- Relationship between online trolling and Dark Tetrad personality traits: A meta-analysis
Sergio Hidalgo-Fuentes, Elisa González-Pérez, & Isabel Martínez-Álvarez 29-41
- Extended persuasion: Elaboration moderates indirect attitude change
Diego Bustos, Borja Paredes, Lorena Moreno,
Pablo Briñol, & Richard E. Petty 42-51
- Development and validation of a scale to detect late language emergence in Spanish-speaking
children
Verónica Martínez, Paula Solano, & José Carlos Núñez 52-61

Book review

- Morales, J. F. (2025). *Invitación a la psicología social*. Sanz & Torres.
Reviewed by Itziar Fernández 62-63

Article

Development and Validation of a Scale to Detect Late Language Emergence in Spanish-Speaking Children

Verónica Martínez¹ , Paula Solano²  and José Carlos Núñez¹ 

¹ Universidad de Oviedo (Spain)

² Consejería de Educación de la Comunidad Autónoma del Principado de Asturias (Spain)

ARTICLE INFO

Received: 31/12/2024

Accepted: 17/04/2025

Keywords:

Late Language Emergence
Language disorders
Detection
Validation
Scale

ABSTRACT

Background: Detecting late talkers among children is important because they are at greater risk of subsequently manifesting persistent disorders in language development. Therefore, it would be beneficial to have quick measurement tools with adequate psychometric properties, which have not yet been developed. This study set out to develop, analyse and validate the factor structure of an observational scale for detecting Late Language Emergence (LLE) in Spanish-speaking children aged 2-3 years that could be applied not only by language professionals but also teachers in early childhood education centres for children aged 0-3 years. **Method:** The sample comprised 364 children aged 24-38 months ($M = 31.93$; $SD = 3.512$; 49.7% girls). **Results:** The estimated reliability coefficients ranged from $\omega = .77$ to $\omega = .97$. Factor analyses indicated that the best model explaining the scale's assessment criteria was articulated around five factors: phonology, lexical semantics, morphosyntax, comprehension and communication. The correlations between the dimensions of the scale and those of the MacArthur Communicative Development Inventory were statistically significant. **Conclusions:** The scale to detect late talkers among Spanish children appears to be a concise, simple instrument with suitable psychometric properties.

Desarrollo y Validación de una Escala para la Detección de Niños Castellanohablantes Tardíos del Lenguaje

RESUMEN

Antecedentes: La detección de niños hablantes tardíos es relevante por su mayor riesgo a manifestar posteriormente trastornos persistentes en el desarrollo del lenguaje. Sería conveniente disponer de instrumentos de medida breves y con adecuadas propiedades psicométricas, que hasta ahora no han sido elaborados. El objetivo fue desarrollar, analizar y validar la estructura factorial de una escala observacional para la detección de niños castellanohablantes de entre 2 y 3 años con Inicio Tardío aplicable por los profesionales del lenguaje y por las educadoras de los centros de Educación Infantil de 0 a 3 años. **Método:** La muestra está formada por 364 niños de entre 24 y 38 meses ($M = 31.93$; $DT = 3.512$; 49.7% niñas). **Resultados:** Los coeficientes de fiabilidad estimados oscilaron entre $\omega = .77$ y $\omega = .97$. Los análisis factoriales indicaron que el mejor modelo que explica los criterios de evaluación de la escala se articuló en torno a cinco factores: fonología, léxico-semántico, morfosintaxis, comprensión y comunicación. Las correlaciones entre las dimensiones de la escala y las del Inventario Comunicativo MacArthur son estadísticamente significativas. **Conclusiones:** La escala para detectar a los niños españoles hablantes tardíos parece ser un instrumento breve, sencillo y con adecuadas propiedades psicométricas.

Palabras clave:

Inicio Tardío del Lenguaje
Trastornos del Lenguaje
Detección
Validación
Escala

Language development is a fundamental process in early childhood, dependent on communicative and social processes necessary to exchange and construct meanings with others over the course of the child's evolutionary development (Karmiloff & Karmiloff-Smith, 2005). Children also need to attain a certain level of cognitive and communicative development to begin to master skills involved in the development of more formal aspects of language (phonology, semantics, morphology, and syntax). The appropriate development of cognitive, communicative and linguistic competence is necessary for later access to the precursor skills for learning to read and write, and for self-regulation.

Studies on language acquisition propose a division between the different domains of language which has given rise to a wealth of knowledge on how children acquire the structural aspects of language (phonology, morphosyntax, semantics), comprehension and communicative development. On the other hand, it has been pointed out that the structure of language comprises a series of processes that are related to each other thanks to interface systems, which would give rise to a relationship between different domains during development: grammar and lexical development (Devescovi et al., 2005; Serra, 2008), lexical and phonological development (Stoel-Gammon, 2011; Rose & Blackmore, 2018), and the relationship between communicative development and speech comprehension (Arachchige et al., 2021; Colonnese et al., 2010).

Language development occurs similarly in all children as a result of the complex interaction between different biological, psychological and social factors (Cuetos et al., 2015). However, there is a group of children who present persistent difficulties in their linguistic competence between the second and third year of the verbal stage of language development, which may affect all areas of their development, especially social and school development (Llorenç et al., 2021).

These children might show a pattern of late onset language development and communication, characteristic symptoms of which are the late appearance of the first words and/or combination of two or more words in their first sentences, and a scarce and slow vocabulary growth at the age of 24 months in the absence of other difficulties (Nouraey et al., 2021). These are called late talkers or children with late language emergence (Fisher, 2017; Mendoza, 2016; Rescorla, 2011), a term recently proposed by the American Speech-Language-Hearing Association [ASHA] (2018), with an estimated prevalence of 10-15% of children around 24 months old (Collison et al., 2016).

Late language emergence (LLE) is not currently a specific category in any diagnostic criteria manuals. The LLE is labelled by inclusion and specificity criteria (fewer than 50 words at 24 months and/or absence of word combinations) and by exclusion criteria related to the absence of other conditions that may confirm another type of disorder (hearing impairment, intellectual functional diversity, neurological damage, organic malformations or other neurodevelopment disorders that account for it). The ASHA (2018) states that LLE is labelled when language development trajectories are below age expectancies. However, it should be considered as a transitional label, as permanent language difficulties will be determined from the age of five.

Ever since the seminal studies by Thal and Bates (Thal, 1991; Thal & Bates, 1988), Paul (1991), and Rescorla (Rescorla, 1989; Rescorla et al., 1997), there has been broad interest in researching

both the causes and characterisation of LLE, which has led to significant progress in typifying it. Different studies have found that the characteristic signs of LLE are accompanied by other difficulties and particular courses of development in different linguistic-communicative dimensions, such as slower vocabulary acquisition, an absence of the lexical explosion period and a delay in receptive language (Auza & Murata, 2021; Chilosi et al., 2019; Desmarais et al., 2008; Paul, 1991; Rescorla et al., 1997; Thal, 1991; Thal & Bates, 1988).

There is a high percentage of children who present LLE may subsequently enjoy typical development between the ages of three and four (Rice et al., 2008; Sylvestre et al., 2017), the so-called Late Bloomers (Rescorla et al., 2000). Nevertheless, there is another significant percentage of children with LLE who will continue to manifest difficulties permanently beyond age 3 year (Chilosi et al., 2019; Perry et al., 2023). So it is that children with LLE make up a very important pre-clinical group, as they have a greater risk than their typically developing peers of manifesting persistent language development disorders at later ages, as well as difficulties in processes of learning, socialisation, and literacy acquisition (ASHA, 2018; Chilosi et al., 2019; Fisher, 2017; Hammer et al., 2017; Horvath et al., 2019, 2022; Kautto et al., 2021; Morgan et al., 2020; Rescorla & Dale, 2013; Rescorla et al., 2000, 2002; Sylvestre et al., 2017; Thal et al., 2013). However, as several research have noted (Desmarais et al., 2008, 2010; Rescorla, 2011), it is not known exactly which late talkers will become children with Developmental Language Disorder (DLD), a neurodevelopmental disorder that can affect one or more language domains at different levels, in both expressive and receptive language (Bishop et al., 2016).

Because of all of the above, early identification and detection of children with LLE is essential for two reasons. Firstly, it will permit timely monitoring and treatment measures in school, as part of a preventative response to intervention (RTI) model, leaving behind the widely-used *wait and see* approach that has contributed little scientific evidence (Capone, 2018; Moreno & Nieva, 2021); and secondly, individual or family intervention can then begin as early as possible, as in addition to producing immediate benefits in the family and the child, in the long term this is highly profitable in economic terms (Rydz et al., 2006).

In this identification, given that LLE is not a diagnostic category but rather a label that refers to children whose language development is not normotypical, fast and reliable detection instruments are necessary to identify the signs of delay in language acquisition in children between two and three years.

These detection instruments should be simple tests (valid and with little administration time) that make it possible to differentiate between children with and without LLE. They must include growth milestones or development indicators considered critical and predictive in proper language development, and it should be possible for them to be carried out by those people who are most in contact with the child, namely parents and early childhood education teachers.

The instruments most used in English-speaking contexts are language and communication checklists completed by parents or early childhood education teachers, which aim to identify early signs of delay or alteration in the growth milestones in question. These include the Language Development Survey (LDS; Rescorla, 1989; Rescorla & Achenbach, 2002) in which parents assess expressive vocabulary and word combinations in children aged 18 to 35

months, but this is not scaled in Spanish; another is the Children's Communication Checklist (CCC-2; Bishop, 2003) which makes it possible to detect difficulties only in pragmatic language use; lastly, there is the Preschool Language Scale (Zimmerman et al., 2011) which assesses language from birth to 6;11 years and is scaled for monolingual Spanish-speaking children.

However, no studies or systematic reviews have been found that focus on the analysis of early-detection screening instruments and tools at solely linguistic level, validated for Spanish speakers aged two to three years. Some authors have also observed that early predictors for LLE have never been analysed in terms of their specific, individual predictive and discriminating power (Sansavini et al., 2021).

Although the MacArthur Communicative Development Inventory (CDI) makes it possible to assess vocalizations, vocabulary and grammar use in Spanish-speaking children aged 16-30 months (López-Ornat et al., 2005; Mariscal et al., 2007), it cannot be considered as a screening test because it is time-consuming to complete. Although a brief CDI and CDIII assessments suitable for evaluating language skills in Mexican Spanish-speaking children aged 30-37 months have been developed, they have not yet been adapted and validated for other Spanish-speaking populations (Jackson-Maldonado et al., 2013, 2024). There are some approaches, such as the Language Observation Protocol for pre-school teachers (Ygual et al., 2011) based on teachers' contributions, but it is aimed at children aged 3;6 to 5;11 years.

Professionals frequently use general development evaluation or screening instruments and development inventories that include questions or items on language development and/or communication at the same time as other areas of development, such as motor, cognitive, socio-emotional, adaptive, etc. These include the Early Detection System for Development Disorders (Sistema de Detección Precoz de Trastornos del Desarrollo, SDPTD; Alcántud et al., 2015) and the Haizea-Llevant scale (Fernández et al., 1991) used in the context of Primary Health Care.

The objective of this study was to develop, analyse and validate the factor structure of an observational scale to detect Spanish-speaking children with LLE that could be applied not only by language professionals but also by early childhood education teachers in preschool education centres from age 0 to 3, to support and bolster the work of paediatricians in detecting this population. In terms of validity, a positive relationship is expected between this questionnaire to detect LLE and the MacArthur Communicative Development Inventory, which would make it possible to quickly and reliably detect the warning signs in language development between two and three years. Based on previous studies, it is assumed that the resulting scale's structure should be made up of five factors (phonology, lexical-semantic, morphosyntax, comprehension and communicative area), as these factors are interrelated. In addition to structural validity, the scale must show concurrent validity (the dimensions of the scale must be significantly correlated with the corresponding dimensions of the MacArthur Communicative Development Inventory). Lastly, the reliability of the scale's dimensions should be appropriate.

Method

Participants

The sample is formed by 364 children with an age range of 24 to 38 months ($M = 31.93$; $SD = 3.512$). Of the total sample, 183 are boys ($M = 31.87$; $SD = 3.452$) and 181 girls ($M = 31.98$; $SD = 3.580$). All of the children are registered in kindergarten for 2-3-year-olds in the first cycle (age 0-3) of early childhood education at different state centres in the Principality of Asturias, Spain. Of the sample, seventeen children attend weekly the early attention service in their areas. Twenty early childhood education teachers participated from the ten early childhood education schools that the children attend.

The selection criteria for the participants were children aged between 24 and 40 months enrolled in the first cycle of kindergarten. The exclusion criteria were that participants did not have diagnostic reports of ASD, hearing or vision disabilities, and/or intellectual disabilities.

Instruments

The Scale for the Detection of Speakers with Late Language Emergence (DHITLE-S, Detección de Hablantes con Inicio Tardío del Lenguaje) is a Likert scale made up of 43 items, in which three possible scores were established (1 = Never; 2 = Sometimes; 3 = Often). It is divided into five sections: phonology, lexical-semantic, morphosyntax, comprehension and communicative area. Each of the sections is formed by a given, different number of items (phonology = 7 items; lexical-semantic = 9 items; morphosyntax = 12 items; comprehension = 9 items; communicative = 6 items).

In the first page of the scale, the participants' personal details are taken, including their name, sex, date of birth as year, month and day, whether they receive any specialist attention, the date when the questionnaire was completed and the assessor's name. Basic instructions are included to explain the purpose of the scale. The early childhood education teacher completing the scale must choose the option that best describes the child's communicative and linguistic competence and evaluate whether the described behaviour was consolidated, in progress, or not yet present by the child. Thus, she/he to choose one of three options: 1 (never), indicating that the behaviour or information described by the item does not occur; 2 (sometimes), indicating that the behaviour or information occurs inconsistently or is still developing; and 3 (often), indicating that the behaviour or information occurs frequently or consistently. Lastly, four prior questions are included related to the prototypical symptoms of late onset with yes/no answers and a multiple-choice question, related to the sounds that the words they produce contain.

Each of the five sections begins with a brief explanation so that the early childhood education teachers know what they must assess at each stage. In the phonological section, the inquiry focuses on whether the child, spontaneously or in response to adult or peer demands, produced onomatopoeias, intoned slang, repeated syllables or words, simplified the structure of words and made

pronunciation errors. In the lexical-semantic section, we asked about the amount of vocabulary and the type of words that the child produced. In the morphosyntactic section, the inquiry focuses on the child's use of irregular verbs, the use of the plural, the description of events that happened to him/her, the use of prepositions, etc. In the comprehension section, the inquiry focuses on words and situations that the child should understand, such as whether he/she responds to his/her name or is able to follow simple instructions. In the communication section, inquiries were made regarding the acquisition of specific communicative competencies, including the use of the point gesture, the effective utilization of play materials, and the employment of pointing to request desired items or actions. The form of the items is always the same, i.e., a sentence, except in item 16 of the lexical-semantic section, which assesses the number of real words or approximations that the child produces, where the early childhood education teachers must choose from four options.

The MacArthur Communicative Development Inventory (CDI), adapted to Spanish (López-Ornat et al., 2005) was given to the parents of participating children to establish the children's level of early communicational and linguistic development in the areas of vocabulary, comprehension and grammar. This inventory, completed by the parents or carers, reflects typical process in early language acquisition and consists of two forms, based on the age group. This research used the CDI: Words & Sentences form, which is aimed at children aged 16 to 30 months, and can also be applied to older children with language delay. In Spanish, this form is made up of three parts: Part 0, Vocalizations, assesses the type of vocalizations that the child produces; Part 1, Words, assesses the early production, vocabulary development, the number of words the child understands and produces, and language use; lastly, Part 2, Grammar, assesses nominal morphology, the use of irregular verbs, overgeneralizations of morphological rules, word combinations and morphosyntactic complexity.

Procedure

To construct the DHITLE-S scale, the following steps proposed by Muñiz and Fonseca-Pedrero (2019) were followed.

The first version of the scale was designed by two of the authors of the study. The items are based on the review of the literature on language development and difficulties that children with developmental language disorder present.

The scale was then reviewed by two experts in the study of language acquisition and DLD, independent of the study. Those items considered by the two experts to have little content validity to detect children with LLE were removed (2 items), leaving the scale with 43 items with three possible answers. The distribution of the items in each of the sectors is unequal and more items assess morphosyntax, semantics and comprehension, making these more relevant (Muñiz et al., 2005), because children with DLD present more difficulties in these areas.

With the scale completed, the authors contacted the education authority for the first cycle kindergartens (age 0-3) in the area where the study was to be conducted. Having obtained authorization, a meeting was held with all the principals of the kindergartens to inform them of the aim of the study, show them the scale for

detecting LLE and request their cooperation. Then, in each of the ten participating schools a meeting was held with the teachers of the 2 to 3-year-olds to show them the goal of the study, request their participation and explain how to complete the questionnaire.

All the parents of the participating children were informed by letter of the aim of the study and were asked to answer the CDI that was sent to them. The purpose of using another measurement test, the CDI, was to confirm the validity of the participants' scores on the scale (Elosúa, 2003), as this test assesses the population under study.

Those families that participated in the study signed an informed consent form, authorizing the use of the data provided and stating the protection of the children's identities. When all the families had signed the informed consent, the early childhood education teachers covered each questionnaire for each child.

The conditions and characteristics of the present study were approved by the Research Ethics Committee of the University of Oviedo.

Data Analysis

The data resulting from this research was processed in various stages. Initially, the descriptive statistics and correlations matrix were analysed. There were few missing values in the scale items (0.74% in total). The maximum likelihood procedure was used to complete the information. To study the scale's factor structure, confirmatory factor analyses were conducted with the Mplus 8.7 program. Three models were fitted: a one-dimensional model (all the items of the scale are explained by a single general factor), a multifactorial model with three first-order factors (phonology, morphology-semantics and comprehension-communication), and a multifactorial model with five first-order factors (phonology, lexical semantics, morphology, comprehension and communication). A three-factor multifactorial model was fitted because it has been suggested that language development can be advanced through the interaction between different interfaces, morphological-semantic, lexical-phonological, and comprehension-communication (Arachchige et al., 2021; Serra, 2008; Stoel-Gammon, 2011). Specifically, it has been observed that the advances that occur in the morphosyntax of Spanish children between 16 and 30 months may be preceded by vocabulary growth (Serrat et al., 2010). It has also been found that there is a relationship between communicative and non-communicative gestures and language comprehension, not only at early ages (Bates et al., 1980) but also at 3 years (Alcock & Connor, 2021).

The results of establishing the factorial models were assessed according to criteria typically used: Chi-square (χ^2), Goodness of Fit (GFI), Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Standardized Root Mean Square Residual (SRMR), and Root Mean Square Error of Approximation (RMSEA). There is evidence of a good fit when χ^2 has $p > .05$, GFI and TLI $\geq .90$, CFI $\geq .95$, SRMR and RMSEA $\leq .06$. The best model is selected based on the Akaike Information Criterion (AIC) statistic (the best model is that which presents lowest values in AIC). Having analysed the scale structure, concurrent validity was studied by analysing correlations and linear regressions (taking subscales as predictive variables and the five dimensions of the MacArthur Inventory as criteria variables). The reliability of the scale, and its dimensions, was estimated through α and ω , and interpreted according to Watkins (2017).

Results

Confirmatory Factor Analyses

Table 1 shows the results of the fit of the three models. As can be seen, taking into account the assessment criteria established, the multifactorial model fits better than the one-dimensional model. However, both the three-factor model and the five-factor model show a similar, moderate (albeit acceptable) fit. Given that we must choose the best fit from both models, the five-factor model shows a slightly better fit than the three-factor. Although almost all the statistics are very similar, AIC informs us that the five-factor model is a better fit: $AIC_{3F} - AIC_{5F} = 9.865, p < .001; d = 0.334$. Small effect.

Table 1
Statistics of Factorial Models Fit

	Unifactorial model	Three-factor model	Five-factor model
$\chi^2(gf)$	4347.933(666)	1578.034(528)	1560.168(528)
p_{χ^2}	.001	.001	.001
GFI	.966	.981	.983
TLI	.654	.890	.891
CFI	.673	.900	.901
SRMR	.113	.081	.075
RMSEA	.127(.124-.131)	.079(.074-.083)	.079(.074-.083)
AIC	16250.034	11137.807	11127.942

Note. Unifactorial model (a general factor), Three-factor model (phonology, morphology-semantics and comprehension-communication), Five-factor model (phonology, lexical-semantics, morphology, comprehension and communication). χ^2 = Chi-square; GFI = Goodness of Fit; TLI = Tucker-Lewis Index; CFI = Comparative Fit Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root Mean Square Error of Approximation; AIC = Akaike Information Criterion.

Table 2 shows the statistics corresponding to the five-factor model (non-standardized factorial weights, standard errors, R^2 , p , standardized factorial weights). It can be seen that the factorial weights (amount of variance of each item explained by factor) are all statistically significant at $p < .001$. Furthermore, the estimation errors of these parameters are low.

The relationship between the five factors is all statistically significant (see Table 3). We can also observe that the phonology factor is negatively related to the other four factors (among which the relationship is positive). This is because the wording of the items in factor 1 (Phonology) is in the opposite direction to that of the other four factors.

Table 3
Factor Covariances

	Estimate	S.E.	z-value	p	95% Confidence Interval	
					Lower	Upper
Factor 1 ↔ Factor 2	-.502	.059	-8.565	< .001	-0.617	-0.387
Factor 1 ↔ Factor 3	-.661	.040	-16.622	< .001	-0.739	-0.583
Factor 1 ↔ Factor 4	-.183	.059	-3.104	.002	-0.298	-0.067
Factor 1 ↔ Factor 5	-.192	.066	-16.622	.003	-0.320	-0.064
Factor 2 ↔ Factor 3	.949	.015	64.957	< .001	0.921	0.978
Factor 2 ↔ Factor 4	.726	.045	16.103	< .001	0.638	0.814
Factor 2 ↔ Factor 5	.842	.044	19.167	< .001	0.756	0.928
Factor 3 ↔ Factor 4	.607	.036	16.994	< .001	0.537	0.677
Factor 3 ↔ Factor 5	.690	.040	17.304	< .001	0.612	0.768
Factor 4 ↔ Factor 5	.947	.035	27.125	< .001	0.878	1.015

Note. Factor 1 (Phonology), Factor 2 (Lexical-Semantic), Factor 3 (Morphology), Factor 4 (Comprehension), Factor 5 (Communication). The wording of the items in Factor 1 (Phonology) is in the opposite direction to that of the other four factors.

The Table 4 shows the final version of the Scale for Detection of Speakers with Late Language Emergence (DHITLE-S).

Reliability

The five factors on the scale show good and/or excellent reliability: Phonology ($\alpha = .82, \omega = .82$; limits: .78 - .84), Lexical-Semantics ($\alpha = .89, \omega = .90$; limits: .87 - .91), Morphosyntax ($\alpha = .97, \omega = .97$; limits: .96 - .97), Comprehension ($\alpha = .93, \omega = .93$; limits: .90 - .93) and Communication ($\alpha = .77, \omega = .77$; limits: .70 - .78).

Concurrent Validity

For the study of concurrent validity, correlation analyses were conducted between the five factors of this scale and the five CDI subscales. In addition, regression analyses were conducted to see the relevance of the five dimensions of the scale in predicting each of the CDI subscales. Table 5 shows the correlations and Table 6 the results of the regression analysis.

Table 6
Regression Analysis Results

	β	t	p	$R^2(p)$
Vocabulary				
PHO	.040	0.550	.583	
SEM	.322	2.776	.006	
MOR	.513	4.249	< .001	
CMP	-.034	-0.367	.714	
CMU	-.209	-2.527	.012	
				.509(< .001)
Word endings				
PHO	-.013	-0.172	.864	
SEM	.229	1.918	.057	
MOR	.621	4.977	< .001	
CMP	-.057	-0.582	.561	
CMU	-.200	-2.350	.020	
				.476(< .001)
Difficult verbs				
PHO	.103	1.194	.234	
SEM	.010	0.077	.939	
MOR	.523	3.679	< .001	
CMP	.081	0.728	.468	
CMU	-.248	-2.553	.011	
				.322(< .001)
Sentences				
PHO	.034	0.376	.707	
SEM	-.055	-0.387	.699	
MOR	.639	4.222	< .001	
CMP	.047	0.431	.667	
CMU	-.204	-2.153	.033	
				.332(< .001)
Complexity				
PHO	-.058	-0.599	.550	
SEM	.215	1.404	.162	
MOR	.424	2.675	.008	
CMP	-.052	-0.433	.666	
CMU	-.142	-1.298	.196	
				.230(< .001)

Note. PHO (Phonology), SEM (Lexical-Semantic), MOR (Morphology), CMP (Comprehension), CMU (Communication).

Table 2
Statistics of the Five-Factor Model

	Non- standardized factorial weights	Standard errors	R^2	p	Standardized factorial weights
F1: Phonology					
PHO1	.466	.033	.419	< .001	.648
PHO2	.491	.032	.605	< .001	.778
PHO3	.504	.031	.565	< .001	.752
PHO4	.527	.034	.537	< .001	.733
F2: Lexical-Semantic					
SEM1	.515	.040	.605	< .001	.778
SEM2	.554	.038	.618	< .001	.786
SEM3	.560	.030	.618	< .001	.786
SEM4	.641	.024	.615	< .001	.785
SEM5	.296	.031	.492	< .001	.701
SEM6	.580	.033	.643	< .001	.802
F3: Morphology					
MOR1	.695	.022	.744	< .001	.862
MOR2	.696	.024	.756	< .001	.869
MOR3	.679	.027	.745	< .001	.863
MOR4	.565	.034	.615	< .001	.784
MOR5	.763	.023	.825	< .001	.908
MOR6	.630	.030	.699	< .001	.836
MOR7	.557	.028	.591	< .001	.769
MOR8	.713	.023	.769	< .001	.877
MOR9	.732	.021	.813	< .001	.902
MOR10	.782	.019	.853	< .001	.924
F4: Comprehension					
COMP1	.309	.035	.644	< .001	.803
COMP2	.309	.032	.745	< .001	.863
COMP3	.294	.036	.734	< .001	.857
COMP4	.319	.036	.692	< .001	.832
COMP5	.211	.032	.564	< .001	.751
COMP6	.210	.025	.409	< .001	.639
COMP7	.248	.030	.534	< .001	.731
COMP8	.389	.040	.550	< .001	.741
COMP9	.356	.044	.373	< .001	.610
F5: Communication					
COMU1	.205	.025	.440	< .001	.663
COMU2	.420	.040	.496	< .001	.704
COMU3	.268	.037	.417	< .001	.646
COMU4	.319	.034	.407	< .001	.638

Note. R^2 = Coefficient of determination.

As can be seen in Table 5, the correlations of the dimensions of the scale with the CDI dimensions are all statistically significant, except the dimension of communication with the subscales of difficult verbs and mean length of utterances. However, only Morphology is observed as a good predictor of the five CDI subscales, while Communication predicts significantly four of the five dimensions (vocabulary, word endings, difficult verbs and mean length of utterances). Comprehension and Phonology do not predict any of the five CDI dimensions, and Lexical-Semantics only predicts Vocabulary. Lastly, in general the scales explain a significant amount of the variability of the five CDI subscales: Vocabulary (50.9%), Word endings (47.6%), Difficult verbs (32.2%), Mean length of utterances (33.2%) and Morphosyntactic complexity (23%) (Table 6).

Discussion

The present study was designed to develop, analyse and validate the factor structure of a scale to detect LLE in Spanish-speaking children, which could be applied not only by language professionals but also early childhood education teachers at preschool learning centres for children aged 0 to 3 years, to support and bolster the work of paediatricians in detecting this population. The results obtained generally indicate that the Scale for Detection of Speakers with Late Language Emergence to detect late talkers among Spanish-speaking children (DHITLE-S) presents suitable psychometric properties that allow it to be presented as a quick, simple and appropriate instrument for detecting this population.

Table 4*Final Version of the Scale for Detection of Speakers with Late Language Emergence (DHITLE-S)*

Item	Phonology
1	Often makes onomatopoeia of animals instead of their names or familiar sounds.
2	Can make approximations to simple words, even if he/she cannot say them correctly.
3	It simplifies the structure of words from a two-syllable word to a single-syllable word.
4	Reduces words of more than two syllables to shorter productions.
Item	Lexical-Semantic
1	Can name a minimum of four animals.
2	Can name at least four colors and four basic concepts.
3	Produces pronouns and determinants.
4	Produce adverbs.
5	Produces basic words of everyday life, names things he/she sees in class, repeats the names of classmates.
6	Choose one of the following options: - Produces less than 20 words. - Produces between 20 and 50 words. - Produces between 50 and 100 words. - Produces more than 100 words.
Item	Morphology
1	Use prepositions and connectives.
2	Uses verbs in the personal form.
3	Can use the plural appropriately.
4	If he/she is already using sentences or combinations of words, use the words in the correct order.
5	Makes combinations of three or more words.
6	Uses the masculine and feminine forms of words.
7	Uses irregular forms of verbs.
8	Can describe and narrate things he/she has done, anecdotes, things that have happened.
9	Can coordinate gender and number correctly when speaking.
10	Can describe what he/she has done when asked (in sentences of up to 4 words).
Item	Comprehension
1	Knows or identifies body parts, toys, food, clothing.
2	Responds to or follows instructions.
3	Is able to identify (point to or give) toys that represent objects, animals or people.
4	Responds verbally or gesturally with 'yes' or 'no' in appropriate contexts.
5	Can follow simple instructions or commands.
6	Responds to the word 'no'.
7	Responds to 'give', 'take', 'look', 'come' without adult pointing or gesturing.
8	Identifies or points to objects described by the adult by their functional use.
9	Understands some basic concepts of size, color or space.
Item	Communication
1	Is able to look where the adult points.
2	Points to an object that catches his/her attention by naming it or saying 'Look!'
3	Uses words or gestures spontaneously to say hello and goodbye at appropriate times.
4	Shares play with other children or adults.

Table 5*Correlation Matrix*

	PHO	SEM	MOR	CMP	CMU	VOC	WDE	VRB	MLU
SEM	.400**	----							
MOR	.597**	.848**	----						
CMP	.227**	.709**	.642**	----					
CMU	.046	.622**	.508**	.746**	----				
VOC	.500**	.626**	.684**	.410**	.240**	----			
WDE	.467**	.590**	.667**	.385**	.227**	.913**	----		
VRB	.449**	.415**	.522**	.280**	.097	.782**	.788**	----	
MLU	.436**	.431**	.547**	.290**	.109	.715**	.657**	.588**	----
MOC	.301**	.424**	.462**	.272**	.178*	.599**	.573**	.449**	.510**

Note. PHO (Phonology), SEM (Lexical-Semantic), MOR (Morphology), CMP (Comprehension), CMU (Communication), VOC (Vocabulary), WDE (Word endings), VRB (Difficult verbs), MLU (Mean length of utterances), MOC (Morphosyntactic complexity). VOC, WDE, VRB, MLU and MOC are scales in the MacArthur test. For this purpose, the phonology calculation was performed by recoding the items so that they all go in the same direction.

* $p < .05$; ** $p < .01$.

The study of the structure of the instrument provided data that support a model with five factors: phonology, lexical-semantics, morphology, comprehension and communication. In addition, the data obtained indicated concurrent validity, as statistically significant correlations were observed among the dimensions of the scale validated in this study and the CDI dimensions. Thus, it was observed that the Morphology dimension is a good predictor of the five CDI subscales, while Communication predicts significantly four of the five dimensions (Vocabulary, Word Endings, Difficult Verbs and Mean Length of Utterances), and Lexical-Semantics only predicts Vocabulary. CDI is a reliable and valid source of information about young children's language and have proven useful in both clinical and research setting (Jackson-Maldonado et al., 2024). Since three dimensions of the DHITL-S scale correlate significantly with the subscales of the CDI, it could be considered adequate to quickly detect children with LLE, despite the fact that both instruments were administered by different informants in different social contexts.

The data provided shows strong evidence of the reliability of the scores, with good or very good alpha and omega coefficients for the five factors.

This type of instrument is vital due to its practical implications, as in educational contexts it is essential to work from a preventative approach, focusing on the detection of possible risks that can be addressed with a response to intervention (RTI) model and an inclusive approach focusing on eliminating barriers and implementing facilitators.

The use of this instrument makes it possible to detect early risks in communicative-linguistic development that permit the implementation of clinical and educational intervention measures before or at the start of schooling in the second cycle (age 4-5) of preschool education. This type of tool is essential for two related reasons: firstly, it enables the implementation of proactive actions centred on the stimulation of linguistic competence systematically and consciously (Moreno & Nieva, 2021); and secondly, it minimizes medium- to long-term implications (ASHA, 2018; Collison et al., 2016; Morgan et al., 2020; Rydz et al., 2006), as those pupils who present risks or changes in their language development face important barriers for participating and progressing successfully in teaching and learning processes, given that this sustains the acquisition of knowledge and access to the curricular demands of each education stage (Auza & Murata, 2021; Hammer et al., 2017).

Although the results of the study are consistent, it is necessary to bear in mind some limitations. Specifically, the most important limitation of the study is the sample bias, as it was not done at random (it is not a probabilistic sample) but rather participants were selected from a single city, belonging to different socioeconomic classes. However, a broad sample size is used which would comply with the recommendation that there be 5 to 10 people for each item administered (Ferrando & Anguiano, 2010). Future research should address the following limitation: using multiple informants (parents, teachers in early childhood education, and educators) to triangulate information, especially when constructs are measured indirectly, as in this study with 2- and 3-year-olds. Finally, another limitation of the study, as well as a future line of research, would be to determine the level of specificity and sensitivity of the scale. This would allow reliable detection of LLE, thus reducing the occurrence of false positives and false negatives. Thus, future research should focus on gathering additional validity evidence for the intended uses of the

scale, particularly regarding test content, cognitive processes, and test consequences, as well as examining the scale's specificity and sensitivity for diagnostic purposes, in alignment with the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) and Sireci & Benítez (2023).

In short, the present study developed, determined the factor structure of, and validated the Scale for Detection of Speakers with Late Language Emergence (DHITL-S) for the early detection of Spanish-speaking children with LLE. This scale can be administered by early childhood education teachers in preschool centers serving children aged 0 to 3 years. It would be interesting to carry out future studies on the subject in different autonomous communities in Spain in other languages and also broaden the sample to assess the test. It may also be useful to follow up those children in whom LLE was detected and see how they evolve in linguistic competence. It is possible that this new and promising line of research may help to continue progressing in this field to pin down the predictors of language development disorder.

Author Contributions

Verónica Martínez: Conceptualization, Methodology, Investigation, Data Curation, Writing - Original Draft, Writing-Review and Editing. **Paula Solano:** Conceptualization, Methodology, Investigation, Data Curation, Writing - Original Draft. **José Carlos Núñez:** Methodology, Formal Analysis, Writing-Review.

Acknowledgements

We thank the children who formed part of this research for their collaboration in all the study sessions.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Interest

The authors state that there is no conflict of interests in this study.

Data Availability Statement

The research data related to this article are available in the research team's databases that conducted the study. Interested parties must contact the team to request access.

References

- Alcantud, F., Alonso, Y., & Rico, D. (2015). *Sistema de Detección Precoz de Trastornos del Desarrollo (SDPTD). Construcción, validación y manual de uso (versión 1.0)* [Early Detection System for Developmental Disorders (SDPTD). Construction, validation and user manual (version 1.0)]. Madrid Real Patronato sobre Discapacidad.
- AERA, APA, & NCME (2014). *Standards for educational and psychological testing: National council on measurement in education*. American Educational Research Association.

