

Psicothema



Volumen 38, no. 4

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
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

Managing Editor:
Gloria García-Fernández,
Univ. de Oviedo, Spain

Editorial Office:
Leticia García, COPPA,
Spain
M.ª Ángeles Gómez,
COPPA, Spain

Honorary Editor: José Muñiz, Univ. Nebrija, Spain

ADVISORY BOARD

Olivia Afonso. Oxford Brookes Univ., UK
Leandro Almeida. Univ. do Minho, Portugal
David Álvarez. Univ. de Oviedo, Spain
Rui Alves. Univ. of Porto, Portugal
Antonio M. Amor González. Univ. de Salamanca,
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
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
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
Mª Estrella Fernández Alba. 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
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
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
Christa Labouliere. Columbia University, USA
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
Verónica Marina Guillén. Univ. de Cantabria, Spain
Eduardo Martín Cabrera. Univ. de La Laguna, Spain
Víctor Martínez Loreda. Univ. de Oviedo, Spain
Ana Miranda. 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 Páino. Univ. de Oviedo, Spain
Alicia Pérez de Albéniz. Univ. de La Rioja, Spain
Celestino Rodríguez. Univ. de Oviedo, Spain
Mª Fe Rodríguez Muñoz. UNED, Spain
Sonia Romero. Univ. Aut. de Madrid, Spain
Pedro Rosário. Univ. do Minho, Portugal
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
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
Jianzhong Xu. Mississippi State Univ., USA
Jin H. Yoon. Univ. of Texas, USA

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.

The publication guidelines and instructions for authors are permanently available and kept up to date on the Psicothema website: <https://www.psicothema.com/norms>

Print ISSN (historical): 0214 - 9915

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>

Articles

- Digital Therapeutic Interventions for Suicide Prevention: A Systematic Review and Meta-Analysis of Their Effectiveness in Reducing Suicidal Ideation and Suicide Attempts
Intervenciones Terapéuticas Digitales para la Prevención del Suicidio: una Revisión Sistemática y Metaanálisis de su Eficacia en la Reducción de la Ideación Suicida y los Intentos de Suicidio
Ainize Sarrionandia, Maria Ruiz-Iriondo, Irene Izquierdo-Sanchez, Andrea Gabilondo, Craig J. Bryan, Jon García-Ormaza & Alexander Muela 174-184
- University Students' Coping Profiles and their Relationship with the Intention to Drop Out
Perfiles de Afrontamiento y su Relación con la Intención de Abandono en Estudiantes Universitarios
Ana B. Bernardo, Elena Blanco, Carlos Freire and María del Mar Ferradás 185-194
- PROVAL: A Comprehensive Protocol for Developing Validation Studies
PROVAL: un Protocolo Integral para el Desarrollo de Estudios de Validación
Isabel Benítez, Andrés González-Gómez and José Luis Padilla 195-203
- Psychometric Properties of the Negative and Positive Affect Scales (NAPAS) in Patients with Breast Cancer
Propiedades Psicométricas de las Escalas de Afecto Negativo y Positivo (NAPAS) en Pacientes con Cáncer de Mama
M. Victoria Cerezo, Lucía Sánchez-Espíldora, Rafael Alarcón, Lorena M. Soria-Reyes and María J. Blanca 204-212
- Psychometric Evaluation and Normative Framework for the Movie for the Assessment of Social Cognition in Spanish-Speaking Adults
Propiedades psicométricas del Montreal Cognitive Assessment (MoCA) en adultos españoles
Aina Llompart-Casasnovas, Adoración Castro, Margalida Gili, Guillem Navarra-Ventura, Roberto Rodríguez-Jimenez, Miguel A. Ortega, Leticia Boada, Mercè Jodar, Narcís Cardoner, Albert Sesé, Miquel Roca, and Guillermo Lahera..... 213-223

Psicothema

Volumen 38, no. 4

PUBLISHED BY



MEMBER OF



SPONSORED BY



Article

Digital Therapeutic Interventions for Suicide Prevention: A Systematic Review and Meta-Analysis of Their Effectiveness in Reducing Suicidal Ideation and Suicide Attempts

Ainize Sarrionandia¹, Maria Ruiz-Iriondo², Irene Izquierdo-Sanchez¹, Andrea Gabilondo^{2,3}, Craig J. Bryan⁴, Jon García-Ormaza^{5,6} and Alexander Muela^{1,2}

¹ University of the Basque Country UPV/EHU (Spain)

² Biogipuzkoa Health Research Centre (Spain)

³ Gipuzkoa OMH Network, Osakidetza (Spain)

⁴ University of Vermont (USA)

⁵ Biobizkaia Health Research Institute (Spain)

⁶ Osakidetza Basque Health Service, Zamudio Hospital (Spain)

ARTICLE INFO

Received: 21/02/2026

Accepted: 23/03/2026

Keywords:

Suicide
Digital intervention
Meta-analysis
Prevention
Mental health

ABSTRACT

Background: Although digital interventions show potential for reducing suicidal ideation and behavior, the evidence base is mixed and characterized by methodological variability. This study synthesizes randomized clinical trials that have evaluated their effectiveness. **Method:** We conducted a systematic review and meta-analysis in accordance with PRISMA guidelines. Seventeen studies ($N = 5,089$) met the inclusion criteria. Effect sizes were estimated using random-effects models, and we conducted analyses for type of control group (traditional vs. active), moderators (age, delivery setting, and intervention format), and follow-up. Heterogeneity and publication bias were also examined. **Results:** Digital interventions significantly reduced suicidal ideation ($g = -0.61$), with effects maintained at follow up ($g = -0.77$). Significant effects were observed with respect to both traditional and active controls. There was a moderate reduction in relative risk of suicide attempts, although it was not statistically significant. Moderator analysis indicated a greater effect for clinical populations, self-guided interventions, and younger participants. There was no robust evidence of publication bias. **Conclusions:** Digital interventions may represent a promising complementary approach for reducing suicidal ideation and potentially lowering suicide risk. However, the findings should be interpreted cautiously given the heterogeneity of the studies and the limited evidence regarding suicide attempts.

Intervenciones Terapéuticas Digitales para la Prevención del Suicidio: una Revisión Sistemática y Metaanálisis de su Eficacia en la Reducción de la Ideación Suicida y los Intentos de Suicidio

RESUMEN

Antecedentes: Las intervenciones digitales han mostrado potencial para reducir la ideación y la conducta suicida, pero la evidencia disponible es heterogénea. Este estudio sintetiza los hallazgos de ensayos clínicos aleatorizados que han evaluado su eficacia. **Método:** Se realizó una revisión sistemática y metaanálisis siguiendo las directrices PRISMA. Se incluyeron 17 estudios ($N = 5,089$). Los tamaños del efecto se estimaron mediante modelos de efectos aleatorios y se realizaron análisis por tipo de grupo control, moderadores (edad, contexto y modalidad), seguimiento, heterogeneidad y sesgo de publicación. **Resultados:** Las intervenciones digitales redujeron significativamente la ideación suicida ($g = -0.61$), manteniéndose los efectos en el seguimiento ($g = -0.77$). Se observaron efectos significativos frente a controles tradicionales y activos. En los intentos de suicidio se observó una reducción del riesgo de magnitud moderada, aunque sin significación estadística. Los efectos fueron mayores en muestras clínicas, intervenciones autoguiadas y participantes jóvenes. No se detectó sesgo de publicación robusto. **Conclusiones:** Las intervenciones digitales pueden constituir una estrategia complementaria prometedora para reducir la ideación suicida y potencialmente disminuir el riesgo suicida. No obstante, los resultados deben interpretarse con cautela debido a la heterogeneidad de los estudios y a la limitada evidencia disponible sobre los intentos de suicidio.

Palabras clave:

Suicidio
Intervenciones digitales
Metaanálisis
Prevención
Salud mental

Suicide is a major public health problem worldwide. The [World Health Organization \(2021\)](#) estimates that more than 700,000 people die by suicide each year, making it one of the leading causes of death. Suicidal behavior is currently understood as a complex and multifactorial phenomenon that must be interpreted within individuals' biographical, social, and cultural contexts, as well as broader social determinants shaping trajectories of suicide risk across the life course ([Al-Halabi & Fonseca-Pedrero, 2024](#); [Hawton & Pirkis, 2024](#)). From a psychological perspective, this view is broadly consistent with the Fluid Vulnerability Theory of suicide, which proposes that suicide risk emerges from the interaction between relatively stable vulnerabilities and dynamic suicidal states that fluctuate over time ([Rudd, 2006](#)). In this regard, it is important to distinguish between suicidal ideation, which refers to thoughts about death or killing oneself, and suicidal behavior, which includes suicide attempts. Although these phenomena are closely related, they are not interchangeable and may involve partly distinct mechanisms and intervention targets ([Al-Halabi & Fonseca-Pedrero, 2024](#)).

Despite advances in identifying risk and protective factors and in developing effective psychological interventions ([Fox et al., 2020](#); [Mann et al., 2021](#)), important barriers to accessing care remain ([Bryan et al., 2025](#)). Many individuals experiencing suicidal ideation or behaviors do not seek or receive timely help due to factors such as self-reliance, low perceived need for treatment, and stigma toward mental health and help-seeking ([Han et al., 2017](#)). These barriers highlight the need for innovative approaches that complement traditional services and reach individuals who are reluctant or unable to engage in face-to-face care.

In this context, digital interventions have emerged as a promising alternative ([Yosep et al., 2024](#)). Mobile applications, online programs, self-guided platforms, telehealth, and blended models of care offer several advantages, including immediate access, the possibility of anonymity, scalability, and potentially reduced costs ([Bryan et al., 2025](#)).

Evidence from recent systematic reviews and meta-analyses shows mixed findings regarding the impact of digital interventions on suicidal ideation, with generally small but significant effects and persistent limitations related to adherence and maintenance over time ([Burr et al., 2025](#); [Büscher et al., 2020, 2022](#); [Josifovski et al., 2025](#); [Torok et al., 2020](#)). In one of the earliest meta-analyses, [Torok et al. \(2020\)](#) reported small but significant reductions in suicidal ideation, particularly for interventions directly targeting suicidality. However, subsequent reviews have highlighted that effects vary depending on factors such as comparison group and participant age, with weaker or null findings in adolescents and when compared with active controls ([Burr et al., 2025](#)).

From a related perspective, some studies have examined telehealth interventions that include synchronous human contact. A meta-analysis of telephone- or video-based follow-up interventions found reductions in suicide rates and re-attempts, along with high acceptability, particularly among patients with depressive symptoms ([Shoib et al., 2024](#)). However, the substantial heterogeneity and the specific focus on telehealth limit the generalizability of these findings to other types of digital interventions.

A recent systematic review and meta-analysis of digital interventions targeting suicidality ([Josifovski et al., 2025](#)) concluded that these approaches produce small but significant reductions in suicidal ideation in clinical and community settings, but not consistently in educational contexts. Evidence for effects on suicidal

behaviors remained inconclusive. The authors also reported minimal differences between direct and indirect interventions and slightly stronger effects for self-guided formats in community samples, highlighting the influence of delivery setting and format on outcomes.

Complementing this literature, reviews focusing on specific therapeutic approaches have reported consistent positive effects on suicidal ideation, although evidence for reductions in suicide attempts remains inconclusive ([Büscher et al., 2020, 2022](#)). Similarly, a recent meta-analysis of digital psychotherapy based mainly on CBT and DBT identified moderate effects on suicidal ideation and depressive symptoms, albeit with considerable heterogeneity and a limited number of studies ([Oh et al., 2024](#)).

Overall, prior research suggests that digital interventions have meaningful potential to reduce suicidal ideation, although effect sizes are typically small to moderate. Outcomes appear to vary according to participant characteristics and delivery context, with less consistent findings in adolescents and educational settings. The literature is also marked by substantial methodological heterogeneity and limited evidence regarding suicide attempts. In addition, persistent concerns remain about participant retention, maintenance of effects, and the influence of comparison group type on observed outcomes.

In light of these limitations, the present study aimed to update and extend current knowledge regarding the effectiveness of digital interventions in reducing suicidal ideation and suicide attempts. We conducted a systematic review and meta-analysis of randomized clinical trials published between 2014 and 2025, covering recent technological developments and the post-pandemic period. Specifically, we examined (a) overall intervention effectiveness; (b) differences according to type of comparison group (traditional vs. active); (c) maintenance of effects over time; and (d) the moderating role of participant age, delivery setting, and intervention format. Between-study heterogeneity and potential publication bias were systematically assessed.

Method

This study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines ([Page et al., 2021](#)). The study protocol was registered in PROSPERO (International Prospective Register of Systematic Reviews; registration number: CRD420251266844).

Search Strategy and Inclusion Criteria

A systematic search was carried out in four electronic databases: Web of Science, PubMed, PsycINFO, and Scopus. The search strategy used different combinations of the following MeSh terms: “suicide prevention”, “suicidal attempts”, “suicidal behaviors”, “suicidal ideation”, “suicidality”, “self-harm”, “non-suicidal self-injuries”, “hopelessness”, “thwarted belongingness”, “perceived burdensomeness” AND/OR “telemedicine”, “telepsychiatry”, “telehealth”, “digital interventions”, “combined treatment”, “internet-based interventions (IBIs)”, “ICBT”, “IDBT”, AND/OR “systematic review”, “meta-analysis”.

The inclusion criteria were: (a) randomized clinical trials evaluating digital interventions aimed specifically at reducing suicidal ideation or suicide attempts; (b) reports published in English or Spanish; and (c) publication between January 2014 and September 2025.

The following exclusion criteria were applied: (a) studies with non-experimental, observational or qualitative designs; (b) studies that did not report results related to suicidal ideation or suicide attempts; and (c) protocols, narrative reviews or any other type of study reporting non-empirical data.

Extraction and Coding of Data

Data were extracted using a standardized matrix comprising: (a) bibliographic and contextual information—authors, year of publication, and country where the research was conducted; (b) characteristics of the digital intervention—format (web, mobile app, hybrid program), duration, therapeutic content, and type of comparison group; (c) outcome variables and instruments used to evaluate them; (d) sample characteristics—size, delivery setting, age, gender, and details of follow-up assessments; and (e) main results reported.

For this meta-analysis we distinguished between two types of comparison group: (a) traditional (i.e., no treatment or waitlist controls); and (b) active comparators, where participants received an alternative intervention that had been previously evaluated and considered to be effective.

Data Analysis

Effect sizes were calculated based on the available data reported in each study (pre–post means and standard deviations, differences in means, comparative statistics, and—when possible—pre–post correlations). Suicidal ideation effect sizes were based on pre–post changes within each arm, whereas suicide attempts were analyzed using ratio-based comparisons between intervention and control groups. For suicidal ideation, we calculated standardized mean differences using Hedges' g so as to correct for small sample bias. For suicide attempts, we used ratio-based effect sizes (odds ratios or incidence rate ratios), transformed onto a logarithmic scale for the purposes of analysis.

Meta-analyses were estimated with random-effects models and the restricted maximum likelihood (REML) estimator, using the Comprehensive Meta-Analysis software (version 4; Borenstein et al., 2022) and the metafor package (Viechtbauer, 2010) in R (version 4.5.2). For each outcome we calculated the overall effect size with 95% confidence intervals, as well as Cochran's Q , P , and τ^2 to estimate heterogeneity. Heterogeneity was interpreted in accordance with the criteria of Higgins et al. (2003), such that values of P close to 25%, 50%, and 75% were considered to indicate low, moderate, and high heterogeneity, respectively.

For studies which did not report pre–post correlations, we imputed a value of 0.50, based on evidence suggesting high temporal stability in psychological scales used in clinical interventions (Balk et al., 2012; Cuijpers et al., 2017). This imputation, consistent with the recommended range in previous meta-analyses (0.50–0.70), accounts for the dependence between repeated measures when estimating the standard error (Hesser et al., 2011; Hofmann et al., 2012). To examine the robustness of this assumption, sensitivity analyses were conducted using an alternative correlation value ($r = 0.70$), yielding comparable results.

Possible publication bias was assessed visually using funnel plots and statistically with Egger's regression test (Egger et al., 1997),

with an intercept significantly different from zero being considered indicative of asymmetry.

For suicide attempts, we only included those studies that reported dichotomous measures or rates (e.g., total number of attempts, proportion of participants with at least one attempt or rates of attempts) and which provided sufficient information to estimate odds ratios with their standard errors. Following logarithmic transformation of effect sizes, we applied random-effects models, weighting each study by the inverse variance. Given the small number of studies and the limited variability across samples, we did not perform sub-group analyses for this outcome.

Methodological Quality of Included Studies

The methodological quality and risk of individual bias in the studies reviewed were assessed using Version 2 of the Cochrane Risk-of-Bias Tool for Randomized Trials (RoB 2; Sterne et al., 2019). This tool comprises five domains: (1) bias arising from the randomization process; (2) bias due to deviation from intended interventions; (3) bias due to missing outcome data; (4) bias in measurement of the outcome; and (5) bias in selection of the reported result. For each domain there are a series of signaling questions, the answers to which allow a judgment to be made as to the corresponding risk of bias for the domain in question, classified as low, some concerns or high risk. The overall risk of bias for a study is determined based on the judgments made across the five domains.

Results

Study Selection

Figure 1 shows the PRISMA flow diagram for study selection. A total of 370 records were identified through database searches and 7 through other sources. After removing duplicates, 255 records were screened and 96 full texts assessed for eligibility. When necessary, authors were contacted for missing data. Finally, 17 randomized clinical trials published between 2014 and 2025 were included in the review and meta-analysis.

Descriptive Characteristics of the Studies

Table 1 summarizes the main characteristics of the included studies.

Evaluation of the Risk of Bias

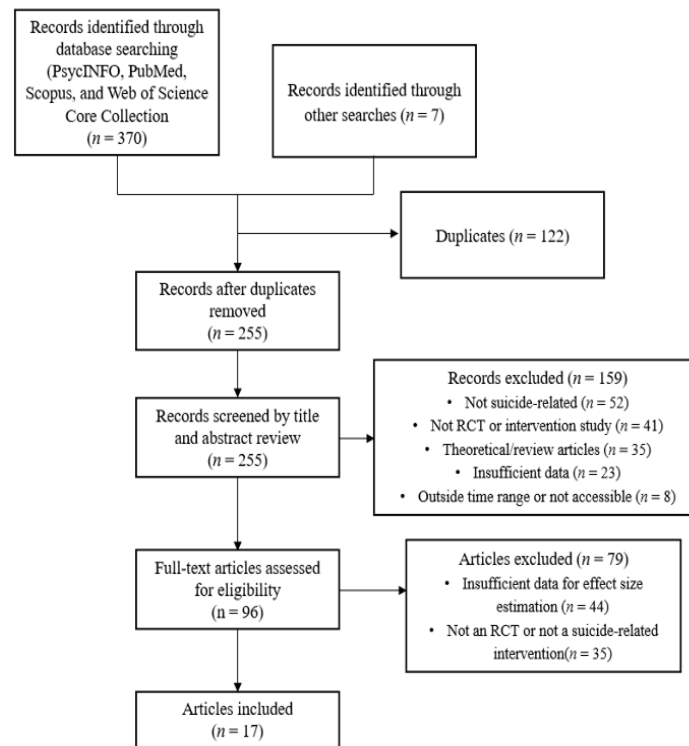
Based on the RoB 2 assessment (Sterne et al., 2019), most studies showed moderate or high risk of bias, with only a few trials judged to have low risk across all domains. Although randomization procedures were generally adequate, limitations included lack of masking, follow-up difficulties, high attrition rates, and reliance on self-reported outcomes. Overall, the findings should therefore be interpreted with caution.

Effect of Digital Interventions on Suicidal Ideation

The meta-analysis indicated a significant effect of digital interventions on suicidal ideation ($g = -0.61$, 95% CI $[-0.81, -0.40]$).

Figure 1

Flow Diagram of the Search and Selection Process

**Table 1**

Main Characteristics of Included Studies

Study	Country	n	Population	Age group	Format	Guidance	Comparison	Outcomes
Baker et al. (2024)	USA	96	Clinical	Adults	Web	Guided	Active	SI, SA
Batterham et al. (2018)	Australia	418	Community	Adults	Web	Self	Active/Traditional	SI
Batterham et al. (2021)	Australia	1986	Community	Adults	Web	Self	Active	SI
Bryan et al. (2025)	USA	339	Clinical	Adults	App	Self	Active	SI, SA
Burback et al. (2024)	Canada	42	Clinical	Adults	Web	Guided	Active	SI, SA
Bush et al. (2016)	USA	118	Clinical	Adults	App	Guided	Active	SI
De Jaegere et al. (2019)	Belgium	724	Clinical	Adults	Web	Self	Traditional	SI, SA
Eylem et al. (2021)	Netherlands	18	Clinical	Adults	Web	Guided	Traditional	SI, SA
Hetrick et al. (2017)	Australia	50	School	Adolescents	Web	Self	Active	SI, SA
Hill & Pettit (2019)	USA	80	Clinical	Adolescents	Web	Guided	Active	SI, SA
Kaskow et al. (2015)	USA	25	Clinical	Adults	Web	Guided	Active	SI
Li et al. (2019)	China	300	Clinical	Adults	Web	Self	Active	SI/SA
Lohani et al. (2024)	USA	82	Community	Adults	Web	Guided	Active	SI
Tighe et al. (2017)	Australia	61	Clinical	Adults	App	Guided	Traditional	SI
Torok et al. (2022)	Australia	455	Community	Young adults	App	Self	Active	SI, SA
van Spijker et al. (2014)	Netherlands	236	Community	Adults	Web	Self	Traditional	SI, SA
Wilks et al. (2018)	USA	59	Clinical	Adults	Web	Self	Traditional	SI

Notes. SI = suicidal ideation; SA = suicide attempts; Self = self-guided; Guided = intervention with human support; Traditional = waitlist/no treatment; Active = active comparator.

This effect size, of moderate magnitude, was the largest observed among the different clinical domains analyzed. Figure 2 displays the corresponding forest plot, showing the individual effect sizes, along with the 95% confidence intervals. Baker et al. (2024) and Li et al. (2019) assessed suicidal ideation but did not report sufficient data to calculate an independent effect size for this outcome and were therefore excluded from the quantitative synthesis.

Meta-analyses were estimated using a random-effects model, which is appropriate when true effects are expected to vary across studies. The analysis showed substantial heterogeneity ($Q(15) = 173.14, p < .001; I^2 = 91.35\%$). Despite this heterogeneity, the analysis of publication bias, using Egger's regression test, did not indicate significant asymmetry ($t(14) = 0.59, p = .56$), which suggests a low probability of bias in the distribution of effect sizes (Figure 3).

When studies were stratified by type of control condition, the magnitude of the pre-post improvement observed in the intervention groups remained significant both in studies including active comparators ($g = -0.44$; 95% CI $[-0.66, -0.22]$) and in those using traditional control conditions ($g = -0.34$; 95% CI $[-0.49, -0.18]$). Heterogeneity across studies was high in both cases (active controls: $Q(11) = 102.09, p < .01, I^2 = 83.85\%$; traditional controls: $Q(5) = 13.19, p = .02, I^2 = 62.10\%$), justifying the use of a random-effects model. Egger's regression test indicated no significant evidence of publication bias (active comparators: $t(10) = 1.74, p = .11$; traditional controls: $t(4) = 0.90, p = .41$), suggesting an absence of systematic asymmetry in estimates.

Follow-up analyses showed that improvements in suicidal ideation were maintained in the medium term in experimental

Figure 2

Forest Plot of Suicidal Ideation in The Experimental Group: Pretest-Posttest

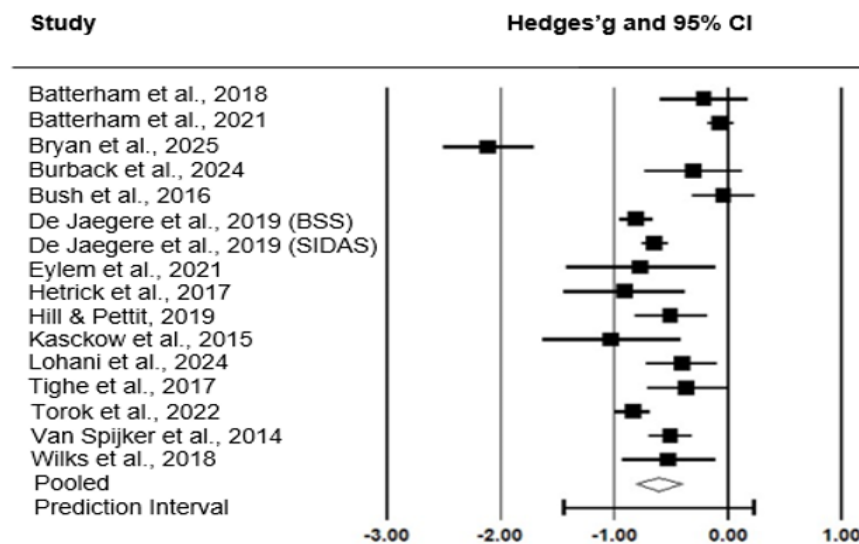
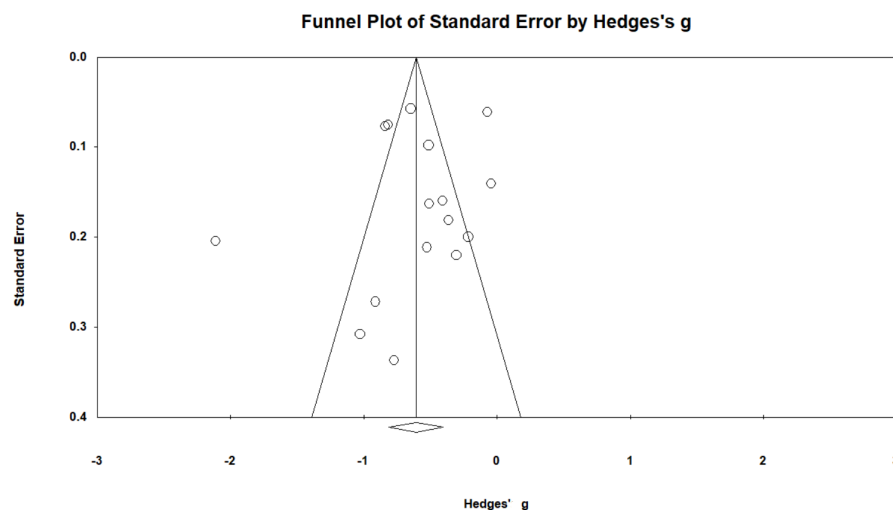


Figure 3

Funnel Plot of Suicidal Ideation in The Experimental Group: Pretest-Posttest



groups ($g = -0.75$; 95% CI $[-0.95, -0.55]$), and also in both traditional controls ($g = -0.57$, 95% CI $[-0.81, -0.34]$) and active comparators ($g = -0.49$, 95% CI $[-0.68, -0.29]$). Heterogeneity across studies was significant in all three contrasts (experimental group: $Q(9) = 54.00$, $p < .01$, $I^2 = 83.33\%$; traditional controls: $Q(3) = 17.24$, $p < .01$, $I^2 = 82.0\%$; active comparators: $Q(6) = 13.47$, $p = .01$, $I^2 = 55.45\%$), indicating moderate to high between-study variability, possibly associated with differences between interventions or in sample characteristics. Egger's regression test showed no evidence of publication bias in any of the analyses (experimental group: $t(8) = 0.61$, $p = .56$; traditional controls: $t(2) = 0.60$, $p = .60$; active comparators: $t(5) = 0.39$, $p = .71$).

Suicidal Ideation: Analysis of Moderator Variables

Digital interventions had an effect of moderate magnitude on suicidal ideation in samples comprising both adolescents/youth ($g = -0.65$, 95% CI $[-0.91, -0.39]$) and adults ($g = -0.59$, 95% CI $[-0.85, -0.34]$). There was significant heterogeneity across studies involving adolescents and youth ($Q(3) = 8.42$, $p = .04$; $I^2 = 64.36\%$), indicating moderate between-study variability. Heterogeneity was also high for studies in adult populations ($Q(11) = 153.26$, $p < .01$; $I^2 = 92.82\%$), suggesting substantial variability that was possibly related to differences in the type of intervention, clinical severity or the duration of follow up. Egger's regression test indicated no evidence of publication bias in either of the two contrasts (adolescents/youth: $t(2) = 0.96$, $p = .43$; adults: $t(10) = 0.66$, $p = .51$).

Effects were also greater in clinical samples ($g = -0.71$, 95% CI $[-1.01, -0.42]$) than in the general population ($g = -0.43$, 95% CI $[-0.73, -0.12]$). Heterogeneity was significant in both cases (clinical samples: $Q(8) = 81.22$, $p < .01$; $I^2 = 90.15\%$; general population: $Q(5) = 64.46$, $p < .01$; $I^2 = 92.24\%$), which suggests that the magnitude of effects may be influenced by differences in clinical severity and the content and/or format of interventions. There was no evidence of publication bias (clinical samples: $t(7) = 0.17$, $p = .87$; general population: $t(4) = 0.38$, $p = .72$).

Finally, fully self-guided interventions showed a greater effect ($g = -0.71$, 95% CI $[-0.99, -0.43]$) than did mixed interventions that included human support ($g = -0.41$, 95% CI $[-0.62, -0.20]$). Heterogeneity was high for studies involving self-guided interventions ($Q(8) = 153.83$, $p < .01$; $I^2 = 94.80\%$), and moderate for those describing mixed interventions ($Q(6) = 12.29$, $p = .05$; $I^2 = 51.19\%$). Egger's regression test indicated no evidence of publication bias in studies reporting self-guided interventions ($t(7) = 0.95$, $p = .37$), but there was possible asymmetry in the case of mixed

interventions ($t(5) = 2.57$, $p = .05$), suggesting marginal bias and the need to interpret the results with caution.

Effect of Digital Interventions on Suicide Attempts

The meta-analysis included five studies that assessed the effectiveness of digital interventions in preventing suicide attempts (Table 2). Sample sizes ranged between 42 and 724 participants, and interventions involved both self-guided applications and structured CBT programs in digital format.

The random-effects model, with the restricted maximum likelihood estimator, showed a pooled effect in the expected direction, indicating a reduction in the risk of suicide attempts in groups that received the digital intervention, in comparison with controls (Figure 4). The pooled estimate suggested a reduction in the risk of suicide attempts (ratio = 0.65; 95% CI $[0.40, 1.04]$; $p = .07$), corresponding to an approximate 35% lower likelihood among intervention participants, although the result did not reach statistical significance. Heterogeneity was moderate ($I^2 = 35\%$) and the test of heterogeneity was not significant ($Q(4) = 6.06$, $p = .19$), indicating acceptable between-study variability.

In contrast to the results obtained for suicidal ideation, it was not possible to conduct a stratified analysis for suicide attempts. This limitation is due primarily to the small number of studies that assessed this outcome (only five studies provided sufficient data), reducing substantially the statistical power for sub-group comparisons. In addition, all but one of these studies were focused exclusively on a clinical population, preventing us from examining possible differences between clinical and non-clinical contexts. Furthermore, as four of the five studies (the exception being the study by De Jaegere et al., 2019) used an active comparator, we could not compare effectiveness with respect to traditional (no treatment or waitlist) controls. Neither was it possible to explore differences by age group, as none of the five studies focused specifically on adolescents or youth. Finally, the intervention formats were not sufficiently diverse or balanced to allow for a reliable comparison between self-guided and mixed (self-guided with human support) interventions.