- American Speech-Language-Hearing Association (ASHA). (2018). *Late Language Emergence*. American Speech-Language-Hearing Association. https://www.asha.org/practice-portal/clinical-topics/late-language-emergence/#collapse_8
- Alcock, K., & Connor, S. (2021). Oral motor and gesture abilities independently associated with preschool language skill: Longitudinal and concurrent relationships at 21 months and 3-4 years. *Journal of Speech, Language, and Hearing Research*, 64(6), 1944-1963. https://doi.org/10.1044/2021_JSLHR-19-00377
- Arachchige, K.G.K., Loureiro, I.S., Blekic W., Rossignol, M., & Lefebvre, L. (2021). The role of iconic gestures in speech comprehension: An overview of various methodologies. *Frontiers in Psychology*, 12, 634074. <https://doi.org/10.3389/fpsyg.2021.634074>
- Auza, A., & Murata, C. (2021). ¿Las clases léxicas y gramaticales de los hablantes tardíos podrían predecir a los futuros niños con TEL? [Could lexical and grammatical classes of late talkers predict the future children with SLI?] *Revista de Investigación en Logopedia*, 11(Especial), 39-51. <https://doi.org/10.5209/rlog.70268>
- Bates, E., Bretherton, I., Snyder, L., Shore, C., & Volterra, V. (1980). Vocal and gestural symbols at 13 months. *Merrill-Palmer Quarterly*, 26(4), 407-423.
- Bishop, D.V. (2003). *The children's communication checklist second edition (CCC-2) manual*. Harcourt Assessment.
- Bishop, D., Snowling, M.J., Thompson, P.A., Greenhalgh, T., & CATALISE consortium (2016). CATALISE: a multinational and multidisciplinary delphi consensus study. Identifying language impairments in children. *PLoS ONE*, 11(7), Article e0158753. <https://doi.org/10.1371/journal.pone.0158753>
- Capone, N. (2018). Late talkers: Why the wait-and-see approach is outdated. *Pediatric Clinics of North America*, 65(1), 13-29. <https://doi.org/10.1016/j.pcl.2017.08.018>
- Collisson, B.A., Graham, S.A., Preston, J.L., Rose, M.S., McDonald, S., & Tough, S. (2016). Risk and protective factors for late talking: An epidemiologic investigation. *Journal of Pediatrics*, 172, 168-174. <https://doi.org/10.1016/j.jpeds.2016.02.020>
- Colonnese, C., Stams, G., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, 30(4), 352-366. <https://doi.org/10.1016/j.dr.2010.10.00>
- Cuetos, F., González, J., & de Vega, M. (2015). *Psicología del lenguaje [Psychology of language]*. Editorial Médica Panamericana.
- Chilosi, A.M., Pfanner, L., Pecini, C., Salvadorini, R., Casalini, C., Brizzolara, D., & Cipriani, P. (2019). Which linguistic measures distinguish transient from persistent language problems in Late Talkers from 2 to 4 years? A study on Italian speaking children. *Research in Developmental Disabilities*, 89, 59-68. <https://doi.org/10.1016/j.ridd.2019.03.005>
- Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders*, 43(4), 361-389. <https://doi.org/10.1080/13682820701546854>
- Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2010). Three profiles of language abilities in toddlers with an expressive vocabulary delay: Variations on a theme. *Journal of Speech Language and Hearing Research*, 53(3), 699-709. [https://doi.org/10.1044/1092-4388\(2009/07-0245\)](https://doi.org/10.1044/1092-4388(2009/07-0245)
- Devescovi, A., Caselli, M.C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language*, 32(4), 759-86. <https://doi.org/10.1017/S0305000905007105>
- Elosúa, P. (2003). Sobre la validez de los tests [About test validity]. *Psicothema*, 15, 315-321.
- Fernández, E., Fernández, I., Fuentes, J., & Rueda, J. (1991). *Tabla de desarrollo Haizea-Llevant (0- 5 años) [Haizea-Llevant development table (0- 5 years)]*. Servicio Central de Publicaciones del Gobierno Vasco.
- Ferrando, P.J., & Anguiano, C. (2010). El análisis factorial como técnica de investigación en psicología [Factor analysis as a research technique in Psychology]. *Papeles del Psicólogo*, 31, 18-33.
- Fisher, E.L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *Journal of Speech, Language, and Hearing Research*, 60, 2935-2948. https://doi.org/10.1044/2017_JSLHR-L-16-0310
- Hammer, C. S., Morgan, P., Farkas, G., Hillemeier, M., Bitetti, D., & Maczuga, S. (2017). Late talkers: A population-based study of risk factors and school readiness consequences. *Journal of Speech, Language, and Hearing Research*, 60(3), 607-626. https://doi.org/10.1044/2016_JSLHR-L-15-0417
- Horvath, S., Rescorla, L., & Arunachalam, S. (2019). The syntactic and semantic features of two-year olds' verb vocabularies: A comparison of typically developing children and late talkers. *Journal of Child Language*, 46(3), 409-432. <https://doi.org/10.3390/children8111052>
- Horvath, S., Kueser, J. B., Kelly, J., & Borovsky, A. (2022). Difference or delay? Syntax, semantics, and verb vocabulary development in typically developing and late-talking toddlers. *Language Learning and Development*, 18(3), 352-376. <https://doi.org/10.1080/15475441.2021.1977645>
- Jackson-Maldonado, D., Marchman, V.A. & Fernald, L.C.H. (2013). Short-form versions of the Spanish MacArthur-Bates Communicative Development Inventories. *Applied Psycholinguistics*, 34, 837-868. <https://doi.org/10.1017/S0142716412000045>
- Jackson-Maldonado, D., Friend, M., Marchman, V., Weisleder, A., Auza, A., Conboy, B., Rubio-Codina, M., & Dale, P. (2024). The MacArthur Inventario del Desarrollo de Habilidades Comunicativas III: A measure of language development in Spanish-speaking two- to four-year-olds. *Language Development Research*, 4, 361-398. <https://doi.org/10.34842/12fb-qz93>
- Karmiloff, K. & Karmiloff-Smith, A. (2005). *Hacia el lenguaje [Towards language]*. Morata.
- Kautto, A., Jansson-Verkasalo, E., & Mainela-Arnold, E. (2021). Generalized slowing rather than inhibition is associated with language outcomes in both late talkers and children with typical early development. *Journal of Speech, Language, and Hearing Research*, 64(4), 1222-1234. https://doi.org/10.1044/2020_JSLHR-20-00523
- López-Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S., & Martínez, M. (2005). *Inventarios de desarrollo comunicativo MacArthur. Manual [MacArthur-Bates Communicative Development Inventories. Manual]*. TEA.
- Llorenç, A., Ahufinger, N., Igualada, A., & Sanz-Torrent, M. (2021). Descripción del cambio del TEL al TDL en contexto angloparlante [Description of the change from SLI to DLD in an English-speaking context]. *Revista de Investigación en Logopedia*, 11(Núm. Especial), 9-20. <https://doi.org/10.5209/rlog.70857>
- Mariscal, S., López-Ornat, S., Gallego, C., Gallo, P., Karousou, A., & Martínez, M. (2007). La evaluación del desarrollo comunicativo y lingüístico mediante la versión española de los inventarios MacArthur-Bates [Evaluation of communicative and linguistic development using

- the Spanish version of the MacArthurBates inventories]. *Psicothema*, 19, 190-197.
- Mendoza, E. (2016). *Trastorno específico del lenguaje [Specific language disorder]*. Pirámide.
- Moreno, R., & Nieva, S. (2021). Intervención logopédica naturalista con familias de hablantes tardíos: efectos en el lenguaje infantil y en los intercambios adultos [Language and speech naturalistic intervention with late talkers' families: effects on child language and adult exchanges]. *Revista de Investigación en Logopedia*, 11(1), 61-75. <https://doi.org/10.5209/rlog.68157>
- Morgan, L., Delehanty, A., Dillon, J.C., Schatschneider, C., & Wetherby, A. M. (2020). Measures of early social communication and vocabulary production to predict language outcomes at two and three years in late-talking toddlers. *Early Childhood Research Quarterly*, 51, 366-378. <https://doi.org/10.1016/j.ecresq.2019.12.005>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. [Ten steps for test development]. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Muñiz, J., Fidalgo, A.M., García-Cueto, E., Martínez R., & Moreno, R. (2005). *Análisis de los ítems [Item analysis]*. La Muralla.
- Nouraei, P., Ayatollahi, M. A., & Moghadas, M. (2021). Late language emergence: A literature review. *Sultan Qaboos University Medical Journal*, 21(2), e182-e190. <https://doi.org/10.18295/squmj.2021.21.02.005>
- Paul, R. (1991). Profiles of toddlers with slow expressive language development. *Topics in Language Disorders*, 11(4), 1-13. <https://doi.org/10.1097/00011363-199111040-00003>
- Perry, L. K., Kucker, S. C., Horst, J. S., & Samuelson, L. K. (2023). Late bloomer or language disorder? Differences in toddler vocabulary composition associated with long-term language outcomes. *Developmental Science*, 26, Article e13342. <https://doi.org/10.1111/desc.13342>
- Rescorla, L. (1989). The language development survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, 54, 587-599. <https://doi.org/10.1044/jshd.5404.587>
- Rescorla L. (2011). Late talkers: do good predictors of outcome exist? *Developmental Disabilities Research Reviews*, 17(2), 141-150. <https://doi.org/10.1002/ddrr.1108>
- Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech Language and Hearing Research*, 45(2), 360-371. [https://doi.org/10.1044/1092-4388\(2002/028\)](https://doi.org/10.1044/1092-4388(2002/028))
- Rescorla, L., & Achenbach, T. M. (2002). Use of the language development survey (LDS) in a national probability sample of children 18 to 35 months old. *Journal of Speech, Language, and Hearing Research*, 45(4), 733-743. [https://doi.org/10.1044/1092-4388\(2002/059\)](https://doi.org/10.1044/1092-4388(2002/059))
- Rescorla, L. y Dale, P.S. (2013). *Late talkers. Language, development, interventions, and outcome*. Paul H. Brookes Publishing Co.
- Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language*, 27, 643-664. <https://doi.org/10.1017/s0305000900004232>
- Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language, and Hearing Research*, 40, 556-566. <https://doi.org/10.1044/jshr.4003.556>
- Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech Language and Hearing Research*, 51(2), 394-407. [https://doi.org/10.1044/1092-4388\(2008/029\)](https://doi.org/10.1044/1092-4388(2008/029))
- Rose, Y., & Blackmore, S. (2018). Questioning the role of lexical contrastiveness in phonological development: Converging evidence from perception and production studies. *Canadian Journal of Linguistics*, 63(4), 580-608. <https://doi.org/10.1017/cnj.2018.12>
- Rydz, D., Srouf, M., Oskoui, M., Marget, N., Shiller, M., Birnbaum, R., Majnemer, A., & Shevell, M. I. (2006). Screening for developmental delay in the setting of a community pediatric clinic: A prospective assessment of parent-report questionnaires. *Pediatrics*, 118(4), e1178-e1186. <https://doi.org/10.1542/peds.2006-0466>
- Sansavini, A., Favilla, M.E., Guasti, M.T., Marini, A., Millepiedi, S., Di Martino, M.V., Vecchi, S., Battajon, N., Bertolo, L., Capirci, O., Colatei, M. P., Frioni, C., Marotta, L., Massa, S., Michelazzo, L., Pecini, C., Piazzalunga, S., Pieretti, M., Rinaldi, P., ... Lorusso, M. L. (2021). Developmental language disorder: Early predictors, age for the diagnosis, and diagnostic tools. A scoping review. *Brain Sciences*, 11(5), Article 654. <https://doi.org/10.3390/brainsci11050654>
- Serra, M. (2008). Las interfaces entre los componentes del lenguaje: Tierra de nadie en un viaje obligatorio. In E. Diez-Itza (Ed.), *Estudios de desarrollo del lenguaje y educación [Studies on language development and education]* (pp. 511-51). Publicaciones de la Universidad de Oviedo.
- Serrat, E., Sanz-Torrent, M., Badia, I., Aguilar, E., Olmo R., Lara, F., Andreu, L., & Serra, M. (2010). La relación entre el aprendizaje léxico y el desarrollo gramatical [The relationship between lexical learning and grammatical development]. *Infancia y Aprendizaje*, 33(4), 435-448. <https://doi.org/10.1174/021037010793139590>
- Sireci, S., & Benítez, I. (2023). Evidence for test validation: A guide for practitioners. *Psicothema*, 35(3), 217-226. <https://doi.org/10.7334/psicothema2022.477>
- Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, 38(1), 1-34. <https://doi.org/10.1017/S0305000910000425>
- Sylvestre, A., Desmarais, C., Meyer, S. F., Bairati, I., & Leblond, J. (2017). Prediction of the outcome of children who had a language delay at age 2 when they are aged 4: Still a challenge. *International Journal of Speech & Language Pathology and Audiology*, 2, 1-14. <https://doi.org/10.1080/17549507.2017.1355411>
- Thal, D. (1991). Language and cognition in normal and late-talking toddlers. *Topics in Language Disorders*, 11(4), 33-42.
- Thal, D., & Bates, E. (1988). Language and gesture in late talkers. *Journal of Speech, Language, and Hearing Research*, 31, 115-123. <https://doi.org/10.1044/jshr.3101.115>
- Thal, D., Marchman, V.A., & Tomblin, J.B. (2013). Late-talking toddlers: Characterization and prediction of continued delay. In L. A. Rescorla, & P. S. Dale (Eds.). *Late talkers. Language, development, interventions and outcome* (pp. 169-201). Paul H. Brookes Publishing Co.
- Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *Clinical Neuropsychologist*, 31(6-7), 1113-1126. <https://doi.org/10.1080/13854046.2017.1317364>
- Ygual, A., Cervera, J.F., Baixauli, I., & Meliá, A. (2011). Protocolo de observación del lenguaje para maestros de educación infantil. Eficacia en la detección de dificultades semánticas y morfosintácticas [Language observation protocol for teachers in pre-school education. Effectiveness in the detection of semantic and morphosyntactic difficulties]. *Revista de Neurología*, 52, 127-134.
- Zimmerman, I. L., Steiner, V. G., & Pond, E. (2011). *Preschool Language Scales-fifth edition (PLS-5)*. Pearson Assessments Inc.

Article

Extended Persuasion: Elaboration Moderates Indirect Attitude Change

Diego Bustos¹ , Borja Paredes¹ , Lorena Moreno² , Pablo Briñol¹  and Richard E. Petty³ 

¹ Universidad Autónoma de Madrid (Spain)

² Universidad Complutense de Madrid (Spain)

³ Ohio State University (United States)

ARTICLE INFO

Received: 27/11/2024

Accepted: 25/04/2025

Keywords:

Indirect change
Elaboration
Attitude change
Persuasion

ABSTRACT

Background: Indirect attitude change effect occurs when a change in attitudes toward an object (i.e., focal attitude) is accompanied by a change in attitudes toward other related objects (i.e., distal attitude). The current research examines to what extent the amount of thinking (elaboration) involved in changing focal attitudes can moderate this indirect change effect. We expect that indirect change will be more likely to occur under high elaboration conditions (i.e., elaboration-indirect change effect). **Method:** Across two studies, participants' thought valence toward a healthy diet (Study 1) and toward coffee (Study 2) was manipulated to be positive or negative. The amount of elaboration accompanying the thought generation tasks was either measured (Study 1) or manipulated (Study 2). Finally, focal and distal attitudes were assessed in both studies to test the proposed elaboration-indirect change effect. **Results:** Focal attitudes were generally associated with distal attitudes. More importantly, the amount of thinking moderated this effect, with increased elaboration leading to more indirect change from focal to distal attitudes. **Conclusions:** Elaboration moderates indirect attitude change.

El Efecto de la Elaboración en el Cambio Indirecto de Actitudes

RESUMEN

Introducción: El cambio indirecto de actitudes ocurre cuando un cambio en las actitudes hacia un objeto focal (i.e., actitudes focales) va acompañado de un cambio en las actitudes hacia otros objetos potencialmente relacionados con los primeros (i.e., actitudes distales). La presente investigación estudia en qué medida la cantidad de pensamiento (elaboración) involucrada en el cambio de actitudes focales puede moderar este efecto de cambio indirecto. Esperamos que el cambio indirecto sea más probable cuanto más elaboración haya (i.e., efecto de elaboración sobre el cambio indirecto). **Método:** En dos estudios, se manipuló la valencia de los pensamientos hacia una dieta saludable (Estudio 1) y hacia el café (Estudio 2) para que fuera positiva o negativa. La elaboración que acompañaba esas inducciones fue medida (Estudio 1) o manipulada (Estudio 2). Finalmente, se midieron las actitudes focales y distales en ambos estudios. **Resultados:** Las actitudes focales se asociaron con las actitudes distales en ambos estudios. Los resultados también revelaron que una mayor elaboración produjo un mayor cambio indirecto desde las actitudes focales a las distales. **Conclusiones:** La elaboración modera el efecto de cambio indirecto de actitudes.

Palabras clave:

Cambio indirecto
Elaboración
Cambio de actitudes
Persuasión

Persuasion treatments can change attitudes that are not directly related to the specific persuasive proposal advocated. This phenomenon has been called indirect change (Alvaro & Crano, 1997), lateral change (Brannon et al., 2019; Linne et al., 2024), attitude generalization (Shook et al., 2007), spillover (Paredes et al., 2019), attitude displacement (Linne et al., 2020) and a secondary transfer effect (Pettigrew, 2009; Vezzali et al., 2023). Although there are important differences between these paradigms (see Bergh & Brandt, 2023; Glaser et al., 2015; Linne et al., 2024; Vezzali et al., 2021, for reviews), we conceptualize indirect attitude change as instances in which a persuasive treatment produces the intended change in an attitude targeted by the persuasive appeal (i.e., the focal attitude), but also leads to change in some other distal attitudes. Therefore, the indirect change approach of the present manuscript implies changes in both focal and distal attitudes as a result of a persuasive treatment.

Indirect change effect has been explored in a number of important research domains. As an illustrative example, Wolstenholme et al. (2020) examined the extent to which a treatment impacting consumption of red and processed meat also affected the willingness to eat any other type of meat and dairy. In this research, participants read an article about the environmental and/or health impacts of eating red and processed meat. Results indicated that the message reduced their red and processed meat consumptions (focal event). This behavior predicted then a reduced willingness to eat meat and dairy (distal event).

Indirect change effects have also been studied as a relatively subtle approach for reducing prejudiced attitudes (Pettigrew, 1997; Ratliff & Nosek, 2011; Schmid et al., 2012), for consumer attitudes (Ahluwalia et al., 2001; Kirchoff et al., 2018) and for environmentally relevant attitudes (Brügger & Höchli, 2019; Cruz, 2019; Evans et al., 2012). The present work contributes to this literature by examining indirect change effects in the context of previously unexplored, health-related attitudes, with important social consequences. Of greater conceptual relevance, the current research examines for the first time the amount of thinking about the persuasive message (elaboration) as a new potential moderator of indirect change effects.

Elaboration refers to the amount of thinking a person engages in when processing information. Beyond processing, elaboration involves adding something of their own to the information externally provided (Petty & Cacioppo, 1986). Elaboration is a key element of dual-process models like The Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986; Petty & Briñol, 2012). According to the ELM, the more an attitude is based on elaboration, the more it tends to persist over time, resist attempts at change, and perhaps most importantly, have consequences for other judgments and behavior (Petty et al., 1995; Krosnick & Petty, 1995). Put simply, the more an attitude change is based on extensive thought, the stronger that attitude is. Thus, even if high and low thinking processes resulted in the same degree of attitude change, the consequences of this influence in terms of stability and further impact on behavior can differ (e.g., Haugtvedt & Petty, 1992). Just as attitude changes induced through high elaboration are more persistent, resistant to change, and predictive of behavior than those induced through low elaboration, we propose in the present research that attitude changes of focal attitudes based on high elaboration are more likely to indirect change distal attitudes.

We propose that attitude change processes that require elaboration about the attitude object are likely to result in attitude representations that are well integrated and connected with other material in memory (McGuire, 1981; Tesser, 1978). Because of the strong linkage among constructs associated with elaboration, activating one mental representation should activate related ones relatively easily (Horcajo et al., 2010; Petty et al., 2008).

In sum, given that elaboration strengthens associations among mental constructs, and increases the accessibility and relevance of those associations, we propose that it would be likely to lead to change on distal attitudes as well as focal ones. This hypothesis is consistent with other research on indirect change effects, suggesting that stronger associations between mental representations and greater accessibility of a construct increase the likelihood of indirect change (Blankenship et al., 2012; Evans et al., 2013; Fazio et al., 2015). In the current research, we propose that the extent of elaboration is important not only because it leads to stronger attitudes (as shown in previous research) but also because it might be more consequential for indirect change from focal to distal attitudes. Thus, the current research has the potential to reconcile apparently contradictory results of prior research showing that sometimes attitudes toward an object indirectly change other attitudes toward distal objects whereas other times they do not (e.g., Spaccatini et al., 2023; Sütterlin & Siegrist, 2014).

In Study 1, participants were asked to generate positive or negative thoughts about a healthy diet. In this first study, the extent of elaboration was assessed objectively by counting the number of thoughts toward the diet listed by each participant. After assessing focal attitudes toward that diet (focal attitudes), participants were also asked to rate the social group of overweight people as part of an unrelated study (distal attitudes). In this initial study, we examined the extent to which focal attitudes about diets would be more predictive of distal attitudes regarding overweight people as a function of measured elaboration.

After testing the moderating role of elaboration by measuring it in Study 1, Study 2 enhanced internal and construct validity by manipulating participants' elaboration using an ability induction based on cognitive load. This study used novel materials for the attitude objects. Both the focal (i.e., coffee) and the distal (i.e., chocolate) attitudes were new to increase generalization across topics, materials, inductions, and measures. Thus, Study 2 sought to replicate and extend the effect found in Study 1 using different materials and inductions. In sum, the general prediction across studies is that elaboration will moderate indirect change from focal to distal attitudes.

Study 1

The goal of this study was to provide a first examination of the effect of elaboration on indirect attitude change from a focal to a distal attitude object. We began this study by randomly assigning participants to generate either positive or negative thoughts about the Mediterranean diet. This diet has been associated to healthier life, longer life expectancy, lower rates of obesity and a number of positive outcomes (Guasch-Ferré & Willett, 2021). Next, attitudes toward the diet were measured and served as focal attitudes. Elaboration was assessed by counting the number of thoughts listed by each

participant. Finally, participants were also asked to rate the social group of overweight people as part of an ostensibly unrelated study. Attitudes toward overweight people served as the distal attitudes.

We expected attitudes toward the diet (focal attitudes) to be more favorable in the positive (vs. negative) thoughts condition. Because of the association between healthy diets and weight (e.g., Mohammadbeigi et al., 2018; Puhl et al., 2012; Puhl et al., 2013; Tognon et al., 2014), we expected to find a link between focal attitudes toward the Mediterranean diet and distal attitudes toward overweight people. Most importantly, we expected this link to be stronger as the number of thoughts listed increased. In other words, we expected focal attitudes to predict distal attitudes to a greater extent for participants whose elaboration was relatively higher than lower.

Method

Participants

Two hundred and thirteen undergraduate students (184 women, 28 men, and 1 unidentified; $M_{age} = 19.89$; $SD = 3.28$) participated in this study in exchange for course credit. Participants were randomly assigned to conditions in a 2 Thought Valence (Negative vs. Positive thoughts) $\times$ Extent of Elaboration (continuous variable) design. Focal attitudes toward the diet and distal attitudes toward overweight individuals (continuous variables) served as the dependent measures. A *sensitivity* power analysis was conducted using *G*Power* (Faul et al., 2007). Results indicated that our final sample size for a two-tailed test ($\alpha = .05$) had .80 power to detect an effect size of Cohen's $f^2 = .037$ for the predicted two-way interaction between focal attitudes and elaboration on distal attitudes. Our final sample size was determined by the number of participants that we were able to collect during the two weeks in which the study was posted.

Procedure

Participants were told that they would be taking part in two studies designed to test different materials for future studies. First, as part of a study on life habits, participants were asked to generate either negative or positive thoughts concerning a healthy diet (Mediterranean diet). To assess the extent of elaboration, the number of thoughts listed by each participant during this initial task was registered. After listing their thoughts, participants reported their attitudes toward the diet (focal attitude measure). Finally, participants were also asked to rate the social group of overweight people as part of the control measures of an ostensibly unrelated study. Thus, attitudes toward overweight people were the distal attitudes in this study. After participants completed the dependent measures, they were debriefed, thanked and dismissed. Participants' debriefing followed standard ethical guidelines for research, adhering to the American Psychological Association [APA] (2017) ethical standards for research with human participants. This study

was approved by the Institutional Ethics Committee of Universidad Autónoma de Madrid (Approval Code: UAM-CEI-120- 2426).

Instruments

Predictor Variables

Thought Valence. Participants were randomly assigned to list either positive or negative thoughts concerning a healthy diet (Mediterranean diet). In the positive (negative) thoughts condition, participants were told to list as many positive (negative) aspects about the diet as they could. Participants could take as long as they needed and stop whenever they wanted. This manipulation of thought direction has been successfully used in previous research to influence attitudes toward diets (e.g., Briñol et al., 2013; Gascó et al., 2018; see Requero et al., 2021, for a review on persuasion and healthy eating).

Extent of Elaboration. Elaboration was assessed by counting the number of thoughts listed by each participant. Specifically, after reading about the diet, participants were given several boxes in which they could write down their thoughts. Participants could list up to 7 thoughts, with a higher number of thoughts indicating more elaboration. Previous research has shown that this measure can be used to classify participants according to their extent of thoughtful processing of a persuasive message (Barden & Petty, 2008; Burnkrant & Howard, 1984; Petty & Wegener, 1998).

Dependent Variables

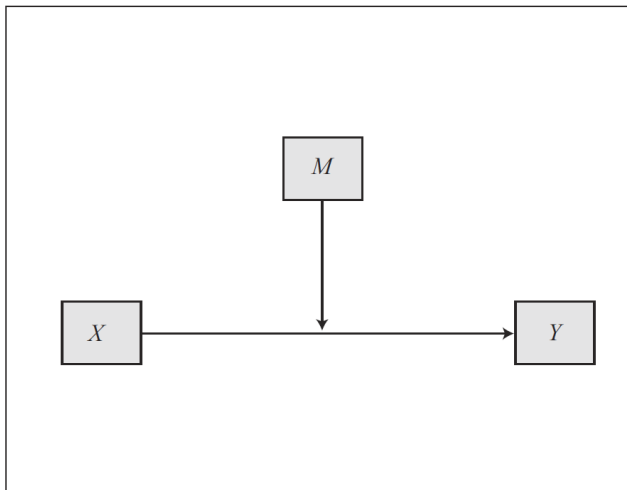
Focal Attitudes. Attitudes toward the diet were assessed using four semantic differential scales. Specifically, participants indicated to what extent the Mediterranean diet seemed: *bad vs. good*, *negative vs. positive*, *unfavorable vs. favorable*, and *harmful vs. beneficial*, on scales from 1 (e.g., *extremely bad*) to 9 (e.g., *extremely good*). Prior research has used similar items to assess attitudes (Gandarillas et al., 2018; Moreno et al., 2021). Item-ratings were highly correlated ($\alpha = .82$), thus averaged to create a merged attitude index. Responses to this attitude scale were scored so that higher values represented more favorable attitudes toward the assigned diet. Focal attitudes were reverse-coded for the distal attitudes' analysis because both attitudes were negatively correlated.

Distal Attitudes. Participants were indicated to what extent they had a positive opinion about overweight people, anchored by (1) *Not at all* and (9) *Totally*. Responses to this question were scored so that higher values represented more favorable attitudes toward overweight people. As noted, the object of distal attitudes (overweight people) was selected because of its conceptual relation to diet, as well as for its importance for reducing prejudiced attitudes as it has been used in prior research (Vezzali et al., 2023). Indeed, diet and body image are regularly considered together when making health-related choices (e.g., Breines et al., 2014; Cazzato et al., 2016; Requero et al., 2020).

Data Analysis

Pearson correlations and multiple linear regressions were run for this research. Focal attitudes were submitted to a multiple lineal regression analysis tested by using the PROCESS add-on for SPSS (Model 1; Hayes, 2013; see Figure 1). Thought valence (-1 = negative; 1 = positive), extent of elaboration (centered), and their interactions were entered as predictor variables. Distal attitudes were also submitted to a multiple lineal regression analysis with extent of elaboration (centered), focal attitudes (centered), and their interactions entered as predictor variables.

Figure 1
Conceptual Diagram of PROCESS Model 1 (Hayes, 2013)



Results

A non-significant negative correlation was observed between focal and distal attitudes $r(211) = -.073, p = .291$.

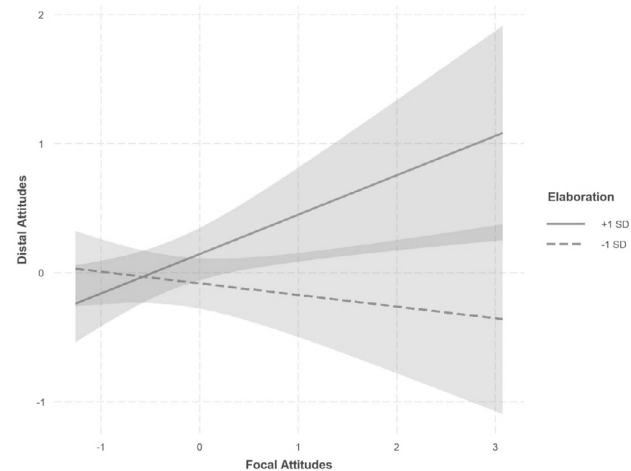
Focal Attitudes. This measure was submitted to a multiple lineal regression analysis tested by using the PROCESS add-on for SPSS (Model 1; Hayes, 2013). Thought valence (manipulated), extent of elaboration (centered), and their interaction were entered as predictor variables. As expected, results revealed a significant main effect of thought valence on focal attitudes such that those who listed positive thoughts reported significantly more favorable attitudes toward the diet, $B = 0.685, t(209) = 4.218, p < .001, 95\% \text{ CI: } [0.365, 1.006]$. Results also revealed a marginal main effect of elaboration, $B = -0.170, t(209) = -1.404, p = .162, 95\% \text{ CI: } [-0.409, 0.069]$, such that those who engaged in low elaboration (-1 *SD*) reported more favorable attitudes toward the diet than those who engaged in high elaboration (+1 *SD*). The interaction did not reach statistical significance ($p = .533$).

Distal Attitudes. The measure of distal attitudes was also submitted to a multiple lineal regression analysis. Extent of elaboration (centered), focal attitudes (centered), and their interactions were entered as our predictor variables. Results revealed a non-significant main effect of focal attitudes $B = 0.093, t(209) = 1.351, p = .178, 95\% \text{ CI: } [-0.043, 0.227]$. As predicted, results also revealed a significant two-way interaction between focal attitudes and elaboration, $B = 0.170, t(209) = 2.359, p = .019, f^2 = .026$ (small effect; Cohen, 1988), $95\% \text{ CI: } [0.028, 0.312]$ (see

Figure 2). Specifically, focal attitudes were positively associated with distal attitudes for those participants who engaged in high elaboration (+1 *SD*), $B = 0.262, t(209) = 2.508, p = .013, 95\% \text{ CI: } [0.056, 0.469]$. However, for those participants whose extent of elaboration was relatively low the effect of focal attitudes on distal attitudes did not reach significance (-1 *SD*), $B = -0.078, t(209) = -0.826, p = .410, 95\% \text{ CI: } [-0.262, 0.107]$. No other effects reached significance ($ps. > .325$).

Other research on attitudes indirect change has used message direction as a predictor for both focal and distal attitudes (Brannon et al., 2019; Linne et al., 2020). Conducting this analysis including thought valence (manipulated), extent of elaboration (centered), and their interactions as predictors of distal attitudes, results showed a marginal main effect of thought direction, $B = 0.242, t(209) = 1.574, p = .117, 95\% \text{ CI: } [-0.061, 0.546]$. No other effects reached significance ($ps. > .377$), meaning that the impact of the thought valence on distal attitudes did not vary as a function of elaboration.

Figure 2
Standardized Distal Attitudes Toward Overweight People as a Function of the Extent of Elaboration and Focal Attitudes in Study 1



The first study revealed that generating positive (vs. negative) thoughts about a diet resulted in more favorable (vs. unfavorable) attitudes toward that diet. More relevant to the present concerns, this study also revealed that elaboration moderated the subsequent link between focal and distal attitudes. That is, attitudes toward a healthy diet were associated with distal attitudes toward overweight people to a greater extent for those participants generating many (vs. few) thoughts about the diet.

Given that elaboration in the first study was measured rather than manipulated, there may be potential alternative interpretations, as confounding variables might have co-varied with elaboration (e.g., knowledge, experience, need for consistency, etc.). To address this potential issue and in order to establish a causal role in the effect of elaboration on indirect attitude change, study 2 employed an experimental approach by manipulating participants' elaboration. This manipulation also serves to prevent potential differences in thought content across conditions. In the next study we also varied the focal and distal attitude objects to generalize across domains. Lastly, given that focal attitudes were just as affected by the thought valence induction regardless of their higher (vs. lower) elaboration

conditions, it stands to reason that the indirect change effect found in Study 1 is due to focal attitudes being *stronger* under higher elaboration, and not due to focal attitudes being more *extreme* under higher elaboration. Therefore, the next study aims at replicating this effect by generating similar attitude change for high and low elaboration, via assigning participants to different directions of a message that uses both strong arguments and heuristic cues (thus keeping focal attitude extremity constant for both high and low elaboration conditions; see [Haugtvedt & Petty, 1992](#)). And then have that change affect distal attitudes more under high (vs. low) elaboration conditions.

Study 2

Study 1 revealed that measured elaboration moderates the subsequent link between focal and distal attitudes. Study 2 was designed to replicate and generalize to other attitude objects the findings of the prior study and gain greater internal and construct validity using a fully experimental design in which both predictors of the elaboration-indirect change effect (focal attitudes, and elaboration) were manipulated rather than measured.

Participants were first assigned to a high vs. low elaboration condition. Elaboration was manipulated through cognitive load, specifically, participants were asked to memorize a short vs. long number ([Block et al., 2010](#)). Next, participants were randomly assigned to read a message either in favor or against coffee. Both conditions of the message contained strong arguments that participants in the high elaboration condition were more likely to process, as well as heuristic cues (e.g., titles with large, bolded fonts) that participants in the low elaboration condition could rely on. The technique of using both strong arguments with positive cues to produce similar degrees of attitude change in both high and low elaboration groups has proven successful in prior research (e.g., [Haugtvedt & Petty, 1992](#)). Participants were then asked to report their attitudes toward coffee (i.e., focal attitude) as well as their attitudes toward chocolate (i.e., distal attitude). These objects were selected because there is room for people to link them, considering that both have caffeine, both have a dark color, both are associated with pleasure and common ingredients in some famous recipes, such as tiramisu, and both share connotations of pleasure, comfort, and sometimes health concerns. Although they can be related when thinking about their potential similarities, the association can be remote enough for some people not to make that link if they do not think about it.

We predicted that the message direction would impact focal attitudes both for high and for low elaboration conditions. We also expected a similar effect of message direction regardless of elaboration on distal attitudes. However, we expected such focal attitudes to predict distal attitudes significantly more for those who were assigned to the high elaboration conditions relative to those who were assigned to the low elaboration conditions.

Method

Participants

One hundred and fifty-seven people (76 women and 81 men; $M_{age} = 42.40$; $SD = 12.87$) participated in this study online via CloudResearch. All participants were randomly assigned to

conditions in a 2 Elaboration (Low vs. High) $\times$ 2 Message Direction (in favor vs. against coffee) design. Due to having a higher level of experimental control, we expected the effects to be potentially larger than those obtained in the previous study. A *sensitivity* power analysis was conducted using *G*Power* ([Faul et al., 2007](#)). Results indicated that our final sample size for a two-tailed test ($\alpha = .05$) had .80 power to detect an effect size of Cohen's $f^2 = .051$ for the predicted interaction between focal attitudes and elaboration on distal attitudes. Our final sample size was determined by the number of participants that we were able to collect during the two weeks in which the study was posted.

Procedure

First, participants were told that they were taking part in a study to know their opinion toward coffee. Then, as part of an ostensibly different study, they were asked to memorize a number. Half of them were asked to memorize a three digit one (low mental load), and the other half an eleven-digit number (high mental load). Next, participants were randomly assigned to read a message in favor or a message against coffee, with three arguments supporting or opposing coffee intake. The messages were created based on previously used messages about coffee consumption (e.g., [Block & Williams, 2002](#); [Petty et al., 1993](#); [Philipp-Muller et al., 2022](#)). Then, participants reported their attitudes toward coffee, and after doing so, they were asked to provide their opinion toward chocolate as part of a separate study. Finally, after participants completed the dependent measures, they were debriefed, thanked, and dismissed, as in Study 1.

Instruments

Predictor Variables

Elaboration. Participants were told that as a part of the study they needed to memorize a short list of numbers. Half of them were told to remember a three-digit number (low cognitive load), and the other half were told to remember an eleven-digit number (high cognitive load). The longer number was expected to distract and increase mental load, thus impairing participants' mental capacity. Previous research has revealed that this procedure is useful to vary ability to elaborate, as people have fewer cognitive resources available under high-load conditions and therefore they have less capacity to think extensively ([Block et al., 2010](#); [Cacioppo & Petty, 1989](#); [Moreno et al., 2024](#)).

Message Direction. Participants were presented with a persuasive message which contained either arguments in favor or against coffee. The manipulation was designed to influence the opinions of participants toward coffee, such as that those in the first condition would hold more favorable attitudes toward coffee than those on the second condition. The gist of some strong arguments in favor of coffee were that coffee drinkers are more likely to resist development of dementia and Alzheimer's later in life and caffeine in coffee is a well-known stimulant, coffee promotes alertness, attention and wakefulness. The gist of some strong arguments against coffee were that coffee can induce stomach ulcers and impair digestion by raising stomach acidity levels and caffeine is a drug, a mild central nervous system stimulant, and it produces dependence.

Dependent Variables

Elaboration Check. Participants reported to what extent they were very distracted vs. not distracted at all regarding the task of remembering the numbers using a 9-point scale ranging from 1 (not at all) to 9 (extremely). Responses of this item were scored so that higher values represented being less distracted while remembering those numbers.

Focal Attitudes. Attitudes toward coffee were assessed using three out of the four items in Study 1 (*bad vs. good, negative vs. positive and unfavorable vs. favorable*). Item-ratings were highly correlated ($\alpha = .98$), thus averaged to create a merged attitude index. Responses to this attitude scale were scored so that higher values represented more favorable attitudes toward coffee.

Distal Attitudes. Participants reported their distal attitudes toward chocolate using the same three 9-point scales as for the focal attitude. Ratings were highly inter-correlated ($\alpha = .94$), thus averaged to create one overall distal attitude index. Higher values represented more favorable evaluations toward chocolate. As noted, this particular distal object (chocolate) was selected because this study aims to show an object to object elaboration-indirect change effect.

Data Analysis

The conducted analyses were identical to those conducted in Study 1.

Results

A significant and positive correlation was observed between the attitudes, $r(155) = .512, p < .001$.

Elaboration Check. Results of the 2 (Elaboration: high or low) $\times$ 2 (Message Direction: in favor or against coffee) ANOVA conducted on the elaboration check measure revealed a significant main effect of elaboration, such that participants assigned to the high elaboration condition reported being less distracted ($M = 8.62, SD = 1.15$) than those assigned to the lower elaboration condition ($M = 7.44, SD = 2.53$), $F(1, 153) = 12.762, p < .001$. Results also revealed a significant main effect of message direction, such that participants who received a message in favor of coffee reported being less distracted ($M = 8.36, SD = 1.76$) than did those who received the message against it ($M = 7.71, SD = 2.28$), $F(1, 153) = 4.188, p = .042$. The two-way interaction did not reach significance ($ps. > .367$).

Focal Attitudes. Results of the 2 (Elaboration: high or low) $\times$ 2 (Message Direction: in favor or against coffee) ANOVA conducted on the focal attitudes measure revealed a main effect of message direction, such that participants who received a message in favor of coffee held more favorable attitudes toward coffee ($M = 7.66, SD = 1.75$) than did those who received the message against it ($M = 6.62, SD = 2.27$), $F(1, 153) = 9.164, p = .003$. No other effects reached significance ($ps. > .203$).

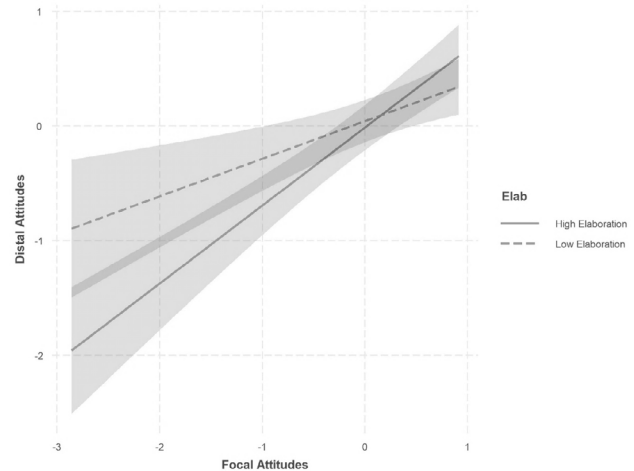
Distal Attitudes. Results revealed a significant main effect of focal attitudes on attitudes toward chocolate, $B = 0.504, t(153) = 7.377, p < .001, 95\% \text{ CI: } [0.369, 0.639]$, such that reporting more favorable attitudes toward coffee was associated with reporting more favorable attitudes toward chocolate. As predicted, results also revealed a significant two-way interaction between focal attitudes and elaboration, $B = 0.176, t(153) = 2.578, p = .011, f^2 = .043, 95\% \text{ CI: } [0.041, 0.311]$ (see Figure 3). Specifically, the effect of focal attitudes on distal attitudes was significantly greater for those participants assigned to the high elaboration condition, $B = 0.680, t(153) = 7.149, p < .001, 95\% \text{ CI: } [0.492, 0.868]$, than for those assigned to the low elaboration condition, $B = 0.328, t(153) = 3.343, p = .001, 95\% \text{ CI: } [0.134, 0.522]$. No other effects reached significance ($ps. > .683$).