Discussion

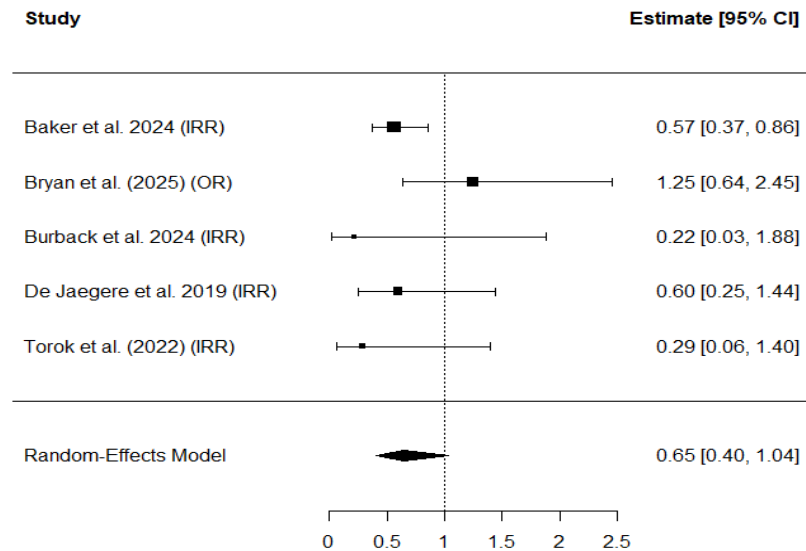
Overall, results of this meta-analysis suggest that digital interventions may be effective in reducing suicidal ideation and may, to a lesser extent, help to reduce the risk of suicide attempts. Accordingly, digital technologies may be promising tools within the spectrum of suicide prevention strategies, especially in a global context characterized by persistent barriers to accessing mental health services (Bryan et al.,

Table 2
Characteristics and Effect Sizes of Included Studies Examining the Effect of Digital Interventions On Suicide Attempts

Study	Experimental <i>n</i>	Control <i>n</i>	Metric	Effect size (IRR/OR)	95% CI	Weight (%)
Baker et al. (2024)	51	45	IRR	0.57	0.37 – 0.86	41
Bryan et al. (2025)	127	139	OR	1.25	0.64 – 2.45	27
Burback et al. (2024)	20	22	IRR	0.22	0.03 – 1.88	4
De Jaegere et al. (2019)	365	359	IRR	0.60	0.25 – 1.44	19
Torok et al. (2022)	228	232	IRR	0.29	0.06 – 1.40	8

Pooled effect: 0.65 $[0.40 - 1.04]$; $p = .07$; $I^2 = 35\%$.

Notes. IRR = Incidence Rate Ratio; OR = Odds Ratio; CI = Confidence Interval; I^2 = percentage of between-study heterogeneity.

Figure 4*Forest Plot of the Random-Effects Meta-Analysis Examining the Effect of Digital Interventions On Suicide Attempts*

Notes. Each square represents the study-specific effect size (rate ratio or odds ratio) and its 95% confidence interval. The diamond at the bottom represents the pooled estimate under a random-effects model (REML). Values below 1 indicate a lower odds/rate of suicide attempts in the digital intervention group.

2025) and the urgent need to develop scalable, accessible, and evidence-based therapeutic modalities (Büscher et al., 2022).

The most consistent finding relates to a reduction in suicidal ideation, with moderate pooled effect sizes being observed in studies that compare digital interventions with baseline. Effectiveness was maintained in both the overall analysis and when conducting subgroup analyses by participants' age, the delivery setting, and the intervention format. However, these moderator analyses should be considered exploratory given the small number of studies in some subgroups, and the larger effects observed in self-guided interventions should therefore be interpreted cautiously. Variability in the level of therapeutic guidance may also be relevant, as alliance-related processes have been shown to influence outcomes in both traditional and digital psychotherapy (Berger, 2017; Flückiger et al., 2018). This finding is in line with previous reviews that identified significant—albeit modest—effects on suicidal ideation (Burr et al., 2025; Büscher et al., 2020; Oh et al., 2024; Sutori et al., 2024; Torok et al., 2020), although it brings clarity to the conditions under which digital interventions appear to be most effective. Although the effect size observed in the present review appears somewhat larger than those reported previously, this comparison should be interpreted cautiously given methodological differences across studies and the substantial heterogeneity observed. From a clinical perspective, even moderate reductions in suicidal ideation may be meaningful, as ideation is a proximal risk factor for suicide attempts (Jobes et al., 2024). In addition to symptom reduction, some digital interventions may also strengthen protective processes such as reasons for living, hope, or sources of meaning and connectedness (Bryan & Bozzay, 2026; Rain et al., 2024). Particularly noteworthy is the temporal stability of the observed effects on suicidal ideation. The analyses of follow-up data showed that post-intervention reductions in suicidal ideation were not only maintained over time but tended, in the experimental group, to increase in the medium term. This

pattern suggests that digital interventions may help to promote processes of sustained change, beyond any immediate effect in terms of symptom relief. From a clinical and preventive perspective, the persistence of effects is especially relevant insofar as suicidal ideation is a dynamic and fluctuating risk factor (Jobes et al., 2024), and its sustained reduction may help to decrease the likelihood of a transition to more serious suicidal behaviors. This interpretation is consistent with the Fluid Vulnerability Theory of suicide (Rudd, 2006), which conceptualizes suicide risk as the interaction between relatively stable vulnerabilities and fluctuating suicidal states that may be modified through targeted interventions.

An important contribution of the present review relates to the comparison of results according to the type of control group used in the pre-test vs. post-test analyses. Although both active comparators and traditional control groups showed an improvement in suicidal ideation, the magnitude of change followed a descending pattern, with the largest effect sizes being observed in experimental groups ($g = -0.61$), followed by active comparators ($g = -0.44$) and traditional controls ($g = -0.34$). This pattern is consistent with the literature on clinical trials in mental health, which show that treatment effects tend to be smaller when interventions are compared with active comparators, as the latter can also lead to important clinical improvements (Cuijpers et al., 2024; Patterson et al., 2016). Although active comparators varied in their specific content and intensity, they were grouped together because all represented interventions intended to provide some form of therapeutic input beyond no-treatment or waitlist conditions. From a clinical perspective, these results suggest that the therapeutic potential of digital interventions extends beyond their consideration as low-threshold-resources.

The analysis by sub-groups adds further nuance to our findings. Digital interventions appeared to be more effective in adolescents and youth than in adults, although both groups showed significant benefits.

This finding is relevant given that suicidal ideation is one of the main risk factors for suicide behavior during adolescence (Tabares et al., 2025). It should be noted that adolescents and youth are generally adept in using digital technologies and may therefore experience fewer barriers to engaging with interventions that use this medium (Forte et al., 2021). From the public health perspective, the possibility of implementing scalable digital programs aimed specifically at young people represents a strategic opportunity for early and universal prevention (Potts et al., 2025; Wanniarachchi et al., 2025).

Our analysis also identified differences depending on the delivery setting, with larger effect sizes being observed in clinical samples ($g = -0.71$) than in the general population ($g = -0.43$). This pattern is in line with the meta-analysis conducted by Josifovski et al. (2025), although the respective effect sizes they found were more modest. The present review therefore extends the previous literature by reporting larger effect sizes, especially in clinical samples, which may reflect both a greater activation of mechanisms of change in populations with clinical symptoms and the application of more specific digital interventions in recent studies.

Another important finding of the present review relates to the intervention format. Self-guided interventions were associated with larger pretest–post-test effect sizes than were mixed interventions that included human support, which raises questions as to the mechanisms of change involved. One possible explanation is that the self-guided format offers greater anonymity, flexibility, and autonomy, which may foster therapeutic engagement, whereas mixed interventions may be characterized by greater variability in their implementation. However, this finding should be interpreted with caution given the marginal evidence of publication bias in studies involving mixed interventions. Furthermore, and in line with a point made by Sutori et al. (2024), these results do not mean that self-guided interventions are sufficient as a sole treatment for people with suicidal ideation; rather they should be considered as useful components within a broader framework of care.

In contrast to the robust results obtained for suicidal ideation, the evidence relating to suicide attempts remains limited. Although the five studies reviewed consistently pointed towards a protective effect of digital interventions, this did not reach statistical significance, a finding in line with previous broad scope reviews (Büscher et al., 2020, 2022; Josifovski et al., 2025; Sutori et al., 2024; Witt et al., 2017). Given the small number of studies and the low base rate of suicide attempts in clinical trials, the statistical power of this analysis is limited and the results should therefore be interpreted cautiously. These findings suggest that the current evidence is inconclusive rather than indicative of the absence of an effect, and highlights the need for larger trials specifically designed to examine suicide attempts as an outcome. This pattern is consistent with the greater methodological and clinical complexity associated with an infrequent outcome, one characterized by a low base rate and strong contextual influences. In this respect, while the available evidence suggests that digital interventions (especially when self-guided) may be capable of reducing suicidal ideation, the prevention of suicide attempts likely requires more intense interventions that include more specific face-to-face input from one or more professionals.

A relevant benchmark for interpreting our findings is the large meta-analysis by van Ballegooijen et al. (2025), which synthesized the effects of direct and indirect psychotherapies on suicidal ideation and suicide attempts, and reported treatment effects by specific therapeutic

protocols. In their overall estimates, direct psychotherapy showed a pooled effect of $g = -0.39$ for suicidal ideation and $RR = 0.72$ for suicide attempts, while indirect psychotherapy showed $g = -0.30$ for suicidal ideation and $RR = 0.68$ for suicide attempts.

Against this backdrop, the effect of digital interventions observed in our meta-analysis appears clinically meaningful. Specifically, we found a moderate reduction in suicidal ideation ($g = -0.61$), which is comparable to—or slightly larger than—the overall effects reported for direct psychotherapy in face-to-face settings. Moreover, subgroup estimates from van Ballegooijen et al. (2025) indicate that several evidence-based protocols can achieve similar magnitudes of change in ideation, including CBT ($g = -0.51$) and CAMS ($g = -0.46$), while DBT showed a more uncertain estimate for ideation outcomes.

Regarding suicide attempts, van Ballegooijen et al. (2025) reported a pooled RR of 0.65 for CBT and $RR = 0.49$ for DBT, with direct psychotherapy overall showing $RR = 0.72$. In our meta-analysis, the pooled effect for suicide attempts ($RR = 0.65$) was highly similar in magnitude to CBT in traditional psychotherapy, but did not reach statistical significance, likely reflecting the smaller number of included trials, limited statistical power, and the low base rate of suicide attempts in most samples. Taken together, these comparisons suggest that digital interventions may provide effects on suicidal ideation that are broadly comparable to established face-to-face modalities, while evidence for effects on attempts remains promising but less precise and warrants further high-powered trials with longer follow-up and sufficient event rates. Digital interventions should therefore not be interpreted as substitutes for established psychotherapeutic approaches, but rather as accessible and scalable tools that can complement evidence-based treatments within stepped-care or blended models of suicide prevention.

This review has several limitations. First, between-study heterogeneity was high in several analyses, reflecting the considerable diversity in intervention formats, measurement instruments, therapeutic doses, and sample characteristics. Although moderator analyses were conducted (age group, delivery setting, and intervention format), the relatively small number of studies limited the possibility of more detailed analyses such as meta-regression. Reporting of intervention fidelity and participant adherence was inconsistent, and suicidal ideation was assessed predominantly through self-report measures, which may be influenced by factors such as social desirability or stigma. In addition, several trials showed moderate or high risk of bias according to the RoB 2 assessment, particularly due to attrition and the difficulty of blinding participants in digital interventions. Some analyses relied on pre–post within-group comparisons, which may produce larger effect size estimates than between-group comparisons and should therefore be interpreted cautiously (Cuijpers et al., 2017). Second, the limited available data relating to suicide attempts prevented us from conducting more detailed sub-group analyses, reducing the accuracy of estimates for this outcome. This limitation is compounded by 1) the fact that many digital interventions are not conceived as independent treatments but rather as a complement to treatment as usual, and 2) the frequent exclusion of individuals who are considered high risk for suicide, a decision taken on safety grounds. These aspects reduce the ecological validity of studies and make it difficult to isolate the specific effect of digital interventions (Sutori et al., 2024). Finally, most studies were conducted in high-income Western countries, limiting the generalizability of the findings.

In light of these limitations, future studies should prioritize trials examining the effects of digital interventions on suicide attempts, including designs that allow the safe participation of individuals at elevated risk. Research should also explore mechanisms of change using proximal variables such as hopelessness, perceived burdensomeness, thwarted belongingness, suicide-related beliefs, or mental pain, and evaluate the implementation of digital interventions in educational and community settings.

In conclusion, the findings of this systematic review and meta-analysis suggest that digital interventions may represent a promising complementary approach for reducing suicidal ideation and potentially reducing suicide risk. However, these results should be interpreted cautiously given the heterogeneity of the studies and the limited evidence regarding suicide attempts.

Author Contributions

Ainize Sarrionandia: Methodology, Supervision. **Maria Ruiz-Iriondo:** Data curation, Visualization. **Irene Izquierdo-Sanchez:** Data curation, Formal Analysis. **Andrea Gabilondo:** Funding acquisition, Writing—Review and Editing. **Craig J. Bryan:** Writing—Review and Editing. **Jon García-Ormaza:** Funding acquisition, Writing—Review and Editing. **Alexander Muela:** Conceptualization, Writing—Review and Editing.

Funding

This research was funded by a grant from the Eusko Jaurlaritza (2025333029). 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 there is no conflict 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
- Al-Halabi, S., & Fonseca-Pedrero, E. (2024). Editorial for special issue on understanding and prevention of suicidal behavior: Humanizing care and integrating social determinants. *Psicothema*, 36(4), 309-318. <https://doi.org/10.7334/psicothema2024.341>
- * Baker, J. C., Starkey, A., Ammendola, E., Bauder, C. R., Daruwala, S. E., Hiser, J., Khazem, L. R., Rademacher, K., Hay, J., Bryan, A. O., & Bryan, C. J. (2024). Telehealth brief cognitive behavioral therapy for suicide prevention: A randomized clinical trial. *JAMA Network Open*, 7(11), Article e2445913. <https://doi.org/10.1001/jamanetworkopen.2024.45913>
- Balk, E. M., Earley, A., Patel, K., Trikalinos, T. A., & Dahabreh, I. J. (2012). *Empirical assessment of within-arm correlation imputation in trials of*

- continuous outcomes*. Agency for Healthcare Research and Quality (US).
- * Batterham, P. J., Calear, A. L., Farrer, L., McCallum, S. M., & Cheng, V. W.-S. (2018). FitMindKit: Randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. *Internet Interventions*, 12, 91–99. <https://doi.org/10.1016/j.invent.2017.08.002>
- * Batterham, P. J., Calear, A. L., Farrer, L., Gulliver, A., & Kurz, E. (2021). Efficacy of a transdiagnostic self-help internet intervention for reducing depression, anxiety, and suicidal ideation in adults: Randomized controlled trial. *Journal of Medical Internet Research*, 23(1), Article e22698. <https://doi.org/10.2196/22698>
- Berger, T. (2017). The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. *Psychotherapy Research*, 27(5), 511–524. <https://doi.org/10.1080/10503307.2015.1119908>
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2022). *Comprehensive meta-analysis version 4*. Biostat.
- Bryan, C. J., & Bozzay, M. (2026). Competing motives and forces: A complex systems view of suicidal ambivalence. In R. Baston & M. Weichold (Eds.), *Exploring suicidal ambivalence: Philosophical and psychological perspectives*. Taylor & Francis.
- * Bryan, C. J., Simon, P., Wilkinson, S. T., Allen, M. H., Perez, J., Adler, C., Moon, K., Astorino, L., Carpenter, K. M., Misquitta, L., Brownlowe, K., Khazem, L. R., Hay, J., Starkey, A. G., Tartaglia, J., Winston, H., Simpson, S., Dager, A. D., & Feuerstein, S. (2025). A digital therapeutic intervention for inpatients with elevated suicide risk: A randomized clinical trial. *JAMA Network Open*, 8(8), Article e2525809. <https://doi.org/10.1001/jamanetworkopen.2025.25809>
- * Burback, L., Yap, S., Purdon, S. E., Abba-Aji, A., O'Shea, K., Brémault-Phillips, S., Greenshaw, A. J., & Winkler, O. (2024). Randomized controlled trial investigating web-based, therapist delivered eye movement desensitization and reprocessing for adults with suicidal ideation. *Frontiers in Psychiatry*, 15, Article 1361086. <https://doi.org/10.3389/fpsyt.2024.1361086>
- Burr, S. K., Yu, M., Clark, D., Alonzo, D., & Gearing, R.E. (2025). Digital interventions for suicide prevention: Systematic review and meta-analysis of RCTs examining age, gender, and control group type. *Crisis*, 46(3), 176–186. <http://doi.org/10.1027/0227-5910/a000996>
- Büscher, R., Beisemann, M., Doebl, P., Micklitz, H. M., Kerkhof, A., Cuijpers, P., Batterham, P. J., Calear, A. L., Christensen, H., De Jaegere, E., Domhardt, M., Erlangsen, A., Eylem van Bergeijk, O., Hill, R., Lungu, A., Mühlmann, C., Pettit, J. W., Portzky, G., Steubl, L. S., ... Sander, L. B. (2022). Digital cognitive-behavioral therapy to reduce suicidal ideation and behaviors: a systematic review and meta-analysis of individual participant data. *Evidence-based Mental Health*, 25(e1), e8–e17. <https://doi.org/10.1136/ebmental-2022-300540>
- Büscher, R., Torok, M., Terhorst, Y., & Sander, L. (2020). Internet-based cognitive behavioral therapy to reduce suicidal ideation: A systematic review and meta-analysis. *JAMA Network Open*, 3(4), Article e203933. <https://doi.org/10.1001/jamanetworkopen.2020.3933>
- * Bush, N. E., Smolenski, D. J., Denneson, L. M., Williams, H. B., Thomas, E. K., & Dobscha, S. K. (2016). A Virtual Hope Box: Randomized controlled trial of a smartphone app for emotional regulation and coping with distress. *Psychiatric Services*, 68(4), 330–336. <https://doi.org/10.1176/appi.ps.201600283>
- Cuijpers, P., Miguel, C., Ciharova, M., Harrer, M., Basic, D., Cristea, I. A., de Ponti, N., Driessen, E., Hamblen, J., Larsen, S. E., Matbouriah,

- M., Papola, D., Pauley, D., Plessen, C. Y., Pfund, R. A., Setkowski, K., Schnurr, P. P., van Ballegooijen, W., Wang, Y., ... Karyotaki, E. (2024). Absolute and relative outcomes of psychotherapies for eight mental disorders: A systematic review and meta-analysis. *World Psychiatry*, 23(2), 267–275. <https://doi.org/10.1002/wps.21203>
- Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences*, 26(4), 364–368. <https://doi.org/10.1017/S2045796016000809>.
- *De Jaegere, E., van Landschoot, R., van Heeringen, K., van Spijker, B. A. J., Kerkhof, A. J. F. M., Mokkenstorm, J. K., & Portzky, G. (2019). The online treatment of suicidal ideation: A randomised controlled trial of an unguided web-based intervention. *Behaviour Research and Therapy*, 119, Article 103406. <https://doi.org/10.1016/j.brat.2019.05.003>
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- *Eylem, O., van Straten, A., de Wit, L., Rathod, S., Bhui, K., & Kerkhof, A. J. F. M. (2021). Reducing suicidal ideation among Turkish migrants in the Netherlands and in the UK: The feasibility of a randomised controlled trial of a guided online intervention. *Pilot and Feasibility Studies*, 7(1), Article 30. <https://doi.org/10.1186/s40814-021-00772-9>
- Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, 55(4), 316–340. <https://doi.org/10.1037/pst0000172>
- Forte, A., Sarli, G., Polidori, L., Lester, D., & Pompili, M. (2021). The role of new technologies to prevent suicide in adolescence: A systematic review of the literature. *Medicina*, 57(2), Article 109. <https://doi.org/10.3390/medicina57020109>
- Fox, K. R., Huang, X., Guzmán, E. M., Funsch, K. M., Cha, C. B., Ribeiro, J. D., & Franklin, J. C. (2020). Interventions for suicide and self-injury: A meta-analysis of randomized controlled trials across nearly 50 years of research. *Psychological Bulletin*, 146(12), 1117–1145. <https://doi.org/10.1037/bul0000305>
- Han, J., Batterham, P. J., Calfar, A. L., & Randall, R. (2017). Factors influencing professional help-seeking for suicidality: A systematic review. *Crisis*, 39(3), 175–196. <https://doi.org/10.1027/0227-5910/a000485>
- Hawton, K., & Pirkis, J. (2024). Preventing suicide: a call to action. *The Lancet. Public Health*, 9(10), e825–e830. [https://doi.org/10.1016/S2468-2667\(24\)00159-2](https://doi.org/10.1016/S2468-2667(24)00159-2)
- Hesser, H., Weise, C., Westin, V. Z., & Andersson, G. (2011). A systematic review and meta-analysis of randomized controlled trials of cognitive-behavioral therapy for tinnitus distress. *Clinical Psychology Review*, 31(4), 545–553. <https://doi.org/10.1016/j.cpr.2010.12.006>
- *Hetrick, S. E., Yuen, H. P., Bailey, E., Cox, G. R., Templer, K., Rice, S. M., Bendall, S., & Robinson, J. (2017). Internet-based cognitive behavioural therapy for young people with suicide-related behaviour (Reframe-IT): A randomised controlled trial. *Evidence-Based Mental Health*, 20(3), 76–82. <https://doi.org/10.1136/eb-2017-102719>
- 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>
- *Hill, R. M., & Pettit, J. W. (2019). Pilot randomized controlled trial of LEAP: A selective preventive intervention to reduce adolescents' perceived burdensomeness. *Journal of Clinical Child & Adolescent Psychology*, 48(S1), S45–S56. <https://doi.org/10.1080/15374416.2016.1188705>
- Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, 36(5), 427–440. <https://doi.org/10.1007/s10608-012-9476-1>.
- Jobes, D. A., Mandel, A. A., Kleiman, E. M., Bryan, C. J., Johnson, S. L., & Joiner, T. E. (2024). Facets of suicidal ideation. *Archives of Suicide Research*, 28(4), 1263–1278. <https://doi.org/10.1080/13811118.2023.2299259>
- Josifovski, N., Dit Rochesson, S. E., Wong, Q. J., Han, J., Larsen, M. E., & Torok, M. (2025). What works where and why? A systematic review and meta-analysis of digital interventions addressing suicide-related outcomes in community, education and clinical settings. *International Journal of Social Psychiatry*, 71(8), 1453–1469. <https://doi.org/10.1177/00207640251358109>
- *Kasckow, J., Gao, S., Hanusa, B., Rotondi, A., Chinman, M., Zickmund, S., Gurklis, J., Fox, L., Cornelius, J., Richmond, I., & Haas, G. L. (2015). Telehealth monitoring of patients with schizophrenia and suicidal ideation. *Suicide and Life-Threatening Behavior*, 45(5), 600–611. <https://doi.org/10.1111/sltb.12154>
- *Li, Y., Guo, Y., Hong, Y. A., Zhu, M., Zeng, C., Qiao, J., Xu, Z., Zhang, H., Zeng, Y., Cai, W., Li, L., & Liu, C. (2019). Mechanisms and effects of a WeChat-based intervention on suicide among people living with HIV and depression: Path model analysis of a randomized controlled trial. *Journal of Medical Internet Research*, 21(11), Article e14729. <https://doi.org/10.2196/14729>
- *Lohani, M., Bryan, C. J., Elsej, J. S., Dutton, S., Findley, S. P., Langenecker, S. A., West, K., & Baker, J. C. (2024). Collaboration matters: A randomized controlled trial of patient-clinician collaboration in suicide risk assessment and intervention. *Journal of Affective Disorders*, 360, 387–393. <https://doi.org/10.1016/j.jad.2024.06.004>
- Mann, J. J., Michel, C. A., & Auerbach, R. P. (2021). Improving suicide prevention through evidence-based strategies: A systematic review. *American Journal of Psychiatry*, 178(7), 611–624. <https://doi.org/10.1176/appi.ajp.2020.20060864>
- Oh, J., Ho, J., Lee, S., & Park, J. H. (2024). Effects of digital psychotherapy on suicide: A systematic review and meta-analysis. *Healthcare*, 12(14), Article 1435. <https://doi.org/10.3390/healthcare12141435>
- 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., McGuinness, L. A., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71. <https://doi.org/10.1136/bmj.n71>
- Patterson, E., Boyle, M. H., Kivlenieks, M., & Van Ameringen, M. (2016). The use of waitlists as control conditions in anxiety disorders research. *Journal of Psychiatric Research*, 83, 112–120. <https://doi.org/10.1016/j.jpsychires.2016.08.015>
- Potts, C., Kealy, C., McNulty, J. M., Madrid-Cagigal, A., Wilson, T., Mulvenna, M. D., O'Neill, S., Donohoe, G., & Barry, M. M. (2025). Digital mental health interventions for young people aged 16–25 years: Scoping review. *Journal of Medical Internet Research*, 27, Article e72892. <https://doi.org/10.2196/72892>
- Rain, S., Williams, M., Mitchell, R., Zaheer, R., Bryan, C. J., Schaffer, A., Men, V. Y., Westreich, N., Ellis, J., Goldstein, B. I., Cheung, A.

- H., Selchen, S., & Sinyor, M. (2024). Youth self-harm and the role of reasons for living and hope: A secondary analysis from a randomized controlled trial. *Psychiatry Research Communications*, 4(2), Article 100174. <https://doi.org/10.1016/j.psycom.2024.100174>
- Rudd, M. D. (2006). Fluid vulnerability theory: A cognitive approach to understanding the process of acute and chronic suicide risk. In T. E. Ellis (Ed.), *Cognition and suicide: Theory, research, and therapy* (pp. 355–368). American Psychological Association. <https://doi.org/10.1037/11377-016>
- Shoib, S., Shaheen, N., Anwar, A., Saad, A. M., Akr, L. M., Saud, A. I., Kundu, M., Nahidi, M., Chandradasa, M., Swed, S. & Saeed, F. (2024). The effectiveness of telehealth interventions in suicide prevention: A systematic review and meta-analysis. *International Journal of Social Psychiatry*, 70(3), 415–423. <https://doi.org/10.1177/00207640231206059>
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, 366, Article 14898. <https://doi.org/10.1136/bmj.14898>
- Sutori, S., Hadlaczk, G., Eliasson, E., Wasserman, D., & Carli, V. (2024). Systematic review and meta-analysis: Effectiveness of stand-alone digital suicide preventive interventions for the self-management of suicidality. *Journal of Technology in Behavioral Science*, 9(3), 440–451. <https://doi.org/10.1007/s41347-023-00374-7>
- Tabares, J. V., Muela, A., Ammendola, E., & García-Ormaza, J. (2025). Simulation networks elucidate the preliminary structure of suicide risk factors among adolescents. *Journal of Child and Adolescent Psychiatric Nursing*, 38(4), Article e70037. <https://doi.org/10.1111/jcap.70037>
- * Tighe, J., Shand, F., Ridani, R., Mackinnon, A., De La Mata, N., & Christensen, H. (2017). Ibobly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. *BMJ open*, 7(1), Article e013518. <https://doi.org/10.1136/bmjopen-2016-013518>
- Torok, M., Han, J., Baker, S., Werner-Seidler, A., Wong, I., Larsen, M. E., & Christensen, H. (2020). Suicide prevention using self-guided digital interventions: A systematic review and meta-analysis of randomized controlled trials. *The Lancet. Digital Health*, 2(1), Article e25–e36. [https://doi.org/10.1016/S2589-7500\(19\)30199-2](https://doi.org/10.1016/S2589-7500(19)30199-2)
- * Torok, M., Han, J., McGillivray, L., Wong, Q., Werner-Seidler, A., O'Dea, B., Calear, A., & Christensen, H. (2022). The effect of a therapeutic smartphone application on suicidal ideation in young adults: Findings from a randomized controlled trial in Australia. *PLOS Medicine*, 19(5), e1003978. <https://doi.org/10.1371/journal.pmed.1003978>
- van Ballegooijen, W., Rawee, J., Palantza, C., Miguel, C., Harrer, M., Cristea, I., de Winter, R., Gilissen, R., Eikelenboom, M., Beckman, A., & Cuijpers, P. (2025). Suicidal ideation and suicide attempts after direct or indirect psychotherapy: a systematic review and meta-analysis. *JAMA Psychiatry*, 82(1), 31–37. <https://doi.org/10.1001/jamapsychiatry.2024.2854>
- * van Spijker, B. A. J., van Straten, A., & Kerkhof, A. J. F. M. (2014). Effectiveness of online self-help for suicidal thoughts: Results of a randomised controlled trial. *PLOS ONE*, 9(2), Article e90118. <https://doi.org/10.1371/journal.pone.0090118>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Wanniarachchi, V. U., Greenhalgh, C., & Warren, J. (2025). Adolescents' and youths' perceived barriers and facilitators to engaging with digital mental health interventions for depression and anxiety: A scoping review. *Internet Interventions*, 42, Article 100884. <https://doi.org/10.1016/j.invent.2025.100884>
- * Wilks, C. R., Lungu, A., Ang, S. Y., Matsumiya, B., Yin, Q., & Linehan, M. M. (2018). A randomized controlled trial of an internet-delivered dialectical behavior therapy skills training for suicidal and heavy episodic drinkers. *Journal of Affective Disorders*, 232, 219–228. <https://doi.org/10.1016/j.jad.2018.02.053>
- Witt, K., Spittal, M. J., Carter, G., Pirkis, J., Hetrick, S., Currier, D., Robinson, J., & Milner, A. (2017). Effectiveness of online and mobile telephone applications ('apps') for the self-management of suicidal ideation and self-harm: A systematic review and meta-analysis. *BMC Psychiatry*, 17, 297. <https://doi.org/10.1186/s12888-017-1458-0>
- World Health Organization. (2021). *Suicide worldwide in 2019: global health estimates*. World Health Organization
- Yosep, I., Hikmat, R., Mardhiyah, A., & Hernawaty, T. (2024). A scoping review of digital-based intervention for reducing risk of suicide among adults. *Journal of Multidisciplinary Healthcare*, 17, 3545–3556. <https://doi.org/10.2147/JMDH.S47226>

Article

University Students' Coping Profiles and their Relationship with the Intention to Drop Out

Ana B. Bernardo¹, Elena Blanco¹, Carlos Freire² and María del Mar Ferradás²

¹ University of Oviedo (Spain)

² University of A Coruña (Spain)

ARTICLE INFO

Received: 25/12/2025

Accepted: 30/03/2026

Keywords:

Coping profiles
University dropout
Latent profiles

ABSTRACT

Background: University dropout is a persistent problem in higher education, closely linked to psychological processes that affect student adaptation. Among these processes, coping strategies has received little attention, particularly from person-centred approaches. The objective of this study is to explore coping profiles in university students and analyse their relationship with the intention to dropout, as well as the distribution by gender and knowledge area. **Method:** 1,632 first-year students from a Spanish university participated. A latent profile analysis was performed to identify coping profiles, along with multivariate analyses to examine differences in dropout intention and sociodemographic variables. **Results:** Three profiles were identified: Low Coping Profile (LCP), Mixed Engagement-Disengagement Profile (MEDP) and High Emotion-Engagement Profile (HEEP). A clear gradient emerged in dropout intention (LCP > HEEP > MEDP). No statistically significant differences in profile membership were found by gender or field of study. **Conclusions:** The results show that students' coping profiles differ in their vulnerability to dropout intention, underscoring the importance of strengthening adaptive coping resources from the outset of university.