As in Study 1, we conducted the same analysis with message direction as an additional predictor. We ran a multiple linear regression with message direction (manipulated), elaboration (manipulated), and their interactions as predictors of distal attitudes. Results showed significant main effects of message direction, $B = 0.319, t(153) = 2.317, p = .022, 95\% \text{ CI: } [0.047, 0.591]$. As in Study 1, no other effects reached significance ($ps. > .346$).

When we collapse both studies ($n = 370$), the two-way interaction between focal attitudes and elaboration remains significant, $B = 0.178, t(366) = 3.265, p = .001, f^2 = .029, 95\% \text{ CI: } [0.071, 0.286]$. A sensitivity analysis indicated that the sample size of the collapsed studies had a .80 power to detect an effect size of $f^2 = .022$ and above.

Figure 3

Standardized Distal Attitudes Toward Chocolate as a Function of Elaboration (Low vs. High Capacity) and Focal Attitudes in Study 2



Study 2 replicated the impact of elaboration on indirect attitude change using new materials. Furthermore, this study generated attitude change via randomly assigning participants to a positive vs. negative message, and this change in focal attitudes indirectly changed distal attitudes more under high (vs. low) elaboration conditions. Elaboration was manipulated rather than measured, addressing potential confounds present in the first study (e.g., knowledge, experience, need for consistency, etc.).

Discussion

Taken together, this research revealed that attitudes may indirectly change other relevant attitudes, especially under high thinking conditions. Across two studies, changes in focal attitudes were associated with changes in distal attitudes to a greater extent when participants were higher in elaboration (both measured and manipulated; see Table 1). Among other implications, these findings are important because they can help to specify when indirect change effect is more likely to emerge, with the potential for shedding light on reconciling past findings showing that sometimes there is an indirect change following attitude change whereas at other times the effect does not emerge (Spaccatini et al., 2023; Sütterlin & Siegrist, 2014).

Table 1
Summary of the Interactions Found in Studies 1 and 2

Variable	Study 1 (n=213)				
	<i>B</i>	<i>t</i>	<i>p</i>	IC 95%	
Focal Attitudes	.093	1.351	.178	-.043	.227
Focal Attitudes x Elaboration	.170	2.359	.019	.028	.312
Variable	Study 2 (n=157)				
	<i>B</i>	<i>t</i>	<i>p</i>	IC 95%	
Focal Attitudes	.504	7.377	<.001	.369	.639
Focal Attitudes x Elaboration	.176	2.578	.011	.041	.311

This finding extends the Elaboration Likelihood Model by showing that the impact of elaboration goes beyond the strength and persistence of attitude change to influence the indirect change across different domains. It could also mean the possibility of reinterpreting previous moderators as if they were cases in which elaboration might have co-varied with the proposed moderator. For instance, change in attitudes that are based on unexpected arguments (Glaser et al., 2015) would spread relatively more because such arguments might have been perceived as surprising, thus increasing elaboration (Petty et al., 2001).

Results from this research have significant real-life applications that could be applied to public health campaigns or marketing. Interventions designed to promote healthy behaviors (e.g., regular exercise) could have unintended positive effects for other distal attitudes (e.g., reduced smoking), especially for audiences in higher elaboration conditions. Also, marketing campaigns that successfully change consumer attitudes toward a specific product (e.g., eco-friendly packaging) could potentially be affecting other desirable distal attitudes, such as increased recycling or purchasing other sustainable products (see Horcajo et al., 2010), particularly for those are highly motivated and/or capable to process such marketing campaigns. Finally, results from this research also have potential implications for reducing prejudiced attitudes toward a social group in a way that direct contact with the group is not needed (Pettigrew & Tropp, 2006). As noted, this research has the potential to advance equity, inclusion and anti-racism by explaining how and when evaluations unrelated to prejudice can be consequential for discrimination toward stigmatized groups, even when prejudiced attitudes toward one collective might indirectly change attitudes to other disadvantaged collectives, and even when attitudes toward one particular individual might indirectly change attitudes over the entire social category for which that person belongs. Future studies can benefit from exploring the role of elaboration in those prejudice-relevant paradigms.

Future research should explore the boundaries and mechanisms of the elaboration-indirect change effect further. For example, the extent to which participants perceive the link between focal and distal attitudes may vary this effect, and prior knowledge can influence how this link is perceived. It is known that focal and distal attitudes must be related (or to be perceived to be associated) in some way for indirect change effect to occur (Glaser et al., 2015), and elaboration makes that effect stronger. Future studies should examine to what extent elaboration can also contribute to creating new linkages with other distal attitudes for which prior associations do not exist. Also, if the focal-distal attitude link is too obvious, people might assume that the indirect change to distal attitudes is an unwanted bias potentially contaminating judgments, in which case people might engage in correcting effects to reduce this perceived influence (Wegener & Petty, 1995). Additionally, there is evidence that

motivational variables are also important in affecting the likelihood of message elaboration (Petty & Cacioppo, 1986). Therefore, future studies could explore how motivational variables, such as personal relevance, can also influence indirect change. Future research can also benefit from achieving larger and more diverse samples, as well as replicating these results under more ecological approaches. Moreover, examining the long-term persistence of indirect attitude change moderated by elaboration would provide insights into the durability of the changes obtained over time. These directions could offer a deeper understanding of how elaboration functions across different contexts and how it can be optimally utilized to foster widespread and enduring attitude change.

Author Contributions

Diego Bustos: Formal analysis, Investigation, Methodology, Writing – Original draft. **Borja Paredes:** Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – Review & editing. **Lorena Moreno:** Conceptualization, Formal analysis, Methodology, Writing – Review & editing. **Pablo Briñol:** Conceptualization, Funding acquisition, Supervision, Writing - Review and editing. **Richard Petty:** Conceptualization, Funding acquisition, Writing - Review and editing.

Funding

This research was supported by the Ministerio de Ciencia, Innovación y Universidades (Spain) (PID2020-116651GB-C31; PID2023-148050NA-I00; PID2023-148146NB-I00) and the Consejería de Educación, Ciencia y Universidades, Comunidad de Madrid (Spain) (SI4/PJI/2024-00244). This work was supported also by the Templeton World Charity Foundation under Grant TWCF-2023-32571. The funding source has not been involved in the design of the study, the collection, analysis or interpretation of data, the writing of the manuscript, or the decision to submit it for publication.

Declaration of Interest

The authors declares that there is no conflict of interest.

Data Availability Statement

All data materials can be made available upon request.

References

- Ahluwalia, R., Unnava, H. R., & Burnkand, R. (2001). The moderating role of commitment on the spillover effect of marketing communications. *Journal of Marketing*, 38(4), 458-470. <https://doi.org/10.1509/jmkr.38.4.458.18903>
- American Psychological Association [APA]. (2017). *Ethical principles of psychologists and code of conduct*. American Psychological Association. <https://www.apa.org/ethics/code/>
- Alvaro, E. M., & Crano, W. D. (1997). Indirect minority influence: Evidence for leniency in source evaluation and counter argumentation. *Journal of Personality and Social Psychology*, 72(5), 949-964. <https://doi.org/10.1037/0022-3514.72.5.949>

- Barden, J., & Petty, R. E. (2008). The mere perception of elaboration creates attitude certainty: Exploring the thoughtfulness heuristic. *Journal of Personality and Social Psychology*, 95(3), 489-509. <https://doi.org/10.1037/a0012559>
- Bergh, R., & Brandt, M. J. (2023). Generalized prejudice: Lessons about social power, ideological conflict, and levels of abstraction. *European Review of Social Psychology*, 34(1), 92-126. <https://doi.org/10.1080/10463283.2022.2040140>
- Blankenship, K. L., Wegener, D. T., & Murray, R. A. (2012). Circumventing resistance: Using values to indirectly change attitudes. *Journal of Personality and Social Psychology*, 103(4), 606-621. <https://doi.org/10.1037/a0029226>
- Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, 134(3), 330-343. <https://doi.org/10.1016/j.actpsy.2010.03.006>
- Block, L. G., & Williams, P. (2002). Undoing the effects of seizing and freezing: Decreasing defensive processing of personally relevant messages. *Journal of Applied Social Psychology*, 32(4), 803-830. <https://doi.org/10.1111/j.1559-1816.2002.tb00243.x>
- Brannon, S. M., DeJong, A., & Gawronski, B. (2019). Determinants of lateral attitude change: The roles of object relatedness, attitude certainty, and moral conviction. *Social Cognition*, 37(6), 624-658. <https://doi.org/10.1521/soco.2019.37.6.624>
- Breines, J., Toole, A., Tu, C., & Chen, S. (2014). Self-compassion, body image, and self-reported disordered eating. *Self and Identity*, 13(4), 432-448. <https://doi.org/10.1080/15298868.2013.838992>
- Briñol, P., Gascó, M., Petty, R. E., & Horcajo, J. (2013). Treating thoughts as material objects can increase or decrease their impact on valuation. *Psychological Science*, 24(1), 41-47. <https://doi.org/10.1177/0956797612449176>
- Brügger, A. & Höchli, B. (2019). The role of attitude strength in behavioral spillover: attitude matters—but not necessarily as a moderator. *Frontiers in Psychology*, 10, 160-183. <https://doi.org/10.3389/fpsyg.2019.01018>
- Burnkrant, R. E., & Howard, D. J. (1984). Effects of the use of introductory rhetorical questions versus statements on information processing. *Journal of Personality and Social Psychology*, 47(6), 1218-1230. <https://doi.org/10.1037/0022-3514.47.6.1218>
- Cacioppo, J. T., & Petty, R. E. (1989). Effects of message repetition on argument processing, recall, and persuasion. *Basic and Applied Social Psychology*, 10(1), 3-12. https://doi.org/10.1207/s15324834basp1001_2
- Cazzato, V., Mian, E., Mele, S., Tognana, G., Todisco, P., & Urgesi, C. (2016). The effects of body exposure on self-body image and esthetic appreciation in anorexia nervosa. *Experimental Brain Research*, 234(3), 695-709. <https://doi.org/10.1007/s00221-015-4498-z>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Cruz, S. M. (2019). Lateral attitude change on environmental issues: implications for the climate change debate. *Climatic Change*, 156, 151-169. <https://doi.org/10.1007/s10584-019-02474-x>
- Evans, C. E., Christian, M. S., Cleghorn, C. L., Greenwood, D. C., & Cade, J. E. (2012). Systematic review and meta-analysis of school-based interventions to improve daily fruit and vegetable intake in children aged 5 to 12 y. *The American journal of clinical nutrition*, 96(4), 889-901. <https://doi.org/10.3945/ajcn.111.030270>
- Evans, L., Maio, G. R., Corner, A., Hodgetts, C. J., Ahmed, S., & Hahn, U. (2013). Self-interest and pro-environmental behaviour. *Nature Climate Change*, 3, 122-125. <https://doi.org/10.1038/nclimate1662>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. <https://doi.org/10.3758/BF03193146>
- Fazio, R. H., Pietri, E. S., Rocklage, M. D., & Shook, N. J. (2015). Chapter three - Positive versus negative valence: asymmetries in attitude formation and generalization as fundamental individual differences. *Advances in Experimental Social Psychology*, 51, 97-146. <https://doi.org/10.1016/bs.aesp.2014.09.002>
- Gandarillas, B., Briñol, P., Petty, R. E., & Díaz, D. (2018). Attitude change as a function of the number of words in which thoughts are expressed. *Journal of Experimental Social Psychology*, 74, 196-211. <https://doi.org/10.1016/j.jesp.2017.09.012>
- Gascó, M., Briñol, P., Santos, D., Petty, R. E., & Horcajo, J. (2018). Where did this thought come from? A self-validation analysis of the perceived origin of thoughts. *Personality & Social Psychology Bulletin*, 44(11), 1615-1628. <https://doi.org/10.1177/0146167218775696>
- Glaser, T., Dickel, N., Liersch, B., Rees, J., Süßenbach, P., & Böhner, G. (2015). Lateral attitude change. *Personality and Social Psychology Review*, 19(3), 257-276. <https://doi.org/10.1177/1088868314546489>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Guasch-Ferré, M., & Willett, W. C. (2021). The Mediterranean diet and health: a comprehensive overview. *Journal Of Internal Medicine*, 290(3), 549-566. <https://doi.org/10.1111/joim.13333>
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.
- Haugtvedt, C. P., & Petty, R. E. (1992). Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes. *Journal of Personality and Social Psychology*, 63(2), 308-319. <https://doi.org/10.1037/0022-3514.63.2.308>
- Horcajo, J., Briñol, P., & Petty, R. E. (2010). Consumer persuasion: Indirect change and implicit balance. *Psychology & Marketing*, 27(10), 938-963. <https://doi.org/10.1002/mar.20367>
- Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty, & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1-24). Lawrence Erlbaum Associates, Inc.
- Kirchoff J., Nichols B., Stolze H., & Brown C. (2018). The spillover effects of negative supply chain information on consumers' perceptions of product attributes. In N. Krey, & P. Rossi (Eds.), *Back to the future: Using marketing basics to provide customer value. AMSAC 2017. Developments in Marketing Science: Proceedings of the Academy of Marketing Science*. Springer, Cham. https://doi.org/10.1007/978-3-319-66023-3_242
- Linne, R., Glaser, T. & Böhner, G. (2024). Lateral attitude change: displacement effects as a reaction to blatant social influence. *Current Psychology*, 43(23), 1-7. <https://doi.org/10.1007/s12144-024-05906-9>
- Linne, R., Glaser, T., Pum, K., & Böhner, G. (2020). Lateral attitude change: Stalking the elusive displacement effect. *Social Cognition*, 38(4), 324-353. <https://doi.org/10.1521/soco.2020.38.4.324>
- McGuire, W.J. (1981). The probabiological model of cognitive structure and attitude change. In R.E. Petty, T.M. Ostrom, & T.C. Brock (Eds.), *Cognitive responses in persuasion*. Erlbaum.
- Mohammadbeigi, A., Asgarian, A., Moshir, E., Heidari, H., Afrashteh, S., Khazaei, S., & Ansari, H. (2018). Fast food consumption and overweight/

- obesity prevalence in students and its association with general and abdominal obesity. *Journal of preventive medicine and hygiene*, 59(3), E236-E240. <https://doi.org/10.15167/2421-4248/jpmh2018.59.3.830>
- Moreno, L., Requero, B., Santos, D., Paredes, B., Briñol, P., & Petty, R. E. (2021). Attitudes and attitude certainty guiding pro-social behaviour as a function of perceived elaboration. *European Journal of Social Psychology*, 51(6), 990-1006. <https://doi.org/10.1002/ejsp.2798>
- Moreno, L., Briñol, P., Paredes, B., & Petty, R. E. (2024). Scientific identity and STEM-relevant outcomes: Elaboration moderates use of identity certainty. *Journal of Experimental Social Psychology*, 115, 1-16. <https://doi.org/10.1016/j.jesp.2024.104663>
- Paredes, B., Guyer, J. J., Briñol, P., & Petty, R. E. (2019). Subtle priming of subtraction versus addition: A spill-over effect of math / La inducción sutil de sumar vs. restar: un efecto indirecto de las matemáticas. *International Journal of Social Psychology*, 34(3), 590-622. <https://doi.org/10.1080/02134748.2019.1649890>
- Pettigrew, T. F. (1997). Generalized intergroup contact effects on prejudice. *Personality and Social Psychology Bulletin*, 23(2), 173-185. <https://doi.org/10.1177/0146167297232006>
- Pettigrew, T. F. (2009). Secondary transfer effect of contact. *Social Psychology*, 40(2), 55-65. <https://doi.org/10.1027/1864-9335.40.2.55>
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751-783. <https://doi.org/10.1037/0022-3514.90.5.751>
- Petty, R. E., & Briñol, P. (2012). The elaboration likelihood model. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 224-245). Sage Publications Ltd. <https://doi.org/10.4135/9781446249215.n12>
- Petty, R. E., & Cacioppo, J. T. (1986). *The elaboration likelihood model of persuasion*. Springer.
- Petty, R. E., DeMarree, K. G., Briñol, P., Horecjo, J., & Strathman, A. J. (2008). Need for cognition can magnify or attenuate priming effects in social judgment. *Personality and Social Psychology Bulletin*, 34(7), 900-912. <https://doi.org/10.1177/0146167208316692>
- Petty, R. E., Fleming, M. A., Priester, J. R., & Feinstein, A. H. (2001). Individual versus group interest violation: Surprise as a determinant of argument scrutiny and persuasion. *Social Cognition*, 19(4), 418-442. <https://doi.org/10.1521/soco.19.4.418.20758>
- Petty, R. E., Haugtvedt, C. P., & Smith, S. M. (1995). Elaboration as a determinant of attitude strength: Creating attitudes that are persistent, resistant, and predictive of behavior. In R. E. Petty, & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 93-130). Lawrence Erlbaum Associates, Inc.
- Petty, R. E., Schumann, D. W., Richman, S. A., & Strathman, A. J. (1993). Positive mood and persuasion: Different roles for affect under high- and low-elaboration conditions. *Journal of Personality and Social Psychology*, 64(1), 5-20. <https://doi.org/10.1037/0022-3514.64.1.5>
- Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 323-390). McGraw-Hill.
- Philipp-Muller, A., Teeny, J. D., & Petty, R. E. (2022). Do consumers care about morality? A review and framework for understanding morality's marketplace influence. *Consumer Psychology Review*, 5(1), 107-124. <https://doi.org/10.1002/arcp.1072>
- Puhl, R., Luedicke, J., & Lee Peterson, J. (2013). Public reactions to obesity-related health campaigns: a randomized controlled trial. *American Journal of Preventive Medicine*, 45(1), 36-48. <https://doi.org/10.1016/j.amepre.2013.02.010>
- Puhl, R., Peterson, J. L., & Luedicke, J. (2012). Fighting obesity or obese persons? Public perceptions of obesity-related health messages. *International Journal of Obesity* (2005), 37(6), 774-782. <https://doi.org/10.1038/ijo.2012.156>
- Ratliff, K. A., & Nosek, B. A. (2011). Negativity and outgroup biases in attitude formation and transfer. *Personality and Social Psychology Bulletin*, 37(12), 1692-1703. <https://doi.org/10.1177/0146167211420168>
- Requero, B., Briñol, P., Moreno, L., Paredes, B., & Gandarillas, B. (2020). Promoting healthy eating by enhancing the correspondence between attitudes and behavioral intentions. *Psicothema*, 32(1), 60-66. <https://doi.org/10.7334/psicothema2019.154>
- Requero, B., Santos, D., Cancela, A., Briñol, P., & Petty, R. E. (2021). Promoting healthy eating practices through persuasion processes. *Basic and Applied Social Psychology*, 43(4), 239-266. <https://doi.org/10.1080/01973533.2021.1929987>
- Schmid, K., Hewstone, M., Küpper, B., Zick, A., & Wagner, U. (2012). Secondary transfer effects of intergroup contact: A cross-national comparison in Europe. *Social Psychology Quarterly*, 75(1), 28-51. <https://doi.org/10.1177/0190272511430235>
- Shook, N. J., Fazio, R. H., & Eiser, J. R. (2007). Attitude generalization: Similarity, valence, and extremity. *Journal of Experimental Social Psychology*, 43(4), 641-647. <https://doi.org/10.1016/j.jesp.2006.06.005>
- Spaccatini, F., Riva, P., Richetin, J., Porcelli, E., Pancani, L., Capellini, R., & Sacchi, S. (2023). From past to present (for a better future): The moderating role of cognitive mindset on spillover effects in environmental behaviors. *Current Psychology*, 42, 15858-15873. <https://doi.org/10.1007/s12144-022-02917-2>
- Sütterlin, B., & Siegrist, M. (2014). The reliance on symbolically significant behavioral attributes when judging energy consumption behaviors. *Journal of Environmental Psychology*, 40, 259-272. <https://doi.org/10.1016/j.jenvp.2014.07.005>
- Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 289-338). [https://doi.org/10.1016/s0065-2601\(08\)60010-6](https://doi.org/10.1016/s0065-2601(08)60010-6)
- Tognon, G., Hebestreit, A., Lanfer, A., Moreno, L. A., Pala, V., Siani, A., Tornaritis, M., De Henauw, S., Veidebaum, T., Molnár, D., Ahrens, W., & Lissner, L. (2014). Mediterranean diet, overweight and body composition in children from eight European countries: Cross-sectional and prospective results from the IDEFICS study. *Nutrition, Metabolism and Cardiovascular Diseases*, 24(2), 205-213. <https://doi.org/10.1016/j.numecd.2013.04.013>
- Tormala, Z. L., Briñol, P., & Petty, R. E. (2004). Hidden effects of persuasion. *Advances in Consumer Research*, 31, 75-76. <https://pablobrinol.com/wp-content/uploads/papers/2004%20ACR%20-hidden%20effects-.pdf>
- Tormala, Z. L., & Petty, R. E. (2002). What doesn't kill me makes me stronger: the effects of resisting persuasion on attitude certainty. *Journal of Personality and Social Psychology*, 83(6), 1298-1313. <https://doi.org/10.1037/0022-3514.83.6.1298>
- Vezzali, L., Di Bernardo, G. A., Cocco, V. M., Stathi, S., & Capozza, D. (2021). Reducing prejudice in the society at large: A review of the

- secondary transfer effect and directions for future research. *Social and Personality Psychology Compass*, 15(3), e12583. <https://doi.org/10.1111/spc3.12583>
- Vezzali, L., Pagliaro, S., Di Bernardo, G. A., McKeown, S., & Margherita Cocco, V. (2023). Solidarity across group lines: secondary transfer effect of intergroup contact, perceived moral distance, and collective action. *European Journal of Social Psychology*, 53, 450-470. <https://doi.org/10.1002/ejsp.2914>
- Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, 68(1), 36-51. <https://doi.org/10.1037/0022-3514.68.1.36>
- Wolstenholme, E., Poortinga, W., & Whitmarsh, L. (2020). Two birds, one stone: the effectiveness of health and environmental messages to reduce meat consumption and encourage pro-environmental behavioral spillover. *Frontiers in Psychology*, 11, 577111. <https://doi.org/10.3389/fpsyg.2020.577111>

Article

Relationship Between Online Trolling and Dark Tetrad Personality Traits: A Meta-Analysis

Sergio Hidalgo-Fuentes¹ , Elisa González-Pérez²  and Isabel Martínez-Álvarez² 

¹ Universidad de Valencia (Spain)

² Universidad a Distancia de Madrid (UDIMA) (Spain)

ARTICLE INFO

Received: 14/05/2025

Accepted: 08/07/2025

Keywords:

Online trolling
Machiavellianism
Narcissism
Psychopathy
Sadism

ABSTRACT

Background: Online trolling refers to a specific form of disruptive behavior in digital environments, aimed at interrupting interactions, provoking other users, and drawing them into fruitless arguments. The Dark Tetrad personality traits (narcissism, Machiavellianism, psychopathy and sadism) have been linked to various forms of antisocial behavior in virtual contexts. **Method:** A systematic literature review was conducted across the PsycINFO, PubMed, Scopus, ERIC, and Web of Science Core Collection databases to identify relevant studies examining these variables. A total of 24 relevant articles from 11 countries were identified, comprising a combined sample size of 14,044 participants. Four random-effects meta-analyses were performed using Pearson's r coefficients. **Results:** The results reveal a positive association between all four personality traits and greater tendency to engage in online trolling. Meta-regression analyses show that the assessment instruments used moderated some of the observed associations. No evidence of publication bias was detected. **Conclusions:** In general, the relationship between Dark Tetrad traits and online trolling was found to be positive, helping to explain individual differences in it.

Relación Entre el Trolling Online y los Rasgos de Personalidad de la Tétrada Oscura: un Metaanálisis

RESUMEN

Antecedentes: El trolling online se refiere a una forma específica de comportamiento disruptivo en entornos digitales, cuyo propósito es interrumpir las interacciones, provocar a otros usuarios y arrastrarlos a discusiones estériles. Los rasgos de la Tétrada Oscura de la personalidad (narcisismo, maquiavelismo, psicopatía y sadismo) se han asociado con diversas formas de conducta antisocial en contextos virtuales. **Método:** Se llevó a cabo una revisión bibliográfica de estudios relevantes sobre las variables de interés en las bases de datos PsycINFO, PubMed, Scopus, ERIC y Web of Science Core Collection. Se incluyeron 24 artículos procedentes de 11 países, con un tamaño muestral combinado de 14,044 participantes. Se realizaron cuatro metaanálisis de efectos aleatorios utilizando los coeficientes r . **Resultados:** Los resultados revelaron una relación positiva entre los cuatro rasgos de personalidad y la propensión al trolling online. Los análisis de meta-regresión indicaron que los instrumentos de evaluación utilizados moderan algunas de las asociaciones observadas. No se detectó evidencia de sesgo de publicación. **Conclusiones:** En general, la relación entre los rasgos de la Tétrada Oscura y el trolling online fue positiva, lo que ayuda a explicar las diferencias individuales en su perpetración.

Palabras clave:

Trolling online
Maquiavelismo
Narcisismo
Psicopatía
Sadismo

The exponential rise in internet access and usage in recent decades has brought about numerous benefits and opportunities. However, this digital environment has also given rise to disruptive and harmful behaviors, such as online trolling (Demsar et al., 2021). This phenomenon, characterized by posting provocative or malicious comments intended to elicit negative emotional reactions in others, has become particularly important in contemporary society. Previous research indicates that more than one-third of millennials have engaged in online trolling, highlighting the need for further investigation (March & Marrington, 2019; Ortiz, 2020). Despite the lack of consensus regarding its precise definition, most researchers agree that trolling involves intentional behaviors aimed at creating conflict in online interactions (Coles & West, 2016; Craker & March, 2016).

Online trolling not only disrupts the dynamics of digital platforms but also has serious psychological consequences for its victims, contributing to mental health issues (Kircaburun et al., 2020). Various typologies of trolling have been identified; these differ according to the perpetrator's motivations, which range from amusement to aggression or the promotion of political ideologies (Komaç & Çagiltay, 2019). However, these motivations are not randomly distributed; some individuals are more likely to engage in trolling behaviors frequently and systematically. This suggests that individual factors, such as specific personality traits, may predispose individuals to such conduct (Buckels et al., 2014). Understanding these differences not only helps explain why certain users are more prone to trolling but also provides valuable insights for designing interventions tailored to specific psychological profiles (March, 2019).

One of the most relevant and emerging approaches in the study of dysfunctional aspects of personality is the investigation of the Dark Tetrad (Paulhus, 2014). This evolved from the original Dark Triad model (Paulhus & Williams, 2002) comprising three personality traits: narcissism, Machiavellianism, and psychopathy. More recently, everyday sadism has been proposed as a fourth component, forming the Dark Tetrad. Machiavellianism is characterized by manipulateness, cynicism, and an instrumental view of interpersonal relationships (Rauthmann & Will, 2011). Narcissism involves an inflated sense of self-importance, a constant need for admiration, and a lack of empathy (Thomaes et al., 2008). Psychopathy is associated with impulsivity, emotional callousness, and a tendency toward antisocial behavior (Hare, 1998). Finally, everyday sadism refers to the tendency to derive pleasure from the suffering of others (Buckels et al., 2013). Recent studies have confirmed the relationship between the Dark Tetrad and disruptive behaviors in digital contexts, such as cyberbullying, cyberstalking, and digital technology addiction (Craker & March, 2016; Johnson et al., 2019). Specifically, trolling exhibits a significant association with sadism and psychopathy (Buckels et al., 2014), due to the low empathy and high disinhibition these individuals exhibit in online environments where the consequences of their actions often appear less tangible or immediate (March et al., 2024). Several studies have found that the relationship between the Dark Tetrad and online trolling may be mediated by contextual factors such as normative beliefs about online aggression, introducing significant variability into research findings (Hilvert-Bruce & Neill, 2020). Understanding this relationship is essential for developing preventive strategies and promoting healthier interactions in cyberspace. Certain strategies implemented by digital platforms have shown effectiveness in

reducing the prevalence of online trolling and mitigating its impact. These include automated moderation using toxic language detection algorithms (Gorwa et al., 2020), proactive human intervention in content management (Jhaver et al., 2019), the redesign of interfaces to discourage impulsive behavior (Matias, 2019), and the application of graduated sanctions, such as feature limitations or account suspension for repeat offenders.

The present meta-analysis aims to address two research questions to provide a comprehensive understanding of the links between online trolling and the Dark Tetrad: (1) What are the overall correlations between online trolling and the personality traits of the Dark Tetrad? and; (2) Do the sample characteristics (gender and age) and methodological features of the included studies (methodological quality, the instrument used to assess online trolling, and the instruments used to assess Dark Tetrad personality traits), moderate the association between the studied variables?

Method

This study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al., 2021) guidelines, and its protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO, CRD420250655916).

Search Strategy and Inclusion Criteria

A systematic search for relevant studies on online trolling and the Dark Tetrad was conducted in February 2025 across the databases PsycINFO, PubMed, Scopus, ERIC, and Web of Science Core Collection. The following search terms were used in the title, abstract, and keyword fields: ("online trolling" OR "internet trolling" OR "internet troll" OR "online troll") AND ("dark triad" OR "dark tetrad" OR psychopathy OR narcissism OR machiavellianism OR sadism). The search was focused on articles published in peer-reviewed journals, with no restrictions regarding publication date. In addition, a manual search was conducted by reviewing the reference lists of the selected articles.

To be included in the present meta-analysis, studies had to meet several inclusion criteria. Included studies had to: 1) be empirical research published in peer-reviewed journals; 2) be written in either Spanish or English; 3) use valid and reliable instruments, defined as those with prior psychometric validation, to assess online trolling and at least one of the Dark Tetrad personality traits; 4) report a correlation coefficient (Pearson or Spearman) between online trolling and at least one of the Dark Tetrad traits; 5) provide access to the full text; and 6) report the sample size.

Methodological Quality of the Included Studies

The methodological quality and risk of individual bias of the included studies were assessed using the abbreviated version of the Newcastle-Ottawa Scale developed by Deng et al. (2020). This version consists of five items: (1) sample representativeness (inclusion of the entire population or random sampling); (2) justification of sample size through methods such as power analysis; (3) response rate above 80%; (4) use of valid measures to assess online trolling and Dark Tetrad traits; and (5) appropriate and clearly

described statistical analyses. Each item receives one point if the criterion is met, and zero points if it is not met or if the information is unavailable. The total score ranges from zero to five points, with studies scoring three or more points considered to have a low risk of individual bias, and those scoring fewer than three points considered to have a high risk of individual bias. Evaluations were conducted independently by two authors, and discrepancies were resolved through group discussion. The inter-rater agreement was 96.8%.

Data Coding

The following variables were recorded: study identification (author[s] and year of publication), country (if the sample was reported to come from multiple countries and the percentage of participants per country was specified, the country with the highest representation was coded; if countries were mentioned without specifying percentages, or if the country of origin was not reported, the country of affiliation of the first author was coded), sample size, mean age of participants, participant gender (as the percentage of women in the sample), instrument used to assess online trolling, instrument used to assess Dark Tetrad personality traits, methodological quality of the study (high or low), and correlation between online trolling and Dark Tetrad traits. Data were independently coded by two of the study's authors, and discrepancies were resolved by consensus. The inter-rater agreement was 94.96%.

To complete missing information on the characteristics of the included studies, the corresponding authors of three studies were contacted via email to request data on participants' mean age (Buckels et al., 2019; Gylfason et al., 2021; Schade et al., 2021). All three authors responded, but only two were able to provide the requested data (Buckels et al., 2019; Schade et al., 2021).

Data Analysis

Four meta-analyses were conducted to evaluate the relationship between online trolling and the Dark Tetrad personality traits using Pearson correlations as effect sizes. First, to normalize sample distributions, Pearson correlations were transformed into Fisher's Z scores (Hedges & Olkin, 2014). After the analyses, the average effect size and its confidence interval were back-transformed to Pearson correlations for ease of interpretation (Borenstein et al., 2021). A random-effects model was used due to the expected heterogeneity among the included studies. Parameter estimation for the random-effects model was performed using the restricted maximum likelihood method (Viechtbauer, 2005). Estimated correlations were interpreted according to the criteria proposed by Gignac and Szodorai (2016), who classified correlations of .10 as small, .20 as moderate, and .30 or higher as large, based on an empirical analysis of psychology studies.

Heterogeneity among studies was assessed using Cochran's Q statistic, I^2 (Higgins & Thompson, 2002), and the prediction interval. A significant Q value suggests variability between studies, while the I^2 statistic reflects the percentage of observed variability not attributable to sampling error. According to Higgins et al. (2003), I^2 values of 25%, 50%, and 75% can be interpreted as indicating low, moderate, and high heterogeneity, respectively. Finally, the

prediction interval represents the range within which effect sizes of a new study are expected to fall, based on the analyzed dataset (Borenstein, 2023).

Publication bias was assessed through multiple methods: visual inspection of funnel plots, Egger's regression test (Egger et al., 1997), and Begg and Mazumdar's rank correlation test (Begg & Mazumdar, 1994). In the absence of publication bias, the funnel plot should appear symmetrical around the average effect size, and both Egger's test and Begg and Mazumdar's test should yield non-significant results.

To evaluate the robustness of the results and examine the potential excessive influence of any single study, a sensitivity analysis was conducted using the leave-one-out technique, which involves removing one study at a time and recalculating the combined effect size.

Potential moderator variables influencing the relationship between online trolling and the Dark Tetrad traits were explored using meta-regression analyses for continuous variables and subgroup analyses for categorical variables, estimating the average effect and heterogeneity separately within each category. For subgroup analyses, following the recommendation of Fu et al. (2011), each subgroup was required to include a minimum of four studies. When this criterion could not be met due to a limited number of studies, the remaining studies were grouped into an "other" category and included in the analyses under this label, provided this group totaled at least four studies.

All analyses were performed using the meta (Schwarzer, 2007) and metafor (Viechtbauer, 2010) packages in R Studio.

Results

Study Selection

Figure 1 presents a flowchart depicting the literature search. The electronic search yielded a total of 113 records. In the first stage, duplicates were identified and removed. Subsequently, initial screening was conducted by reviewing titles and abstracts, excluding references that were not relevant. Finally, a full-text review of the remaining records was performed to determine their eligibility for inclusion. Additionally, a manual search was carried out by reviewing the reference lists of the included articles, resulting in the inclusion of two additional articles that met the inclusion criteria.

Study Characteristics

The analysis included data from 24 studies conducted in 11 countries, with a combined total sample size of 14,044 participants (Table 1). The mean age of participants was 27.71 years, and most studies' samples comprised a higher percentage of women. The countries with the greatest representation were Australia (nine studies), the United States (four studies), and Japan (two studies). To assess online trolling, the Global Assessment of Internet Trolling (nine studies) and its revised version (nine studies) were predominantly used. Regarding the Dark Triad traits, the most commonly employed instruments were the Short Dark Triad (14 studies) and the Dirty Dozen (six studies). For sadism, the most frequently used scales were the Short Sadistic Impulse Scale (10 studies) and the Comprehensive Assessment of Sadistic Tendencies (six studies).