Perfiles de Afrontamiento y su Relación con la Intención de Abandono en Estudiantes Universitarios

RESUMEN

Palabras clave:

Abandono universitario
Perfiles de afrontamiento
Perfiles latentes

Antecedentes: El abandono universitario constituye un problema persistente en la Educación Superior, estrechamente vinculado a procesos psicológicos que condicionan la adaptación del alumnado. Entre ellos, las estrategias de afrontamiento han recibido escasa atención, especialmente desde enfoques centrados en la persona. El objetivo es explorar los perfiles de afrontamiento en estudiantes universitarios y analizar su relación con la intención de abandono, así como su distribución según género y área de conocimiento. **Método:** Participaron 1.632 estudiantes de primer curso de una universidad española. Se realizó un análisis de perfiles latentes para identificar perfiles de afrontamiento, junto con análisis multivariados para examinar diferencias en intención de abandono y variables sociodemográficas. **Resultados:** Se identificaron tres perfiles: Perfil de Bajo Afrontamiento (LCP), Perfil Mixto de Engagement-Disengagement (MEDP) y Perfil de Alto Engagement Emocional (HEEP). Emergió un gradiente claro en la intención de abandono (LCP > HEEP > MEDP). No se encontraron diferencias estadísticamente significativas en la pertenencia a los perfiles por género y área de estudio. **Conclusiones:** Los resultados muestran que los perfiles de afrontamiento difieren en su vulnerabilidad a la intención de abandono, lo que subraya la importancia de fortalecer los recursos adaptativos de afrontamiento del alumnado desde el inicio de la vida universitaria.

University dropout has significant psychological implications for students, who frequently experience reduced motivation and engagement, and increased emotional distress (Buizza et al., 2024; Mâta et al., 2024). The magnitude of the issue is evident: in Spain, 22.1% of students do not continue after their first university year (Ministerio de Ciencia, Innovación y Universidades, 2026), and there is similar variability across Europe, where first-year dropout rates range from 5.4% in Finland, to over 18% in Belgium, Slovenia and Poland (European Commission, 2025). Conceptualising dropout as a psychologically and contextually shaped phenomenon underscores the need for frameworks explaining how students decide whether to remain on their courses or dropout.

The *Psychological Model of College Student Retention* (Bean & Eaton, 2001), provides a comprehensive framework for understanding the psychological mechanisms underlying students' decisions to persist or dropout. According to this model, students enter university with personal expectations, beliefs, and prior experiences that shape how they interpret the university environment. These interactions activate three key psychological processes—self-efficacy, causal attributions, and coping strategies—which influence students' attitudes toward the institution and ultimately give rise to their intention to persist or dropout.

Within this framework, dropout intention is identified as the most proximal antecedent of actual dropout. It can be interpreted as a student's subjective evaluation that continuing to attend university may no longer be viable or desirable, and it is one of the strongest predictors of subsequent dropout behaviour (Maluenda-Albornoz et al., 2022). This makes dropout intention particularly important when looking at dropout during the first year—a period characterised by the increased demands of academic, social, and emotional adjustment that mean students are more vulnerable to doubts about their academic careers (Linden et al., 2025; Parsons et al., 2023). According to national and European data, a student's first academic year is where dropout is most concentrated, underscoring the need to understand the psychological responses that shape students' persistence during this transition.

Lazarus and Folkman's *Transactional Theory of Stress* (1986) is particularly useful for understanding coping in first-year students because it conceptualizes stress as the result of an interaction between the individual and the environment. Accordingly, when faced with academic stressors—such as workload, teaching methodologies, assessment systems, or the pressure to perform (Cabanach et al., 2016)—students first engage in primary appraisal, evaluating whether these demands represent a threat or challenge. Then during secondary appraisal, they assess the resources available for coping with the situation.

From this theoretical standpoint, coping strategies play a central role in shaping dropout intention. Coping has traditionally been classified according to its focus (problem-focused vs. emotion-focused) and orientation (engagement vs. avoidance). Problem-focused coping emphasizes managing the stress-inducing situation itself, whereas emotion-focused coping prioritizes regulation of the affective response to the problem (Tous-Pallarés et al., 2022). In terms of orientation, engagement strategies are efforts aimed at directly managing the situation or the emotions underlying it, whereas avoidance strategies involve attempts to limit exposure to the stressor by avoiding threatening situations altogether (Zimmer-Gembeck & Skinner, 2016).

Empirical evidence suggests that engagement-oriented coping strategies (whether problem-focused or emotion-focused) are more adaptive, since they promote persistence and well-being; on the contrary, disengagement-oriented coping strategies (problem-focused and emotion-focused) are associated with dysfunctional outcomes, such as poorer academic adjustment and greater likelihood of dropping out (López-Aguilar et al., 2022; Martín-Arbós et al., 2024; Reyes-Pérez et al., 2021; Rogers et al., 2016). These findings highlight a key mechanism: coping shapes students' appraisals of their ability to continue at university. When disengagement strategies predominate, academic demands are more likely to be interpreted as overwhelming, reducing self-efficacy or academic integration (Pi et al., 2024; Schladitz et al., 2024; Torquati et al., 2025)—factors identified as precursors of dropout intention (Bean & Eaton, 2001). Conversely, engagement-focused strategies support adaptive appraisal by supporting emotional regulation and activating problem-solving (Alves et al., 2022; Mize, 2024).

Despite this, focusing on individual strategies may obscure the complexity of students' responses. There is growing evidence that first-year students rarely rely on a single coping style. Instead, they often combine engagement and disengagement strategies, yielding a hybrid pattern in which the same student may be confronting, regulating, and avoiding stress (Smart et al., 2025; Thomas et al., 2024). A person-centred approach, such as Latent Profile Analysis (LPA), allows these patterns of co-occurrence to be identified and offers a more nuanced understanding of the diversity of student trajectories (González-Casas et al., 2025).

Moreover, coping styles do not emerge in isolation but are shaped by personal and contextual factors (Vidal-Conti et al., 2018). For example, women tend to use more emotion-focused strategies, whereas men are more likely to adopt avoidant or problem-focused styles (Coppola et al., 2024; Martínez et al., 2019). Similarly, a student's field of study may influence their perceptions of stressors and the responses to them, given the diversity of academic demands, teaching methodologies, and disciplinary cultures across academic domains (Bernardo et al., 2025; González-Casas et al., 2025). Understanding how these factors intersect with coping profiles is essential for identifying groups who may be more vulnerable to disengagement-based coping and, consequently, to stronger dropout intentions.

In line with this evidence, the present study adopts a person-centred approach to explore how coping strategies shape distinct profiles in first-year university students and how these profiles are related to university dropout intention. This involves the following specific objectives: (1) To identify different latent profiles based on the coping strategies students use, (2) To examine university dropout intention in the identified latent profiles, and (3) To analyse the sociodemographic characteristics (gender and field of study) associated with each of the identified latent profiles.

Method

Participants

The sample consisted of first-year undergraduate students from a public university in northern Spain with approximately 21,000 enrolled students, of whom 4,416 were first-year undergraduates. To achieve the best population representativeness, all first-year students enrolled in the academic programs offered by the university

were invited to participate, resulting in the inclusion of 66 degree programs. First-year students who were minors (age < 18) at the time of data collection, and those with atypical ages ($Z > 3$), were excluded from the study, yielding 33 removals.

The final sample comprised 1,632 students aged between 18 and 25 years old ($M = 18.45$; $SD = 1.01$). Just over half (904; 55.4%) of the participants were women, 704 (43.1%) were men, and 5 (0.3%) identified as non-binary. The remaining participants (19; 1.2%) chose not to disclose this information. With respect to fields of study, 658 (40.3%) were enrolled in degrees related to Social and Legal Sciences, 335 (20.5%) in Engineering, 262 (16.1%) in Sciences, 252 (15.4%) in Health Sciences, and 125 (7.7%) in Arts and Humanities.

Instruments

Coping strategies were assessed using the *Coping Strategy Indicator – Short Form* (CSI-SF; Tous-Pallarés et al., 2022). It consists of 16 items that are rated on a 5-point Likert scale (1 = *never*; 5 = *almost always*) and is structured into four factors: Emotion-Focused Engagement (EFE, e.g., “*I try to talk to a friend or family member*”), Problem-Focused Engagement (PFE, e.g., “*I make an action plan and I follow it*”), Emotion-Focused Disengagement (EFD, e.g., “*I’m waiting for a miracle*”), and Problem-Focused Disengagement (PFD, e.g., “*I hope the problem solves itself*”). In the present study, the psychometric properties were as follows: EFE (reliability: $\alpha = .69$; $\omega = .73$, 95% IC [.71, .75], construct validity: $\chi^2(2) = 58.88$; CFI = .96; GFI = .98; RMSEA = .13; SRMR = .05); PFE (reliability: $\alpha = .57$; $\omega = .57$, 95% IC [.54, .61], construct validity: $\chi^2(2) = 26.20$; CFI = .95; GFI = .99; RMSEA = .08; SRMR = .03); EFD (reliability: $\alpha = .62$; $\omega = .68$, 95% IC [.65, .70], construct validity: $\chi^2(2) = 77.34$; CFI = .94; GFI = .98; RMSEA = .15; SRMR = .06); PFD (reliability: $\alpha = .63$; $\omega = .62$, 95% IC [.59, .65], construct validity: $\chi^2(2) = 164.85$; CFI = .83; GFI = .95; RMSEA = .32; SRMR = .08).

Although some factors exhibited questionable internal consistency values, they are comparable to those reported in the original validation of the instrument (Addison et al., 2007). According to Loewenthal and Lewis (2001), in scales with a limited number of items, Cronbach’s alpha values around .60 may be considered acceptable. In this regard, for the three factors with alpha values close to .60, the mean inter-item correlations (Briggs & Cheek, 1986) fell within the optimal range ($r = .20 - r = .40$): PFE ($r = .25$), PFD ($r = .30$), EFD ($r = .28$). These results indicate adequate item covariation despite the brevity of the scale, supporting the instrument’s internal consistency. Likewise, the CFA fit indices seem to indicate that, except for some parameters (especially the RMSEA), in general terms the model demonstrated acceptable construct validity.

Dropout intention was assessed using two complementary formats: a dichotomous measure and a Likert-type scale. First, a dichotomous item: “*Have you ever thought about dropping out of your current course of study?*” (answer: *yes/no*) was used to assess the presence or absence of dropout intention. This item is part of the Early University Dropout Intention Questionnaire –EUDIQ–R– (Bernardo et al., 2022). Second, five items were designed ad hoc to capture different types of dropout: discontinuing university education altogether (“*I have sometimes thought about dropping out of university and not continuing my studies at all*”); entering employment (“*I have sometimes thought about dropping out of university to join the workforce*”); changing degree course (“*I*

have sometimes thought about changing the degree I am studying and enrolling in another university degree program”); continuing to study at another university (“*I have sometimes thought about continuing my university studies at another university*”); and enrolling in vocational education and training programs (“*I have sometimes thought about dropping out of university and continuing my studies with a vocational education and training program*”). Responses to these questions were recorded on a four-point Likert-type scale (1 = *strongly disagree*; 4 = *strongly agree*).

Procedure

The study was conducted with the collaboration of the administrative teams in the different faculties at the participating university, who selected the teaching staff involved in data collection. Teaching staff received brief training on the study objectives and in the online administration of the instruments. Data were collected during the second semester of the 2024–2025 academic year.

The study was approved by the Ethics Committee at the participating university (Ref.: 40_RRI_2023). Participants were provided with written information regarding the study objectives, the voluntary nature of participation, and the confidential and anonymous treatment of data. Participation was formalized through the provision of informed consent.

Data Analysis

Coping profiles were identified via Latent Profile Analysis –LPA– (Lanza et al., 2003), conducted using Mplus version 7.11 (Muthén & Muthén, 2012). The maximum likelihood with robust standard errors (MLR) criterion was used, as well as full information maximum likelihood (FIML) as a method for handling missing data. In specifying the models, variances were allowed to differ across indicators within each profile, while being constrained to be equal across profiles. In addition, local independence among indicators was imposed as a model constraint, both within and across profiles. To avoid local maxima, 500 random starts and 100 final iterations were established (Starts = 500 100). In addition, the variables were standardized using the command OUTPUT = STDXY.

Model selection was based on the following indices of fit: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), sample-size adjusted BIC (SSA-BIC), the Lo–Mendell–Rubin adjusted likelihood ratio test (LMRT), the parametric bootstrap likelihood ratio test (PBLRT), and class sample sizes. Lower values of AIC, BIC, and SSA-BIC indicate better model fit; however, final model selection relied primarily on the LMRT and PBLRT. The p values associated with these tests indicate whether a solution with more classes ($p \leq .05$) or fewer classes ($p > .05$) provides a better fit to the data. Additionally, solutions containing spurious classes ($\leq 5\%$ of the sample) were excluded, as that would suggest overextraction of profiles (Hipp & Bauer, 2006).

Once the optimal model was selected, classification accuracy was evaluated using posterior probabilities and the entropy statistic, which ranges from 0 to 1, with values closer to 1 indicating greater classification accuracy. Model validity was further examined through MANOVA, assessing differences between classes across the four coping variables (EFE, PFE, EFD, and PFD). Statistically significant differences between classes were interpreted as evidence that the latent profiles were meaningfully distinct.

Differences between the identified profiles in the various modalities of dropout intention (permanently stopping university study, dropping out to enter employment, enrolling in another degree program, transferring to another university, or enrolling in vocational education and training) were examined using MANCOVA, with gender and age as covariates. Effect sizes were estimated using partial eta squared (η_p^2) and Cohen's *d*, applying the following criteria: no effect ($\eta_p^2 < .01$; $d < 0.09$), small effect ($\eta_p^2 = .01-.058$; $d = 0.10-.49$), medium effect ($\eta_p^2 = .059-.137$; $d = 0.50-.79$), and large effect ($\eta_p^2 \geq .138$; $d \geq 0.80$).

To further explore the identified profiles, dropout intention was compared between profiles. In addition, differences in profile distribution as a function of field of study and gender were examined using Pearson's chi-square test of independence, with Cramer's *V* used as an effect size estimator ($V \leq .20$, small effect; $V = .21-.59$, medium effect; $V \geq .60$, large effect). These analyses were conducted using SPSS version 26.0 (IBM Corp., 2019).

Results

Identification of Stress Coping Profiles

Once the existence of missing values was ruled out, the fit of several latent profile models ($k = 2$ to $k = 5$) was examined.

Table 1 presents the results of the model fit analyses. All reported models converged on a stable solution, confirmed by the stability of the log-likelihood (LL) value across multiple random starts. Model estimation was discontinued at the five-class solution because this solution yielded non-significant values for LMRT and PBLRT ($p > .05$), indicating that the fit of this model was no better than the four-class model. Likewise, the four-class solution was discarded because, despite producing significant values for LMRT and PBLRT ($p \leq .05$), it included one class comprising less than 5% of the total sample (0.55%, $n = 9$). Moreover, the values of other parameters (AIC, BIC, SSA-BIC, entropy) were very similar to those for the three-class model.

Table 1
Fit Indices for Latent Profile Model Identification and Classification Accuracy

	<i>k</i> = 2	<i>k</i> = 3	<i>k</i> = 4	<i>k</i> = 5
AIC	13355.922	13215.376	13142.477	13093.122
BIC	13426.090	13312.532	13266.621	13244.254
SSA-BIC	13384.791	13255.349	13193.554	13155.303
Entropy	.597	.642	.681	.681
$n \leq 5\%$	0	0	1	1
LMRT	192.950*	150.546*	82.899*	59.355
PBLRT	187.870*	146.583*	80.717*	57.792

Note. AIC = Akaike Information Criterion; BIC = Schwarz Bayesian Information Criterion; SSA-BIC = BIC adjusted for the sample size; LMRT = adjusted Lo-Mendell-Rubin maximum likelihood ratio test; PBLRT = Parametric Bootstrap likelihood ratio test.

* $p < .01$.

Overall, the three-class model exhibited the most favourable indices of fit, with statistically significant values for LMRT and PBLRT, as well as lower AIC, BIC, and SSA-BIC values compared to the two-class model. In addition, the MANOVA results support the adequacy of the selected model, showing statistically significant differences between the three classes across all criterion variables: EFE ($F(2,1629) = 75.445$; $p < .001$; $\eta_p^2 = 0.369$), PFE ($F(2,1629) = 225.489$; $p < .001$; $\eta_p^2 = 0.217$), EFD ($F(2,1629) = 261.399$; $p < .001$; $\eta_p^2 = 0.242$), PFD ($F(2,1629) = 399.775$; $p < .001$; $\eta_p^2 = 0.329$).

The classification accuracy of the three-class model was examined using both the entropy statistic and the average posterior probabilities. Although the value of entropy for the three-class model was modest (0.642), this parameter tends to decrease when sample sizes are large (Wang et al., 2017). Furthermore, as Table 2 shows, the average posterior probabilities indicated that the classification accuracy in the three classes was similar and adequate (over .75).

Table 2
Characterization of Latent Profiles and Individual Classification Accuracy for Each Profile

	1	2	3	<i>n</i>	%
1. LCP	.788	.175	.037	259	15.87
2. MEDP	.089	.865	.046	1,215	74.45
3. HEPP	.052	.200	.748	158	9.68

Note. LCP = Low Coping Profile; MEDP = Mixed Engagement–Disengagement Profile; HEPP = High Emotion-Engagement Profile.

Further evidence that the selected model is relevant lies in its theoretical significance, supported by the fact that it is consistent with previously identified coping profiles in the literature. Table 3 shows the mean scores from participants belonging to the latent classes in the selected model (both raw and standardized scores), and Figure 1 provides a graphical representation of these profiles:

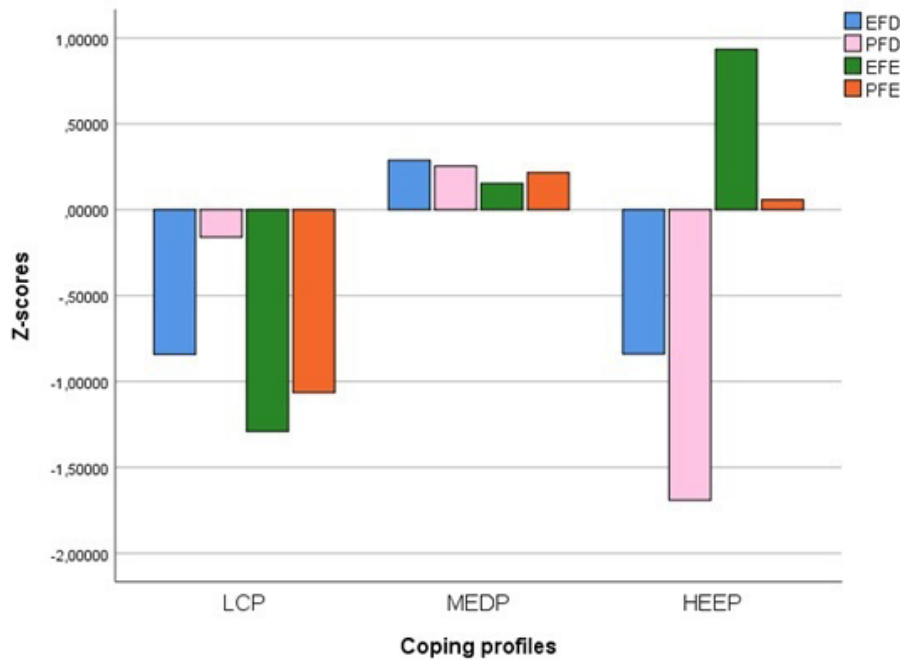
- *Low Coping Profile* (LCP; $n = 259$; 15.87%): Students with low scores in all strategies, both engagement (both problem- and emotion- focused) and disengagement (both problem- and emotion-focused).
- *Mixed Engagement-Disengagement Profile* (MEDP; $n = 1,215$; 74.45%): Students with moderately high scores across all four types of coping strategies.
- *High Emotion-Engagement Profile* (HEPP; $n = 158$; 9.68%): Students with high scores in emotion-focused engagement strategies, very moderate scores in problem-focused engagement strategies and low scores in disengagement strategies (both problem- and emotion-focused).

Table 3
Means, Standard Errors, and Confidence Intervals for the Three-class Latent Solution

		95% CI			
		<i>M</i>	<i>SE</i>	LL	UL
LCP	PFE	2.81 (–1.07)	.09	2.66	2.97
	EFE	2.07 (–1.29)	.09	1.93	2.22
	PFD	2.96 (–0.16)	.11	2.77	3.14
	EFD	3.17 (–0.84)	.13	2.96	3.38
MEDP	PFE	3.57 (0.22)	.04	3.51	3.63
	EFE	2.70 (0.15)	.03	2.65	2.76
	PFD	3.64 (0.25)	.03	3.59	3.68
	EFD	3.44 (0.29)	.04	3.38	3.50
HEPP	PFE	3.97 (0.06)	.11	3.79	4.15
	EFE	2.02 (0.93)	.11	1.84	2.20
	PFD	3.50 (–1.69)	.15	3.25	3.75
	EFD	2.20 (–0.84)	.11	2.02	2.38

Note. LCP = Low Coping Profile; MEDP = Mixed Engagement–Disengagement Profile; HEPP = High Emotion-Engagement Profile; PFE = Problem-Focused Engagement; EFE = Emotion-Focused Engagement; PFD = Problem-Focused Disengagement; EFD = Emotion-Focused Disengagement. Standardized mean scores (*z*) are shown in parentheses.

Figure 1
Graphical Representation of Coping Profiles



Note. LCP = Low Coping Profile; MEDP = Mixed Engagement–Disengagement Profile; HEED = High Emotion-Engagement Profile; PFE = Problem-Focused Engagement; EFE = Emotion-Focused Engagement; PFD = Problem-Focused Disengagement; EFD = Emotion-Focused Disengagement.

Differences in Dropout Intentions between Coping Profiles

After controlling for the effects of gender and age, the results of the MANCOVA indicated statistically significant differences between profiles ($\lambda_{\text{Wilks}} = .977$, $F(10,3246) = 3.820$, $p < .001$, $\eta_p^2 = .012$), with a small effect size. Additionally, the covariate age was statistically significant ($F(5,1623) = 11.502$, $p < .001$, $\eta_p^2 = .034$), although the effect size was small.

Statistically significant differences were also found between coping profiles in four of the five modalities of dropout intention: stopping university altogether ($F(2,1629) = 7.317$, $p < .001$, $\eta_p^2 = .01$); entering employment ($F(2,1629) = 35.718$, $p < .01$, $\eta_p^2 = .02$); enrolling in another degree program ($F(2,1629) = 9.922$, $p < .001$, $\eta_p^2 = .01$); and enrolling in vocational education and training ($F(2,1629) = 11.984$, $p < .001$, $\eta_p^2 = .02$). In all cases, the effect sizes were small. No significant differences were found between coping profiles in the variable continuing to study at another university ($F(2,1629) = 2.375$, $p = .093$).

Table 4 shows the descriptive statistics from the MANCOVA. Across the five dimensions analysed, the LCP group exhibited the highest scores in dropout intention, whereas the MEDP group exhibited the lowest scores in all five dimensions. Differences between these two profiles were statistically significant in four of the five intention to dropout variables: stopping university altogether ($d = 0.26$), entering employment ($d = 0.24$), enrolling in another degree program ($d = 0.28$), and enrolling in vocational education and training ($d = 0.29$). The effect sizes were small in all cases.

There were also statistically significant differences between the MEDP and the HEED groups in enrolling in vocational education and training, although the effect size was small ($d = 0.19$).

Table 4

Descriptive Statistics for Coping Profiles across Dimensions of Dropout Intention

	Coping profiles					
	LCP (<i>n</i> = 259)		MEDP (<i>n</i> = 1,215)		HEED (<i>n</i> = 158)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Stopping university altogether ^a	1.48	0.69	1.31	0.64	1.37	0.68
Entering employment ^b	1.74	0.87	1.55	0.74	1.65	0.85
Changing degree programs ^c	2.24	1.07	1.95	1.01	2.15	1.10
Continuing to study at another university ^d	1.97	0.97	1.83	0.94	1.91	1.06
Enrolling in vocational education and training programs ^e	1.69	0.93	1.44	0.77	1.60	0.90
Significant comparisons	^a LCP-MEDP ($d = 0.26$) ^b LCP-MEDP ($d = 0.24$) ^c LCP-MEDP ($d = 0.28$) ^d LCP-MEDP ($d = 0.29$) ^e MEDP-HEED ($d = 0.19$)					

Note. LCP = Low Coping Profile; MEDP = Mixed Engagement–Disengagement Profile; HEED = High Emotion-Engagement Profile.

Characterization of University Dropout Intention

A statistically significant association was found between profile membership and dropout intention, although the effect size was small ($\chi^2 = 47.34$, $df = 2$, $p < .001$, $V = .17$).

The results indicate that in the three coping profiles, 27.03% of students ($n = 441$) reported dropout intentions (Table 5). The profile

with the largest proportion of students expressing dropout intention was LCP (43.24%), followed by HEEP (32.28%). The lowest percentage was in the MEDP profile (22.88%).

Table 5
Percentages of Dropout Intention across Coping Profiles

		LCP	MEDP	HEEP	Total
Dropout intention	<i>n</i>	112	278	51	441
	%	43.24	22.88	32.28	27.03
No dropout intention	<i>n</i>	147	937	107	1,191
	%	56.76	77.12	67.72	72.97

Note. LCP = Low Coping Profile; MEDP = Mixed Engagement–Disengagement Profile; HEEP = High Emotion-Engagement Profile.

Distribution of Field of Study and Gender Variables

The association between latent profile membership and field of study was not statistically significant ($\chi^2 = 12.88$, $df = 8$, $p = .12$, $V = .06$). This finding seems to indicate that the proportion of students in each of the identified coping profiles is very similar across the five knowledge areas.

As Table 6 shows, the most widely represented profile was MEDP, with 74.4% of students in the five fields of knowledge belonging to this group. Only Arts and Humanities students were slightly below this percentage (MEDP, 68.8%). Around 15% (15.9%) of students in each of the fields of knowledge would fall into the LCP category. This is not an insignificant proportion, given that this profile of students seems to lack strategies (either engagement- or disengagement-focused) when it comes to dealing with academic demands. In this regard, the percentage of students (19.85%) who fit this profile in the Sciences is particularly noteworthy.

Finally, the HEEP profile was the least represented of the three profiles in the five knowledge areas (9.7%). The highest percentages were among Arts and Humanities students (13.6%), while the lowest representation of the profile was in students studying Sciences (7.63%).

Table 6
Percentage Representation of Coping Profiles across Different Fields of Study

		LCP	MEDP	HEEP	Total
Arts and Humanities	<i>n</i>	22	86	17	125
	%	17.6	68.8	13.6	100
Sciences	<i>n</i>	52	190	20	262
	%	19.85	72.52	7.63	100
Health Sciences	<i>n</i>	34	197	21	252
	%	13.5	78.2	8.3	100
Social and Legal Sciences	<i>n</i>	92	493	73	658
	%	13.98	74.92	11.1	100
Engineering	<i>n</i>	59	249	27	335
	%	17.61	74.33	8.06	100
Total	<i>n</i>	259	1,215	158	1,632
	%	15.87	74.45	9.68	100

Note. LCP = Low Coping Profile; MEDP = Mixed Engagement–Disengagement Profile; HEEP = High Emotion-Engagement Profile.

The association between coping profile membership and gender was analysed after excluding participants who chose not to disclose their gender and non-binary participants due to the low sample representation ($n = 24$). This relationship was not statistically significant ($\chi^2 = 3.46$, $df = 2$, $p = .18$, $V = .05$), suggesting that the three profiles identified are homogeneously represented by women and men.

As Table 7 shows, both women and men were primarily concentrated in the MEDP profile (74.89% and 74.43%, respectively). Men were slightly more represented in LCP (17.05%) than women (14.49%), whereas women were slightly more represented in HEEP (10.62%) than men (8.52%).

Table 7
Percentage Representation of Coping Profiles by Gender

		LCP	MEDP	HEEP	Total
Women	<i>n</i>	131	677	96	904
	%	14.49	74.89	10.62	100
Men	<i>n</i>	120	524	60	704
	%	17.05	74.43	8.52	100

Note. LCP = Low Coping Profile; MEDP = Mixed Engagement–Disengagement Profile; HEEP = High Emotion-Engagement Profile.

Discussion

The present study provides insight into how coping strategies are organised into different patterns among university students and how these relate to dropout intention. Adopting a person-centred perspective is particularly important in a context where dropout is a multifactorial phenomenon and psychological processes play a key role in students' adaptation to academic demands (Bean & Eaton, 2001; Diaz-Mujica et al., 2019).

Regarding the first objective, three coping profiles were identified: a Low Coping Profile (LCP, 15.87%), a Mixed Engagement-Disengagement Profile (MEDP, 74.45%), and a High Emotion-Engagement Profile (HEEP, 9.68%). The LCP includes students who mobilise few coping resources, neither engaging with academic difficulties nor resorting to avoidance, a pattern that may reflect uncertainty or a still-developing regulatory repertoire (Aldridge & Roesch, 2008). The HEEP describes students who rely predominantly on emotion-focused engagement, combining strong affective involvement and efforts to regulate emotions with only moderate use of problem-focused strategies. This configuration suggests a tendency to remain emotionally connected to academic demands, while mobilising instrumental action in a more limited manner (Smart et al., 2025). Finally, the MEDP, the largest group, represents a heterogeneous configuration in which engagement and disengagement strategies coexist, indicating a tendency to alternate between confronting stressors and distancing oneself from them, depending on situational demands (Kato, 2023).

This differentiation is consistent with recent international evidence. For instance, Thomas et al. (2024) identified two broad orientations—approach-oriented patterns, characterised by active, problem-focused and engagement-based strategies, and avoidance-oriented patterns, defined by behavioural and cognitive withdrawal. Although the present study yielded a three-profile structure, these profiles map onto this distinction: the HEEP profile reflects a predominantly emotional form of engagement, LCP shows minimal mobilisation of either type of strategy, and MEDP represents an intermediate or mixed configuration where engagement and disengagement co-occur.

While these empirical distinctions help interpret the profiles, they also intersect with the students' developmental stage. Most participants were in late adolescence or emerging adulthood, a period characterised by ongoing consolidation of self-regulatory abilities and greater variability in coping responses (Folker et al., 2025; Wesarg-Menzel et al., 2023). This is particularly salient

during the transition to university, which requires renegotiating autonomy, identity and academic competence. Evidence from Takács et al. (2021) reinforces this interpretation, showing more variable coping patterns among younger university entrants. Such developmental variability is reflected in the predominance of the MEDP, which may represent an exploratory configuration, and in the smaller proportions of LCP and HEPP, profiles that may reflect more consolidated tendencies—either limited mobilisations of coping resources (LCP) or a predominantly emotional form of engagement (HEPP).