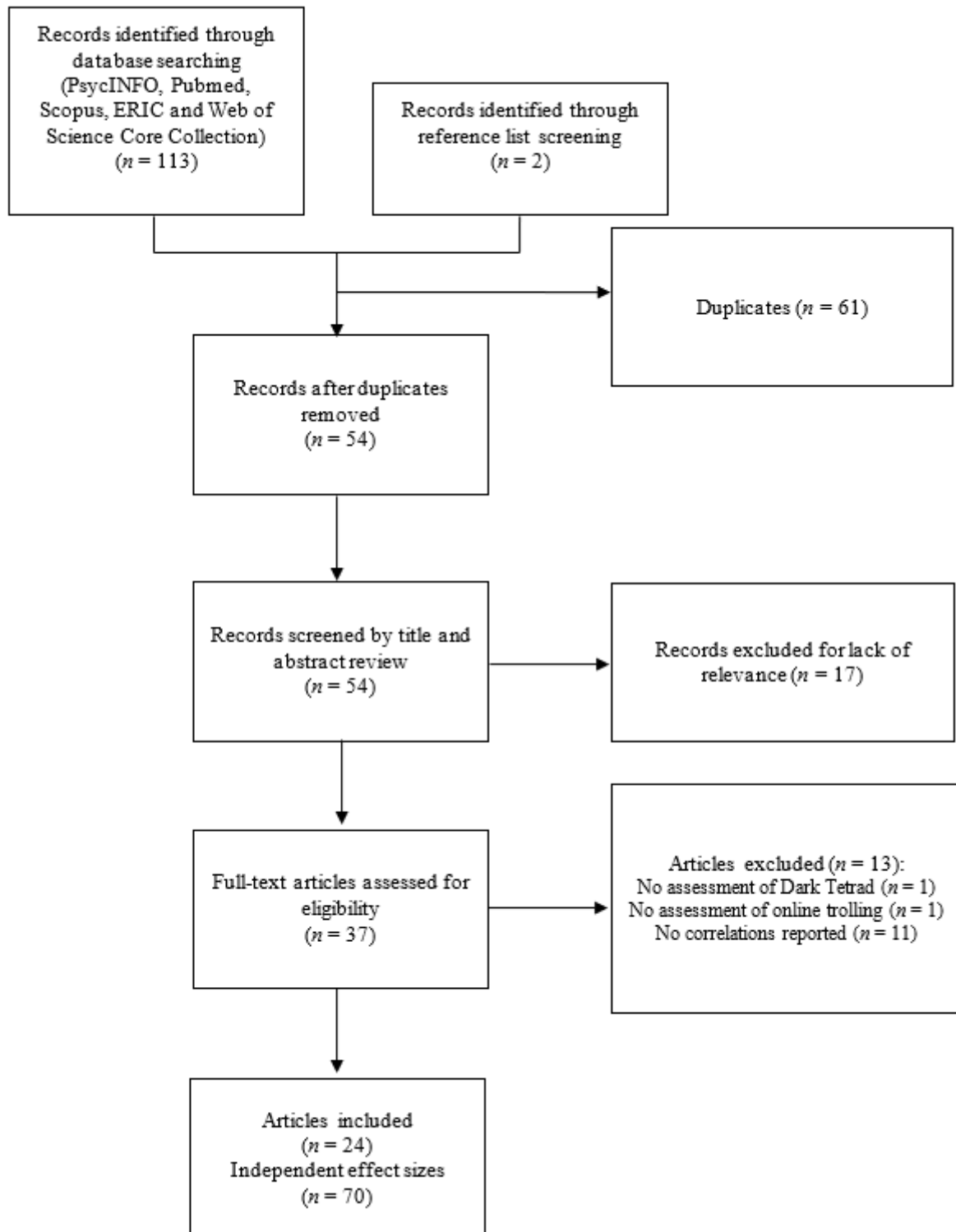
Figure 1
Flow Diagram of the Search and Selection Process


Table 1
Characteristics of the Included Studies

Study	Country	n	Mean age	Gender (% women)	Online trolling measure	Machiavellianism measure	Narcissism measure	Psychopathy measure	Sadism measure
Buckels et al., 2014	USA	797	31,78	45,8	GAIT	SD3	SD3	SD3	CAST
Buckles et al., 2019(1)	USA	345	34,4	51,8	GAIT	SD3	SD3	SD3	CAST
Buckles et al., 2019(2)	Canada	1134	19,71	70,55	iTroll	N.E.	N.E.	N.E.	CAST
Craker & March, 2016	Australia	396	34,41	75,9	GAFT	DD	DD	DD	SSIS
Gylfason et al., 2021	Iceland	139	N.R.	85,6	GAFT	DD	DD	DD	SSIS
Kircaburun et al., 2018	Turkey	761	20,7	63,99	Ad-hoc	DD	DD	DD	SSIS
Lee, 2025	USA	414	38,55	57	GAIT	SD3	SD3	SD3	SSIS
March & Steele, 2020	Australia	400	24,97	67,5	GAIT-R	N.E.	N.E.	SD3	SSIS
March et al., 2017	Australia	357	22,5	71	GAIT (modified)	SD3	SD3	SD3	SSIS
March et al., 2024	Australia	163	27,36	50,3	GAIT-R	N.E.	N.E.	SD3	SSIS
March, 2019	Australia	733	23,53	70,5	GAIT-R	N.E.	HNS	LSRP	VAST
Marrington et al., 2023	Australia	157	15,58	58	GAIT-R	N.E.	N.E.	YPI-SV	SSIS
Masui, 2019	Japan	513	46,8	51,1	GAIT-R	DD	DD	DD	VAST
Masui, 2023	Japan	447	45,7	51,2	GAIT-R	N.E.	N.E.	DD	VAST
Molenda et al., 2022	Poland	1048	22,35	52	ITQ	SD3	SD3	SD3	N.E.
Nitschinsk et al., 2022	Australia	242	21,28	69,01	GAIT (modified)	N.E.	N.E.	SD3	CAST
Nitschinsk et al., 2023	Australia	515	20,47	64,47	iTroll	N.E.	N.E.	SD3	SSIS
Paananen & Reichl, 2019	USA	347	32,67	0	GAIT	N.E.	N.E.	N.E.	CAST
Pineda et al., 2024	Spain	758	31,44	72,8	GAIT	SD3	SD3	SD3	ASP
Resett & González, 2023	Argentina	837	28,4	61	GAIT-R	SD3	SD3	SD3	N.E.
Schade et al., 2021	Austria	743	33,65	54	GAIT	MACH ⁺	NPI-15 and HNS	SRP-III	N.E.
Sest & March, 2017	Australia	415	23,37	63	GAIT-R	N.E.	N.E.	SD3	SSIS
Türk Kurtça & Demirci, 2023	Turkey	234	20	79,1	iTroll	N.E.	N.E.	DD	N.E.
Volkmer et al., 2023	Germany	1026	26,46	77,2	GAIT-R	SD3	SD3	SD3	CAST
Wu et al., 2023	China	1123	19,06	49,1	GAIT	N.E.	N.E.	SD3	N.E.

Notes. n = sample size; N.R. = not reported; N.E. = not evaluated; GAIT = Global Assessment of Internet Trolling; GAIT-R = Global Assessment of Internet Trolling Revised; GAFT = Global Assessment of Facebook® Trolling; ITQ = Internet Trolling Questionnaire; SD3 = Short Dark Triad; DD = Dark Triad Dirty Dozen; MACH* = MACH-IV Machiavellianism Scale short version; HNS = Hypersensitive Narcissism Scale; NPI-15 = Narcissistic Personality Inventory-15; LSRP = Levenson Self-Report Psychopathy Scale; YPI-SV = Youth Psychopathic Index short version; SRP-III = Self-Report Psychopathy-III; CAST = Comprehensive Assessment of Sadistic Tendencies; SSIS = Short Sadistic Impulse Scale; VAST = Varieties of Sadistic Tendencies.

Regarding the methodological quality of the included studies (Table 2), assessed using the short version of the Newcastle-Ottawa Scale, 12 studies demonstrated high quality and a low risk of individual bias, while 13 showed low quality and a high risk of bias. None of the studies met the criterion for item 3 (response rate above 80%), whereas 16% met item 1 (sample representativeness) and 48% met item 2 (sample size justification). All studies fulfilled items 4 and 5 (valid assessment tools and appropriate statistical analyses).

Association Between Online Trolling and Dark Tetrad Personality Traits

The estimated effect sizes and heterogeneity indices for the correlations between online trolling and the Dark Tetrad personality traits are presented in Table 3, while the forest plots are shown in Figures 2, 3, 4, and 5. Among the Dark Tetrad traits, sadism exhibited the strongest association with online trolling ($r = .49, p < .001$), followed by psychopathy ($r = .43, p < .001$), Machiavellianism ($r = .31, p < .001$), and lastly narcissism ($r = .20, p < .001$). According to the criteria proposed by Gignac and Szodorai (2016), the observed effect sizes for sadism, psychopathy, and Machiavellianism may be considered large, whereas the effect size for narcissism is of moderate magnitude.

Regarding effect heterogeneity, Cochran's Q statistic was significant in all cases, indicating underlying heterogeneity among

effect sizes. Additionally, I^2 values ranged between 73.92% and 93.50%, suggesting high heterogeneity. Concerning prediction intervals, which estimate the probable range of effect sizes in a new study, it was observed that the associations of Machiavellianism and sadism with online trolling tend to fall between moderate and high magnitude. In contrast, for narcissism and psychopathy potential effects span from low to high magnitude associations, indicating greater heterogeneity for these traits.

Publication Bias and Sensitivity Analysis

Visual inspection of the funnel plots revealed no indications of publication bias in the cases of psychopathy and sadism. However, the plots for Machiavellianism and narcissism exhibited signs of asymmetry (Figure 6).

However, the results of Egger's regression test and Begg and Mazumdar's rank correlation tests were not significant regarding the association between online trolling and any of the Dark Tetrad traits, thereby ruling out the risk of publication bias. Specifically, the results for Machiavellianism were $t = 0.66, p = .525$ and $\tau = 0.08, p = .765$; for narcissism, $t = -1.69, p = .118$ and $\tau = -0.21, p = .331$; for psychopathy, $t = -0.33, p = .742$ and $\tau = -0.05, p = .754$; and for sadism, $t = 0.43, p = .671$ and $\tau = 0.11, p = .542$.

The leave-one-out analysis showed that none of the included studies exerted an undue influence on the estimated effect sizes, as the obtained

Table 2
Methodological Quality of the Studies

	Item 1	Item 2	Item 3	Item 4	Item 5	Total
Buckels et al., 2014	0	0	0	1	1	2
Buckels et al., 2019(1)	0	0	0	1	1	2
Buckels et al., 2019(2)	0	0	0	1	1	2
Craker & March, 2016	0	0	0	1	1	2
Gylfason et al., 2021	0	1	0	1	1	3
Kircaburun et al., 2018	0	1	0	1	1	3
Lee, 2025	0	0	0	1	1	2
March & Steele, 2020	0	0	0	1	1	2
March et al., 2017	0	0	0	1	1	2
March et al., 2024	0	1	0	1	1	3
March, 2019	0	0	0	1	1	2
Marrington et al., 2023	0	1	0	1	1	3
Masui, 2019	1	0	0	1	1	3
Masui, 2023	0	1	0	1	1	3
Molenda et al., 2022	1	0	0	1	1	3
Nitschinsk et al., 2022	0	1	0	1	1	3
Nitschinsk et al., 2023	0	0	0	1	1	2
Paananen & Reichl, 2019	0	1	0	1	1	3
Pineda et al., 2024	0	1	0	1	1	3
Resett & González, 2023	0	0	0	1	1	2
Schade et al., 2021	0	0	0	1	1	2
Sest & March, 2017	0	0	0	1	1	2
Türk Kurtça & Demirci, 2023	0	0	0	1	1	2
Volkmer et al., 2023	0	1	0	1	1	3
Wu et al., 2023	1	0	0	1	1	3

correlations did not vary significantly. Specifically, the correlation coefficients in the successive meta-analyses excluding one study at a time ranged between .30 and .33 for Machiavellianism; .19 and .21 for narcissism; .43 and .44 for psychopathy; and .48 and .51 for sadism.

Moderation Analysis

Meta-regression analyses were conducted to examine the role of participants' mean age and sex in the relationship between online trolling and the Dark Tetrad personality traits (Table 4). Neither variable emerged as a significant moderator of this association.

Regarding categorical moderators (Table 5), the methodological quality of the included studies did not significantly moderate the association between online trolling and the Dark Tetrad. The measure used to assess online trolling moderated its association with sadism, accounting for 20.1% of the observed heterogeneity. The instrument employed to evaluate Machiavellianism moderated the relationship between this trait and online trolling, with a significant difference in results depending on the measure used; the Short Dark Triad exhibited a stronger association ($r = .38$) compared to the Dirty Dozen ($r = .29$), explaining 16% of the observed heterogeneity. Similarly, the measure used to assess psychopathy moderated the relationship with online trolling, explaining 49.4% of the variance, with the Short Dark Triad exhibiting the strongest correlation ($r = .49$). Finally, the instrument used to assess sadism moderated the association between online trolling and

this personality trait, accounting for 27% of the heterogeneity found. Specifically, the use of the Comprehensive Assessment of Sadistic Tendencies showed a stronger correlation ($r = .60$) compared to the Short Sadistic Impulse Scale ($r = .44$) and other instruments employed ($r = .44$), suggesting that the choice of instrument may influence the magnitude of the observed association.

Table 4
Results of the Moderation Analyses for Continuous Variables

	Coefficient (Standard Error)	CI95%	<i>p</i>	R ²
Machiavellianism				
Mean age	0.0035(0.0033)	[-0.0028, 0.0099]	.277	N.A.
Gender (% women)	0.0011(0.0020)	[-0.0029, 0.0051]	.580	N.A.
Narcissism				
Mean age	-0.0021(0.0034)	[-.0087, 0.0045]	.540	N.A.
Gender (% women)	-0.0015(0.0021)	[-0.0057, 0.0026]	.474	N.A.
Psychopathy				
Mean age	-0.0031(0.0042)	[-0.0114, 0.0051]	.455	N.A.
Gender (% women)	-0.0057(0.0030)	[-0.0115, 0.0001]	.056	N.A.
Sadism				
Mean age	0.0025(0.0050)	[-0.0073, 0.0123]	.620	N.A.
Gender (% women)	-0.0035(0.0022)	[-0.0078, 0.0009]	.117	N.A.

Notes. CI95% = 95% confidence interval; R² = explained variance; N.A. = not applicable.

Table 3*Effect Sizes and Heterogeneity Indices in the Relationship Between Online Trolling and the Dark Tetrad Personality Traits*

Factors	k	N	r	CI95%	PI95%	Q	I ²	τ^2
Machiavellianism	13	8,134	.31***	.27; .35	.18; .43	51.06***	73.92	0.0046
Narcissism	14	8,867	.20***	.16; .25	.05; .34	63.46***	77.33	0.0055
Psychopathy	23	12,563	.43***	.38; .48	.16; .65	257.87***	92.59	0.0233
Sadism	20	10,183	.49***	.43; .55	.20; .71	264.23***	93.50	0.0289

Notes. *k* = number of effect sizes; *N* = combined sample size; *r* = estimated Pearson correlation; CI95% = 95% confidence interval; PI95% = 95% prediction interval; Q = Cochran's Q; *** = *p* < .001; τ^2 = tau squared.

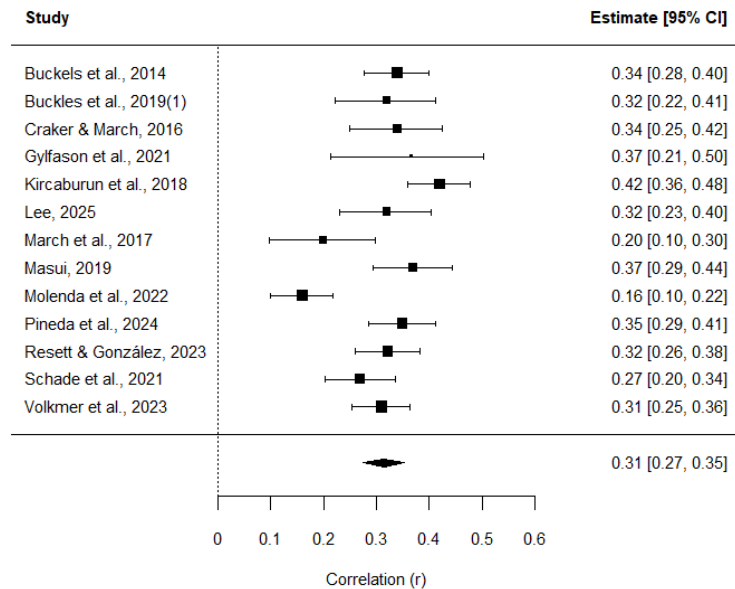
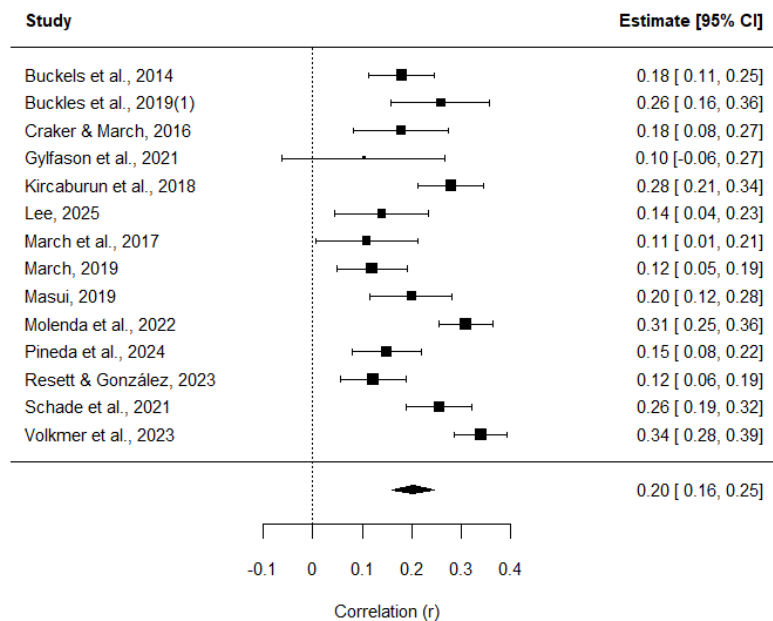
Figure 2*Forest Plot of the Relationships Between Online Trolling and Machiavellianism***Figure 3***Forest Plot of the Relationships Between Online Trolling and Narcissism*

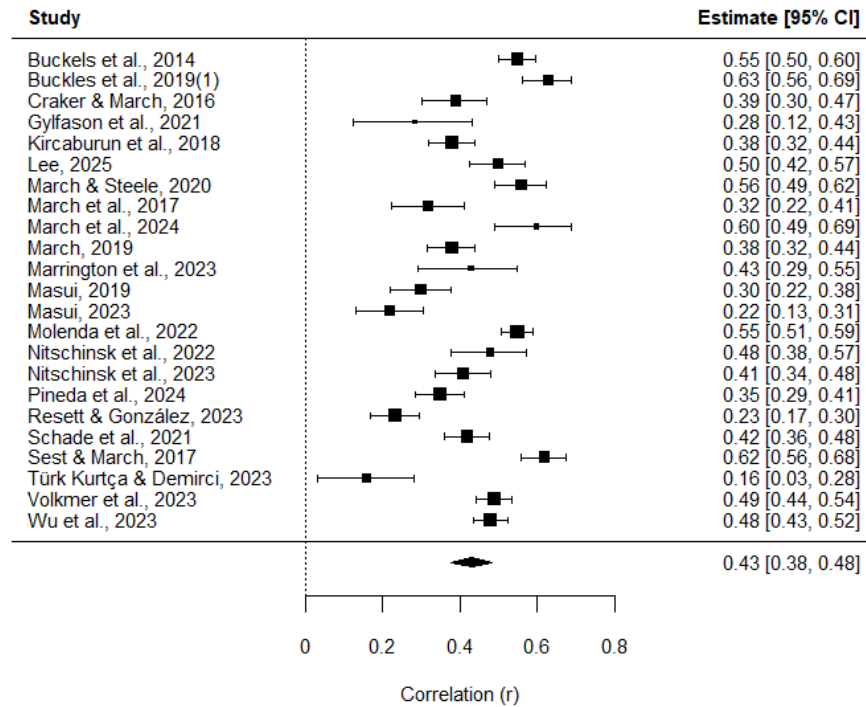
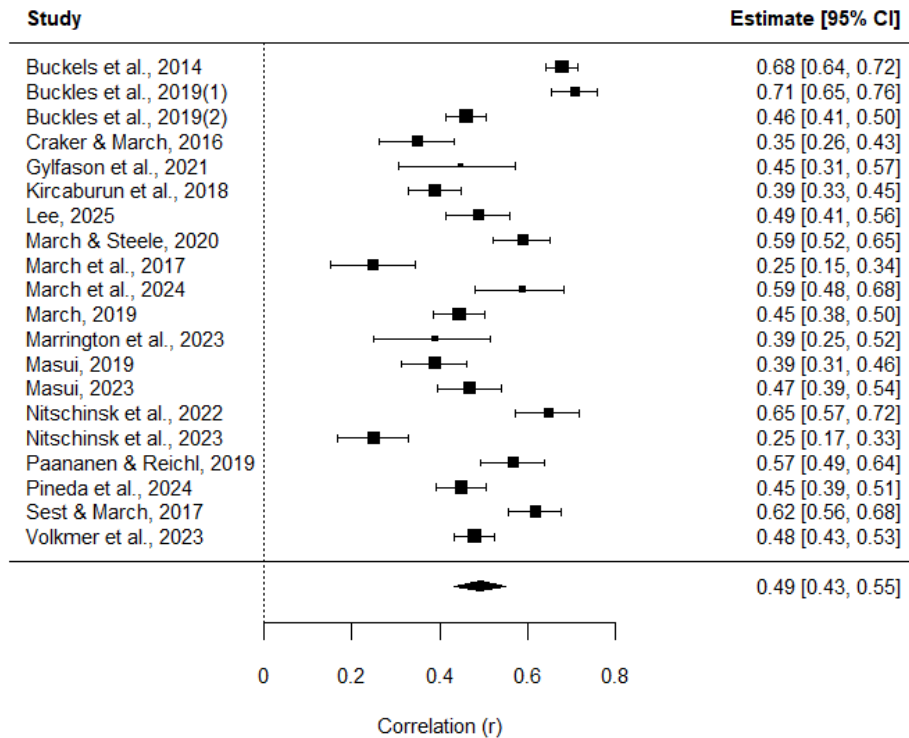
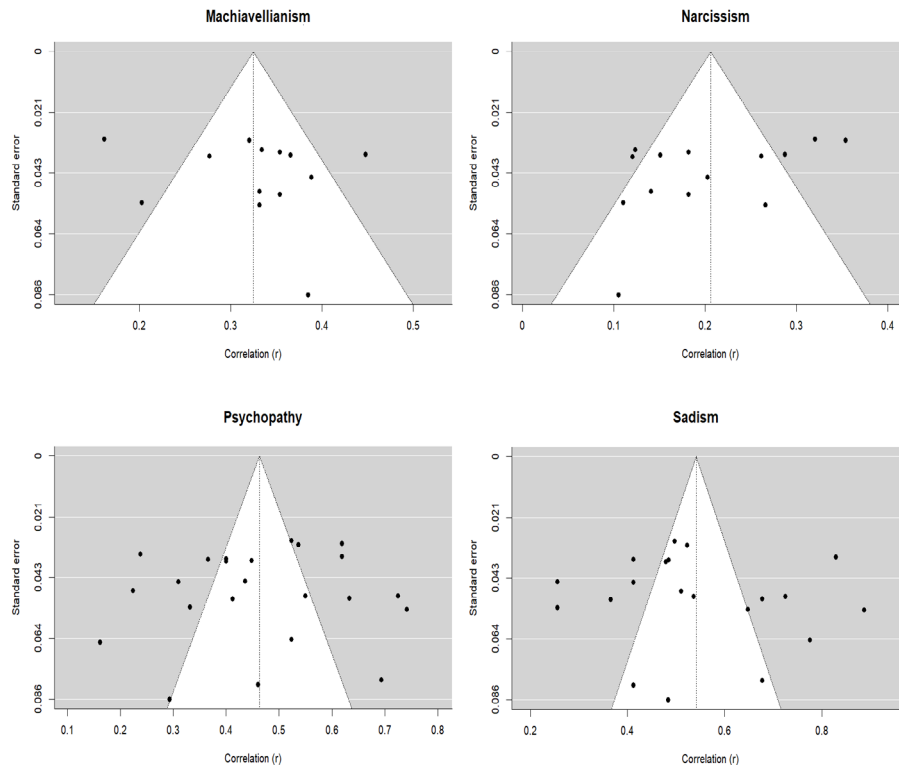
Figure 4*Forest Plot of the Relationships Between Online Trolling and Psychopathy***Figure 5***Forest Plot of the Relationships Between Online Trolling and Sadism*

Figure 6

Funnel Plots of the Meta-Analyses on the Relationship Between Online Trolling and Dark Tetrad Personality Traits



Discussion

The objective of this study was to analyze, through meta-analytic techniques, the relationship between online trolling and the personality traits of the Dark Tetrad. The analyses revealed positive correlations between online trolling and the four Dark Tetrad traits. Specifically, strong associations were found with sadism ($r = .49$), psychopathy ($r = .43$), and Machiavellianism ($r = .31$), as well as a moderate association with narcissism ($r = .20$).

The fact that the strongest association was with sadism reinforces the idea that trolls actively enjoy the suffering of others and seek emotional pleasure through causing humiliation or psychological harm. This result is consistent with previous research indicating that sadism is more strongly related to online trolling than the personality traits of the Big Five model or the other components of the Dark Tetrad (Buckels et al., 2014). This may be because online trolling allows sadistic individuals to exercise their cruelty in a socially unrestricted environment, often protected by the anonymity and feeling of impunity provided by the internet (Nitschinsk et al., 2022). Likewise, psychopathy, characterized by impulsivity, emotional coldness, and lack of empathy, also exhibited a strong relationship with trolling, possibly because these traits facilitate social disinhibition and norm violation in virtual contexts (March, 2019). Moreover, psychopathy has been found to be strongly associated with aggressive behaviors, socioemotional deficits, and interpersonal difficulties (Muris et al., 2017), factors that in turn are related to online trolling (March & Steele, 2020; Marrington et al., 2023). Regarding Machiavellianism, the relationship can be explained

by a tendency toward instrumental manipulation, with individuals high in Machiavellianism using trolling as a strategy to control interactions (Jones & Paulhus, 2014). A recent study found that the relationship between online trolling and Machiavellianism is mediated by the pleasure these individuals experience while trolling, suggesting that the behavior is not limited to an instrumental manipulation strategy but also responds to a hedonistic motivation (Craker & March, 2016). That is, Machiavellian individuals use trolling not only as a means to achieve interpersonal goals, but also for the pleasure inherent in the activity itself. Finally, although narcissism presented the weakest association, it remains significant, especially considering that narcissistic individuals may engage in trolling as a way to protect their self-image or reaffirm their superiority (Casale et al., 2016). Narcissists have a distorted self-view based on exaggerated beliefs about their personal importance; therefore, unlike individuals high in psychopathy or sadism, their motivation for engaging in online trolling does not lie in harming others but in establishing favorable social comparisons (Lopes & Yu, 2017). These findings are consistent with and align with previous studies showing that undesirable behaviors in digital environments are related to dark personality traits (Buckels et al., 2014; Craker & March, 2016; Johnson et al., 2019; Lopes & Yu, 2017; Međedović & Petrović, 2016).

Meta-regression analyses demonstrated that continuous variables such as sex and age are not significant moderators and do not explain the heterogeneity in the relationship between online trolling and the Dark Tetrad. The methodological quality of the included studies also did not moderate this relationship, with no differences found between studies with low or high risk of individual bias. However,

Table 5
Results of the Moderation Analyses for Categorical Variables

Machiavellianism	<i>r</i>	CI95%	<i>p</i>	R²
<i>Methodological quality</i>			.593	N.A.
High	.33	.23, .42		
Low	.31	.26, .35		
<i>Machiavellianism measure</i>			.002	16%
Dirty Dozen	.29	.23, .35		
Short Dark Triad	.38	.32, .44		
<i>Online trolling measure</i>			.618	N.A.
Global Assessment of Internet Trolling	.31	.25, .36		
Others	.32	.25, .40		
Narcissism	<i>r</i>	IC 95%	<i>p</i>	R²
<i>Methodological quality</i>			.084	N.A.
High	.24	.15, .34		
Low	.17	.12, .22		
<i>Narcissism measure</i>			.917	N.A.
Dirty Dozen	.21	.10, .31		
Short Dark Triad	.21	.13, .28		
<i>Online trolling measure</i>			.549	N.A.
Global Assessment of Internet Trolling	.18	.12, .25		
Global Assessment of Internet Trolling Revised	.20	.03, .36		
Others	.24	.10, .37		
Psychopathy	<i>r</i>	IC 95%	<i>p</i>	R²
<i>Methodological quality</i>			.707	N.A.
High	.42	.34, .50		
Low	.44	.34, .53		
<i>Psychopathy measure</i>			<.001	49.4%
Dirty Dozen	.30	.20, .39		
Short Dark Triad	.49	.42, .55		
<i>Online trolling measure</i>			.329	N.A.
Global Assessment of Internet Trolling	.47	.38, .55		
Global Assessment of Internet Trolling Revised	.43	.31, .55		
Others	.38	.23, .51		
Sadism	<i>r</i>	IC 95%	<i>p</i>	R²
<i>Methodological quality</i>			.791	N.A.
High	.48	.41, .55		
Low	.50	.37, .61		
<i>Sadism measure</i>			.007	27%
Comprehensive Assessment of Sadistic Tendencies	.60	.48, .70		
Short Sadistic Impulse Scale	.44	.34, .54		
Others	.44	.39, .49		
<i>Online trolling measure</i>			.012	20.1%
Global Assessment of Internet Trolling	.56	.40, .68		
Global Assessment of Internet Trolling Revised	.50	.42, .57		
Others	.38	.27, .48		

Notes. *r* = estimated Pearson correlation; CI95% = 95% confidence interval; R² = explained variance; N.A. = not applicable.

the instruments used to assess both online trolling and the personality traits of the Dark Tetrad represent a potential source of heterogeneity in some of the associations.

Certain limitations should be considered when interpreting the results of this study. First, this meta-analysis used correlations as the measure of effect size, which prevents the establishment of causal relationships between the variables. Second, only articles published in English and Spanish were considered, which may represent a selection bias. Third, although methodological quality did not moderate the associations, more than half of the studies presented a high risk of individual bias, highlighting the need to strengthen methodological rigor in future research. Fourth, it should be noted that the restricted maximum likelihood estimation method may underestimate both average effect and heterogeneity, as the distribution of parametric effects deviates from normality (Blázquez-Rincón et al., 2023; Suero et al., 2025); this could represent an additional source of bias in the results. Finally, the number of available studies was not very large, especially given the use of subgroup analyses. These analyses usually require a larger number of studies than primary analyses to achieve adequate statistical power (Cuijpers et al., 2021). Therefore, the results obtained should be interpreted with caution, as there is a possibility that significant differences between subgroups were not detected.

Nonetheless, the present meta-analysis adds to the literature demonstrating that Dark Tetrad traits are strongly associated with antisocial behavior in digital environments (Kim, 2023; Van Geel et al., 2017; Xu et al., 2024). This knowledge not only allows for a better theoretical understanding of the underlying mechanisms of online trolling, but also provides a solid empirical basis for the development of early detection and intervention tools. For example, integration of detection algorithms based on linguistic and behavioral patterns characteristic of users with high levels of these traits could be implemented on digital platforms where trolling is particularly prevalent, promoting more effective and proactive moderation (Balakrishnan et al., 2019).

Declaration of Author Contributions

Sergio Hidalgo-Fuentes: Conceptualization, Data curation, Formal Analysis, Methodology, Supervision, Writing – Original draft. **Isabel Martínez-Álvarez:** Conceptualization, Investigation, Validation, Writing – Original draft, Writing – Review & editing. **Elisa González-Pérez:** Conceptualization, Investigation, Visualization, Writing – Original draft, Writing – Review & editing.

Funding

This study did not receive any specific assistance from the public sector, the commercial sector, or non-profit organizations.

Conflict of Interests

The authors declare that there are no conflicts of interest.

Data Availability Statement

The data described in the manuscript and the analytical code will be available upon request to the corresponding author.

References

- References marked with an asterisk (*) were included in the meta-analysis.
- Balakrishnan, V., Khan, S., Fernandez, T., & Arabnia, H. R. (2019). Cyberbullying detection on twitter using Big Five and Dark Triad features. *Personality and Individual Differences*, 141, 252-257. <https://doi.org/10.1016/j.paid.2019.01.024>
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088-1101. <https://pubmed.ncbi.nlm.nih.gov/7786990>
- Blázquez-Rincón, D., Sánchez-Meca, J., Botella, J., & Suero, M. (2023). Heterogeneity estimation in meta-analysis of standardized mean differences when the distribution of random effects departs from normal: A Monte Carlo simulation study. *BMC Medical Research Methodology*, 23(1), 19. <https://doi.org/10.1186/s12874-022-01809-0>
- Borenstein, M. (2023). How to understand and report heterogeneity in a meta-analysis: the difference between I-squared and prediction intervals. *Integrative Medicine Research*, 12(4), 101014. <https://doi.org/10.1016/j.imr.2023.101014>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to Meta-Analysis*. John Wiley & Sons
- Buckels, E. E., Jones, D. N., & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological Science*, 24(11), 2201-2209. <https://doi.org/10.1177/0956797613490749>
- *Buckels, E. E., Trapnell, P. D., Andjelovic, T., & Paulhus, D. L. (2019). Internet trolling and everyday sadism: Parallel effects on pain perception and moral judgment. *Journal of Personality*, 87(2), 328-340. <https://doi.org/10.1111/jopy.12393>
- *Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, 67, 97-102. <https://doi.org/10.1016/j.paid.2014.01.016>
- Casale, S., Fioravanti, G., & Rugai, L. (2016). Grandiose and vulnerable narcissists: who is at higher risk for social networking addiction? *Cyberpsychology, Behavior, and Social Networking*, 19(8), 510-515. <https://doi.org/10.1089/cyber.2016.0189>
- *Craker, N., & March, E. (2016). The dark side of Facebook®: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences*, 102, 79-84. <https://doi.org/10.1016/j.paid.2016.06.043>
- Coles, B. A., & West, M. (2016). Trolling the trolls: Online forum users constructions of the nature and properties of trolling. *Computers in Human Behavior*, 60, 233-244. <https://doi.org/10.1016/j.chb.2016.02.070>
- Cuijpers, P., Griffin, J. W., & Furukawa, T. A. (2021). The lack of statistical power of subgroup analyses in meta-analyses: a cautionary note. *Epidemiology and Psychiatric Sciences*, 30, e78. <https://doi.org/10.1017/S2045796021000664>
- Demsar, V., Brace-Govan, J., Jack, G., & Sands, S. (2021). The social phenomenon of trolling: understanding the discourse and social practices of online provocation. *Journal of Marketing Management*, 37(11-12), 1058-1090. <https://doi.org/10.1080/0267257X.2021.1900335>
- Deng, J., Zhou, F., Hou, W., Silver, Z., Wong, C. Y., Chang, O., Huang, E., & Zuo, Q. K. (2020). The prevalence of depression, anxiety, and sleep disturbances in COVID-19 patients: a meta-analysis. *Annals of the New York Academy of Sciences*, 1486(1), 90-111. <https://doi.org/10.1111/nyas.14506>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629-634. <https://doi.org/10.1136/bmj.315.7109.629>
- Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., Griffith, L., Oremus, M., Raina, P., Ismaila, A., Santaguida, P., Lau, J., & Trikalinos, T. A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of Clinical Epidemiology*, 64(11), 1187-1197. <https://doi.org/10.1016/j.jclinepi.2010.08.010>
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74-78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1-15. <https://doi.org/10.1177/2053951719897945>
- *Gylfason, H. F., Sveinsdóttir, A. H., Vésteinsdóttir, V., & Sigurvinsdóttir, R. (2021). Haters gonna hate, trolls gonna troll: The personality profile of a Facebook troll. *International Journal of Environmental Research and Public Health*, 18(11), 5722. <https://doi.org/10.3390/ijerph18115722>
- Hare, R. D. (1998). Psychopathy, Affect and Behavior. In D. J. Cooke, A. E. Forth, & R. D. Hare (Eds), *Psychopathy: Theory, Research and Implications for Society* (NATO ASI Series, Vol. 88) (pp. 105-137). Springer. https://doi.org/10.1007/978-94-011-3965-6_6
- Hedges, L. V., & Olkin, I. (2014). *Statistical Methods for Meta-Analysis*. Academic Press.
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539-1558. <https://doi.org/10.1002/sim.1186>
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), 557-560. <https://doi.org/10.1136/bmj.327.7414.557>
- Hilvert-Bruce, Z., & Neill, J. T. (2020). I'm just trolling: The role of normative beliefs in aggressive behaviour in online gaming. *Computers in Human Behavior*, 102, 303-311. <https://doi.org/10.1016/j.chb.2019.09.003>
- Jhaver, S., Birman, I., Gilbert, E., & Bruckman, A. (2019). Human-machine collaboration for content regulation: The case of Reddit Automoderator. *ACM Transactions on Computer-Human Interaction*, 26(5), 1-35. <https://doi.org/10.1145/3338243>
- Johnson, A. K., Bevan, J. L., & O'Mara, E. M. (2019). Trolling for sympathy: The Dark Tetrad and online needs-based stress relief. *Computers in Human Behavior*, 96, 44-50. <https://doi.org/10.1016/j.chb.2019.02.008>
- Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, 21(1), 28-41. <https://doi.org/10.1177/1073191113514105>
- Kim, B. (2023). A multilevel meta-analysis of cyberstalking: Domains (and subdomains) of contributing factors. *Journal of Criminal Justice*, 87, 102084. <https://doi.org/10.1016/j.jcrimjus.2023.102084>
- Kircaburun, K., Alhabash, S., Tosuntaş, Ş. B., & Griffiths, M. D. (2020). Uses and gratifications of problematic social media use among university students: A simultaneous examination of the Big Five of personality traits, social media platforms, and social media use motives. *International Journal of Mental Health and Addiction*, 18, 525-547. <https://doi.org/10.1007/s11469-018-9940-6>