Following profile identification, the second objective examined dropout intention within each profile. The pattern of results indicated a clear gradient: the Low Coping Profile (LCP) had the highest proportion of students reporting dropout intention (43.24%), followed by the High Emotion-Engagement Profile (HEPP; 32.27%), whereas the Mixed Engagement-Disengagement Profile (MEDP) had the lowest percentage (22.88%). These proportions were consistent with the MANCOVA results, which showed significant differences between profiles in several dropout modalities—mainly reflecting higher scores in the LCP and lower scores in the MEDP—with small effect sizes.

Moreover, this gradient is consistent with psychological models of student persistence, which argue that coping processes shape appraisals, self-efficacy, and academic/emotional integration—the proximal antecedents of dropout intention (Bean & Eaton, 2001; Tinto, 2006). In person-centred research with university samples, recent latent-profile studies have consistently shown that mixed or intermediate coping configurations tend to be more adaptively positioned than profiles predominantly marked by avoidance or low resource-mobilisation, which are linked to greater distress and poorer academic adjustment (Schlesier & Westphal, 2023; Smart et al., 2025; Thomas et al., 2024).

Interpreted through the transactional framework, MEDP students may benefit from regulatory flexibility—the capacity to alternate between engagement and limited, situational disengagement in line with appraisal of controllability (Leslie-Miller et al., 2025)—thereby buffering stress appraisals, sustaining self-efficacy, and reducing dropout intention. In contrast, the LCP restricts options for problem-solving and emotion regulation, undermining perceived manageability (Smart et al., 2025) and increasing doubts about continuing academically. Finally, although the HEPP reflects adaptive emotional engagement, its moderate problem-focused component may leave instrumental academic demands insufficiently addressed (Alves et al., 2022), which may explain its middle position in dropout intention.

Across the three profiles, the most frequent modality of dropout was the intention to change to another degree program or another university. This pattern is in line with recent findings indicating that for many students, dropout intention reflects a process of academic reorientation rather than a rejection of the university system (Findeisen et al., 2024).

The third objective focused on the distribution of gender and field of study across the identified profiles. No statistically significant differences were found, indicating that women and men, as well as students from different disciplinary areas, were similarly distributed across the three profiles. This contrasts with previous findings reporting gender-linked patterns—such as women's greater use of emotional or avoidant strategies and men's tendency toward problem-

focused coping (Sitinjak et al., 2024; Vidic, 2024). Nonetheless, recent research suggests that gender differences in coping may be highly context-dependent, varying according to cultural, normative, and situational factors (Morales-Rodríguez et al., 2023). Our results support the idea that individual coping tendencies may not manifest uniformly across contexts, and that first-year students may display more heterogeneous and less gender-differentiated patterns.

Similarly, there were no statistically significant associations between coping profiles and knowledge areas. This contrasts with previous studies highlighting discipline-specific patterns in coping and stress perception (Bernardo et al., 2025), such as a predominance of engagement-based strategies in Health Sciences or Social Sciences (Rodríguez-Fernández et al., 2020; Vizoso et al., 2019) and more avoidant styles in Arts and Humanities (Waterhouse & Samra, 2026). The absence of differences may reflect the strong contextual pressures shared by first-year students across degrees, who all face comparable novelty, uncertainty and academic demands, regardless of disciplinary area (Linden et al., 2025; Parsons et al., 2023).

In summary, the results from a Spanish university context highlight the importance of coping strategies as a key psychological mechanism in explaining university dropout intention. The three identified profiles showed distinct patterns of vulnerability, with the MEDP appearing comparatively more protected—consistent with the adaptive value of regulatory flexibility—whereas the LCP and HEPP reflected more constrained or unbalanced regulatory resources. Although gender and field of study did not significantly differentiate profile membership, the overall pattern suggests that during the transition to university, coping may be shaped more strongly by shared developmental and contextual demands than by sociodemographic or disciplinary characteristics. Taken together, these findings underscore the importance of understanding students' coping tendencies within the specific features of the Spanish higher-education context, while offering potentially relevant insights for institutions facing similar first-year adjustment challenges.

Despite the findings, this study has some limitations. First, although the CSI-SF is a commonly used instrument in research on coping, this study did not provide fully satisfactory results regarding its psychometric properties. Consequently, our findings should be interpreted with caution, and they will need to be confirmed by future research.

Second, the cross-sectional design restricts the ability to explore the relationship between coping profiles and dropout intention over time. Future research could incorporate longitudinal methodologies to examine how these coping styles change over time and their impact on academic trajectories.

Finally, the sample was drawn from a single Spanish university. Replicating the study in other institutions and cultural contexts would help assess the generalisability of the identified profiles and clarify the influence of institutional characteristics.

These considerations also provide important context for interpreting the broader implications of the study. At a theoretical level, the identification of three coping profiles reinforces the value of person-centred approaches for understanding how students regulate academic stress during the transition to university. The findings are consistent with psychological models of persistence (Bean & Eaton, 2001), highlighting that coping shapes adaptation through appraisal processes, perceived controllability, and self-efficacy. In this regard, the comparatively more adaptive position of the MEDP

underscores the importance of regulatory flexibility—alternating between engagement and situational disengagement according to controllability appraisals—as a key mechanism in emerging adulthood (Zimmer-Gembeck & Skinner, 2016).

At an applied level, the results point towards strengthening students' regulatory repertoires, enabling complementary use of emotional and problem-focused strategies. As profile membership did not differ by gender or field of study, interventions may be most effective at a general institutional level, targeting first-year cohorts across programmes (Alves et al., 2022). Initiatives that foster constructive interpretations of academic challenges, promote selection of adaptive strategies, and support emotional and academic adjustment may contribute to reducing dropout intention in diverse university contexts.

Author Contributions

Ana B. Bernardo: Conceptualization, Funding Acquisition, Investigation, Project Administration, Supervision, Validation, Writing – Review & Editing. **Elena Blanco:** Conceptualization, Data Curation, Investigation, Visualization, Writing – Original Draft, Writing – Review & Editing. **Carlos Freire:** Data Curation, Formal Analysis, Methodology, Software, Visualization, Writing – Original Draft, Writing – Review & Editing. **María del Mar Ferradás:** Data Curation, Formal Analysis, Methodology, Software, Validation, Writing – Original Draft, Writing – Review & Editing.

Funding

This research was funded by the Spanish Ministry of Science and Innovation (ref. PID2022-141290NB-I00), by the Severo Ochoa Programme of the Principality of Asturias (Ref.: BP21-049. Candidate Elena Blanco) and by the European Union on behalf of the Principality of Asturias GRUPIN (ID2024000713). This funding source had no role in the design of this study, data collection, management, analysis and interpretation of data, writing of the manuscript, or the decision to submit the manuscript for publication.

Declaration of Interests

The authors declare that there are no conflicts of interest.

Data Availability Statement

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

References

- Addison, C. C., Campbell-Jenkins, B. W., Sarpong, D. F., Kibler, J., Singh, M., Dubbert, P., Wilson, G., Payne, T., & Taylor, H. (2007). Psychometric evaluation of a Coping Strategies Inventory Short-Form (CSI-SF) in the Jackson Heart Study Cohort. *International Journal of Environmental Research and Public Health*, 4(4), 289–295. <https://doi.org/10.3390/ijerph200704040004>
- Aldridge, A. A., & Roesch, S. C. (2008). Developing coping typologies of minority adolescents: A latent profile analysis. *Journal of Adolescence*, 31(4), 499–517. <https://doi.org/10.1016/j.adolescence.2007.08.005>
- Alves, S. A., Sinval, J., Lucas, L., Maroco, J., Ferreira, A., & Oliveira, P. (2022). Burnout and dropout intention in medical students: The protective role of academic engagement. *BMC Medical Education*, 22(1), Article 83. <https://doi.org/10.1186/s12909-021-03094-9>
- Bean, J. P., & Eaton, S. (2001). The psychology underling successful retention practices. *Journal College Student Retention: Research, Theory & Practice*, 3(1), 73–89. <https://doi.org/10.2190/6R55-4B30-28XG-L8U0>
- Bernardo, A. B., Blanco, E., Tuero, E., & Herrero, J. (2025). Academic stress and socio-demographic variables as predictors of university dropout intention. *Electronic Journal of Research in Education Psychology*, 23(66), 321–348. <https://doi.org/10.25115/ejrep.v23i66.10127>
- Bernardo, A. B., Esteban, M., Cervero, A., Tuero, E., & Herrero, F. J. (2022). Validation of the Early University Dropout Intentions Questionnaire (EUDIQ-R). *Journal of Higher Education Theory and Practice*, 22(10), 17–29. <https://doi.org/10.33423/jhetp.v22i10.5384>
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54(1), 106–148. <https://doi.org/10.1111/j.1467-6494.1986.tb00391.x>
- Buizza, C., Cela, H., Sbravati, G., Bornatici, S., Rainieri, G., & Ghilardi, A. (2024). The role of self-efficacy, motivation, and connectedness in dropout intention in a sample of Italian college students. *Education Science*, 14(81), Article 67. <https://doi.org/10.3390/educsci14010067>
- Cabanach, R. G., Souto-Gestal, A., & Franco, V. (2016). Stressor Academic Scale for the evaluation of academic stressors in undergraduates. *Revista Iberoamericana de Psicología y Salud*, 7(2), 41–50. <https://doi.org/10.1016/j.rips.2016.05.001>
- Coppola, L., Salvo, M., Sassano, G., del Valle, M., & Arias, C. (2024, December 4–6). *Diferencias de género en el uso de estrategias de afrontamiento en estudiantes universitarios* [Gender differences in the use of coping strategies among university students] [Paper presentation]. X Congreso Marplatense Internacional de Psicología, Buenos Aires, Argentina.
- Díaz-Mujica, A., Pérez-Villalobos, M. J., Bernardo, A. B., Cervero, A., & González-Pianda, J. A. (2019). Affective and cognitive variables involved in structural prediction of university dropout. *Psicothema*, 31(4), 429–436. <https://doi.org/10.7334/psicothema2019.124>
- European Commission. (2025). *Main indicators areas. Tertiary education*. Education and Training Monitor 2025, Publications Office of the European Union. <https://op.europa.eu/webpub/eac/education-and-training-monitor/en/monitor-toolbox/themes/tertiary-education.html>
- Findeisen, S., Brodsky, A., Michaelis, C., Chimmelpenningh, B., & Seifried, J. (2024). Dropout intention: A valid predictor of actual dropout? *Empirical Research in Vocational Education and Training*, 16, Article 10. <https://doi.org/10.1186/s40461-024-00165-1>
- Folker, A., Bertrand, C., Hong, Y. L., Steinberg, L., Duell, N., Chang, L., Giunta, L. D., Dodge, K. A., Gurdal, S., Junla, D., Lansford, J. E., Oburu, P., Pastorelli, C., Skinner, A. T., Sorbring, E., Bornstein, M. H., Tirado, L. M. U., Yotanyamaneewong, S., Alampat, L. P., ... Deater-Deckard, K. (2025). A longitudinal study of adolescent-to-young adult executive function development in seven countries. *Developmental Science*, 28(4), Article e70040. <https://doi.org/10.1111/desc.70040>
- González-Casas, D., Dorado-Barbe, A. I., & Gálvez-Nieto, J. L. (2025). Profiles of coping strategies in university students in relation to personal and academic variables. *Educación XX1*, 28(1), 103–130. <https://doi.org/10.5944/educxx1.39618>

- Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, 11(1), 36–53. <https://doi.org/10.1037/1082-989X.11.1.36>
- IBM Corp. (2019). *SPSS STATISTICS for Windows, Version 26.0*. IBM Corp.
- Kato, T. (2023). Effect of coping flexibility on life satisfaction of college students: A longitudinal study. *Applied Research in Quality of Life*, 18(6), 2981–2991. <https://doi.org/10.1007/s11482-023-10215-7>
- Lanza, S. T., Flaherty, B. P., & Collins, L. M. (2003). Latent class and latent transition analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (pp. 663–685). Wiley.
- Lazarus, R. & Folkman, S. (1986). *Estrés y procesos cognitivos* [Stress and cognitive processes]. Ediciones Martínez Roca.
- Leslie-Miller, C. J., Joormann, J., & Quinn, M. E. (2025). Coping flexibility: Match between coping strategy and perceived stressor controllability predicts depressed mood. *Affective Science*, 6, 94–103. <https://doi.org/10.1007/s42761-024-00275-9>
- Linden, K., Teakel, S., & van der Ploeg, N. (2025). Seen and supported: Implementing a university-wide transition framework to improve first-year success. *Student Success*, 16(3), 67–78. <https://doi.org/10.63608/ssj.3761>
- Loewenthal, K. M., & Lewis, C. A. (2001). *An introduction to psychological tests and scales*. Routledge.
- López-Aguilar, D., Álvarez-Pérez, P. R., & Ravelo-González, Y. (2022). Adaptability skills and the intention to drop out in university students. *Revista de Investigación Educativa*, 40(1), 237–255. <https://doi.org/10.6018/rie.463811>
- Maluenda-Albornoz, J., Infante-Villagrán, V., Galve-González, C., Flores-Oyarzo, G., & Berrios-Riquelme, J. (2022). Early and dynamic socio-academic variables related to dropout intention: A predictive model made during the pandemic. *Sustainability*, 14(2), Article 831. <https://doi.org/10.3390/su14020831>
- Martín-Arbós, S., Castarlenas, E., Morales-Vives, F., & Dueñas, J. M. (2024). Students' thoughts about dropping out: Sociodemographic factors and the role of academic help-seeking. *Social Psychology of Education*, 27, 2079–2092. <https://doi.org/10.1007/s11218-024-09903-5>
- Martínez, I. M., Meneghel, I., & Peñalver, J. (2019). Does gender affect coping strategies leading to well-being and improved academic performance? *Revista de Psicodidáctica*, 24(2), 111–119. <https://doi.org/10.1016/j.psicod.2019.01.003>
- Măta, L., Cirtita-Buzoianu, C., Amalancei, B. M., Cojocariu, V. M., & Mares, G. (2024). Students' perception of the determining factors of university dropout. *Romanian Journal of Multidimensional Education*, 16(4), 127–153. <https://doi.org/10.18662/rrem/16.4/910>
- Ministerio de Ciencia, Innovación y Universidades. (2026). *Datos y cifras del Sistema Universitario Español. Publicación 2025-2026* [Facts and figures on the Spanish University System. Publication 2025-2026]. Secretaría General Técnica, Centro de Publicaciones, Ministerio de Ciencia, Innovación y Universidades. <https://www.ciencia.gob.es/InfoGeneralPortal/detalle-publicacion/MICIU/Datos-y-cifras-del-Sistema-Universitario-Espanol-2025-2026.html>
- Mize, J. L. (2024). Depression, anxiety, and stress symptoms and coping strategies in the context of the sudden course modality shift in the Spring 2020 semester. *Current Psychology*, 43(5), 3944–3955. <https://doi.org/10.1007/s12144-023-04566-5>
- Morales-Rodríguez, F. M., Martínez-Ramón, J. P., Corvasce, C., & Morales Rodríguez, A. M. (2023). Coping strategies in adolescence: A preliminary study of gender differences. *Acción Psicológica*, 20(1), 21–32. <https://doi.org/10.5944/ap.20.1.39179>
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide, 6th ed.* Muthén & Muthén.
- Parsons, J. M., McColl, M. A., Martin, A., & Reynard, D. (2023). Transition and accommodation experiences of students with disabilities: A qualitative study. *Journal of Postsecondary Education and Disability*, 36(2), 179–192. <https://eric.ed.gov/?id=EJ1434801>
- Pi, Y., Wang, Y., Li, X., & Zhang, F. (2024). Reciprocal longitudinal relation between self-efficacy and coping style: The explanatory mechanism of future self-continuity. *Journal of Adolescence*, 96(5), 1078–1090. <https://doi.org/10.1002/jad.12318>
- Reyes-Pérez, V., Muñoz-López, D. B., & Alcázar-Olán, R. (2021). Emotions and coping strategies when university students pass their exams. *Nova Scientia*, 12(25), 1–23. <https://doi.org/10.21640/ns.v12i25.2344>
- Rodríguez-Fernández, A., Maury-Sitjago, E. A., Troncoso-Pantoja, C., Morales-Urzuá, M., & Parras-Flores, J. (2020). Academic stress and coping strategies in students of health degrees in Santiago de Chile. *EduMeCentro*, 12(4), 1-16. <https://dialnet.unirioja.es/servlet/articulo?codigo=9525226>
- Rogers, M. E., Creed, P. A., Searle, J., & Nicholls, S. L. (2016). Coping with medical training demands: Thinking of dropping out, or in it for the long haul. *Studies in Higher Education*, 21(9), 1715–1732. <https://doi.org/10.1080/03075079.2014.999318>
- Schlading, S., Rölle, D., & Drüge, M. (2024). Is intrinsic motivation related to lower stress among university students? Relationships between motivation for enrolling in a study program, stress, and coping strategies. *Education Sciences*, 14(8), Article 851. <https://doi.org/10.3390/educsci14080851>
- Schlesier, J., & Westphal, A. (2023). Exploring student teachers' latent coping profiles and their relations to perceived stress and stressors. *Stress and Health*, 40, Article e3422. <https://doi.org/10.1002/smi.3422>
- Sitinjak, C., Rakhmawati, E., Almasitoh, U. H., & Febrianty, A. (2024). Unveiling gender differences in adolescent coping: Insights from high school students in Jakarta. *Indigenous*, 9(2), 107–120. <https://doi.org/10.23917/indigenous.v9i2.4308>
- Smart, I., McCabe, M., Bird, L. J., Byrne, M. L., & Cornish, K. (2025). Australian university student coping profiles and psychological distress: A latent profile analysis. *Studies in Higher Education*, 50(1), 93–106. <https://doi.org/10.1080/03075079.2024.2329748>
- Takács, R., Takács, S., Kárász, J. T., Horváth, Z., Oláh, A. (2021). Exploring coping strategies of different generations of students starting university. *Frontiers in Psychology*, 12, Article 740569. <https://doi.org/10.3389/fpsyg.2021.740569>
- Thomas, C. L., Nair, M., & Ganske, J. K. (2024). Coping profiles and mental health outcomes in American university students: A latent profile analysis. *Psychological Reports*, 0(0). <https://doi.org/10.1177/00332941241308501>
- Tinto, V. (2006). Research and Practice of Student Retention: What Next? *Journal of College Student Retention: Research. Theory & Practice*, 8(1), 1-19. <https://doi.org/10.2190/4YNU-4TMB-22DJ-AN4W>
- Torquati, J., Lombardi, C., Cohen, R., Matthews, A., & Brophy-Herb, H. (2025). Mindfulness, coping, and disruptions as predictors of college