- *Kircaburun, K., Jonason, P. K., & Griffiths, M. D. (2018). The Dark Tetrad traits and problematic social media use: The mediating role of cyberbullying and cyberstalking. *Personality and Individual Differences*, 135, 264-269. <https://doi.org/10.1016/j.paid.2018.07.034>
- Komaç, G., & Çağiltay, K. (2019). An overview of trolling behavior in online spaces and gaming context. In *2019 1st International Informatics and Software Engineering Conference (UBMYK)* (pp. 1-4). IEEE. <https://10.1109/UB-MYK48245.2019.8965625>
- *Lee, S. Y. (2025). Do bored people engage in online trolling more actively? Exploring the role of dark personalities and boredom proneness. *Personality and Individual Differences*, 234, 112956. <https://doi.org/10.1016/j.paid.2024.112956>
- Lopes, B., & Yu, H. (2017). Who do you troll and why: An investigation into the relationship between the Dark Triad Personalities and online trolling behaviours towards popular and less popular Facebook profiles. *Computers in Human Behavior*, 77, 69-76. <https://doi.org/10.1016/j.chb.2017.08.036>
- *March, E. (2019). Psychopathy, sadism, empathy, and the motivation to cause harm: New evidence confirms malevolent nature of the internet troll. *Personality and Individual Differences*, 141, 133-137. <https://doi.org/10.1016/j.paid.2019.01.001>
- March, E., & Marrington, J. (2019). A qualitative analysis of internet trolling. *Cyberpsychology, Behavior, and Social Networking*, 22(3), 192-197. <https://doi.org/10.1089/cyber.2018.0210>
- *March, E., & Steele, G. (2020). High esteem and hurting others online: Trait sadism moderates the relationship between self-esteem and internet trolling. *Cyberpsychology, Behavior, and Social Networking*, 23(7), 441-446. <https://doi.org/10.1089/cyber.2019.0652>
- *March, E., Grieve, R., Marrington, J., & Jonason, P. K. (2017). Trolling on Tinder® (and other dating apps): Examining the role of the Dark Tetrad and impulsivity. *Personality and Individual Differences*, 110, 139-143. <https://doi.org/10.1016/j.paid.2017.01.025>
- *March, E., McDonald, L., & Forsyth, L. (2024). Personality and internet trolling: a validation study of a representative sample. *Current Psychology*, 43(6), 4815-4818. <https://doi.org/10.1007/s12144-023-04586-1>
- *Marrington, J.Z., March, E., Murray, S., Jeffries, C., Machin, T., & March, S. (2023). An exploration of trolling behaviours in Australian adolescents: An online survey. *PLoS ONE*, 18(4), e0284378. <https://doi.org/10.1371/journal.pone.0284378>
- Matias, J. N. (2019). Preventing harassment and increasing group participation through social norms in 2,190 online science discussions. *Proceedings of the National Academy of Sciences*, 116(20), 9785-9789. <https://doi.org/10.1073/pnas.1813486116>
- *Masui, K. (2019). Loneliness moderates the relationship between Dark Tetrad personality traits and internet trolling. *Personality and Individual Differences*, 150, 109475. <https://doi.org/10.1016/j.paid.2019.06.018>
- *Masui, K. (2023). Interactional effects of adverse childhood experiences, psychopathy, and everyday sadism on Internet trolling. *Personality and Individual Differences*, 214, 112327. <https://doi.org/10.1016/j.paid.2023.112327>
- Mededović, J., & Petrović, B. (2016). Can there be an immoral morality?: Dark personality traits as predictors of moral foundations. *Psihologija*, 49(2), 185-197. <http://www.doiserbia.nb.rs/img/doi/0048-5705/2016/0048-57051602185M.pdf>
- *Molenda, Z. A., Marchlewska, M., Rogoza, M., Michalski, P., Górka, P., Szczepańska, D., & Cislak, A. (2022). What makes an Internet troll? On the relationships between temperament (BIS/BAS), Dark Triad, and Internet trolling. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 16(5), 5-11. <https://doi.org/10.5817/CP2022-5-11>
- Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The malevolent side of human nature: A meta-analysis and critical review of the literature on the dark triad (narcissism, Machiavellianism, and psychopathy). *Perspectives on Psychological Science*, 12(2), 183-204. <https://doi.org/10.1177/1745691616666070>
- *Nitschinsk, L., Tobin, S. J., & Vanman, E. J. (2022). The disinhibiting effects of anonymity increase online trolling. *Cyberpsychology, Behavior, and Social Networking*, 25(6), 377-383. <https://doi.org/10.1089/cyber.2022.0005>
- *Nitschinsk, L., Tobin, S. J., & Vanman, E. J. (2023). A functionalist approach to online trolling. *Frontiers in Psychology*, 14, 1211023. <https://doi.org/10.3389/fpsyg.2023.1211023>
- Ortiz, S. M. (2020). Trolling as a collective form of harassment: an inductive study of how online users understand trolling. *Social Media + Society*, 6(2), 1-9. <https://doi.org/10.1177/2056305120928512>
- *Paananen, A., & Reichl, A. J. (2019). Gendertrolls just want to have fun, too. *Personality and Individual Differences*, 141, 152-156. <https://doi.org/10.1016/j.paid.2019.01.011>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, 105906. <https://doi.org/10.1016/j.ijsu.2021.105906>
- Paulhus, D. L. (2014). Toward a taxonomy of dark personalities. *Current Directions in Psychological Science*, 23(6), 421-426. <https://doi.org/10.1177/0963721414547737>
- Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556-563. [https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)
- *Pineda, D., Rico-Bordera, P., Ballester-Pascual, M., Piqueras, J. A., & Galán, M. (2024). How Spiteful Are We? Validation of the Spitefulness Scale in Spaniards. *Aggressive Behavior*, 50(6), e70008. <https://doi.org/10.1002/ab.70008>
- Rauthmann, J. F., & Will, T. (2011). Proposing a multidimensional Machiavellianism conceptualization. *Social Behavior and Personality: An International Journal*, 39(3), 391-403. <https://doi.org/10.2224/sbp.2011.39.3.391>
- *Resett, S., & González, P. C. (2023). Psychometric Properties of the Revised Trolling Questionnaire in Argentinean Adults. *Acta Colombiana de Psicología*, 26(2), 101-113. <https://doi.org/10.14718/acp.2023.26.2.9>
- *Schade, E. C., Voracek, M., & Tran, U. S. (2021). The nexus of the dark triad personality traits with cyberbullying, empathy, and emotional intelligence: a structural-equation modeling approach. *Frontiers in Psychology*, 12, 659282. <https://doi.org/10.3389/fpsyg.2021.659282>
- *Sest, N., & March, E. (2017). Constructing the cyber-troll: Psychopathy, sadism, and empathy. *Personality and Individual Differences*, 119, 69-72. <https://doi.org/10.1016/j.paid.2017.06.038>
- Schwarzer, G. (2007). meta: An R package for meta-analysis. *R News*, 7(3), 40-45. <https://journal.r-project.org/articles/RN-2007-029/RN-2007-029.pdf>
- Suero, M., Botella, J., Duran, J. I., & Blazquez-Rincón, D. (2025). Reformulating the meta-analytical random effects model of the standardized mean difference as a mixture model. *Behavior Research Methods*, 57(2), 74. <https://doi.org/10.3758/s13428-024-02554-6>

- Thomaes, S., Brummelman, E., Reijntjes, A., & Bushman, B. J. (2008). When narcissism and self-esteem clash: Self-esteem instability predicts increased aggression in narcissistic children. *Development and Psychopathology*, 20(2), 411-422. <https://doi.org/10.1017/S095457940800020X>
- *Türk Kurtça, T., & Demirci, I. (2023). Psychopathy, impulsivity, and internet trolling: Role of aggressive humour. *Behaviour & Information Technology*, 42(15), 2560-2571. <https://doi.org/10.1080/0144929X.2022.2133635>
- Van Geel, M., Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences*, 106, 231-235. <https://doi.org/10.1016/j.paid.2016.10.063>
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3), 261-293. <https://doi.org/10.3102/10769986030003261>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1-48. <https://doi.org/10.18637/jss.v036.i03>
- *Volkmer, S. A., Gaube, S., Raue, M., & Lerner, E. (2023). Troll story: The dark tetrad and online trolling revisited with a glance at humor. *PLoS One*, 18(3), e0280271. <https://doi.org/10.1371/journal.pone.0280271>
- *Wu, B., Xiao, Y., Zhou, L., Li, F., & Liu, M. (2023). Why individuals with psychopathy and moral disengagement are more likely to engage in online trolling? The online disinhibition effect. *Journal of Psychopathology and Behavioral Assessment*, 45(2), 322-332. <https://doi.org/10.1007/s10862-023-10028-w>
- Xu, W., Zhao, B., & Jin, C. (2024). A meta-analysis of the relationship between personality traits and cyberbullying. *Aggression and Violent Behavior*, 79, 101992. <https://doi.org/10.1016/j.avb.2024.101992>

Article

Selective Review of Empirically Supported Psychological Treatments for Pediatric Health Problems

Isabel Cuéllar-Flores¹, Izaskun Basterra², Silvia Cámara³, Borja Esteso-Orduña⁴, Deborah Fuentes-Casany⁵, Laura González-Riesco⁶, Amaia Izquierdo⁷, María del Pilar Mundo-Cid⁸, Estela Salcedo⁹, Raquel Seijas-Gómez¹⁰ and María Serrano-Villar¹¹

¹ Hospital Clínico San Carlos (Spain)

² Hospital Universitario de Navarra (Spain)

³ Hospital General Universitario Gregorio Marañón (Spain)

⁴ Hospital Infantil Universitario Niño Jesús (Spain)

⁵ Institut d'Assistència Sanitària (Spain)

⁶ Hospital Sant Joan de Déu (Spain)

⁷ Hospital Río Hortega (Spain)

⁸ Hospital Universitari Germans Trias i Pujol (Spain)

⁹ Hospital Universitario Clínico San Cecilio (Spain)

¹⁰ Hospital Teresa Herrera (Spain)

¹¹ Hospital de Getafe (Spain)

ARTICLE INFO

Received: 21/12/2024

Accepted: 02/05/2025

Keywords:

Pediatric clinical psychology

Psychological treatments

Revision

Empirical evidence

Childhood and adolescence

ABSTRACT

Background: Significant progress has been made in providing empirical support for psychological interventions designed to improve the health and behavior of children and young people with pediatric health problems. The objective of this study was to carry out a review of empirically supported psychological treatments for various health problems in the child and adolescent population. **Method:** Psychological treatments with empirical support for 12 pediatric health problems were reviewed based on the levels of evidence and how much they are recommended in the Spanish national health system Clinical Practice Guides. **Results:** Psychological treatments to address various health problems in the child and adolescent population have empirical support. The level of evidence ranges from low to high depending on the health problem. The review shows uneven deployment in different health problems. **Conclusions:** This review will help psychology professionals and healthcare managers to make decisions regarding psychological treatments for addressing pediatric health problems.

Revisión Selectiva de los Tratamientos Psicológicos Empíricamente Apoyados Para Problemas de Salud Pediátricos

RESUMEN

Antecedentes: Se han logrado avances significativos en proporcionar apoyo empírico para intervenciones psicológicas diseñadas para mejorar la salud de niños/as y jóvenes que presentan problemas de salud pediátricos. El objetivo de este trabajo ha sido realizar una revisión de los tratamientos psicológicos empíricamente apoyados para el abordaje de diferentes problemas de salud en población infanto-juvenil. **Método:** Se han revisado los tratamientos psicológicos con apoyo empírico para 12 problemas de salud pediátricos en función de los niveles de evidencia y grados de recomendación de las Guías de Práctica Clínica del Sistema Nacional de Salud de España. **Resultados:** Los tratamientos psicológicos para el abordaje de varios problemas de salud en población infanto-juvenil cuentan con apoyo empírico. El nivel de evidencia oscila de bajo a alto en función del problema de salud. La revisión muestra un desarrollo desigual en los diferentes problemas de salud. **Conclusiones:** Esta revisión ayudará a los profesionales de la psicología y gestores en la toma de decisiones con respecto a los tratamientos psicológicos para el abordaje de los problemas de salud pediátricos.

Palabras clave:

Psicología clínica pediátrica

Tratamientos psicológicos

Revisión

Evidencia empírica

Infancia y adolescencia

Health Psychology is the study of psychological and behavioural processes related to health and illness (Johnston, 1997). It focuses on understanding how psychological, behavioural, and cultural factors contribute to physical health and disease (Buela-Casal & Castro, 2008). Paediatric Psychology, or Child and Adolescent Health Psychology (the use of one term or the other depends on institutional traditions—Child and Adolescent Health Psychology is more commonly used in Spain and Europe, while Paediatric Psychology is more typical in the United States), more specifically:

Pediatric psychology is an integrated field of science and practice in which the principles of psychology are applied within the context of pediatric health. The field aims to promote the health and development of children, adolescents, and their families through use of evidence-based methods [...] Areas of expertise within the field include, but are not limited to: psychosocial, developmental and contextual factors contributing to the etiology, course and outcome of pediatric medical conditions; assessment and treatment of behavioral and emotional concomitants of illness, injury, and developmental disorders; prevention of illness and injury; promotion of health and health-related behaviors; education, training and mentoring of psychologists and providers of medical care; improvement of health care delivery systems and advocacy for public policy that serves the needs of children, adolescents, and their families (Society of Pediatric Psychology, n.d.).

The field of paediatric psychology has seen considerable development over the past 20 years (Palermo, 2014). Evidence of this growth includes the establishment of Division 54 (Society of Pediatric Psychology) of the American Psychological Association, its associated journals, *Journal of Pediatric Psychology* and *Clinical Practice in Pediatric Psychology*, and the publication of key reference manuals (Ortigosa et al., 2003; Roberts & Steele, 2018; Roberts et al., 2014; Rosa et al., 2012). In the field of health, significant progress has been made in providing empirical support for psychological interventions aimed at improving the health and behaviour of children, young people, and families facing paediatric health problems (Ortigosa et al., 2003; Palermo, 2014; Roberts & Steele, 2018).

Within child and adolescent health psychology, particular emphasis is placed on addressing chronic and/or severe illnesses and health problems, given their impact on well-being and quality of life, as well as the challenges they pose for care. Epidemiological studies suggest that up to one in four minors has a chronic illness (Bethell et al., 2011; Cortés et al., 2020). Among the most prevalent conditions in this population are obesity, diabetes, and epilepsy (Miller et al., 2016), while others such as cancer, brain injury, heart disease, and chronic pain are associated with high emotional and functional burden (Roberts & Steele, 2018). In paediatric chronic illnesses, adherence remains a frequent challenge (Graves et al., 2010), and in severe conditions, palliative care is essential to improving family quality of life (Grupo de trabajo de la Guía de Práctica Clínica sobre Cuidados Paliativos en Pediatría, 2022). Psychological and behavioural factors also play a central role in the aetiology and maintenance of feeding and eating disorders, as well as functional and somatic symptom disorders (Agarwal et al., 2019; Van Wye et al., 2023).

Despite its importance and the advances of recent years, there is currently no updated review of psychological treatments for child and adolescent health problems that compiles a broad spectrum of conditions and describes their level of evidence and degree of recommendation. Nor are there recent manuals or clinical practice

guidelines available to assist professionals in the selection and application of best-practice treatments for paediatric health problems. Such resources do exist for psychological disorders in childhood and adolescence (Fonseca-Pedrero et al., 2021), but not for chronic and/or severe paediatric health conditions that are not classified as mental disorders. Having empirical evidence for psychological treatments—even while recognising the limitations and biases inherent in evidence-based models (Pérez-Álvarez & Fonseca-Pedrero, 2021)—is essential to understanding the benefits (and risks) of any intervention with quality assurance.

Accordingly, the aim of this study was to conduct a review of empirically supported psychological treatments for addressing a range of health problems in the child and adolescent population, and to establish recommendations based on their level of evidence.

Method

This project was a collaborative selective review, in which the authors were provided with detailed guidance on the review structure, methodology, and presentation of results, following the recommendations of Palermo (2014) and the Grupo de trabajo para la actualización del Manual de Elaboración de Guías de Práctica Clínica (2016). This approach enabled consistency and methodological rigour across the body of evidence pertaining to each paediatric psychology intervention. A four-phase process was followed to conduct the reviews. In the first and second phases, the scope and objectives were defined by formulating the research question (PICO format: Richardson et al., 1995) and establishing the evidence search strategy (through search engines, relevant scientific databases, and recognised websites hosting clinical practice guidelines). In the third phase, the literature search was conducted and data extracted; in the fourth phase, the evidence was analysed and the quality of the findings appraised (Figure 1).

The system employed to analyse levels of evidence and grades of recommendation was that proposed by the Clinical Practice Guidelines of the Spanish National Health System, based on the criteria of the Scottish Intercollegiate Guidelines Network (SIGN) (Table 1). To enhance the reliability of treatment classification, cross-evaluation was conducted by two authors. In cases of disagreement, a reasoned consensus process was applied until agreement was reached. Supplementary Material 1 (<https://repositoriosaludmadrid.es/rest/api/core/bitstreams/d1bdf518-ecaf-4ccf-b0b6-d8365e161a89/content>) include the instructions provided to authors for conducting the review and evaluating the evidence, a synthesis of the reviewed studies, and the results from each phase of the overall review process.

The selection of paediatric health conditions was based on those identified by expert groups according to their prevalence and psychological impact (Ortigosa et al., 2003; Roberts & Steele, 2018). All included conditions were chronic and/or severe health issues, and transversal themes relevant to health (therapeutic adherence, chronic pain, and palliative care) were also considered.

Results

Twelve systematic reviews were conducted to document the empirical support for psychological interventions. A summary of the findings, levels of evidence, and strength of recommendation

Figure 1

Phases of the Review Conducted (adapted from the Working Group for the Update of the Clinical Practice Guidelines Development Manual, 2016)

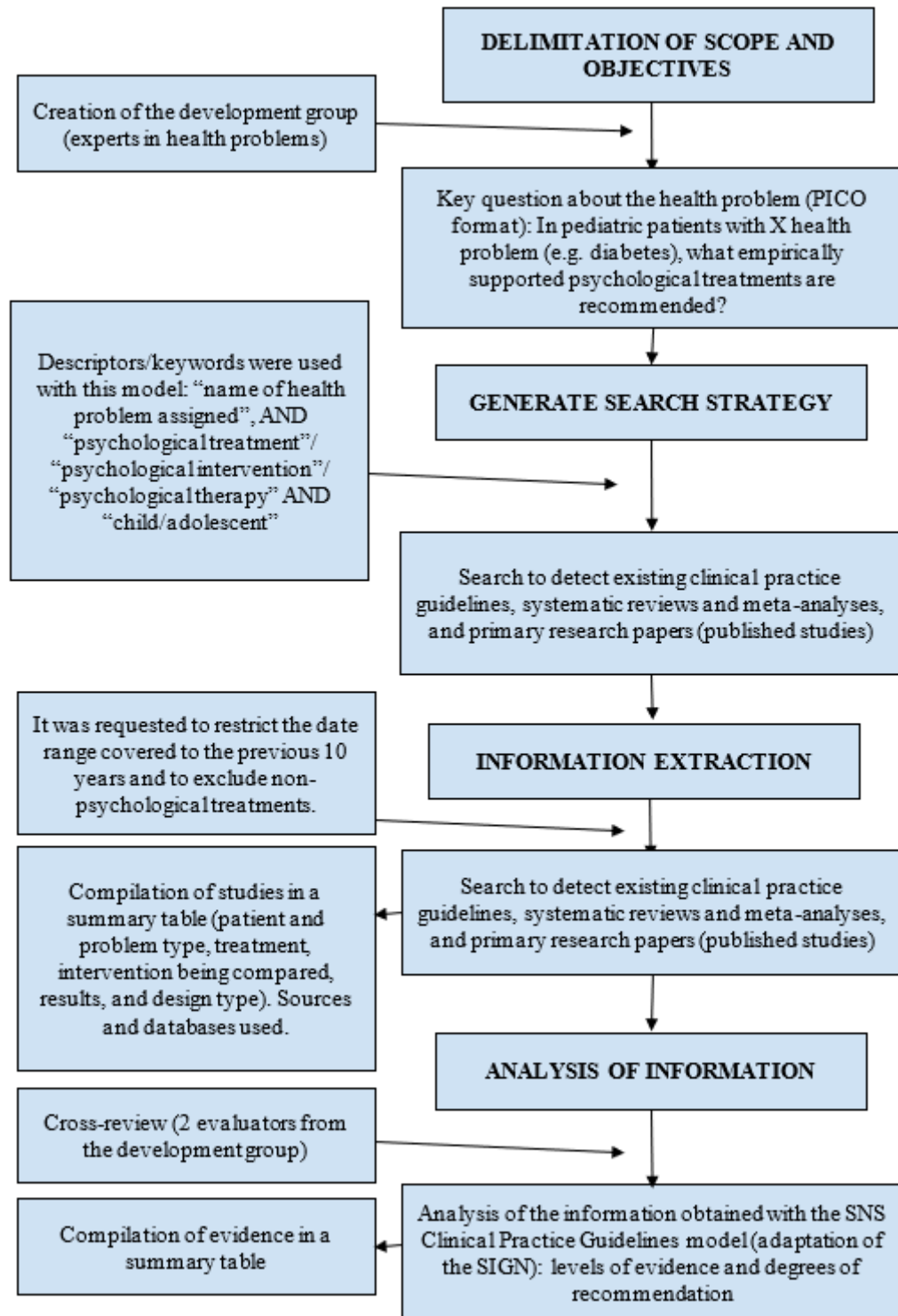


Table 1*Levels of Evidence and Grades of Recommendation from the Clinical Practice Guidelines of the Spanish National Health System*

Levels of evidence	
1++	Meta-analyses, systematic reviews of randomized controlled trials (RCTs), or high-quality RCTs with very low risk of bias.
1+	Meta-analyses, systematic reviews of RCTs, or well-conducted RCTs with low risk of bias.
1-	Meta-analyses, systematic reviews of RCTs, or RCTs with high risk of bias.
2++	Systematic reviews of cohort studies or case-control studies of high quality, cohort studies or case-control studies with very low risk of bias and high probability of establishing a causal relationship.
2+	Well-conducted cohort studies or case-control studies with low risk of bias and moderate probability of establishing a causal relationship.
2-	Cohort studies or case-control studies with high risk of bias.
3	Non-analytical studies, such as case reports and case series.
4	Expert opinion.
Grades of recommendation	
A	At least one meta-analysis, systematic review of RCTs, or RCTs of level 1++, directly applicable to the target population, or sufficient evidence derived from level 1+ studies, directly applicable to the target population, demonstrating overall consistency in the results.
B	Sufficient evidence derived from level 2++ studies, directly applicable to the target population, demonstrating overall consistency in the results. Evidence extrapolated from level 1++ or 1+ studies.
C	Sufficient evidence derived from level 2+ studies, directly applicable to the target population, demonstrating overall consistency in the results. Evidence extrapolated from level 2++ studies.
D	Evidence from level 3 or 4 studies. Evidence extrapolated from level 2+ studies.

for psychological treatments targeting selected pediatric health conditions and cross-cutting health-related issues is presented below. Results reaching an evidence level of 1++, 1+, or 2+ are summarized in [Tables 2 and 3](#). Supplementary material includes detailed information on the reviewed studies and their characteristics (<https://repositoriosaludmadrid.es/rest/api/core/bitstreams/d1bdf518-ecaf-4ccf-b0b6-d8365e161a89/content>).

Pediatric Cancer

Cognitive-behavioral therapy (CBT) has demonstrated efficacy in reducing parental overprotection, anxiety and depressive symptoms in both patients and family members, and in enhancing stress management and coping skills (recommendation A). Combined CBT, distraction techniques, and hypnosis (the latter with lower applicability in younger children) are effective in reducing anxiety and pain during invasive medical procedures (recommendation A). Moreover, neurocognitive interventions have shown improvements in attention, working memory, and academic performance among school-aged children and adolescents (recommendation A).

For post-traumatic stress disorder management in pediatric oncology patients, cognitive-behavioral approaches have been associated with reductions in internalizing symptoms and physiological hyperarousal, although significant heterogeneity in outcomes and study designs leads to a moderate strength of recommendation (B).

Psychoeducation and problem-solving-based interventions, delivered via technological means or targeting families, also have empirical support for improving cancer-related knowledge, treatment adherence, and coping strategies (recommendation B).

Mindfulness-based interventions and motivational interviewing for parents have shown weaker evidence (recommendation C), underscoring the need for further research.

Pediatric Obesity

Family-based behavioral treatment, involving active parental participation in behavior change goals for themselves and their children, has the highest level of empirical support across age

groups (recommendation A). Parent-only behavioral treatment, excluding direct child participation, also shows strong empirical backing for school-aged children and adolescents (recommendation A). Other interventions, such as family-involved behavioral weight-loss treatment without a specific modality and alternative delivery formats (e.g., online, group-based, home-based), are associated with a lower strength of recommendation (C).

Generally, successful treatment combines dietary modification, physical activity, behavioral strategies, and active family involvement to optimize outcomes in Body Mass Index (BMI) reduction.

Pediatric Acquired Brain Injury

The interventions with the strongest empirical support include computerized cognitive training, metacognitive strategy training combined with psychoeducation to enhance cognitive functions, and family-oriented behavior interventions (recommendation A) for school-aged children and adolescents. Other therapies, such as Triple P, goal management training, family problem-solving therapy, and Signposts for disruptive behaviors, are supported with a moderate strength of recommendation (B).

Avoidant/Restrictive Food Intake Disorder (ARFID)

The adapted Family-Based Treatment for ARFID (FBT-ARFID) has the strongest evidence for promoting weight gain and modifying parental behaviors in school-aged children across various clinical profiles (recommendation A). The Parent Training Protocol for ARFID (ARFID-PTP) shows moderate support (recommendation B) for reducing ARFID symptoms and increasing food variety.

For preschoolers, behavioral therapy holds a moderate recommendation (B) for improving BMI, food variety, and ARFID symptoms.

Overall, findings highlight the critical role of family involvement, parent empowerment, and enhancing perceived parental self-efficacy. The heterogeneity of clinical presentations (e.g., selectivity, disinterest, fear of vomiting, swallowing aversion), intervention combinations, and frequent comorbidities limit generalizability.

Table 2*Levels of Evidence and Grades of Recommendation for Psychological Interventions for Pediatric and Adolescent Health Issues*

Reference	Intervention	Evidence level	Grades of Recommendation
Cancer			
Law et al. (2019)Δ Eche et al. (2021)Δ Sánchez-Egea et al. (2019)Δ	Cognitive Behavioural Therapy (CBT) for parents and children (preschoolers, school-age children, adolescents)	1++	A
Brier et al. (2015)Δ	Neurocognitive interventions (school-age children, adolescents)	1++	A
Flowers & Birnie. (2015)Δ	Combined cognitive-behavioral treatments, distraction techniques, and hypnosis for painful invasive medical procedures (preschoolers, school-age children, adolescents)	1++	A
Sánchez-Egea et al. (2019)Δ Eche et al. (2021)Δ	Family psychoeducation (preschoolers, school-age children, adolescents)	1+	B
Kazak et al. (2004)	Cognitive Behavioural Therapy (CBT) for post-traumatic stress disorder (adolescents)	1+	B
Law et al. (2019)Δ	Motivational interviewing in patients undergoing active treatment or survivors (preschoolers, school-age children, adolescents)	2+	C
Tomlinson et al. (2020)Δ	Mindfulness-based interventions (school-age children, adolescents)	2+	C
Acquired Brain Injury			
Corti et al. (2019; 2020; 2023) McKay et al. (2019)	Computerized cognitive training (school-age children, adolescents)	1++	A
Brandt et al. (2021) Hypher et al. (2022)	Training in metacognitive strategies and psychoeducation (school-age children, adolescents)	1++	A
Shen et al. (2023)Δ	Family interventions (school-age children, adolescents)	1++	A
Brown et al. (2014)	Triple P (SSTP) + Acceptance and Commitment Therapy (ACT) (preschoolers, school-age children, adolescents)	1+	B
Sargénus et al. (2024)	Goal Management Training to improve fatigue (school-age children, adolescents)	1+	B
Wade et al. (2019)	Family Problem-Solving Therapy (F-PST): in-person, online guided by a therapist, or online self-guided (adolescents)	1+	B
Chávez Arana et al. (2020)	Parenting program for disruptive behaviors Singpost (school-age children)	1+	B
Obesity			
Davison et al. (2023)Δ	Family-based behavioural therapy (preschoolers, school-age children, adolescents)	1++	A
Davison et al. (2023)Δ	Parent-only behavioural treatment (school-age children, adolescents)	1++	A
Davison et al. (2023)Δ	Family-based behavioural therapy in group format, online, parent-only, low-dose, low-dose with motivational interviewing, educational board game format, peer-supported (school-age children)	2+	C
Davison et al. (2023)Δ	Family-based behavioural therapy at home, parent-only, with motivational interview (adolescents)	2+	C
Davison et al. (2023)Δ	Family-involved behavioural weight-loss treatment (preschoolers, school-age children)	2+	C
Davison et al. (2023)Δ	Cognitive Behavioural Therapy (CBT) (adolescents)	2+	C
Davison et al. (2023)Δ	Motivational interviewing with a family component (adolescents)	2+	C
Avoidant/Restrictive Food Intake Disorder (ARFID)			
Lock et al. (2019) Van Wye et al. (2023)	Family-Based Therapy (FBT-ARFID) (school-age children)	1++	A
Sharp et al. (2016)	Behavioural therapy (preschoolers)	1+	B
Breiner et al. (2024)	Parent training (school-age children)	1+	B
Congenital Heart Diseases			
Hancock et al. (2018)	Early palliative care (psychosocial support) to reduce distress in parents (infants)	1+	B
Li et al. (2018)	Solution-Focused Brief Therapy (SFBT) for parents (school-age children)	1+	B
Kumar et al. (2019)	Brief pre-surgical intervention combining child-focused play and parent-focused family intervention (school-age children, adolescents)	1+	B
McCusker et al. (2012)	Psychoeducational group intervention for parents (school-age children)	2+	C
Tesson et al. (2019)Δ	Parent-oriented psychological therapy (school-age children, adolescents)	2+	C
Kidney Transplant			
Foster et al. (2018)	TAKE-It multicomponent intervention to improve adherence in kidney transplant recipients (adolescents)	2+	C

Reference	Intervention	Evidence level	Grades of Recommendation
Type 1 Diabetes Mellitus			
Resurrección et al. (2021)△ Winkley et al. (2020)△ Rechenberg et al. (2021)△	Cognitive Behavioural Therapy (CBT) (adolescents)	1++	A
Aljawarneh et al. (2020)△	Cognitive Behavioural Therapy (CBT) and stress coping (adolescents)	1+	A
Channon et al. (2007)	Motivational interviewing (adolescents)	2+	C
Salcudean et al. (2024)	Family therapy (adolescents)	2+	C
Ellis et al. (2005)	Intensive Multisystemic Therapy (MST) (adolescents)	2+	v
Epilepsy			
Bennett et al. (2024)△ Michaelis et al. (2021)△	Psychoeducation and Cognitive Behavioural Therapy (CBT) in a modular self-applied programme format (with clinical supervision) for parents and patients - individual format (preschool, school-age, adolescents)	1++	A
Bennett et al. (2024)△ Michaelis et al. (2021)△	Psychoeducation and Cognitive Behavioural Therapy (CBT) for parents and patients - group format (school-age, adolescents)	1++	A
Fleeman et al. (2022)△	Psychoeducation - individual format (school-age, adolescents)	1+	B
Bennett et al. (2024)△ Michaelis et al. (2021)△	Psychoeducation and behavioural treatment - individual format (adolescents)	1+	B
Fleeman et al. (2022)△	Psychoeducation for parents (preschool, school-age, adolescents)	2++	B
Williford et al. (2023)	Problem-solving training (Family Systems Therapy model) applied to parents - group format (preschool, school-age, adolescents)	2+	C
Martinović et al. (2006)	Cognitive Behavioural Therapy (CBT) for the treatment of depressive symptoms (adolescents)	2+	C
Modi et al. (2019) Gutierrez-Colina et al. (2022)	Neuropsychological rehabilitation - executive functioning - web format (adolescents)	2+	C
Functional Disorders/Somatic Symptoms			
Bonvanie et al. (2017)△	Global Psychological Treatments (Functional Abdominal Pain, school-age children, adolescents)	1+	B
Lalouni et al. (2019) Bonnert et al. (2019)	Cognitive-Behavioural Therapy (CBT) in-person and online (Functional Abdominal Pain, school-age children, adolescents)	1+	B
	Gut-directed hypnotherapy/Hypnotherapy (Functional Abdominal Pain, school-age children, adolescents)	1+	B
O'Connell et al. (2020)△ Agarwal et al. (2019)△	Global Psychological Treatments (Functional/Somatic Disorders, school-age children, adolescents)	1+	B
Rutten et al. (2015)△ Gulewitsch et al. (2013, 2017)	Hypnotherapy directed at the intestine (Irritable Bowel Syndrome, school-age children, adolescents)	1+	B
Shah et al. (2020)△	Mind-Body Treatments (MBT) (Irritable Bowel Syndrome, school-age children, adolescents)	1+	B
Velani & Gledhill (2021)△	Multicomponent psychological treatments (Functional Seizures, school-age children, adolescents)	2+	C
Vassilopoulos et al. (2022) (narrative review)	Stepwise Multidisciplinary Treatment based on a Biopsychosocial Approach (Functional Neurological Disorders, school-age children, adolescents)	2+	C
Warner et al. (2011)	Cognitive-Behavioural Therapy (CBT) (Functional/Somatic Disorders, school-age children, adolescents)	2+	C
Kallesoe et al. (2021)	Acceptance and Commitment Therapy (ACT) (Functional/Somatic Disorders, adolescents)	2+	C

Note. △ Systematic review

Congenital Heart Disease

Group-based psychoeducational interventions targeting parental anxiety and depressive symptoms have moderate empirical support across all age groups (recommendation B). Brief solution-focused therapy and pre-surgical preparation programs are also moderately recommended for reducing family distress (recommendation B).

Psychological interventions integrated into multidisciplinary cardiac rehabilitation programs show a lower level of evidence (recommendation C), with benefits identified in only some studies.

In adolescents, individual CBT and mindfulness-based stress reduction programs demonstrate limited evidence for alleviating anxiety and depressive symptoms (recommendation C).

Pediatric Renal Transplantation

Renal transplantation is the most frequent solid organ transplant in pediatric populations (Hart et al., 2016). However, well-designed studies on psychological interventions remain scarce. Multicomponent interventions targeting treatment adherence show the highest level of evidence (2+).

Pediatric Type 1 Diabetes Mellitus

CBT in adolescents shows the best outcomes for psychological measures and quality of life (recommendation A), though no clinically significant changes in glycosylated hemoglobin levels are observed.

Table 3*Levels of Evidence and Grades of Recommendation for Psychological Interventions for Cross-Cutting Issues in Pediatric and Adolescent Health*

Reference	Intervention	Evidence level	Grades of Recommendation
Chronic pain			
Fisher et al. (2022)△	Cognitive-Behavioural Therapy (CBT) (adolescents)	1++	A
Fisher et al. (2022)	Remote Cognitive-Behavioural Therapy (CBT) (adolescents)	1++	A
Fisher et al. (2022)△	Cognitive-Behavioural Therapy (CBT) with a family component (school-age children)	2+	C
Fisher et al. (2022)△	Problem-Solving Therapy (adolescents)	2+	C
Adherence in Chronic Health Conditions			
Graves et al. (2010)△	Cognitive-Behavioural Therapy (CBT) (school-age children, adolescents)	1+ +	A
Pai & McGrady (2014)△			
Kahana et al. (2008)△	Behavioural Interventions (school-age children, adolescents)	1+ +	A
Pai & McGrady (2014)△			
Badawy et al. (2017)△	Behavioural Interventions via Apps and Mobile Messaging (school-age children, adolescents)	1+	B
Graves et al. (2010) △	Educational-Behavioural Treatment (school-age children, adolescents)	2+	C
Kahana et al. (2008)△			
Graves et al. (2010) △	Multicomponent Interventions (including Family Therapy) (school-age children, adolescents)	2+	C
Kahana et al. (2008)△			
Dean et al. (2010)△	Educational-Behavioural Intervention for Adherence to Pharmacological Treatment	1++	A
Pediatric Palliative Care			
Dharmawardene et al. (2016)△	Mindfulness-based intervention (family caregivers and healthcare professionals)	1+	B
Needle et al. (2022)	Family quality of life intervention (FACE) + Pediatric Advance Care Planning (adolescents)	2+	C
Thompkins et al. (2021)			
Akard et al. (2021)	Intervention focused on meaning and legacy creation (adolescents)	2+	C
Rosenberg et al. (2018)	PRISM Promoting Resilience in Stress Management (school-aged children, adolescents)	2+	C
Motlagh et al. (2023)△	Art therapy (school-aged children, adolescents)	2+	C

Note. △ Systematic review

Motivational interviewing improves metabolic regulation and psychological adjustment (recommendation C), while family therapy and multisystemic therapy show benefits for adherence and stress reduction (recommendation C).

Interventions combining individual, family, and group formats, and those increasing session frequency, demonstrate greater effectiveness. More precise descriptions of psychological interventions are needed, given the variability observed across cognitive-behavioral therapy protocols, especially for school-aged children.

Pediatric Epilepsy

The intervention with the highest recommendation grade (A) is a modular adapted CBT program (including psychoeducation), modified from the MATCH-ADTC program and targeting coping skills acquisition for both patients and caregivers.

Standalone psychoeducational interventions for children and adolescents also have some empirical support for improving psychopathology, quality of life, and adherence.

Core components across interventions typically include disease-specific psychoeducation and skills training to manage associated psychological symptoms (e.g., anxiety, depression, behavioral issues), with goals of enhancing patient and caregiver competence, treatment adherence, and promoting healthy lifestyles (e.g., sleep hygiene, social support).

It is important to note that several psychoeducational interventions were not delivered by trained psychologists and often included behavioral or cognitive-behavioral elements.

Pediatric Functional Disorders / Somatic Symptoms

Evidence for psychological treatments of pediatric functional neurological disorders is limited. The strongest support is for multidisciplinary, stepped-care interventions based on a biopsychosocial model (recommendation B).

Specifically, for functional seizures, multicomponent psychological treatments show the highest level of evidence, followed by acceptance and commitment therapy (ACT) and biofeedback (recommendation C).

For functional pain, CBT has a moderate grade of recommendation (B) for symptom reduction.

In functional gastrointestinal disorders, CBT, hypnotherapy, and mind-body interventions show moderate empirical support (recommendation B) for symptom reduction and functional improvement.

Overall, evidence supports psychological interventions for functional symptoms in childhood and adolescence (recommendation B), with CBT and ACT also receiving moderate and weak recommendations, respectively (recommendation C).

Adherence in Pediatric Chronic Health Conditions

Cognitive-behavioral and behavioral interventions consistently demonstrate strong evidence (recommendation A) for improving treatment adherence among children and adolescents with chronic health conditions.

Some studies also support combining educational and behavioral components targeting both patients and families (recommendation

C). Family-centered approaches are particularly emphasized due to the well-established relationship between family dynamics and adherence (Psihogios et al., 2019), as well as the importance of developmental considerations (Pai & McGrady, 2014).

Behavioral interventions delivered through digital media (e.g., apps, mobile messaging) have shown modest empirical support (recommendation B).

Pediatric Chronic Pain

Face-to-face and remote-delivered individual CBT is effective in reducing pain and improving physical functioning in adolescents (recommendation A), although with small effect sizes. In school-aged children, family-involved CBT is recommended (recommendation B). Psychological treatments have a stronger evidence base than pharmacological or physical treatments in this population, despite being less frequently used (Eccleston et al., 2021).

Pediatric Palliative Care

There is strong consensus recommending the integration of psychosocial care within comprehensive, individualized, multidisciplinary, and continuous pediatric palliative care models (Grupo de trabajo de la Guía de Práctica Clínica sobre Cuidados Paliativos en Pediatría, 2022). However, empirical studies remain limited, predominantly focusing on pediatric cancer, with few interventions in neonatal care or complex chronic conditions.

Psychological interventions with the strongest empirical support include mindfulness-based programs for family caregivers and healthcare professionals (recommendation B) for emotional symptom relief and quality of life improvement.

Additional evidence supports interventions focused on family quality of life, art therapy, advance care planning, meaning-making interventions for adolescents, and legacy creation.

The PRISM program (Promoting Resilience in Stress Management) has a weaker evidence base (recommendation C) but integrates techniques from various psychotherapeutic approaches, such as stress management, goal setting, cognitive reframing, and meaning-making.

Discussion

The aim of this work was to review the main empirically supported psychological treatments for addressing various health problems and conditions in children and adolescents, and to provide recommendations based on their level of evidence. Over the past decade, numerous studies have been conducted on psychological interventions in paediatric psychology. Currently, for all the paediatric problems reviewed, there are psychological treatments available with some level of empirical support. However, the growth of high-quality evidence has been uneven. For instance, in paediatric obesity, studies with a high level of evidence and, consequently, strong recommendations are available (Davison et al., 2023), whereas in areas such as palliative care, renal transplantation, or functional disorders, most studies demonstrate a lower level of evidence (Foster et al., 2018; O'Connell et al., 2020). In 9 out of the 12 health conditions reviewed, systematic reviews or meta-analyses published within the past five years were identified, while

for ARFID and renal transplantation, no systematic reviews were found in the past decade.

It can be stated that CBT, adapted to the characteristics of each health condition and developmental stage, remains the therapeutic approach with the strongest empirical support across a range of paediatric pathologies (cancer, diabetes, epilepsy, obesity, somatic symptom disorder, chronic pain). In general, behavioural interventions—defined as sets of techniques aimed at modifying health-related behaviours and integrated into the management of various paediatric conditions (Bennett et al., 2015; Eccleston et al., 2015)—have robust empirical support. In their appropriate implementation, clinical psychologists, as experts in behavioural change, play an essential role.

Nevertheless, other therapeutic approaches are also accumulating evidence regarding their efficacy. Neurocognitive interventions in brain injury and paediatric cancers (Brier et al., 2015) and family-based therapies for obesity, brain injury, ARFID, or paediatric cardiopathies (Davison et al., 2023; Graves et al., 2010; Li et al., 2018; Shen et al., 2023; Van Wye et al., 2023) are gaining support. The latter are conceptualised as treatments that address psychological problems and their management through family interactions, focusing on relational patterns and communication among family members (Society of Clinical Child and Adolescent Psychology – [SCCAP], 2024). They include behavioural family-based interventions, as well as cognitive-behavioural or systemic (e.g., solution-focused) approaches. Integrating parents or caregivers into therapy is a crucial aspect in this population, sometimes positioning them as primary agents of change, at other times as direct recipients of intervention, or as facilitators of improvement in their children.

We agree on the necessity of promoting and disseminating treatments that demonstrate efficacy in reducing symptoms and psychological distress and/or enhancing functioning and quality of life in children, adolescents, and their families facing health problems. This is particularly important given the high prevalence of chronic and/or severe paediatric medical conditions and their profound impact on quality of life and family wellbeing. However, translating research findings into clinical and health psychology practice remains challenging. Psychological treatment is meaningful because it extends beyond theories, methods, techniques, diagnoses, or specific health conditions; it is carried out within a therapeutic relationship and embedded in a broader, necessarily interdisciplinary intervention process within a specific healthcare context, and constrained by available professional resources.

Consequently, although intervention processes in health-related problems need not follow a rigid sequence, they typically include some or all of the following elements (American Psychological Association, 2021): a psychological assessment; a therapeutic plan adapting services to the characteristics of the patient and their family (considering age, culture, and preferences) and organisational resource constraints; ongoing evaluation of patient progress and outcomes; and modification of the clinical approach when necessary. Moreover, collaborative practice is essential—where multiple health professionals deliver comprehensive services working jointly with patients, families, and communities (Roberts et al., 2014). Paediatric clinical and health psychology has a longstanding tradition as an interdisciplinary field, valuing collaboration with a broad range of professionals and requiring the development of interprofessional

competencies within integrated healthcare teams. Nevertheless, for this to occur effectively, clinical and health psychologists working with children and adolescents must implement efficacious and efficient treatments with empirical support and hold a clearly defined role within healthcare organisations (Roberts & Steele, 2018).

In conclusion, a broad range of psychological treatments with empirical support are available for addressing psychosocial factors in the health problems of children and adolescents, particularly in chronic conditions. This selective review provides professionals and healthcare service managers with updated and rigorous information to make informed decisions regarding the implementation of accessible, inclusive, and evidence-based psychotherapeutic procedures for children and adolescents with chronic and/or severe health conditions and their families. The treatments reviewed not only contribute to symptom reduction and emotional distress relief (for both minors and their families) but also help improve quality of life, therapeutic adherence, and physical and social functioning.

Nonetheless, this review has limitations. First, the heterogeneity of the units of analysis across studies (e.g., specific techniques versus multicomponent treatment packages) complicates determining which elements are critical for effectiveness under specific conditions. Second, the term “psychological interventions” encompasses heterogeneous approaches delivered by professionals with diverse qualifications (Law et al., 2019). Third, this is a selective review; although based on a rigorous methodology and homogeneous guidelines, it is not a systematic review (although it builds upon previous and updated systematic reviews) of available psychological treatments for each health condition.

Paediatric and health psychology has made remarkable progress in developing empirically supported psychological treatments. Guided by the solid framework of evidence-based psychological practice (American Psychological Association, 2021), the field must continue to meet challenges and grow as a discipline integrating science and practice to deliver meaningful, impactful care to children, adolescents, and their families facing health challenges.

Author Contributions

Isabel Cuéllar-Flores: Conceptualization, Methodology, Investigation, Formal Analysis, Writing - Original draft, Writing - Review and Editing, Funding acquisition. **Izaskun Basterra:** Writing - Original draft, Investigation. **Silvia Cámara:** Writing - Original draft, Investigation. **María Serrano-Villar:** Writing - Original draft, Investigation, Writing - Review and Editing, Validation. **Borja Estes:** Writing - Original draft, Investigation, Writing - Review and Editing, Validation. **Deborah Fuentes:** Writing - Original draft, Investigation. **Laura Gonzalez-Riesco:** Writing - Original draft, Investigation. **Amaia Izquierdo:** Writing - Original draft, Investigation. **María del Pilar Mundo-Cid:** Writing - Original draft, Investigation. **Estela Salcedo:** Writing - Original draft, Investigation. **Raquel Seijas-Gómez:** Writing - Original draft, Investigation.

Funding

This work was supported by a research project funded by the Programa IDISSC: Investigación Responsable 2024. This funding source had no role in the design of this study, data collection, management, analysis, and interpretation of data, writing of the manuscript, and the decision to submit the manuscript for publication.

Declaration of Interests

The authors declare that they have no conflicts of interest.

Data Availability Statement

The data from the literature review, including the collected studies and their characteristics, are available in the institutional repository of the Health Department of the Community of Madrid. Access it here: <https://repositoriosaludmadrid.es/rest/api/core/bitstreams/d1bdf518-ecaf-4ccf-b0b6-d8365e161a89/content>

References

- Agarwal, V., Srivastava, C., & Sitholey, P. (2019). Clinical practice guidelines for the management of somatoform disorders in children and adolescents. *Indian Journal of Psychiatry*, 61(Supplement 2), 241-246. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_494_18
- Akard, T. F., Dietrich, M. S., Friedman, D. L., Wray, S., Gerhardt, C. A., Hendricks-Ferguson, V., Hinds, P. S., Rhoten, B., & Gilmer, M. J. (2021). Randomized clinical trial of a legacy intervention for quality of life in children with advanced cancer. *Journal of Palliative Medicine*, 24(5), 680-688. <https://doi.org/10.1089/jpm.2020.0139>
- Aljawameh, Y. M., Al-Qaissi, N. M., & Ghunaim, H. Y. (2020). Psychological interventions for adherence, metabolic control, and coping with stress in adolescents with type 1 diabetes: a systematic review. *World Journal of Pediatrics*, 16(5), 456-470. <https://doi.org/10.1007/s12519-020-00352-6>
- American Psychological Association (2021). *Professional practice guidelines for evidence-based psychological practice in health care*. American Psychological Association. <https://www.apa.org/about/policy/evidence-based-psychological-practice-health-care.pdf>
- Badawy, S. M., Barrera, L., Sinno, M. G., Kaviani, S., O'dwyer, L. C., & Kuhns, L. M. (2017). Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. *JMIR mHealth and uHealth*, 5(5), e7798. <https://doi.org/10.2196/mhealth.7798>
- Bennett, S. D., Cross, J. H., Chowdhury, K., Ford, T., Heyman, I., Coughtrey, A. E., Dalrymple, E., Byford, S., Chorpita, B., Fonagy, P., Moss-Morris, R., Reilly, C., Smith, J. A., Stephenson, T., Varadkar, S., Blackstone, J., Quartly, H., Hughes, T., Lewins, A., ... & Shafraan, R. (2024). Clinical effectiveness of the psychological therapy Mental Health Intervention for Children with Epilepsy in addition to usual care compared with assessment-enhanced usual care alone: A multicentre, randomised controlled clinical trial in the UK. *The Lancet*, 403(10433), 1254-1266. [https://doi.org/10.1016/S0140-6736\(23\)02791-5](https://doi.org/10.1016/S0140-6736(23)02791-5)
- Bennett, S., Shafraan, R., Coughtrey, A., Walker, S., & Heyman, I. (2015). Psychological interventions for mental health disorders in children with chronic physical illness: A systematic review. *Archives of disease in childhood*, 100(4), 308-316. <https://doi.org/10.1136/archdischild-2014-307474>
- Bethell, C. D., Kogan, M. D., Strickland, B. B., Schor, E. L., Robertson, J., & Newacheck, P. W. (2011). A national and state profile of leading health problems and health care quality for US children: Key insurance disparities and across-state variations. *Academic pediatrics*, 11(3), 22-S33. <https://doi.org/10.1016/j.acap.2010.08.011>
- Bonnert, M., Olen, O., Lalouni, M., Hedman, E., Sarnholm, J., Serlachius, E., & Ljotsson, B. (2019). Internet-delivered exposure-based cognitive

- behaviour therapy for adolescents with functional abdominal pain or functional dyspepsia: A feasibility study. *Behavior Therapy*, 50(1), 177-188. <https://doi.org/10.1016/j.beth.2018.05.002>
- Bonvanie, I. J., Kallesøe, K. H., Janssens, K. A., Schröder, A., Rosmalen, J. G., & Rask, C. U. (2017). Psychological interventions for children with functional somatic symptoms: a systematic review and meta-analysis. *The Journal of Pediatrics*, 187, 272-281. <https://doi.org/10.1016/j.jpeds.2017.03.017>
- Brandt, A. E., Finnanger, T. G., Hypher, R. E., Rø, T. B., Skovlund, E., Andersson, S., Risnes, K., & Stubberud, J. (2021). Rehabilitation of executive function in chronic paediatric brain injury: a randomized controlled trial. *BMC Medicine*, 19(1), 253. <https://doi.org/10.1186/s12916-021-02129-8>
- Breiner, C. E., Miller, M. L., & Holmes, J. M. (2024). ARFID Parent Training Protocol ("ARFID-PTP"): Results of a randomized pilot trial evaluating a brief parent-training program for avoidant/restrictive food intake disorder. *The International Journal of Eating Disorders*, 57(11), 2306-2317. <https://doi.org/10.1002/eat.24269>
- Brier, M. J., Schwartz, L. A., & Kazak, A. E. (2015). Psychosocial, health-promotion, and neurocognitive interventions for survivors of childhood cancer: A systematic review. *Health Psychology*, 34(2), 130-148. <https://doi.org/10.1037/hea0000119>
- Brown, F. L., Whittingham, K., Boyd, R. N., McKinlay, L., & Sofronoff, K. (2014). Improving child and parenting outcomes following paediatric acquired brain injury: a randomised controlled trial of Stepping Stones Triple P plus Acceptance and Commitment Therapy. *Journal of child psychology and psychiatry*, 55(10), 1172-1183. <https://doi.org/10.1111/jcpp.12227>
- Buela-Casal, G., & Castro, A. (2008). La psicología de la salud en España: pasado, presente y futuro [Health psychology in Spain: Past, present and future]. *Psicología da Saúde*, 16(2), 152-159. <http://doi.org/10.15603/2176-1019/mud.v16n2p152-159>
- Channon, S. J., Huws-Thomas, M. V., Rollnick, S., Hood, K., Cannings-John, R. L., Rogers, C., & Gregory, J. W. (2007). A multicenter randomized controlled trial of motivational interviewing in teenagers with diabetes. *Diabetes Care*, 30(6), 1390-1395. <https://doi.org/10.2337/dc06-2260>
- Chávez Arana, C., Catroppa, C., Yáñez-Téllez, G., Prieto-Corona, B., de León, M. A., García, A., Gómez-Raygoza, R., Hearps, S.J.C., & Anderson, V. (2020). A Parenting program to reduce disruptive behavior in Hispanic children with acquired brain injury: A randomized controlled trial conducted in Mexico. *Developmental Neurorehabilitation*, 23(4), 218-230 <https://doi.org/10.1016/j.conctc.2017.06.007>
- Cortés, J. B., Fernández, C. S., de Oliveira, M. B., Lagos, C. M., Martínez, M. T. B., Hernández, C. L., & del Cura González, I. (2020). Enfermedades crónicas en población pediátrica: comorbilidades y uso de servicios en atención primaria [Chronic diseases in the pediatric population: Comorbidities and use of services in primary care]. *Anales de Pediatría*, 93(3), 183-193. <https://doi.org/10.1016/j.anpedi.2019.12.019>
- Corti, C., Oldrati, V., Oprandi, M. C., Ferrari, E., Poggi, G., Borgatti, R., Urgesi, C., & Bardoni, A. (2019). Remote technology-based training programs for children with acquired brain injury: A systematic review and a meta-analytic exploration. *Behavioural Neurology*, (1), 1346987. <https://doi.org/10.1155/2019/1346987>
- Corti, C., Oldrati, V., Papini, M., Strazzer, S., Poggi, G., Romaniello, R., Borgatti, R., Urgesi, C., & Bardoni, A. (2023). Randomized clinical trial on the effects of a computerized cognitive training for pediatric patients with acquired brain injury or congenital malformation. *Scientific Reports*, 13(1), 14559. <https://doi.org/10.1038/s41598-023-41810-1>
- Corti, C., Urgesi, C., Poggi, G., Strazzer, S., Borgatti, R., & Bardoni, A. (2020). Home-based cognitive training in pediatric patients with acquired brain injury: preliminary results on efficacy of a randomized clinical trial. *Scientific Reports*, 10(1), 139. <https://doi.org/10.1038/s41598-020-57952-5>
- Davison, G. M., Monocello, L. T., Lipsey, K., & Wilfley, D. E. (2023). Evidence-base update on behavioral treatments for overweight and obesity in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 52(5), 589-603. <https://doi.org/10.1080/15374416.2023.2251164>
- Dean, A. J., Walters, J., & Hall, A. (2010). A systematic review of interventions to enhance medication adherence in children and adolescents with chronic illness. *Archives of Disease in Childhood*, 95(9), 717-723. <https://doi.org/10.1136/adc.2009.175125>
- Dharmawardene, M., Givens, J., Wachholtz, A., Makowski, S., & Tjia, J. (2016). A systematic review and meta-analysis of meditative interventions for informal caregivers and health professionals. *BMJ Supportive & Palliative Care*, 6(2), 160-169. <https://doi.org/10.1136/bmjspcare-2014-000819>
- Eche, I. J., Yusuf, M., Isibor, D. A., & Wolfe, J. (2021). A systematic review and meta-analytic evaluation of psychosocial interventions in parents of children with cancer with an exploratory focus on minority outcomes. *Psycho-Oncology*, 30(5), 653-663. <https://doi.org/10.1002/pbc.29328>
- Eccleston, C., Fisher, E., Howard, R. F., Slater, R., Forgeron, P., Palermo, T. M., Birnie, K.A., Anderson, B.J., Chambers, C.T., Crombez, G., Ljungman, G., Jordan, I., Jordan, Z., Roberts, C., Schechter, N., Sieberg, C.B., Tibboel, D., Walker, S.M., ... & Wood, C. (2021). Delivering transformative action in paediatric pain: A Lancet Child & Adolescent Health Commission. *The Lancet Child & Adolescent Health*, 5(1), 47-87. [https://doi.org/10.1016/s2352-4642\(20\)30277-7](https://doi.org/10.1016/s2352-4642(20)30277-7)
- Eccleston, C., Fisher, E., Law, E., Bartlett, J., & Palermo, T. M. (2015). Psychological interventions for parents of children and adolescents with chronic illness. *Cochrane Database of Systematic Reviews*, 4(4), CD009660. <https://doi.org/10.1002/14651858.cd009660.pub3>
- Ellis, D. A., Frey, M. A., Naar-King, S., Templin, T., Cunningham, P. B., & Cakan, N. (2005). The effects of multisystemic therapy on diabetes stress among adolescents with chronically poorly controlled type 1 diabetes: findings from a randomized, controlled trial. *Pediatrics*, 116(6), e826-e832. <https://doi.org/10.1542/peds.2005-0638>
- Fisher, E., Villanueva, G., Henschke, N., Nevitt, S. J., Zempsky, W., Probyn, K., Buckley, B., Cooper, T.E., Sethna, N., & Eccleston, C. (2022). Efficacy and safety of pharmacological, physical, and psychological interventions for the management of chronic pain in children: a WHO systematic review and meta-analysis. *Pain*, 163(1), e1-e19. <https://doi.org/10.1097/j.pain.0000000000002297>
- Fleeman, N., Bradley, P. M., Panebianco, M., & Sharma, A. (2022). Care delivery and self-management strategies for children with epilepsy. *Cochrane database of systematic reviews*, 27(4), CD006245. <https://doi.org/10.1002/14651858.CD006245.pub5>
- Flowers, S. R., & Birnie, K. A. (2015). Procedural preparation and support as a standard of care in pediatric oncology. *Psycho-Oncology*, 24(6), 734-742. <https://doi.org/10.1002/pbc.25813>
- Fonseca-Pedrero, E., Pérez Álvarez, M., Al-Halabi Díaz, S., Inchausti, F., López-Navarro, E. R., Muñoz, J., Lucas-Molina, B., Pérez-Albéniz, A., Baños-Rivera, R., Cano-Vindel, A., Gimeno-Peón, A., Prado-Abril, J., González-Menéndez, A., Valero, A.V., Priede, A., González-Blanch, C., Ruiz-Rodríguez, P., Moriana, J. A., Gómez, L.E., ... & Montoya-Castilla, I. (2021). Tratamientos psicológicos empíricamente apoyados

- para la infancia y adolescencia: estado de la cuestión [Empirically supported psychological treatments for childhood and adolescence: State of the art]. *Psicothema*, 33(3), 386-398. <https://doi.org/10.7334/psicothema2021.56>
- Foster, B. J., Pai, A. L., Zelikovsky, N., Amaral, S., Bell, L., Dharnidharka, V. R., ... & Furth, S. L. (2018). A randomized trial of a multicomponent intervention to promote medication adherence: the teen adherence in kidney transplant effectiveness of intervention trial (TAKE-IT). *American Journal of Kidney Diseases*, 72(1), 30-41. <https://doi.org/10.1053/j.ajkd.2017.12.012>
- Graves, M. M., Roberts, M. C., Rapoff, M., & Boyer, A. (2010). The efficacy of adherence interventions for chronically ill children: A meta-analytic review. *Journal of Pediatric Psychology*, 35(4), 368-382. <https://doi.org/10.1093/jpepsy/jsp072>
- Grupo de trabajo de la Guía de Práctica Clínica sobre Cuidados Paliativos en Pediatría (2022). *Guía de práctica clínica sobre cuidados paliativos en pediatría* [Clinical practice guideline on palliative care in pediatrics]. Ministerio de Sanidad. Instituto Aragonés de Ciencias de la Salud. Guías de Práctica Clínica en el Sistema Nacional de Salud. https://portal.guiasalud.es/wp-content/uploads/2022/09/gpc_618_cuidados_paliativos_pediatría_iacs_compl.pdf
- Grupo de trabajo para la actualización del Manual de Elaboración de Guías de Práctica Clínica (2016). *Elaboración de guías de práctica clínica en el Sistema Nacional de Salud. Actualización del manual metodológico* [Development of clinical practice guidelines in the National Health System. Update of the methodological manual]. Ministerio de Sanidad, Servicios Sociales e Igualdad; Instituto Aragonés de Ciencias de la Salud (IACS). http://portal.guiasalud.es/emanuales/elaboracion_2/?capitulo
- Gulewitsch, M. D., Müller, J., Hautzinger, M., & Schlarb, A. A. (2013). Brief hypnotherapeutic-behavioral intervention for functional abdominal pain and irritable bowel syndrome in childhood: a randomized controlled trial. *European Journal of Pediatrics*, 172(8), 1043-1051. <https://doi.org/10.1007/s00431-013-1990-y>
- Gulewitsch, M. D., & Schlarb, A. A. (2017). Comparison of gut-directed hypnotherapy and unspecific hypnotherapy as self-help format in children and adolescents with functional abdominal pain or irritable bowel syndrome: a randomized pilot study. *European Journal of Gastroenterology & Hepatology*, 29(12), 1351-1360. <https://doi.org/10.1097/meg.0000000000000984>
- Gutierrez-Colina, A. M., Clifford, L., Wade, S. L., & Modi, A. C. (2022). Uncovering key elements of an executive functioning intervention in adolescents: Epilepsy journey. *Clinical Practice in Pediatric Psychology*, 10(2), 150-163. <https://doi.org/10.1037/cpp0000410>
- Hancock, H. S., Pituch, K., Uzark, K., Bhat, P., Fifer, C., Silveira, M., Yu, S., Welch, S., Donohue J., Lowery, R., & Aiyagari, R. (2018). A randomised trial of early palliative care for maternal stress in infants prenatally diagnosed with single-ventricle heart disease. *Cardiology in the Young*, 28(4), 561-570. <https://doi.org/10.1017/S1047951117002761>
- Hart, A., Smith, J. M., Skeans, M. A., Gustafson, S. K., Wilk, A. R., Robinson, A., Wainright J.L., Haynes, C. R., Snyder, J.J., Kasiske, B.L., & Israni A.K. (2016). OPTN/SRTR 2016 annual data report: Kidney. *American Journal of Transplantation*, 18(1), 18-113. <https://doi.org/10.1111/ajt.15672>
- Hypher, R., Brandt, A. E., Skovlund, E., Skarbø, A. B., Barder, H. E., Andersson, S., Rø, T.B., Risnes, K., Finnanger, T.G., & Stubberud, J. (2022). Metacognitive strategy training versus psychoeducation for improving fatigue in children and adolescents with acquired brain injuries: A randomized controlled trial. *Neuropsychology*, 36(7), 579-596. <https://doi.org/10.1037/neu0000845>
- Johnston, M. (1997). How health psychology makes a difference. *The Irish Journal of Psychology*, 18(1), 4-12. <https://doi.org/10.1080/03033910.1997.10558126>
- Kahana, S., Drotar, D., & Frazier, T. (2008). Meta-analysis of psychological interventions to promote adherence to treatment in pediatric chronic health conditions. *Journal of Pediatric Psychology*, 33(6), 590-611. <https://doi.org/10.1093/jpepsy/jsm128>
- Kallesøe, K. H., Schröder, A., Jensen, J. S., Wicksell, R. K., & Rask, C. U. (2021). Group-based acceptance and commitment therapy (AHEAD) for adolescents with multiple functional somatic syndromes: A randomised trial. *JCPP Advances*, 1(4), e12047. <https://doi.org/10.1002/jcv2.12047>
- Kazak, A. E., Alderfer, M. A., Streisand, R., Simms, S., Rourke, M. T., Barakat, L. P., Gallagher, P., Cnaan, A., & Annunziato, R. A. (2004). Treatment of posttraumatic stress symptoms in adolescent survivors of childhood cancer and their families: A randomized clinical trial. *Journal of Family Psychology*, 18(3), 493-504. <https://doi.org/10.1037/0893-3200.18.3.493>
- Kumar, A., Das, S., Chauhan, S., Kiran, U., & Satapathy, S. (2019). Perioperative anxiety and stress in children undergoing congenital cardiac surgery and their parents: Effect of brief intervention—a randomized control trial. *Journal of Cardiothoracic and Vascular Anesthesia*, 33(5), 1244-1250. <https://doi.org/10.1053/j.jvca.2018.08.187>
- Lalouni, M., Ljótsson, B., Bonner, M., Ssegonja, R., Benninga, M., Bjureberg, J., ... & Olén, O. (2019). Clinical and cost effectiveness of online cognitive behavioral therapy in children with functional abdominal pain disorders. *Clinical Gastroenterology and Hepatology*, 17(11), 2236-2244. <https://doi.org/10.1016/j.cgh.2018.11.043>
- Law, E., Fisher, E., Eccleston, C., & Palermo, T. M. (2019). Psychological interventions for parents of children and adolescents with chronic illness. *Cochrane Database of Systematic Reviews*, 18(3), CD009660. <https://doi.org/10.1002/14651858.cd009660.pub4>
- Li, Y., Solomon, P., Zhang, A., Franklin, C., Ji, Q., & Chen, Y. (2018). Efficacy of solution-focused brief therapy for distress among parents of children with congenital heart disease in China. *Health & Social Work*, 43(1), 30-40. <https://doi.org/10.1093/hsw/hlx045>
- Lock, J., Sadeh-Sharvit, S. & L'Insalata, A. (2019). Feasibility of conducting a randomized clinical trial using family-based treatment for avoidant/restrictive food intake disorder. *The International Journal of Eating Disorders*, 52(6), 746-751. <https://doi.org/10.1002/eat.23077>
- Martinović, Ž., Simonović, P., & Djokić, R. (2006). Preventing depression in adolescents with epilepsy. *Epilepsy & Behavior*, 9(4), 619-624. <https://doi.org/10.1016/j.yebeh.2006.08.017>
- McCusker, C. G., Doherty, N. N., Molloy, B., Rooney, N., Mulholland, C., Sands, A., Craig, B., Stewart, M., & Casey, F. (2012). A randomized controlled trial of interventions to promote adjustment in children with congenital heart disease entering school and their families. *Journal of Pediatric Psychology*, 37(10), 1089-1103. <https://doi.org/10.1093/jpepsy/jss092>
- McKay, E., Richmond, S., Kirk, H., Anderson, V., Catroppa, C., & Cornish, K. (2019). Training attention in children with acquired brain injury: a study protocol of a randomised controlled trial of the TALI attention training programme. *BMJ Open*, 9(12), e032619. <https://doi.org/10.1136/bmjopen-2019-032619>
- Michaelis, R., Tang, V., Nevitt, S. J., Wagner, J. L., Modi, A. C., LaFrance Jr, W. C., Goldstein, L.H., Gandy, M., Bresnahan, R., Valente, K.,

- Donald, K.A., & Reuber, M. (2021). Cochrane systematic review and meta-analysis of the impact of psychological treatment on health-related quality of life in people with epilepsy: an update by the ILAE Psychology Task Force, highlighting methodological changes. *Epileptic Disorders*, 23(6), 803-811. <https://doi.org/10.1684/epd.2021.1357>
- Miller, G. F., Coffield, E., Leroy, Z., & Wallin, R. (2016). Prevalence and costs of five chronic conditions in children. *The Journal of School Nursing*, 32(5), 357-364. <https://doi.org/10.1177/1059840516641190>
- Modi, A. C., Mara, C. A., Schmidt, M., Smith, A. W., Turnier, L., Glaser, N., & Wade, S. L. (2019). Epilepsy Journey: A proof of concept trial of a web-based executive functioning intervention for adolescents with epilepsy. *Epilepsia*, 60(9), 1895-1907. <https://doi.org/10.1111/epi.16317>
- Motlagh, E. G., Bakhshi, M., Davoudi, N., Ghasemi, A., & Moonaghi, H. K. (2023). The physical and psychological outcomes of art therapy in pediatric palliative care: A systematic review. *Journal of Research in Medical Sciences*, 28(1), 13. https://doi.org/10.4103/jrms.jrms_268_22
- Needle, J. S., Friebert, S., Thompkins, J. D., Grosseohme, D. H., Baker, J. N., Jiang, J., Wang, J., & Lyon, M. E. (2022). Effect of the family-centered advance care planning for teens with cancer intervention on sustainability of congruence about end-of-life treatment preferences: A randomized clinical trial. *JAMA Network Open*, 5(7), e2220696-e2220696. <https://doi.org/10.1001/jamanetworkopen.2022.20696>
- O'Connell, C., Shafraan, R., & Bennett, S. (2020). A systematic review of randomised controlled trials using psychological interventions for children and adolescents with medically unexplained symptoms: A focus on mental health outcomes. *Clinical Child Psychology and Psychiatry*, 25(1), 273-290. <https://doi.org/10.1177/1359104519855415>
- Ortigosa, J. M., Quiles, M. J. & Méndez, F. X. (Eds.). (2003). *Manual de psicología de la salud con niños, adolescentes y familia* [Manual of health psychology with children, adolescents and families]. Ediciones Pirámide.
- Palermo, T. M. (2014). Evidence-based interventions in pediatric psychology: Progress over the decades. *Journal of Pediatric Psychology*, 39(8), 753-762. <https://doi.org/10.1093/jpepsy/jsu048>
- Pérez-Álvarez, M., & Fonseca-Pedrero, E. (2021). Para repensar los problemas psicológicos de la infancia y la adolescencia: hacia un enfoque contextual [Rethinking psychological problems in childhood and adolescence: Towards a contextual approach]. In E. Fonseca-Pedrero (Ed.), *Manual de tratamientos psicológicos. Infancia y adolescencia*. (pp. 824-856). Ediciones Pirámide.
- Pai, A. L., & McGrady, M. (2014). Systematic review and meta analysis of psychological interventions to promote treatment adherence in children, adolescents, and young adults with chronic illness. *Journal of Pediatric Psychology*, 39(8), 918- 931. <https://doi.org/10.1093/jpepsy/jsu038>
- Psihogios, A. M., Fellmeth, H., Schwartz, L. A., & Barakat, L. P. (2019). Family functioning and medical adherence across children and adolescents with chronic health conditions: A meta analysis. *Journal of Pediatric Psychology*, 44(1), 84- 97. <https://doi.org/10.1093/jpepsy/jsy044>
- Rechenberg, K., & Koerner, R. (2021). Cognitive behavioral therapy in adolescents with type 1 diabetes: an integrative review. *Journal of Pediatric Nursing*, 60, 190-197. <https://doi.org/10.1016/j.pedn.2021.06.019>
- Resurreccion, D. M., Navas-Campana, D., Gutierrez-Colosia, M. R., Ibanez-Alfonso, J. A., & Ruiz-Aranda, D. (2021). Psychotherapeutic interventions to improve psychological adjustment in type 1 diabetes: A systematic review. *International Journal of Environmental Research and Public Health*, 18(20), 10940. <https://doi.org/10.3390/ijerph182010940>
- Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. (1995). The well-built clinical question: a key to evidence-based decisions. *ACP Journal Club*, 123(3), A12-3.
- Roberts, M. C., Aylward, B. S., & Wu, Y. P. (Eds.) (2014). *Clinical practice of pediatric psychology*. Guilford Publications.
- Roberts, M.C., & Steele, R.G. (2018). *Handbook of pediatric psychology*. (5th ed.). The Guilford Press.
- Rosa, A. I., Olivares, P. J. O., & Rodríguez, J. O. (2012). *Psicología de la salud en la infancia y adolescencia: Casos prácticos* [Health psychology in childhood and adolescence: Practical cases]. Ediciones Pirámide.
- Rosenberg, A. R., Bradford, M. C., McCauley, E., Curtis, J. R., Wolfe, J., Baker, K. S., & Yi-Frazier, J. P. (2018). Promoting resilience in adolescents and young adults with cancer: Results from the PRISM randomized controlled trial. *Cancer*, 124(19), 3909-3917. <https://doi.org/10.1002/cncr.31666>
- Rutten, J. M., Korterink, J. J., Venmans, L. M., Benninga, M. A., & Tabbers, M. M. (2015). Nonpharmacologic treatment of functional abdominal pain disorders: A systematic review. *Pediatrics*, 135(3), 522-535. <https://doi.org/10.1542/peds.2014-2123>
- Salcudean, A., & Lica, M. M. (2024). The Role of Systemic Family Psychotherapy in Glycemic Control for Children with Type 1 Diabetes. *Children*, 11(1), 104. <https://doi.org/10.3390/children11010104>
- Sánchez-Egea, R., Rubio-Aparicio, M., Sánchez-Meca, J., & Rosa-Alcázar, A. I. (2019). Psychological treatment for family members of children with cancer: A systematic review and meta-analysis. *Psycho-oncology*, 28(5), 960-969. <https://doi.org/10.1002/pon.5052>
- Sargénus, H. L., Hypher, R. E., Finnanger, T. G., Brandt, A. E., Andersson, S., Risnes, K., Rø, T.B., & Stubberud, J. E. (2024). Goal management training for improving fatigue in children and adolescents with acquired brain injuries: A 2-year follow-up of a randomised controlled trial. *Neuropsychological Rehabilitation*, 35(3), 1-21. <https://doi.org/10.1080/09602011.2024.2353395>
- Shah, K., Ramos-Garcia, M., Bhavsar, J., & Lehrer, P. (2020). Mind-body treatments of irritable bowel syndrome symptoms: An updated meta-analysis. *Behaviour Research and Therapy*, 128, 103462-103469. <https://doi.org/10.1016/j.brat.2019.103462>
- Sharp, W. G., Stubbs, K. H., Adams, H., Wells, B. M., Lesack, R. S., Criado, K. K., Simon, E. L., McCracken, C. E., West, L. L., & Scchill, L. D. (2016). Intensive, manual-based intervention for pediatric feeding disorders: Results from a randomized pilot trial. *Journal of Pediatric Gastroenterology and Nutrition*, 62(4), 658-663. <https://doi.org/10.1097/MPG.0000000000001043>
- Shen, J., Zhao, S., Horn, T., Benkart, R., Busch, T., Vrabec, A., & Taylor, H. G. (2023). Family matters: A systematic review and meta-analysis on the efficacy of family-oriented interventions for children with acquired brain injuries. *Clinical Psychology Review*, 99, 102218. <https://doi.org/10.1016/j.cpr.2022.102218>
- Society of Clinical Child and Adolescent Psychology [SCCAP]. (2024). *What is family therapy?* Effective Child Therapy. <https://effectivechildtherapy.org/therapies/what-is-family-therapy/>
- Society of Pediatric Psychology (n.d.). *Who we are*. Society of Pediatric Psychology. <https://pedpsych.org/about-us/>
- Tesson, S., Butow, P. N., Sholler, G. F., Sharpe, L., Kovacs, A. H., & Kasparian, N. A. (2019). Psychological interventions for people affected by childhood-onset heart disease: A systematic review. *Health Psychology*, 38(2), 151. <https://doi.org/10.1037/hea0000704>

- Thompkins, J. D., Needle, J., Baker, J. N., Briggs, L., Cheng, Y. I., Wang, J., Friebert, S., & Lyon, M. E. (2021). Pediatric advance care planning and families' positive caregiving appraisals: an RCT. *Pediatrics*, 147(6), e2020029330. <https://doi.org/10.1542/peds.2020-029330>
- Tomlinson, D., Sung, L., Vettese, E., Murphy, S., & Plenert, E. (2020). Mindfulness-based interventions for symptom management in children and adolescents with cancer: A systematic review. *Journal of Pediatric Oncology Nursing*, 37(6), 423-430. <https://doi.org/10.1177/1043454220944126>
- Van Wye, E., Matheson, B., Citron, K., Yang, H.-J., Datta, N., Bohon, C., & Lock, J. D. (2023). Protocol for a randomized clinical trial for Avoidant Restrictive Food Intake Disorder (ARFID) in low-weight youth. *Contemporary Clinical Trials*, 124, 107036. <https://doi.org/10.1016/j.cct.2022.107036>
- Vassilopoulos, A., Mohammad, S., Dure, L., Kozłowska, K., & Fobian, A. D. (2022). Treatment approaches for functional neurological disorders in children. *Current Treatment Options in Neurology*, 24(2), 77-97. <https://doi.org/10.1007/s11940-022-00708-5>
- Velani, H., & Gledhill, J. (2021). The effectiveness of psychological interventions for children and adolescents with non-epileptic seizures. *Seizure: European Journal of Epilepsy*, 93, 20-31. <https://doi.org/10.1016/j.seizure.2021.09.006>
- Wade, S. L., Cassedy, A. E., McNally, K. A., Kurowski, B. G., Kirkwood, M. W., Stancin, T., & Taylor, H. G. (2019). A randomized comparative effectiveness trial of family-problem-solving treatment for adolescent brain injury: Parent outcomes from the Coping with Head Injury through Problem Solving (CHIPS) study. *The Journal of Head Trauma Rehabilitation*, 34(6), E1-E9. <https://doi.org/10.1097/htr.0000000000000487>
- Warner, C. M., Colognori, D., Kim, R. E., Reigada, L. C., Klein, R. G., Browner-Elhanan, K. J., Saborsky, A., Petkova, E., Reiss, P., & Benkov, K. (2011). Cognitive-behavioral treatment of persistent functional somatic complaints and pediatric anxiety: An initial controlled trial. *Depression and Anxiety*, 28(7), 551-559. <https://doi.org/10.1002/da.20821>
- Williford, D. N., Guilfoyle, S. M., & Modi, A. C. (2023). Demystifying a family-based epilepsy adherence problem-solving intervention: Exploring adherence barriers and solutions. *Clinical Practice in Pediatric Psychology*, 11(1), 66-73. <https://doi.org/10.1037/cpp0000436>
- Winkley, K., Upsher, R., Stahl, D., Pollard, D., Kasera, A., Brennan, A., ... & Ismail, K. (2020). Psychological interventions to improve self-management of type 1 and type 2 diabetes: a systematic review. *Health technology assessment (Winchester; England)*, 24(28), 1. <https://doi.org/10.3310/hta24280>

Article

Likert Scales: A Practical Guide to Design, Construction and Use

Pere J. Ferrando¹ , Fabia Morales-Vives¹ , José M. Casas¹  and José Muñiz² 

¹ Universitat Rovira i Virgili, Departament de Psicologia, Research Center for Behavior Assessment (Spain)

² Nebrija University (Spain)

ARTICLE INFO

Received: 16/05/2025

Accepted: 09/07/2025

Keywords:

Likert
Guidelines
Scales
Psychometrics

ABSTRACT

Background: Likert-type scales, first introduced by Rensis Likert in 1932, have become one of the most widely used assessment tools across a range of scientific and professional domains, owing to their simplicity and effectiveness. The purpose of the present study is to critically review their use and to propose a set of practical guidelines aimed at optimizing their construction, analysis, and application. **Method:** A systematic literature review of guidelines focused on the development, analysis, scoring, use, and interpretation of Likert scales was carried out. **Results:** Several key areas for improvement in the construction and use of Likert-type scales were identified, including the operational definition of constructs, item formulation, selection of the number of response categories, response analysis, collection of validity evidence, item calibration, and score interpretation. **Conclusions:** Based on the findings, a practical guide comprising fifteen recommendations is proposed: ten focused on the appropriate design, construction, and analysis of Likert scales, and five aimed at guiding appropriate use of pre-existing scales by researchers and practitioners.

Escalas Likert: Una Guía Práctica para su Diseño, Construcción y Uso

RESUMEN

Palabras clave:

Likert
Directrices
Escalas
Psicometría

Antecedentes: Las escalas tipo Likert fueron propuestas por Rensis Likert en 1932. Dada su sencillez y eficacia son uno de los instrumentos de evaluación más utilizados en muchas áreas científicas y profesionales. El objetivo del presente trabajo es revisar su utilización y proponer unas directrices prácticas para guiar su construcción, análisis y uso adecuados. **Método:** Se llevó a cabo una revisión crítica y sistemática de los trabajos y directrices publicados sobre la construcción, análisis, puntuación, uso e interpretación de las escalas Likert. **Resultados:** Se identificaron distintos aspectos de la construcción y del uso de las escalas tipo Likert que son susceptibles de mejora, como son la definición de los constructos a medir, la formulación de los ítems, el número de categorías, los análisis de las respuestas, las evidencias de validez aportadas, la calibración de los ítems, y la interpretación de los resultados. **Conclusiones:** Los resultados obtenidos se sintetizan en una guía práctica para investigadores y profesionales, compuesta por quince recomendaciones, diez centradas en el diseño, la construcción y el análisis adecuado de las escalas, y cinco encaminadas a guiar a los usuarios en la utilización adecuada de las escalas ya existentes.

Within the social sciences (and related) domains, the history of the Likert scale is a story of success. Initially proposed by [Likert \(1932\)](#) as a technique for attitude measurement, over the years it transcended its roots and expanded to domains as diverse as agriculture, tourism, electronics and robotics (to name a few), and ended up becoming part of popular culture, as aptly depicted in a cartoon by Chas Addams, published in 1982 by the legendary *The New Yorker*, which, incidentally, is celebrating its centennial this year: A warrior pollster approaches a peasant in his humble dwelling and asks: Would you say Attila is doing an excellent job, a good job, a fair job, or a poor job? ([Addams, 1982](#)).

In our view, Likert's proposal had two main ingredients for success: first, it was intended to be practical and as simple and cost-effective as possible, second, it had intuitive appeal, possibly more than any other scaling model. The literature accumulated over its first 92 years is overwhelming: more than a million papers, of which about 15.000 are monographs and user guides (and, in both cases, with a clear upward trend over time). So, a first consideration should be what is the point of making another guide. Our defense rests on two points. The first is transversality. The Likert scale no longer belongs to specific measurement fields and its use is much more general. Unlike domain-oriented tutorials, our guide should be useful for any researcher using the technique, whatever their field of study. The second point is about practicality, clarity, and relevance. We regard the Likert scale as a "Misunderstood Giant", the number of circulating misinterpretations, unfounded recommendations, and urban legends being proportional to its success ([Carifio & Perla, 2007](#); [Uebersax, 2006](#)). And, even in the case of sound research, we believe that an inordinate amount of effort has been devoted to "secondary" issues with the result that the few key points that are relevant for designing and/or using a Likert scale have remained overlooked. Quoting [Box \(1976\)](#), "It is inappropriate to be concerned about mice when there are tigers abroad", and it is well known that researchers (including us) can tend to make mountains out of molehills ([Sijtsma et al., 2024](#)). Being able to separate tigers from mice is a basic aim here. To sum up, we aim to provide a clear, well-founded guide, aimed at the practical researcher or user, and capable of emphasizing what is really important and relativizing what is not so.

Some final remarks are in order. First, certain recommendations that should be included here are quite general and have been discussed in previous guidelines published in this journal (e.g., [Ferrando et al., 2022](#); [Muñiz & Fonseca-Pedrero, 2019](#)). We shall not discuss them again but provide only appropriate references. Second, our approach is construct-oriented, model-based, and deductive, because we believe it to be the best founded and the one that works best. However, alternative approaches exist ([Burisch, 1984](#)). Finally, in no case do we intend to dictate unchangeable rules that must be followed but only to present constructive recommendations aimed at improving the measurement with this type of instrument.

Background and Framework

A Likert scale is a multi-item scale in which the scale scores are obtained as a composite of the scores on the individual items that compose it. Originally ([Likert, 1932](#)) it was defined as a summated

scale, in which scale scores were obtained by simple sum of the item scores. This definition can be broadened (see below point 9), but the most basic defining characteristic is the same: a composite score obtained from the item scores. So, a single item within a Likert scale is not a Likert scale, and neither is its response format. If one cares to re-read [Likert \(1932\)](#), one will see that the item format is considered secondary and the individual items are not taken very seriously. To adopt a precise terminology ([Uebersax, 2006](#)), we shall use here the terms: (a) "Likert scale" or Likert-type scale"; (b) "Likert-type item"; and (c) "Likert response format" for referring to the elements so far discussed, and reserve the term "Likert scaling" for referring to the technique in general.

What Type of Variables does a Likert Scale Measure?

A Likert scale is intended to measure dimensional constructs, i.e. abstractions which are inferred from real observations ([Nunnally, 1978](#)), and which can be conceived as continuous or dimensions along which individuals can be placed in terms of the amount or level in the construct they possess. Within this view, two most basic distinctions as far as the scale design is concerned are construct breadth and construct polarity. Starting with breadth: A narrow-bandwidth construct is specific and has relatively few possible manifestations, whereas a broad-bandwidth construct corresponds to very global phenomena and has multiple possible manifestations or facets ([Bagozzi & Edwards, 1998](#); [Cooper, 2019](#); [John & Soto, 2007](#); [Reise et al., 2000](#)). Indeed, we are defining extremes, and medium-breadth constructs also exist.

The concept of construct polarity refers to how the endpoints of the dimension can be interpreted ([Jebb et al., 2021](#); [Tay & Jebb, 2018](#)). In a conceptually bipolar construct, each end of the dimension can be univocally considered as the logical opposite of the other while in a unipolar construct the construct is defined at the upper end, there is no a univocal opposite for the lower end, and this lower end means, in most cases, only absence of construct manifestations. A construct as Extraversion, for example, is conceptually bipolar, as it is composed by two opposite poles (extraversion vs introversion), each pole describing different extremes of thinking, feeling and behaving. A construct such as positive-negative Mood can be considered bipolar, and so can most attitudinal constructs measured in terms of disapproval-approval ([Malhotra et al., 2009](#)). In contrast, many clinical constructs, such as depression, suicidal ideation or drug addiction, can be considered as unipolar. In the case of suicidal ideation, for example, the upper end of the trait continuum refers to the presence of suicidal ideation, with varying degrees of severity, while the lower end only refers to the absence of suicidal ideation, which does not necessarily imply emotional well-being ([Morales-Vives et al., 2023](#)). Constructs such as Virtue or Perfectionism have no univocal opposite lower end (Vice and Carelessness are not) and can also be conceptualized as unipolar ([Tay & Jebb, 2018](#)). While the Likert technique can be used with both types, it was initially designed for measuring bipolar constructs, and works best with this type (see below points 2, 3 and 4).

Likert scales are designed to be unidimensional, all their items measuring a single common construct. Multidimensional extensions, however, are possible, and are discussed below in point 5.

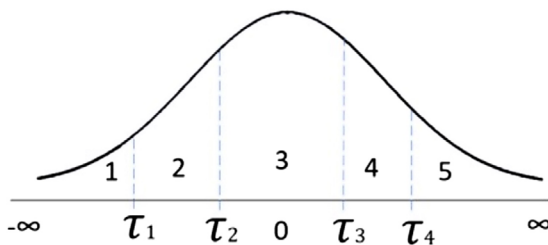
Assumptions and Functioning

The items in a Likert scale are intended to be “effect indicators” of the construct they measure (Bollen & Lennox, 1991; DeVellis, 2003). So, the basic framework is that it is the construct that (partly) determines the response to the Likert-type item whereas the remaining part reflects non-content related item specificities as well as measurement error. The scores derived from a Likert scale, both scale scores and individual item scores, are assumed to function according to a dominance or “monotonicity” relation in which the expected score increases as construct levels increase (Torgerson, 1958). The expected functioning of the original technique is simple (McIver & Carmines, 1981; Nunnally, 1978). First, the scale items measure a single construct. Second, item scores are expected to increase monotonically with construct levels. Third, scale scores obtained by summing the item scores are expected to increase (approximately) linearly with construct levels.

A pragmatic spirit also has his limitations, and Likert never proposed a theoretical basis for his technique. This can, however, be obtained by an item response modeling that we shall denote as Underlying-Variables-Approach (UVA; Muthén, 1984), and which is the basis for a calibration approach we shall propose here.

The UVA assumes that the ordered-categorical observed responses on a Likert-type item are the manifest expression of a latent continuous variable of response intensity, which is distributed as a standard normal variable. Along this latent variable there are a series of thresholds, and the observed categorical responses arise as a result of a division of the latent response continuous as determined by these thresholds (Hernández et al., 2004). Figure 1 illustrates the UVA functioning for an item with 5 response categories

Figure 1
Graphical Representation of the UVA



In Figure 1 the threshold values are approximately symmetrical around the mean zero point and are (also approximately) evenly spaced. These are the ideal conditions for a Likert-type item. If all the scale items function like this, then, the relations of the item scores with the construct are already essentially linear and the relations of the summated scale scores and the construct are almost perfectly linear. If an item behaves in this way, the distribution of its scores should be unimodal and approximately symmetrical.

The worst scenario for Likert scaling is when the items thresholds are neither symmetrical nor evenly spaced, for example, when most of them are piled up at one end of the response continuum. In this case, the relations between the item scores and the construct will continue to be monotonic but markedly nonlinear, and the sum scores will have a much harder time achieving linearity. The

observed item scores will now be asymmetrical, with strong positive or negative skewness.

A Practical Guide for Developing and Analyzing Likert Scales

Presented below is a practical guide comprising 15 recommendations aimed at the development and analysis of Likert-type scales. The first 10 focus on the creation of Likert scales from scratch, whereas the final 5 provide guidelines for the appropriate use of pre-existing Likert scales.

Building a Likert Scale from Scratch: A Decalogue

The 10 recommendations for developing a Likert scale are organized in three blocks. The first (points 1 to 5) refers to the design and construction of the items. The second and third are data-analytic: The purposes of second, calibration block (points 6, 7 and 8) are to evaluate the dimensionality and structure of the set of items, to assess their properties and quality as measures of the corresponding construct, and to select the most appropriate set that will form the final scale. The purposes of the third block (points 9 and 10) are to determine the most appropriate scoring schema for the scale, assess the appropriateness of the chosen scores, and obtain further validity evidence.

1. Be Specific! Specify Clearly the Type of Construct that is to be Measured and the Target Population

So as to organize the recommendations in this section, we shall consider that a prototypical Likert-type item consists of two parts: a declarative statement referred to a situation that is indicative of the construct and a response format that expresses a range of ordered and mutually exclusive categories from which the respondent has to choose one.

Before starting to write the statements for a new scale, the two characteristics of the construct discussed above: bandwidth and polarity, must be considered, as they will determine the appropriateness of the type and number of items. A broad bandwidth construct includes many different facets, and requires a larger number of items for each facet to be adequately represented in the scale (Loevinger, 1957). Otherwise, the scale will assess only part of the construct, which may affect its interpretability and ability to predict relevant outcomes. In contrast, a very specific, narrow-bandwidth construct may be adequately assessed with a small number of items, as otherwise the statements would become too repetitive.

As for the polarity distinction, it becomes particularly relevant when deciding the most appropriate response format (see point 3), and whether reverse items should be included or not (see point 4).

Specifying the target population is crucial, mainly with regards to their language competence, reasoning and reading skills, and abstraction abilities. Short and easy to understand statements are always preferable, but even more so if they are intended for people with certain comprehension difficulties.

2. Design the Statements Rigorously

Likert-type items are characterized by two main properties. First is their quality as indicators of the construct: A high-quality item is strongly and directly influenced by the construct it indicates and very little by

error, either specific or random (Cronbach & Gleser, 1964). The second property is extremeness, i.e. the items' positions in the continuous. Both properties mostly depend on how the statements are formulated.

Sound and consistent guidelines for writing good Likert statements are provided, among others, in Clark & Watson (2019), Fink (2003), Johnson & Morgan (2016), Mellenbergh (2011), or Spector (1992), and we strongly recommend studying them. As an aid, Table 1 provides a summary of the main points to consider.

Table 1
Recommendations for Writing Likert Statements

1. Write the statement as specific and direct as possible.
2. Use brief statements: Ideally fewer than 20 words.
3. Make complete sentences and avoid abbreviations.
4. One statement contains only one complete idea.
5. Avoid compound or double-barreled sentences.
6. Put the situational or conditional part of the statement at the beginning.
7. Use clear and comprehensible wording.
8. Avoid jargon and technicisms.
9. Take into account the reading skills of the target population.
10. Avoid the use of negatives, particularly double negatives.
11. Avoid biasing and sensitive wording.
12. Minimize redundancies in content and in form.
13. The statements must be logically related to the construct.
14. Items should indicate the same construct to most people evaluated.
15. Items should elicit different responses at different construct levels.

The first 12 recommendations from Table 1 are aimed at reducing error, but they do not determine the strength and directness of the construct influence. This goal requires taking into account the last 3 guidelines: The statements must (a) be logically related to the construct (13), (b) indicate the same construct to most people from the target population (McCrae et al., 1993) (14), and (c) elicit different responses at different construct levels (Mellenbergh, 2011) (15).

Buss & Craik (1981) distinguished between “prototypical” and “peripheral” statements. The first ones are those that “hit the core” of, or define the construct. The latter refer to behaviors or situations that are related to the construct but that are not part of its definition (Clark & Watson, 2019). Peripheral items might be of interest for a more complete sampling of construct manifestations, but most of the statements in a good Likert scale should be prototypical.

An assessment of content validity by expert judgement is highly recommended to determine the appropriateness of the statements (DeVellis, 2003). Experts must be able to determine whether (a) their content is suitable for assessing the construct, or a specific facet, and (b) their vocabulary, length, and grammatical complexity make the statements easy to read and understand for the target population (Muñiz & Fonseca-Pedrero, 2019).

Likert statements are written to elicit a response on a specific format. In the original proposal, the format measured in terms of agreement /disagreement (Likert, 1932). Here we shall take a broader view and assume that it can measure in terms of agreement, endorsement, intensity, frequency or amount (Bass et al., 1974; Fink, 2003; Spector, 1992).

The original disagreement-agreement format is bipolar and has two logically opposite endpoints which are antonyms. Statements that elicit a response in this format are clearly meaningful when the construct is also bipolar. However, they are not generally so when it is unipolar. In this case, using statements that elicits a

response in terms of intensity, frequency or amount is generally more appropriate. For example, if we are interested in measuring an addiction or perhaps a belief (note that the low end of the dimension in these cases is absence of addiction or no belief respectively) then a unipolar statement formulation eliciting a response in terms of “rarely vs. most of the time” or “I don’t believe in this vs. I believe in this very much” seems more natural. Statements eliciting unipolar or bipolar responses regardless of the construct polarity, can be indeed formulated, but when polarities do not match, the functioning of the item is generally less efficient and, in many cases, it is perceived as unnatural (Spector, 1992).

We shall finally discuss extremeness as related to the purpose of attaining accurate measurement across the widest possible range of the construct continuum (other purposes can be considered but we cannot address them in this guide). The standard recommendation is to use a “scalability” approach, and develop statements that spreads across most of the continuum range (Henrysson, 1971). This means that we should need “medium evocativeness” or moderate statements that cover the central region, as well as more extreme statements designed to cover both ends of the continuum. This strategy agrees with common sense and is a reasonable approach in the case of binary items, which is where it was initially proposed.

As counterintuitive as it may seems, the strategy Likert recommended is quite the opposite. The idea now is to design all the statements at a “medium” degree of extremeness, and leave the task of covering the widest possible range of construct levels to the response format. In other words, it is not the statement that tries to capture the respondent’s levels, but the respondent who manifests it by using the provided format (see point 3). This strategy agrees with the principles of Likert scaling and makes the calibration and scoring processes simpler.

Our recommendation? First, the breadth of the continuum that the items are able to span can and should be assessed empirically (see point 8). Second, without being as radical as Likert, we believe that writing moderate statements mostly located around medium levels is the best strategy. People with extreme construct levels will tend to score high in most of these items, which would result in a high overall score, without the need to include extreme items to identify them (Cronbach & Warrington, 1952; Henrysson, 1971). It is acceptable to extend somewhat the range of extremeness, but we do not recommend going too extreme. Apart from not being necessary, extreme items make the calibration process more complex and unstable. Furthermore (in our experience) they generally sound bizarre, and are perceived almost as caricatures.

3. Choose the Appropriate Response Format

The multi-category response format in Likert-type items is scored with consecutive integers, and is anchored with verbal labels expected to reflect gradations. According to the literature, there are two main topics related to this issue. The first is about the anchoring labels or category qualifiers: how should they be worded and how many of them should be used, including the convenience or not of using a middle category. The second is about the most appropriate number of categories. Existing evidence and recommendations are consistent with regards to this second topic, but no so to the first one. For this reason, we shall rely primarily in our experience as a guide for the first topic, provide only three basic recommendations

here, and thoroughly discuss it in point 12 below. The basic recommendations are: (a) the labels must be consistent with the terms in which the statements are formulated (e.g. agreement, frequency, amount...); (b), they must conceptually suggest equally spaced categories (see Figure 1), and (c) the label of the middle option (if used) must clearly indicate a neutral position rather than inability or unwillingness to respond.

Turning now to the stellar topic in Likert scaling: The “optimal” number of categories. We shall consider here two different points of view: that of the scale designer and that of the respondent (Preston & Colman, 2000). The former is mostly interested in these issues in terms of (a) maximizing the amount of score reliability and validity and (b) attaining a strong, clear, and stable calibration structure (all of this, of course, at the lowest possible cost). What is most appropriate for the respondent, however, according to the cognitive miser approach (Fiske & Taylor, 2020), is that responding would require minimal cognitive effort, and that the response format would be appropriate to the way in which she/he would have expressed the response.

Starting with the designer. With regards to score reliability (including test-retest), the accumulated results are clear: reliability increases with the number of categories but according (almost) to a law of diminishing returns. The consensus range at which the increases are clear is between four to seven categories (Lee & Paek, 2014; Lissitz & Green, 1975; Lozano et al., 2008; Weng, 2004). From seven points onwards, the consensus disappears (Nunnally, 1978), but what is clear is that the gains either stale or, if there were any, are minimal. These results, however, should be taken critically. Most reliability estimates increase with the observed variance, which tends to increase as the number of categories increases (Lozano et al., 2008), and, above all, with systematic variance, which is composed of the true and specific variance. And it could well be that as the number of categories increases, specific variance would increase (e.g. systematic trends in response scale usage) but not true variance (Cronbach, 1950; Lee & Paek, 2014). An additional consideration, based on our experience, is the potential advantage of employing scales that are already familiar to the target population. For instance, in the academic context of many countries, such as Spain, the 0–10 grading scale is commonly used. Utilizing this scale in such contexts offers the benefit of participant familiarity, while also enabling the omission of verbal labels for response categories. Evidence based on external validity relations finally, is far scarcer (see point 10), but, in any case, the differences within the consensus range seem to be minimal (Hubatka et al., 2024; Sancerni et al., 1990; Speer et al., 2016). As for the relevance of the central category, finally, provided that is well designed, whether or not it is included does not makes much difference in terms of score reliability (DuBois & Burns, 1975; Mariano et al., 2024). Nevertheless, if only three response categories are employed, which, as previously discussed, is not advisable, the central category may exert a disproportionate attraction effect, potentially introducing bias and distorting the results.

Evidence based on the strength clarity and stability of factorial solutions is more well-founded and compelling (Comrey, 1988; Muñoz et al., 2005; Tomás & Oliver, 1998), but the consensus results are quite similar to those above: the structural properties of interest appear to increase with the number of categories and reach a maximum at seven. Again, however, these results need to be qualified. They clearly hold when the model used for item calibration is the linear model. However, the non-linear model would

be expected to perform better with fewer than seven categories (see point 6). As for the role of the central category, again, its inclusion or not does not seem to lead to appreciable structural differences (Mariano et al., 2024; Muñoz et al., 2005).

Turning now to the long-suffering respondent. Responding appropriately to a Likert scale is a relatively complex cognitive task that also requires a certain level of motivation and reading skills. And not all the profiles of the respondents fulfill the requirements or are willing to devote the necessary effort to the task (Krosnick, 1999). As for the agreement between the “respondent-constructed” vs. the “designer-provided” response-format finally, results obtained by the “discovery” method, in which respondents organize their responses along a continuous line, suggest that they: (a) divide the continuous in a number of discrete number of clusters, usually between 5 and 11 (7 being the most common); (b) clearly make use of the two ends of the continuous line; and (d) naturally use a central category (Ferrando, 2003; Mariano et al., 2024; Munshi, 1990).

Taken into account all this information, our reflections and recommendations are as follows. First, there is not a universal “optimal” number of categories. Rather this number should be selected within a reasonable range taking into account the characteristics of the target population, the type of construct, and the cognitive demands of the task (see point 12). Second, the issues we shall discuss in point 12 are far more relevant than the number of categories (as long as they are within a reasonable range). As for recommendations, if the cognitive and motivational levels of the target population are reasonable, an appropriate range is 5 to 7. If they were very low (e.g. cognitively impaired samples, substantial comprehension difficulties or low introspective capacity) then, we suggest going down even as low as binary.

4. Make an Informed Decision About to Balance or not to Balance Statements

The convenience or not of balancing Likert statements is a controversial topic that has given rise to opposite recommendations (Spector, 1992; Suárez-Álvarez et al., 2018; Vigil-Colet et al., 2020). In our opinion, both positions are partly correct and sound recommendations can be made if certain basic conditions are first clearly defined.

The original Likert recommendation was to write statements oriented towards each of both poles of the construct continuum. This is reasonable in the initial formulation, in which a bipolar construct is measured by using a bipolar format. In these conditions, positively-worded statements that fulfill the conditions in Table 1 can be written in a natural way. Furthermore, if a fully balanced scale is obtained with items of this type, some useful information will be gained (mainly proneness to acquiescent responding) and cleaner and more interpretable score estimates can be obtained (Hernández-Dorado et al., 2025). It should be taken into account, however, that the expected improvements in terms of score interpretability additional information and external validity (point 10) are modest.

The problems arise when the recommendations above are attempted to be applied to unipolar constructs measured with unipolar items. To start with, the recommendation of writing statements oriented towards both poles has little meaning, since, in fact, there is only one meaningful pole. As a consequence, in most cases, statements oriented toward the lower end can only

be achieved by using negative wording (sometimes even double negatives) and/or unnatural “forced” statements. For example, in the case of addictions, a statement may ask whether, or how often, you use a particular illegal drug, and it will sound natural. However, if you are asked how much you agree that you do not use that drug, you are likely to find it more difficult to answer. Writing elements of this type has no advantages but serious disadvantages (Suárez-Álvarez et al., 2018; Tay & Jebb, 2018).

Our recommendations on this issue are as follows. First, the option of balancing the statements is only fully feasible and meaningful when you are measuring a bipolar construct with a bipolar format. Otherwise, it is better to orient all the statements in the same direction, preferably towards the meaningful pole of the construct. If balancing is feasible and you decide to do it, then you have to do it well: the scale has to be fully balanced (half of the statements oriented towards one pole and the other half towards the opposite pole) and the statements have to be all positively worded and fulfil the writing rules in Table 1.

In summary, we recommend considering the inclusion of reverse items only in the case of bipolar constructs. However, this inclusion alone is not expected to eliminate automatically the impact of acquiescence. Consequently, if we want to assess a bipolar construct but there is no intention to implement any procedures that control this response bias, consider that adding reverse items may not be beneficial, but in some cases even counterproductive (Suárez-Álvarez et al., 2018; Vigil-Colet et al., 2020), especially if they are not well formulated or sound strange or artificial. In fact, in low-stakes settings, rather than including reverse items, it may be more useful to reduce acquiescence by including a few short, easily understandable items with vocabulary adapted to the target population in order to avoid fatiguing respondents and maintain their attention. Similarly, using appropriate wording for items (e.g. avoiding a judgmental tone) may help reduce social desirability bias in low-stakes settings. In high-stakes assessments, however, statistical procedures designed to mitigate the impact of social desirability would be advisable (e.g. Ferrando et al., 2009). As mentioned in point 2, however, this last setting falls outside the objectives of our proposal.

5. Including Likert-type Items in Multidimensional Instruments can be Done

There is no problem in using Likert-type items in multidimensional instruments. In fact, it is an increasingly widespread practice. Constructing meaningful Likert subscales based on this type of instruments, however, is not so simple, and requires that the items can be univocally assigned to non-overlapping subscales. In more detail: each item assigned to a single subscale and each subscale made up of a different set of items. In an ideal world, this assignment would be directly obtained from a factorially simple structure, which is also known as an independent-cluster structure (ICS; McDonald, 2000). In an ICS, each item behaves as a “marker”, having a substantial loading on only one factor, of which it is an indicator, and zero loadings in the rest of the factors.

Back to the real world. Most of the constructs that are measured using Likert-type items do not allow fully ICSs to be obtained (Clark & Watson, 2019; Ferrando, 2021; Lucke, 2005). Rather, the items are generally (and inherently) complex and tend to load on more than one factor. This fact, however, should not be taken as an excuse

for designing poor multidimensional measures. On the contrary, the designer should strive for attaining the “cleanest” structure possible and, in this line, we dare to propose two goals that the final scale should aim to attain. First, an independent-cluster basis (ICB; McDonald, 2000) consisting on, at least, three markers per factor should be obtained. Second, each of the remaining, complex, items should have a clear dominant loading on a single factor (Comrey & Lee, 1992). Fulfilment of these conditions still allows an almost univocal assignment of the items to the subscales, which means that composite scores obtained for each separate subscale could be validly interpreted as measures of a single dimension.

6. Choosing the Most Appropriate Model for Calibrating the Items

The two most common modeling approaches for calibrating Likert-type items are (a) the linear FA model for continuous responses, and (b) the non-linear FA model for ordered-categorical responses. Both are used with a more general class of items and have been discussed in depth in previous guides (Ferrando et al., 2022; Muñiz & Fonseca-Pedrero, 2019). So, we shall focus here on its use as related to the specific characteristics of Likert items. A previous consideration, however, is needed: Neither model is “the correct” model, there is not such a thing. Rather, both are convenient approximations, and the key point for the developer is to assess his/her data and decide which of them is the most appropriate.

In the linear FA model, item scores are treated as continuous-unlimited and the item-construct relations are assumed to be linear. If the reasonable range of categories recommended in point 3 is used, discreteness by itself does not represent a big problem, but non-linearity sometimes can. Building from the discussion around Figure 1, essentially linear item-construct relations can be expected when: (a) the items are non-extreme, with about equally-spaced thresholds, and (b) the item discriminations are not too high. In more practical terms, these conditions can be expected when the items are designed according to the Likert strategy (point 3), measure normal-range broad constructs (points 2 and 3), and the design has minimized redundancies or correlated-specificities (e.g. Ferrando & Morales-Vives, 2023). When these conditions are met, the use of the simple linear FA model is quite defensible. Furthermore, calibration based on this model is very robust, which is an advantage when the sample is small to medium (say, below 200), the number of items is large (say, more than 20) and the number of categories is also large (seven or more).

Non-linear FA calibration is based on the UVA approach described above, and so, it is more aligned with the foundations of Likert scaling but at the cost of some strong assumptions that are difficult to be tested. Furthermore, the nonlinear UVA-FA model can be viewed as an alternative parameterization of the Item Response Theory (IRT) Graded Response Model (GRM; see e.g. Ferrando, 2021). On the positive side, its most important advantage is that it provides more information from the data (see points 7, 8 and 9). On the negative side, calibration becomes more demanding and potentially unstable when the data is sparse. Overall, the conditions in which the nonlinear model is expected to work well are: large samples, not too many items, and not too many categories (with more than seven it is practically unfeasible). If these conditions are attained, nonlinear calibration is a more informative alternative to linear calibration.

7. Assessing the Appropriateness of the Chosen Solution

Guidelines for assessing the adequacy of structural solutions in item analysis have been previously proposed in this journal (Ferrando et al., 2022; Muñiz & Fonseca-Pedrero, 2019) and shall not be repeated here. Appropriateness is mostly assessed via goodness of model-data fit (Bollen & Long, 1993; Jebb et al., 2021; Maydeu-Olivares et al., 2017), and this is, indeed, a first basic requirement. However, we shall emphasize here a more practical view that focuses on two main additional sources of evidence. First, that the calibration results have to be strong, stable, and replicable, which means that the scale is expected to function well not only in the calibration sample, but in any sample belonging to the target population. Second, that the scores derived from the calibration results have a univocal interpretation as measures of the corresponding construct. Indices such as the *H* index (Hancock & Mueller, 2001) and the single-sample Expected Cross validation indices (Browne, 2000) are good measures of the first group of properties. Indices based on the amount of explained common variance, either absolute or relative (Ferrando et al., 2024), or marginal reliability estimates (see Table 2 in point 9) are of the second.

Even in the case of acceptable solutions, the single-sample results are not sufficient to establish that the items are working properly in a more general sense. At the end, evidence of across-sample replicability is an empirical matter that requires at least two samples. The simplest option is to randomly split the sample into two subsamples, and verify the invariance of the results (see Browne, 2000 for extensions and a detailed treatment of the issue).

8. Optimal Item Selection: Taking the Main Purposes Into Account

For the most part, the model-based item selection process in Likert scaling is common to that used with noncognitive items in general, and has been discussed in previous guides (Ferrando et al., 2022; Muñiz & Fonseca-Pedrero, 2019). However, two distinctive features can be derived from the recommendations so far. First, the basis solution that is sought: either unidimensional or near independent-clusters-basis (ICB; see point 5), is somewhat more restricted than those commonly used in general applications. Second, the recommended process of item design is deductive and rigorous, which means that the initial stages of “cleaning” and discarding inappropriate items are expected to be simplified (i.e. less garbage in; e.g. Wrigley, 1976).

The process of item selection aims at two general goals of which the first is requisite for the second. The first is about achieving a well-fitting, appropriate solution, that agrees with the expected structure (see points 6 and 7) and that is strong stable and replicable. This is necessary but not sufficient. Beyond that, the second goal requires that the items in the final set cover a broad range of construct levels, are of good quality, represent appropriately the different construct manifestations, and their number is sufficient to attain accurate measurement (see points 3 and 9).

We shall now get more specific. As for the item location, apart from the descriptive statistics recommended in previous guides, the main indicators here are the item thresholds (see Figure 1), which can be obtained regardless of the type of solution that is fitted.

So, whether using a linear or a nonlinear solution, we recommend always examining the thresholds (Sideridis et al., 2023; Wakita et al., 2012). In accordance with the discussions in points 3 and 6, we should aim for threshold estimates that are more or less evenly distributed around the zero point and that cover a broad range of the response continuum (Muthén & Kaplan, 1985; Wakita et al., 2012). Items with a very narrow threshold range or in which all thresholds are of the same sign should be better discarded. Threshold estimates can be directly obtained using non-commercial R programs such as GRShiny (Lee et al., 2023).

Item quality is operationalized by the item discrimination index, which, in the present proposal, can be provided in two metrics: standardized factor loadings (in both the linear FA and in the non-linear FA parameterization) or IRT slopes (non-linear FA with IRT parameterization). In our opinion, an appropriate range of values would be between .3 and .7 in loading metric, which translates to .3 to 1.00 in slope metric (Ferrando & Morales-Vives, 2023). Values below .3 would indicate that the item is too noisy, whereas values above .85 (1.70 in slope metric) would possibly indicate design problems, or redundancy.

Overall, the item selection process for obtaining the best possible final scale is a balancing act and an art that requires practice. We need enough items to achieve accurate measurement, but not too many so as not to annoy or demotivate the participant. We need to sample appropriately the construct, but also to avoid almost irrelevant items that are too far removed from its core. And we need consistent, good-quality items but without falling into redundancy.

Arriving to the optimal final solution, requires all the previous steps in scale development to be carried out thoroughly, which requires time and effort. Firstly, an adequate review of previous literature is needed to obtain a specific and accurate definition of the construct to be assessed (point 1). Secondly, a sufficiently large pool of statements that cover all the different facets of the construct are needed. Third, several pilot studies are usually needed to get a preliminary idea of which statements should be discarded or rewritten, following a qualitative and quantitative perspective, as they provide complementary information. Thus, participants may be asked to indicate to what extent they consider each item clear, for example with a 3-point scale (1 = I don't understand this sentence at all, 2 = I have a vague idea of what this sentence means, but not a full understanding, 3 = I understand this sentence completely). Furthermore, they may be asked to explain the meaning of each sentence, as well as to indicate anything they do not understand. This kind of pilot studies can be carried out with small samples. However, some pilot studies should be carried out with sufficient sample size to support preliminary factor analyses. Fourth, evidence about the appropriateness of the chosen final solution is needed. This final solution can be attained through both restricted (or CFA) and unrestricted (or EFA) analysis, or using even both FA models in tandem. In fact, EFA may be especially helpful in the preliminary steps, and also in the cross-validation with different samples or subsamples, in order to determine the number of factors underlying the data, if there are poor working items and if the solution is stable. The CFA can be used as a final verification of the appropriateness of the solution.

Another, sometimes, overlooked issue is the (possibly) differential functioning of items in specific sub-populations. When new scales are developed, or adapted, community samples are often used for

validation. Sometimes, however, these scales are used in specific settings on the assumption that the properties of the instrument will remain stable across specific sub-populations. However, this is not always the case, as language skills, levels of education, behavioral patterns, moral values, interpretations of everyday issues, etc. may widely vary from one population to another (Spector, 1992). It should also be noted that even the factor structure may not be the same in specific sub-populations, with even some items defining a particular factor in some populations but not in others (Casas et al., 2025). It is common practice to test whether factor structures are invariant across gender, age, and even ethnicity (Benson et al., 2020), but the same should be considered for specific sub-populations as compared to the community population.

9. Scoring the Likert Scale and Assessing Score Appropriateness

In the two-stage strategy recommended in this guide, the individual Likert score estimates are obtained on the basis of the final calibration solution, and for them to have a univocal interpretation, the basis solution should attain the first goal in point 8. If so, we have reasonable evidence that the scale (or subscale) scores do not reflect a mixture of unknown determinants but a common dimension. In addition, for these estimates to be accurate, the second goal in point 8 must be also met.

As scoring is dependent on previous calibration results, guidelines can be provided both, when a new scale is developed or when scores are to be obtained from an existing scale. In this respect, we have decided to provide the needed background here and discuss the more practical recommendations in point 14 below.

There are three main scoring choices in Likert scaling which are summarized in Table 2.

Table 2
Main Scoring Options in Likert Scaling

Basis Model	Score estimate	Measures of score accuracy
Linear FA	Sum scores (unweighted composites)	Standard reliability estimates (α and ω mainly)
	Factor score estimates	Marginal reliability estimates
Nonlinear FA-IRT-GRM	Sum scores	Standard reliability estimates (α and ω mainly)
	Score estimates based on the response pattern	Conditional reliability estimates
		Marginal reliability estimates
		Amount of Information

Note. FA: Factor Analysis. IRT: Item Response Theory. GRM: Graded Response Model.

We shall start by discussing the Likert-original, simplest and most general type of scores: the sum scores, which can be used with both, the linear and the nonlinear model. When based on the calibration results (either linear or nonlinear), the only information they use is “configurational”: i.e. which are the items that indicate the construct (see point 8). From here, sum scores assign equal unit weight to all the scale (or subscale) indicators regardless of their quality. So, because of the amount of information they do not use, sum scores are, theoretically, sub-optimal measures of the construct. However, in the case of scales designed according to the conditions recommended here, this theoretical disadvantage might be not that relevant in practice (Speer et al., 2016).

Sum scores are not directly interpretable in terms of the relative meaning of the score with respect to the reference population. So, if this information is required, norms must be compiled. Spector (1992) provides a good summary for compiling norms specifically intended for Likert scales.

If the linear FA model was the most appropriate calibration choice, the factor score estimates or predictors would be, again in theory, the most appropriate choice. Although there are many different types (Grice, 2001), all of them are, essentially, weighted composites of the item scores in which the weights reflect the quality of the item as indicator of the construct. So, factor score estimates use more information from the data than sum scores, and, therefore, are expected to be more accurate. Whether this theoretical advantage is realized in practice, however, depends on the stability of the calibration results (Wainer, 1976). Finally, in terms of interpretation, factor score estimates in Likert applications are almost always scaled in standard metric (zero mean and unit standard deviation). So, they are directly interpretable in terms of relative standing with respect to the reference population.

Scores based on non-linear FA, particularly when using the GRM-IRT parameterization (GRM-IRT-based scores), are (again theoretically) the most informative and accurate choice in Likert scaling. The requirements for these advantages to hold in practice, however, are the same as discussed above: strength and stability of the calibration results, which, in this modeling, are more difficult to be obtained (see point 6). IRT scores provide, for each individual, a score estimate based on his/her full response pattern, and use, virtually, all the information available from the calibration results. Usually, as in the linear-FA-based scores, IRT score estimates are scaled in standard metric. GRM-IRT scores and the corresponding accuracy measures discussed below can be obtained with non-commercial programs such as Factor (Lorenzo-Seva & Ferrando, 2013) and R programs such as GRShiny (Lee et al. 2023).

We turn now to the measures of score accuracy in the third column of Table 2, which usually, are provided in the form of reliability estimates. There are two basic points to emphasize here. First, accuracy is a property of the scores, so, each reliability estimate in Table 2 is intended to be used with a specific type of score. Second, the main difference between the accuracy measures is that, in the case of IRT scores, the reliability varies at different construct levels (i.e. conditional reliability) whereas, for the remaining scores, accuracy is assumed to be the same at all construct levels (i.e. marginal reliability; see Muñiz, 2018).

10. The Importance of Being Valid: Provide External Validity Evidence!

While Validity is discussed at length in many Likert-related tutorials (DeVellis, 2003), we believe that some practices are improvable, and some types of evidence are too scarce. The theatrical title of this section, partly borrowed from John and Soto (2007), is a critical warning regarding this situation.

We have discussed content evidence in point 3, and evidence based on internal structure (American Educational Research Association [AERA], 2014), which is the one that usually receives the most attention in applications, in points 6, 7, and 8. Furthermore, reported practices are clearly improving in this respect. This is correct as a basis, but we cannot solely rely on this source.

External Validity evidence includes that based on relations with other variables (convergent evidence; [AERA, 2014](#)) as well as criterion-related evidence ([Sireci & Benitez, 2023](#)). The first is becoming a requisite and is increasingly used. However, in the Likert applications we have revised, we believe that there is room for improvement. To start with, in most cases the reported evidence is simply a matrix containing the correlations between the scale scores and other measures expected to be related to the construct. This setting serves as a starting point, but can be improved.

A first conceptual problem we have frequently detected is that the related measures seem to be measuring practically the same thing but under a different name ([Furnham, 1990](#)). In fact, some of the intervening items could be in both the scale that is validated and the one that serves as validity source. If external evidence is sought, this is bad practice. To be really external, the chosen measures of the other variables should be expected to be related to the one that is validated but should tape clearly differentiable constructs.

At a more technical level, when reporting the correlation matrix, we recommend to report the point estimated correlations together with their confidence intervals as well as the disattenuated correlations. The disattenuated correlations are theoretical validity estimates that, if correctly obtained, give us an idea of the ‘true’ relationships between the constructs involved ([Lord & Novick, 1968](#)).

Beyond the recommendations above, we believe that the type of evidence we are discussing should be far more elaborated. First, as [Spector \(1992\)](#) recommends, the assessment of the relations should be based on a set of hypotheses derived from well-supported theory, and this implies clearly stating the expected strength of the relations as well as which of them are considered central.

As for criterion-related evidence, the first obvious weakness is that this type of evidence is very scarce in Likert-scaling applications and that its usage should be increased. A second, and quite usual, limitation is the use of some type of test scores as if they were a proper criterion. This again is questionable, because, quoting [Wainer \(1993\)](#), “Nothing predicts a test like another test” (p.2). We know by experience that obtaining proper and suitable criteria is a very hard task and, furthermore, that the results are generally not very rewarding. However, this type of evidence is highly relevant. [Wainer \(1993\)](#) and [Spector \(1992\)](#) provide good guidelines for obtaining meaningful criterion-related evidence. Finally, as in the convergent case, we believe that criterion-related evidence should be also well grounded in theory and based on explicit hypotheses.

Going a step further, we dare to propose the applied researcher or practitioner to try to improve standard validity practices by using structural equation modeling (see e.g. [Bollen, 1989](#)). This recommendation is consistent with the foundations of this guide: we have strongly recommended so far that the “internal” development of a Likert scale should be model-based. Well, we believe that the assessment of external evidence should also be. The advantage of fitting a structural equation model rather than analyzing first-order correlations is that it allows different sources of validity, including convergent and criterion, to be jointly assessed in a single model.

It allows to determine whether the scale scores have the expected relationships with other test-score-based related variables, but also with external criteria, such as academic performance assessed through student grades. It therefore gives an idea of the scale’s predictive capacity and, depending on the variables included in the model, its incremental validity in relation to other relevant variables. It also makes it possible to test general models based on previous theories or studies, providing additional evidence of the scale’s performance in this context.

Using an Existing Likert Scale: A Quintet in Two Acts

The final part of this guide is aimed at the practitioner or applied researcher who needs to use a Likert scale but is not planning to develop one. In this scenario there are two main areas of concern. First is about critically assessing the available options, selecting the most appropriate instrument, and, in some cases, making modifications. The second is about applying the instrument, and estimating and interpreting the scores.

11. Caveat Emptor! Check Thoroughly the Features and Existing Information About the Instrument

Obviously, in the process of selecting one scale or another, it is necessary to check which of them better addresses the construct of interest. Unfortunately, sometimes different scales that seem to assess the same construct according to their name, assess actually different constructs, either totally or partially ([Furnham, 1990](#)). It is therefore necessary to check how the construct is defined and on what models or theories is based on. Furthermore, if a broad-bandwidth construct is to be assessed, one should be wary of scales with very few items, as they are likely to assess only part of it. And if a clearly unipolar construct is to be assessed, one should also be wary of scales with reversed statements, as they are likely to be unnatural, possibly leading to undesirable results (see point 4). Regarding the statements, their appropriateness for the characteristics of the respondents we aim to assess must be checked, especially with regards to length, grammatical complexity, and vocabulary sophistication when the scale is intended to be applied to respondents with a low educational level or with comprehension and reasoning problems. And, with regards to the response format, please, carefully review point 3 above.

12. Examine the Psychometric Properties of the Scale

Once the scale appropriateness with regards to the issues above have been assessed, it is necessary to examine its psychometric properties, and not only in terms of the calibration results: evidence about the stability of the structure in different samples, and also about the predictive ability of the scores, at convergent and criterion validity levels is needed. These recommendations are summarized in [Table 3](#).

Table 3
Issues to Consider When Determining Whether the Features of an Existing Scale are Appropriate

Features	Issues to be considered
Type of construct assessed	Does the definition of the construct provided by the authors of the scale correspond to what it is intended to assess? In unipolar constructs, are there any “unnatural” reversed statements? In broad bandwidth constructs, are there enough items to assess its different facets?
Statements	Are the items too long and grammatically complex to be easily understood by the individuals to be assessed? Is the vocabulary suitable for the individuals to be assessed?
Response format	Are the category labels appropriate for the statements? Is the number of categories suitable for the individuals to be assessed?
Psychometric properties	Can the calibration results (e.g., dimensionality, goodness of fit, simple structure, etc.) be considered as adequate? Is there any evidence about the stability of the factor solution in different samples and populations? Is there any evidence about the relations with other variables?

13. Adjust and Improve the Scale if Possible and Necessary

It is quite usual to get a scale that appropriately assesses the intended construct, but with some features that are not entirely appropriate for the application at hand. The most typical cases are the following: (a) An inappropriate number of response categories; (b) response labels that are not sufficiently aligned with the statements; (c) unequal conceptual distances among categories; and (d) over-labelling.

In the first case, if a scale has too many or too few response categories for the intended population, and no alternative scales are available, the adjustments would consist of adding or deleting categories, according to the recommendations in point 3 above.

In the second case, the recommended adjustment is to change the labels of the categories, to make them more consistent with the wording of the statements. There is a tendency to use by default the strongly disagree to strongly agree format, and this works reasonably well in many cases (Goretzko et al., 2019; Hühne & Krebs, 2018; Spector, 1992). However, if the statements refer to whether a particular thought, situation, emotion, symptom, etc., has been experienced recently, it would be advisable to use labels ranging from, for example, never or almost never to almost always. Furthermore, if the statement already refers to frequency (for example, “I rarely feel happy”), the response categories should not be labelled in frequency terms (Clark & Watson, 2019).

The third case is common in unbalanced response formats in which some kind of response categories are over-represented, while others are underrepresented. Despite being a relatively frequent problem, it usually does not get the attention it deserves. The following set is an example:

1. Never, 2. Rather infrequently, 3. Quite often, 4. Very often, 5. Always

In this case, it cannot be assumed that there is an equivalent psychological distance between the categories, as the jump from category 2 to 3 is clearly conceptually greater than the jump between 3 and 4 or between 4 and 5. Furthermore, it is an unbalanced scale, with three positive labels that involve a high frequency (3-5) and only two negative labels that express low or null frequency. An

appropriate Likert format is characterized by similar conceptual distances between the categories. So, it is advisable here to adjust the labels. Furthermore, if the odd number of categories is maintained, a middle point will be required to achieve a balance between positive and negative categories. Our recommendation is to include it provided that fits well with the response terms and is appropriately labelled (Wang & Krosnick, 2019). In fact, the middle point should not be seen as a problem if it is really part of the gradation, representing one of the possible positions that individuals can take. So, it may be useful that the instructions make it clear that this category is part of the gradation of answers and is not the same as giving no answer or expressing uncertainty. In fact, it is advisable to give instructions on how to answer items with a Likert response format, using some dummy items as examples.

Returning to our example regarding frequency, a balanced alternative with 5 categories could be 1. Never, 2. Rather infrequently, 3. Some of the time, 4. Quite often, 5. Always, as suggested by Casper et al., (2020). According to these authors, a balanced alternative referring to agreement could be 1. Disagree, 2. Somewhat disagree, 3. Neither agree nor disagree, 4. Moderately agree, 5. Very much agree. Other good examples referring to amount, similarity and judgement can be seen in Casper et al., (2020) and Bass et al., (1974).

The fourth case refers to over-labelling, which should also be avoided. Because the Likert response format is so well known, it is sometimes sufficient to indicate the labels of the endpoints, especially when the number of response categories is very large. With 7 or more response categories, each with its own label, the large number of labels may lead to confusion and or excessive cognitive effort (Frary, 2003; Krosnick, 1999; Willits et al., 2016). On the other hand, up to 5 or even 6 categories, it may be advantageous to provide full labelling, because it may help to better organize the response (e.g. Krosnick, 1999). Finally, if the designer wishes to increase the number of categories beyond 7, our recommendation is to use a continuous or visual-analogue format (Frary, 2003; García-Pérez & Alcalá-Quintana, 2023).

14. An Ounce of Prevention: Conduct a Pilot Study

The adjustments described in the previous point may change the performance of the scale, especially if they are substantial. For this reason, it may be advisable to provide some evidence that the functioning of the instrument is maintained or even improved. If the adjustments were very minor, gathering further evidence would not be indispensable. If they are not so minor but still do not involve a major change, a pilot study with a limited, representative sample to determine whether these changes make the scale more understandable and easier to use would suffice. If they were more substantial, and to demonstrate that the scale retains its structural properties would be required, the sample should be larger (see point 8 above). This is especially advisable when the modification involves changes in the statements (for example, reversing items that sound unnatural).

15. Scoring: Make the Most of the Appropriate Choice

We shall assume that the scoring-related recommendations in point 11 (see Table 3) have been followed, that the information needed for obtaining model-based scores is available, and, when

needed, that norms for interpreting the scores in the population on which the scale will be used, are also available. In these conditions, the choice of the scoring approach would mostly depend on the information the user wants to obtain from the scores and the properties of them she/he considers most relevant.

If the main interest of the application is to rank-order the respondents, and the top priorities are: computational simplicity, communicating the scoring results in an easily understandable and transparent way, relate them to those obtained in other studies, and ensure that they are stable under cross-validation, then, in our view, the simple sum scores are the most defensible choice (Sijtsma et al., 2024; Speer et al., 2016; Wainer, 1976, 1993).

Factor score estimates or IRT scores are the most appropriate choice when accurate individual measurement is required, for example for diagnostic, classification or selection purposes. In particular, well-based IRT scores, not only are generally more accurate than the remaining scoring schemas, but provide also “tailored” reliability estimates for each individual.

An appropriate reliability estimate for the chosen scores (see Table 2) should be always reported, as it will allow the user to judge the extent to which the score estimates can be trusted and the inferences from them that are warranted. In our view, however, the main use of this reliability estimate is to provide confidence intervals for each individual score estimate. So, for whatever type of score estimate that has been chosen, we encourage the practitioner to provide not only the marginal or conditional reliability estimates but also these confidence intervals.

The 15 points described so far are summarized in Table 4.

Table 4
Practical Guide for Developing and Analyzing Likert Scales

Guidelines		
A. Building a Likert scale	1	Specify clearly the type of construct that is to be measured and the target population
	2	Design the statements carefully and rigorously
	3	Choose the appropriate response format
	4	Make an informed decision about to balance or not to balance statements
	5	Including Likert-type items in multidimensional instruments can be done
	6	Choosing the most appropriate model for calibrating the items
	7	Assess the appropriateness of the chosen solution
	8	Optimal item selection: Taking the main purposes into account
	9	Scoring the Likert scale and assessing score appropriateness
	10	Provide external validity evidence
B. Using an existing Likert scale	11	Check thoroughly the features and existing information about the instrument
	12	Examine the psychometric properties of the scale
	13	Adjust and improve the scale if possible and necessary
	14	Conduct a pilot study
	15	Scoring: Make the most of the appropriate choice

Looking to the Future: Trusting in the Lindy Effect

Here we conclude our reflections and recommendations on the development and use of Likert scales in the hope that the proposed guidelines will be useful for researchers and practitioners who

develop and apply them. Now, an unavoidable further question arises: will Likert scales survive the radical changes currently taking place in the field of assessment, largely driven by the advent of new technologies? We do not know, the future, as Seneca wisely taught us, lies in uncertainty. However, Likert scales have been with us for ninety-three years, ever since Rensis Likert introduced them in his famous 1932 article. All indications suggest that they will remain with us for many more years, resilient to technological upheavals. As the Lindy effect, popularized by Nassim Taleb, predicts, the longevity of any idea or institution is positively correlated with how long it has already existed: the longer its history, the longer its future life expectancy. The term “Lindy effect” apparently derives from a New York restaurant of the same name, where actors would gather to discuss the future of their careers.

It is true that new information and communication technologies (ICT), and more recently artificial intelligence (AI), are transforming assessment and professional practice across all fields (Elosua et al., 2023; Fonseca et al., 2025; Hao et al., 2024; Santamaría & Sánchez, 2022). ICTs are opening up new forms of assessment and analysis of human behavior. Immersive virtual reality, augmented reality, telepsychology, interactive websites, adaptive testing, and smartphone applications are just a few examples. AI-driven tools to assist psychology and other professionals are becoming increasingly available, helping with administrative tasks, psychological interventions, and patient monitoring (De la Fuente & Armayones, 2025). These AI tools are welcome, but they must be approached with caution, as they are still far from perfect regarding key aspects such as explainability, veracity, generalizability, output consistency, safety, validity, reliability, fairness and equity, privacy, and copyright issues, to name just a few (Hao et al., 2024). These technologies are influencing all aspects of psychological assessment, from test design, item construction and presentation, automated item generation, to scoring and remote assessment.

While new forms of assessment are emerging, psychometric tests in general, and Likert scales in particular, will remain fundamental tools due to their objectivity, efficiency in terms of time and resources, and ease of use (Brown & Zhao, 2023). Of course, it is necessary to continue developing and consolidating a broader range of measurement methods that go beyond self-reports, thus surpassing introspective biases. Examples include multi-informant assessments, situational judgment tests, implicit association tests, neurocognitive evaluations, and computerized adaptive testing. Smartphones and other mobile devices make possible what is known as Ambulatory Assessment, which encompasses approaches such as the Experience Sampling Methodology and Ecological Momentary Assessment. These new methods allow for the evaluation of individuals’ behavior in their daily contexts and in real-time, with all the advantages that entails, representing a radical shift in how human behavior is understood, analyzed, assessed, and intervened upon. The multivariate data collected in this way demands flexible models for analysis, such as Network Models (NM), which have gained increasing attention in recent years. NMs enable alternative ways of analyzing data, modeling relationships between variables, and designing new forms of intervention. It is therefore not surprising that they have sparked growing interest in the psychological and broader scientific communities (Borgatti et al., 2009; Borsboom, 2017, 2022; Fonseca, 2018; Fonseca & Muñoz, 2025; Goyal, 2023; Newman, 2010). In parallel, another clear feature of these advances is

the integration of qualitative and quantitative approaches, enabling a deeper and more realistic understanding of human behavior through so-called mixed methods (Fonseca et al., 2025; Levitt et al., 2018).

In short, we are witnessing major advances in the field of assessment within the social and health sciences, most of them driven by the development of new technologies. However, this does not imply that the psychometric approach in general, nor Likert-type items and scales in particular, will lose relevance in measurement practices. At present, there are no more parsimonious and efficient alternatives in sight. It is difficult to imagine a future without Likert-type items and scales, true basic units of assessment: simple, direct, quick, cost-effective, and efficient. The key will be to use them properly and to combine them complementarily with other approaches. We hope that our modest contribution will help toward that goal.

Author Contributions

Pere J. Ferrando: Conceptualization, Writing - Original draft. **Fabia Morales-Vives:** Conceptualization, Writing - Original draft. **José M. Casas:** Writing - Review & Editing. **José Muñiz:** Supervision, Writing - Original draft, Writing - Review & Editing

Funding

This study did not receive any specific assistance from the public sector, the commercial sector, or non-profit organizations.

Declaration of Interests

The authors declare that there are no conflicts of interest.

Data Availability Statement

There is no data associated to this manuscript.

References

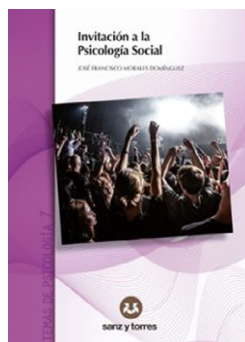
- Addams, C. (1982, November 29). Would you say Attila is doing an excellent job, a good job, a fair job, or a poor job? *The New Yorker*. <https://www.newyorker.com/magazine/1982/11/29/a-normal-tuesday>
- American Educational Research Association [AERA] (2014). *Standards for educational and psychological testing*. American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education.
- Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, 1(1), 45-87. <https://doi.org/10.1177/109442819800100104>
- Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology*, 59(3), 313-320. <https://doi.org/10.1037/h0036653>
- Benson, N., Kranzler, J. H., & Floyd, R. G. (2020). Exploratory and confirmatory factor analysis of the Universal Nonverbal Intelligence Test - Second Edition: Testing dimensionality and invariance across age, gender, race, and ethnicity. *Assessment*, 27(5), 996-1006. <https://doi.org/10.1177/1073191118786584>
- Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons. <https://doi.org/10.1002/9781118619179>
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305-314. <https://doi.org/10.1037//0033-2909.110.2.305>
- Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Sage.
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916), 892-895. <https://doi.org/10.1126/science.1165821>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5-13. <https://doi.org/10.1002/wps.20375>
- Borsboom, D. (2022). Possible futures for network psychometrics. *Psychometrika*, 87(1), 253-265. <https://doi.org/10.1007/S11336-022-09851-Z>
- Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356), 791-799. <https://doi.org/10.2307/2286841>
- Brown, G. T. L., & Zhao, A. (2023). In defense of psychometric measurement: A systematic review of contemporary self-report feedback inventories. *Educational Psychologist*, 58(3), 178-192. <https://doi.org/10.1080/00461520.2023.2208670>
- Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, 44(1), 108-132. <https://doi.org/10.1006/jmps.1999.1279>
- Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*, 39(3), 214-227. <https://doi.org/10.1037//0003-066x.39.3.214>
- Buss, D. M., & Craik, K. H. (1981). The act frequency analysis of interpersonal dispositions: Aloofness, gregariousness, dominance and submissiveness. *Journal of Personality*, 49(2), 175-192. <https://doi.org/10.1111/j.1467-6494.1981.tb00736.x>
- Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, 3(3), 106-116. <https://doi.org/10.3844/jssp.2007.106.116>
- Casas, J. M., Dueñas, J.-M., Ferrando, P. J., Castarlenas, E., Vigil-Colet, A., Hernández-Navarro, J. C., & Morales-Vives, F. (2025). Measuring the callous-unemotional traits in juvenile offenders: Properties and Functioning of the INCA Questionnaire in This Population. *Psychiatry, Psychology and Law*, 1-22. <https://doi.org/10.1080/13218719.2025.2497785>
- Casper, W. C., Edwards, B. D., Wallace, J. C., Landis, R. S., & Fife, D. A. (2020). Selecting response anchors with equal intervals for summated rating scales. *Journal of Applied Psychology*, 105(4), 390-409. <https://doi.org/10.1037/apl0000444>
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, 31(12), 1412-1427. <https://doi.org/10.1037/pas0000626>
- Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, 56(5), 754-761. <https://doi.org/10.1037//0022-006x.56.5.754>
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Psychology Press.
- Cooper, C. (2019). Pitfalls of personality theory. *Personality and Individual Differences*, 151, 109551. <https://doi.org/10.1016/j.paid.2019.109551>
- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10(1), 3-31. <https://doi.org/10.1177/001316445001000101>
- Cronbach, L. J., & Gleser, G. C. (1964). The signal/noise ratio in the comparison of reliability coefficients. *Educational and Psychological Measurement*, 24(3), 467-480. <https://doi.org/10.1177/001316446402400303>

- Cronbach, L.J., & Warrington, W. G. (1952). Efficiency of multiple-choice tests as a function of spread of item difficulties. *Psychometrika*, 17(2), 127–147. <https://doi.org/10.1007/bf02288778>
- De la Fuente, D., & Armayones, M. (2025). AI in psychological practice: what tools are available and how can they help in clinical psychology? *Psychologist Papers*, 46(1), 18–24. <https://doi.org/10.70478/pap.psicol.2025.46.03>
- DeVellis, R. F. (2003). *Scale development: Theory and applications*. Sage Publications.
- DuBois, B., & Burns, J. A. (1975). An analysis of the meaning of the question mark response category in attitude scales. *Educational and Psychological Measurement*, 35(4), 869–884. <https://doi.org/10.1177/001316447503500414>
- Elosua, P., Aguado, D., Fonseca-Pedrero, E., Abad, F. J., & Santamaría, P. (2023). New trends in digital technology-based psychological and educational assessment. *Psicothema*, 35, 50–57. <https://doi.org/10.7334/psicothema2022.241>
- Ferrando, P. J. (2021). Seven decades of factor analysis: From Yela to the present day. *Psicothema*, 33(3), 378–375. <https://doi.org/10.7334/psicothema2021.24>
- Ferrando, P. J. (2003). A Kernel density analysis of continuous typical-response scales. *Educational and Psychological Measurement*, 63(5), 809–824. <https://doi.org/10.1177/0013164403251323>
- Ferrando, P. J., Lorenzo-Seva, U., & Chico, E. (2009). A general factor-analytic procedure for assessing response bias in questionnaire measures. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(2), 364–381. <https://doi.org/10.1080/1070551090275137>
- Ferrando P. J., Lorenzo-Seva, U., Hernández-Dorado A., & Muñiz, J. (2022). Decalogue for the factor analysis of test items. *Psicothema*, 34(1), 7–17. <https://doi.org/10.7334/psicothema2021.456>
- Ferrando, P. J., & Morales-Vives, F. (2023). Is it quality, is it redundancy, or is model inadequacy? Some strategies for judging the appropriateness of high-discrimination items. *Anales de Psicología*, 39(3), 517–527. <https://doi.org/10.6018/analesps.535781>
- Ferrando, P. J., Navarro-González, D., & Lorenzo-Seva, U. (2024). A relative normed effect-size difference index for determining the number of common factors in exploratory solutions. *Educational and Psychological Measurement*, 84(4), 736–752. <https://doi.org/10.1177/00131644231196482>
- Fink, A. (2003). *The survey handbook*. Sage. <https://doi.org/10.4135/9781412986328>
- Fiske, S. T., & Taylor, S. E. (2020). Social cognition evolves: Illustrations from our work on intergroup bias and on healthy adaptation. *Psicothema*, 32(3), 291–297. <https://doi.org/10.7334/psicothema2020.197>
- Fonseca, E. (2018). Network analysis in psychology. *Papeles del Psicólogo*, 39(1), 1–12. <https://doi.org/10.23923/pap.psicol.2018.2852>
- Fonseca, E., Falcó, R., Al-Halabí, S., & Muñiz, J. (2025). Evaluación de la salud mental en contextos educativos [Mental health assessment in educational settings]. In E. Fonseca, & S. Al-Halabí (Eds.), *Salud mental en contextos educativos* (pp. 181–235). Editorial Pirámide.
- Fonseca, E., & Muñiz, J. (2025). Análisis de Redes en la Medición Psicológica: Fundamentos [Network Analysis in Psychological Measurement: Fundamentals]. *Acción Psicológica*, 22(1), 87–100. <https://doi.org/10.5944/ap.22.1.43296>
- Frary, R. B. (2003). *A brief guide to questionnaire development*. Virginia Polytechnic Institute & State University. <https://medrescon.tripod.com/questionnaire.pdf>
- Furnham, A. (1990). The development of single trait personality theories. *Personality and Individual Differences*, 11(9), 923–929. [https://doi.org/10.1016/0191-8869\(90\)90273-t](https://doi.org/10.1016/0191-8869(90)90273-t)
- García-Pérez, M. A., & Alcalá-Quintana, R. (2023). Accuracy and precision of responses to visual analog scales: Inter-and intra-individual variability. *Behavior Research Methods*, 55(8), 4369–4381. <https://doi.org/10.3758/s13428-022-02021-0>
- Goretzko, D., Pargent, F., Sust, L. N., & Bühner, M. (2019). Not very powerful. *European Journal of Psychological Assessment*, 36(4), 563–572. <https://doi.org/10.1027/1015-5759/a000539>
- Goyal, S. (2023). *Networks: An economics approach*. The MIT Press.
- Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6(4), 430–450. <https://doi.org/10.1037/1082-989X.6.4.430>
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. *Structural Equation Modeling: Present and Future*, 195(216), 60–70.
- Hao, J., von Davier, A., Yaneva, V., Lottridge, S., von Davier, M., & Harris, D. (2024). Transforming assessment: The impacts and implications of large language models and generative AI. *Educational Measurement: Issues and Practice*. 43(2), 16–29. <https://doi.org/10.1111/emip.12602>
- Henrysson, S. (1971). Gathering, analyzing and using data on test items. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 130–159). America Council on Education.
- Hernández, A., Drasgow, F., & González-Romá, V. (2004). Investigating the functioning of a middle category by means of a mixed-measurement model. *Journal of Applied Psychology*, 89(4), 687–699. <https://doi.org/10.1037/0021-9010.89.4.687>
- Hernández-Dorado, A., Ferrando, P. J., & Vigil-Colet, A. (2025). The impact and consequences of correcting for acquiescence when correlated residuals are present. *Psicothema*, 37(1), 11–20. <https://doi.org/10.70478/psicothema.2025.37.02>
- Höhne, J. K., & Krebs, D. (2018). Scale direction effects in agree/disagree and item-specific questions: A comparison of question formats. *International Journal of Social Research Methodology*, 21(1), 91–103. <https://doi.org/10.1080/13645579.2017.1325566>
- Hubatka, P., Cigler, H., Elek, D., & Tancoš, M. (2024). *The length and verbal anchors do not matter: The influence of various Likert-like response formats on scales' psychometric properties*. PsyArXiv. <https://doi.org/10.31234/osf.io/bjs2c>
- Jebb, A. T., Ng, V., & Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Frontiers in Psychology*, 12, 637547. <https://doi.org/10.3389/fpsyg.2021.637547>
- John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 461–494). The Guilford Press.
- Johnson, R. L., & Morgan, G. B. (2016). *Survey scales: A guide to development, analysis, and reporting*. Guilford Publications.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1), 537–567. <https://doi.org/10.1146/annurev.psych.50.1.537>
- Lee, J., & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment*, 32(7), 663–673. <https://doi.org/10.1177/0734282914522200>
- Lee S., Whittaker T., & Stapleton L. (2023). *GRShiny: Graded Response Model. R package version 1.0.0*. cran.r-project.org. <https://doi.org/10.32614/CRAN.package.GRShiny>

- Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA publications and communications board task force report. *American Psychologist*, 73, 26-46. <https://doi.org/10.1037/amp0000151>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55.
- Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60(1), 10-13. <https://doi.org/10.1037/h0076268>
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3(3), 635-694. <https://doi.org/10.2466/pr0.1957.3.3.635>
- Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Applied psychological measurement*, 37(6), 497-498. <https://doi.org/10.1177/0146621613487794>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. IAP.
- Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73-79. <https://doi.org/10.1027/1614-2241.4.2.73>
- Lucke, J. F. (2005). The α and the ω of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement*, 29(1), 65-81. <https://doi.org/10.1177/0146621604270882>
- Malhotra, N., Krosnick, J. A., & Thomas, R. K. (2009). Optimal design of branching questions to measure bipolar constructs. *Public Opinion Quarterly*, 73(2), 304-324. <https://doi.org/10.1093/poq/nfp023>
- Mariano, L. T., Phillips, A., Estes, K., & Kilburn, M. R. (2024). *Should survey Likert scales include neutral response categories? Evidence from a randomized school climate survey*. RAND Corporation. <https://doi.org/10.7249/WRA3135-2>
- Maydeu-Olivares, A., Fairchild, A. J., & Hall, A. G. (2017). Goodness of fit in item factor analysis: Effect of the number of response alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 495-505. <https://doi.org/10.1080/10705511.2017.1289816>
- McIver, J., & Carmines, E. G. (1981). *Unidimensional scaling*. Sage Publications. <https://doi.org/10.4135/9781412986441>
- McCrae, R. R., Costa Jr., P. T., & Piedmont, R. L. (1993). Folk concepts, natural language, and psychological constructs: The California Psychological Inventory and the five-factor model. *Journal of Personality*, 61(1), 1-26. <https://doi.org/10.1111/j.1467-6494.1993.tb00276.x>
- McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement*, 24(2), 99-114. <https://doi.org/10.1177/01466210022031552>
- Mellenbergh, G. J. (2011). *A conceptual introduction to psychometrics: Development, analysis and application of psychological and educational tests*. Eleven International Publishing.
- Morales-Vives, F., Ferrando, P. J., & Dueñas, J.-M. (2023). Should suicidal ideation be regarded as a dimension, a unipolar trait or a mixture? A model-based analysis at the score level. *Current Psychology*, 42(25), 21397-21411. <https://doi.org/10.1007/s12144-022-03224-6>
- Munshi, J. (1990). *A method for constructing Likert scales*. Sonoma State University. <http://munshi.sonoma.edu/likert.html>
- Muñiz, J. (2018). *Introducción a la psicometría* [An introduction to psychometrics]. Pirámide.
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Ten steps for test development. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Muñiz, J., García-Cueto, E., & Lozano, L. M. (2005). Item format and the psychometric properties of the Eysenck Personality Questionnaire. *Personality and Individual Differences*, 38(1), 61-69. <https://doi.org/10.1016/j.paid.2004.03.021>
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115-132. <https://doi.org/10.1007/bf02294210>
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189. <https://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.
- Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1-15. [https://doi.org/10.1016/s0001-6918\(99\)00050-5](https://doi.org/10.1016/s0001-6918(99)00050-5)
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). *Factor analysis and scale revision*. *Psychological Assessment*, 12(3), 287-297. <https://doi.org/10.1037/1040-3590.12.3.287>
- Sancerni, M. D., Meliá, J. L., & González Romá, V. (1990). Formato de respuesta, fiabilidad y validez, en la medición del conflicto de rol [Response format, reliability, and validity in the measurement of role conflict]. *Psicológica*, 11(2), 167-175.
- Santamaría, P., & Sánchez, F. (2022). Open questions in the use of new technologies in psychological assessment. *Psychological Papers*, 43(1), 48-54. <https://doi.org/10.23923/pap.psicol.2984>
- Sireci, S., & Benítez, I. (2023). Evidence for test validation: A guide for practitioners. *Psicothema*, 35(3), 217-226. <https://doi.org/10.7334/psicothema2022.477>
- Sijtsma, K., Ellis, J. L., & Borsboom, D. (2024). Recognize the value of the sum score, psychometrics' greatest accomplishment. *Psychometrika*, 89(1), 84-117. <https://doi.org/10.1007/s11336-024-09964-7>
- Sideridis, G., Tsaousis, I., & Ghamdi, H. (2023). Equidistant response options on Likert-type instruments: Testing the interval scaling assumption using Mplus. *Educational and Psychological Measurement*, 83(5), 885-906. <https://doi.org/10.1177/00131644221130482>
- Spector, P. E. (1992). *Summated rating scale construction: an introduction*. Sage Publications. <https://doi.org/10.4135/9781412986038>
- Speer, A. B., Robie, C., & Christiansen, N. D. (2016). Effects of item type and estimation method on the accuracy of estimated personality trait scores: Polytomous item response theory models versus summated scoring. *Personality and Individual Differences*, 102, 41-45. <https://doi.org/10.1016/j.paid.2016.06.058>
- Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta Izquierdo, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 2(30), 149-158. <https://doi.org/10.7334/psicothema2018.33>
- Tay, L., & Jebb, A. T. (2018). Establishing construct continua in construct validation: The process of continuum specification. *Advances in Methods and Practices in Psychological Science*, 1(3), 375-388. <https://doi.org/10.1177/2515245918775707>

- Tomás, J. M., & Oliver, A. (1998). Response format and method of estimation effects on confirmatory factor analysis. *Psicothema*, 10(1), 197-208.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. Wiley.
- Uebersax, J. S. (2006). *Likert scales: dispelling the confusion*. Statistical Methods for Rater Agreement. <https://john-uebersax.com/stat/likert.htm>
- Vigil-Colet, A., Navarro-González, D., & Morales-Vives, F. (2020). To reverse or to not reverse Likert-type items: That is the question. *Psicothema*, 32(1), 108-114. <https://doi.org/10.7334/psicothema2019.286>
- Wainer, H. (1993). Measurement problems. *Journal of Educational Measurement*, 30(1), 1-21. <https://doi.org/10.1111/j.1745-3984.1993.tb00419.x>
- Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 83(2), 213-217. <https://doi.org/10.1037//0033-2909.83.2.213>
- Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological distance between categories in the Likert scale: Comparing different numbers of options. *Educational and Psychological Measurement*, 72(4), 533-546. <https://doi.org/10.1177/0013164411431162>
- Wang, R., & Krosnick, J. A. (2019). Middle alternatives and measurement validity: A recommendation for survey researchers. *International Journal of Social Research Methodology*, 23(2), 169-184. <https://doi.org/10.1080/13645579.2019.1645384>
- Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956-972. <https://doi.org/10.1177/0013164404268674>
- Willits, F. K., Theodori, G. L., & Luloff, A. E. (2016). Another look at Likert scales. *Journal of Rural Social Sciences*, 31(3), 126-139. <https://egrove.olemiss.edu/jrss/vol31/iss3/6>
- Wrigley, J. (1976). Pitfalls in educational research. *Research Intelligence* 2(2), 2-4. <https://doi.org/10.1080/0141192760020201>

Revisión de libros



Invitación a la Psicología Social

Morales, J. F.

Sanz & Torres

Invitación a la Psicología Social, del catedrático emérito de la UNED José Francisco Morales Domínguez, es una obra rigurosa y cercana, fiel al estilo que caracteriza su trayectoria científica. Reconocido por editar manuales fundamentales de Psicología Social, Morales ofrece esta nueva publicación que, como es habitual en su producción, no deja indiferente al lector.

Académico de Número de la Real Academia de Doctores de España (Medalla 92) y Vicepresidente de la Academia de Psicología de España (Medalla 6), su prestigio se refleja tanto en su claridad expositiva como en su compromiso científico y social. Presenta una mirada actualizada de la disciplina, apoyándose en fuentes bibliográficas desconocidas incluso entre especialistas.

El libro se estructura en tres bloques temáticos y ocho capítulos, brindando un recorrido progresivo y articulado desde las bases conceptuales hasta aplicaciones concretas: visiones exteriores de la Psicología Social, procesos sociales y campo de actuación.

Capítulo 1, *La Mirada Psicosocial*, aporta una introducción epistemológica, la Psicología Social como una ciencia dinámica, anclada en una “mirada” propia que se distingue por estudiar los fenómenos sociales desde su nivel específico de análisis. Esta perspectiva, inspirada en Kurt Lewin, se convierte en hilo conductor del libro.

Se profundiza en la teoría de campo, los sistemas de tensión, la investigación-acción y la ingeniería social, destacando la figura de Lewin como pionero del activismo científico aplicado y del pensamiento interdisciplinar. Su influencia se extiende a prácticas actuales como la gestión democrática de grupos y la transformación social. Morales traza así un perfil de la Psicología Social

comprometida con el contexto y los procesos de cambio. Introduce la metáfora del “lecho de Procusto” para referirse críticamente a los intentos de subsumir los contenidos psicosociales en teorías generales, como la del intercambio o la de la racionalidad.

Capítulo 2, *Intercambio*, analiza el intercambio social a partir de las aportaciones de Homans, Thibaut, Kelley y Blau. Homans propone explicaciones basadas en recompensas y autoridad dentro del grupo, mientras Thibaut y Kelley desarrollan la teoría de la interdependencia, con una visión funcional y dinámica sobre las relaciones sociales, que se extiende al ámbito educativo a través del trabajo de Johnson y Johnson.

Blau introduce una perspectiva más estructural y económica del intercambio. Morales realiza una reflexión crítica sobre la visión economicista, incorporando el emergentismo y las críticas de Heath, que abogan por una comprensión de las motivaciones humanas. Además, se menciona el trabajo de van Lange sobre el atlas de las relaciones interpersonales.

Capítulo 3, *Psicología Social y Racionalidad*, muestra una visión compleja de la racionalidad, abordando sus distintos tipos, límites e interacción con lo cultural y evolutivo. Se examinan sesgos cognitivos, dilemas sociales y dinámicas de justicia e interdependencia.

Morales introduce términos como arrogancia y “realismus”, así como el atomismo y la falsa racionalidad. Este recorrido culmina con el análisis de la coherencia y la correspondencia, la nueva visión de la racionalidad en el enfoque cognitivo-ecológico, que rompe con los modelos economicistas.

Capítulo 4, *Identidad Social*, analiza la construcción de la identidad personal y colectiva, explorando la categorización del yo y el valor del

grupo, así como los modelos SIDE y SAMI. Se exponen conceptos como prototipo y anonimato, examinándose sus efectos sobre la conducta.

La identidad es una construcción dinámica, influenciada por la interacción social, las normas y la audiencia. Aunque con capacidad de transformación.

Capítulo 5, *La Psicología de las Relaciones entre Grupos*, parte del trabajo de Sherif y se enriquece con el papel de las diferencias entre grupos. Morales analiza el etnocentrismo, la dominancia social, la infrahumanización y la ceguera al color. Se abordan procesos de moralización, odio colectivo y alofilia, claves para entender tanto el conflicto como la posibilidad de reconciliación intergrupal.

Se defiende una visión integrada del comportamiento intergrupal, que incorpora aspectos estructurales, psicológicos y culturales (véase el modelo de Schwartz). Se proponen la identidad global interdependiente, los rituales colectivos y la empatía para transformar la percepción grupal y construir comunidades solidarias.

Capítulo 6, *La Exclusión social*, presenta un marco que vincula la exclusión social, los derechos, la exclusión moral y la grupalidad. Se analizan sus dimensiones estructurales y consecuencias sobre la salud, la participación y la ciudadanía.

A través de escalas como la IEG se evalúan actitudes ante la inclusión/exclusión. Morales sugiere aquí la necesidad de una Psicología Social comprometida con la justicia y la inclusión.

Capítulo 7, *Liderazgo y Psicología Social*, examina el fenómeno del liderazgo desde cinco pilares: emergencia, motivación, eficacia, estilos y evaluación. Se discuten estilos contrapuestos, el papel del liderazgo prototípico y el intergrupal como estrategias de cohesión.

Especial atención merece la necesidad del liderazgo femenino, sus obstáculos, estereotipos y el “precipicio de cristal”. También se

aborda el liderazgo destructivo y sus implicaciones, realizando una reflexión sobre el liderazgo ético y responsable.

Capítulo 8, *Campo de Actuación de la Psicología Social: Perfiles*, reflexiona sobre la disciplina. A partir de autores como Toulmin, Lévinas y Schachter se plantea la necesidad de una ciencia ética, plural, contextual y aplicada.

Morales subraya que la interdisciplinaridad es con frecuencia más una aspiración discursiva que una práctica real. Ahora bien, los trabajos de los Filósofos Gabriel Bello y Gustavo Bueno evidencian que una perspectiva integradora puede generar aportes significativos al desarrollo teórico y metodológico de la Psicología Social.

El lector tiene ante sí una obra compartida y vital. La dedicatoria a María Ángeles Alonso Serna ilustra esta cuestión, situándola como coautora por su implicación constante y comprometida a lo largo de todo el texto.

En suma, *Invitación a la Psicología Social* logra cumplir con creces su objetivo. Despertar el interés por la disciplina, fomentar el pensamiento crítico y facilitar la comprensión de muchos de los comportamientos sociales del Siglo XXI.

Recomendada para aquellos que se acercan por primera vez a la Psicología Social o para quienes buscan redescubrirla desde una visión actual. Y, siguiendo el propósito que inspira su título, os invito a que la estudiéis, la reviséis, la disfrutéis y aprendáis del Profesor Morales a través de cada una de las 211 páginas que conforman esta obra.

Revisado por:

Itziar Fernández

Catedrática de Psicología Social en la Universidad Nacional de Educación a Distancia (UNED)



Psicothema

Volumen 37, no. 4

PUBLISHED BY



MEMBER OF



SPONSORED BY