- student distress at the beginning of the COVID-19 pandemic. *Journal of American College Health*, 73(4), 1406–1417. <https://doi.org/10.1080/07448481.2023.2277184>
- Tous-Pallarés, J., Espinoza-Díaz, I. M., Lucas-Mangas, S., Valdivieso-León, L., & Gómez-Romero, M. R. (2022). CSI-SF: Psychometric properties of Spanish version of the coping strategies inventory - short form. *Annals of Psychology*, 38(1), 85–92. <https://doi.org/10.6018/analesps.478671>
- Vidal-Conti, J., Muntaner-Mas, A., & Palau, P. (2018). Differences of stress and coping according to gender and how it affects the academic performance of university students. *Contextos Educativos*, 22, 181–195. <https://doi.org/10.18172/con.3369>
- Vidic, Z. (2024). Gender differences on coping, stress, resilience and mindfulness within an academic course intervention with a mindfulness meditation component. *Current Psychology*, 34, 28241–28251. <https://doi.org/10.1007/s12144-024-06395-6>
- Vizoso, C., Arias-Gundín, O., & Rodríguez, C. (2019). Exploring coping and optimism as predictors of academic burnout and performance among university students. *Educational Psychology*, 39(6), 768–783. <https://doi.org/10.1080/01443410.2018.1545996>
- Wang, M. C., Deng, Q., Bi, X., Ye, H., & Yang, W. (2017). Performance of the entropy as an index of classification accuracy in latent profile analysis: A Monte Carlo simulation study. *Acta Psychologica Sinica*, 49(11), 1473–1482. <https://doi.org/10.3724/SP.J.1041.2017.01473>
- Waterhouse, P., & Samra, R. (2026). University students' coping strategies to manage stress: A scoping review. *Educational Review*, 78(2), 127–167. <https://doi.org/10.1080/00131911.2024.2438888>
- Wesarg-Menzel, C., Ebbes, R., Hensums, M., Wagemaker, E., Zaharieva, M. S., Staaks, J. P. C., van den Akker, A. L., Visser, I., Hoeve, M., Brummelman, E., Dekkers, T. J., Schuitema, J. A., Larsen, H., Calonnesi, C., Jansen, B. R. J., Overbeek, G., Huizenga, H. M., & Wiers, R. W. (2023). Development and socialization of self-regulation from infancy to adolescence: A meta-review differentiating between self-regulatory abilities, goals, and motivation. *Developmental Review*, 69, Article 101090. <https://doi.org/10.1016/j.dr.2023.101090>
- Zimmer-Gembeck, M. J., & Skinner, E. A. (2016). The development of coping: Implications for psychopathology and resilience. In D. Cicchetti (Ed.), *Developmental psychology: Risk, resilience, and intervention* (pp. 485–545). John Wiley & Sons. <https://doi.org/10.1002/9781119125556.devpsy410>

Article

PROVAL: A Comprehensive Protocol for Developing Validation Studies

Isabel Benítez^{1,2}, Andrés González-Gómez^{1,2} and José Luis Padilla^{1,2}

¹ University of Granada (Spain)

² Mind, Brain and behavior Research Center (CIMCYC) Granada (Spain)

ARTICLE INFO

Received: 23/12/2025

Accepted: 05/05/2026

Keywords:

Validity evidence

PROVAL

Validation protocol

Sources of validity evidence

ABSTRACT

Introduction: Validity is a fundamental aspect of psychological assessment because it examines the extent to which interpretations of the scores produced by an instrument support its intended purpose. While validation studies are frequently conducted, they often emphasize specific phases or validation methods, such as pilot testing or expert judgments, while neglecting current additional sources of validity evidence that could improve an instrument's psychometric properties. This limitation is partly due to a lack of detailed easy-to-use guidelines, which make validation processes inaccessible to non-expert researchers and practitioners. Additionally, the traditional focus on quantitative methods often overlooks the potential benefits of complementary approaches to validation that will help people design more comprehensive validation studies. **Method:** To address these challenges, the present study introduces PROVAL, a nine-step protocol designed to guide researchers in conducting validation studies. **Results:** PROVAL promotes a reflective approach to help researchers and practitioners identify and prioritize the most important sources of validity evidence for their specific validation requirements. **Conclusions:** An illustrative application of PROVAL is presented, demonstrating its usefulness in optimizing resources and ensuring instruments effectively fulfill their intended purposes.

PROVAL: un Protocolo Integral para el Desarrollo de Estudios de Validación

RESUMEN

Introducción: La validez es un aspecto fundamental en la evaluación psicológica, ya que examina en qué medida las interpretaciones de las puntuaciones obtenidas mediante un instrumento respaldan su uso previsto. Aunque los estudios de validación son frecuentes, habitualmente reflejan las mismas fases y métodos (estudios piloto o evaluación de personas expertas), descuidando otras fuentes de evidencia que podrían mejorar las propiedades psicométricas de los instrumentos. Esta limitación se debe, en parte, a la falta de directrices, lo que a su vez dificulta que profesionales en áreas aplicadas y en investigación puedan desarrollar estudios de validación. Además, el enfoque tradicional centrado en métodos cuantitativos suele ignorar los beneficios potenciales de enfoques complementarios que contribuyan al diseño de estudios de validación más integrales. **Método:** Para abordar estos desafíos, el presente estudio introduce PROVAL, un protocolo de nueve pasos diseñado para guiar a los investigadores en la realización de estudios de validación. **Resultados:** PROVAL promueve un enfoque reflexivo que ayuda a identificar y priorizar las fuentes de evidencia de validez más relevantes según las necesidades específicas. **Conclusiones:** Se presenta una aplicación ilustrativa de PROVAL, que demuestra su utilidad para optimizar recursos y garantizar que los instrumentos cumplan eficazmente con sus propósitos previstos.

Palabras clave:

Evidencias de validez

PROVAL

Protocolo de validación

Fuentes de evidencia de validez

Validity is one of the most extensively discussed topics in psychometrics. Many authors have proposed definitions, theoretical proposals, and validation methods over the years (Crocker & Algina, 1986; Cronbach & Meehl, 1955; Lissitz & Samuels, 2007; Messick, 1989; Sireci, 2009; Zumbo, 2009). However, several long-standing issues and unresolved problems remain salient today. Validity is still mainly regarded as a property of an instrument, although Messick (1989) conceptualized it as an integrated and ongoing process that requires multiple perspectives to support the intended use of test and questionnaire scores. Likewise, most of validation studies follow the same routine process and structure, despite Kane (1992) introducing the argument-based approach to put the focus on the specific arguments supporting the creation of a test. This author also proposed practical guidelines for researchers and practitioners (Kane, 2006, 2013a), but these are not habitually considered. Despite the key role of the construct, the focus of validation studies should be the intended purpose and the interpretations of test scores, as Sireci (1998, 2009, 2013) emphasizes. However, the intended purpose is not even identified in most validation studies. Finally, researchers approach validity as an isolated issue, but it should incorporate cognitive and contextual variables of individuals responding to test items because, as Zumbo and colleagues (Chen & Zumbo, 2017; Zumbo & Hubley, 2017) noted, the “ecology” (i.e., personal, family, and societal factors), influences how people respond. The inability of psychometricians to eliminate these traditional and limiting ideas and practices has led to validity being treated as a formality, causing it to lose its primary purpose: ensuring that an instrument provides adequate measures of a construct that are suitable for the intended purpose. Addressing this challenge involves getting back to the concept and reviewing the main foundations of validity.

Starting with the concept, there appears to be a broad current consensus about the definition provided in the latest edition of the *Standards for educational and psychological testing* (hereafter, *Standards*) (American Educational Research Association [AERA] et al., 2014), where validity is defined as “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (AERA et al., 2014, p. 11). The two most recent editions of the *Standards* (AERA et al., 1999, 2014), conceptualize validity as a unified process centered on the construct, thereby burying the traditional division in types of validity. This unified perspective frames validity as a continuous process in which multiple sources of “data”, such as psychometrics and expert judgments, serve as “evidence” supporting (or failing to support) the interpretations derived from test scores. Specifically, the *Standards* describe five potential sources of evidence: test content, response processes, internal structure, relationships with other variables, and testing consequences. Depending on the validity argument formulated to guide the validation study and the specific purposes of the assessment tool, evidence will be collected from the most relevant and appropriate sources; however, there is still no clear agreement on which sources should be prioritized.

Among the five sources, evidence based on test content examines the degree to which a test’s content adequately represents the intended construct (for specific approaches, see Sireci & Faulkner-Bond, 2014). Evidence based on response processes evaluates the alignment between the construct being measured and participants’ actual response processes, aiming to identify factors that may elicit responses unrelated to the construct (see, for example, Engelhardt

& Goldhammer, 2019; Neuert & Lenzner, 2016; Padilla & Benítez, 2014; Rushton et al., 2015). Evidence based on internal structure focuses on whether the relationships among items and dimensions of the test are consistent with the theoretical interpretation of the scores (for a review of relevant procedures, see Rios & Wells, 2014). Evidence based on relationships with other variables assesses the extent to which associations between test scores and external measures align with the theoretical expectations derived from the construct (for an illustration, see Oren et al., 2014). Finally, evidence based on testing consequences concerns the evaluation of the intended and unintended outcomes of testing and assessment processes for all stakeholders starting by the examinees themselves. Several methodological approaches have been proposed to address this aspect (Ilescu & Greiff, 2021; Lane, 2014), particularly to support the formulation of interpretation and use argument (IUA) that “explicitly links the inferences from tasks to decisions and uses, and then to the resulting consequences” (Lane & Marion, 2025, p. 260). In fact, these authors expand the definition of validity presented in the *Standards* to explicitly incorporate the role of the consequences, referring to validity as “the extent to which theory and evidence support or refute proposed and actual or enacted test score interpretations, uses, and consequences...” (p. 192).

Despite numerous advances, guidance on how to gather and prioritize different sources of validity evidence remains relatively limited. To address this gap, several practical frameworks have been proposed. Both Kane and Sireci, in their previously cited works, developed models and general recommendations to guide validation by outlining steps and recommendations for identifying threats to validity and designing studies to examine them. More recently, Sireci and Benítez (2023) introduced a practical guide for conducting validation studies, detailing procedures for addressing the five sources of validity evidence and proposing a structured sequence of steps to support the validation process. This guide represents a valuable resource for approaching validation from a comprehensive perspective, offering detailed information on procedures for collecting different sources of validity evidence. However, both this guide and other existing frameworks often remain inaccessible to researchers and practitioners who are not experts in psychometrics. This limitation highlights the need to disentangle the validity process, make it more user-friendly, and present it in ways that facilitate rigorous yet practical application. Specifically, there is a need to support researchers in making decisions throughout the design and development of validity studies.

With this goal in mind, several studies have sought to illustrate how proposed frameworks can be applied in specific contexts. For example, Cook et al. (2015) developed a practical introduction to Kane’s framework to make it “accessible and applicable to a wide range of assessment tools and activities” (p. 561). Other authors have focused on outlining the essential elements of validation processes to establish a basic level of confidence in what an instrument measures. Yasir (2016), for instance, summarizes the processes of adaptation and validation, providing a detailed description of the tasks and steps involved in developing assessment instruments suitable for evaluating public health variables in cross-cultural studies. Although Yasir’s framework relies on definitions of validity that differ from those in the *Standards*, it nonetheless seeks to clarify the relationship between the concept of validity and the contributions of specific statistical techniques. Additionally, other tools illustrate

the necessary steps for creating or analyzing the psychometric properties of assessment instruments in specialized domains. For example, [Tsang et al. \(2017\)](#) outline the procedures for developing, adapting, and validating instruments in the field of pain medicine. Their guide offers step-by-step recommendations for producing sound measurement instruments and includes a glossary to make technical terminology more accessible. However, these proposals overlook some of the more established aspects of validity, such as identifying the intended purpose or the integration of multiple sources of validity evidence.

These specific applications offer valuable guidance for researchers seeking to develop or employ psychometrically sound instruments. However, many of these guides present “automatic” or overly standardized procedures that discourage reflection on the unique requirements of each instrument and the assessment context. Consequently, to promote a deeper and more comprehensive understanding of the constructs being measured it is essential to design validation processes that move beyond the mere fulfillment of formal requirements. This aim underscores the responsibility of psychometricians to provide resources that facilitate the rigorous development of validation practices in both research and applied contexts. As emphasized in the *Standards*, “validation should be understood as the process of constructing and evaluating arguments for and against the intended interpretations of scores and their relevance for the proposed use” ([AERA et al., 2014](#), p. 11). The statement highlights the importance of engaging in thoughtful, context-sensitive validation rather than following routine and predetermined steps. It also calls for the continuous analysis of psychological and educational phenomena, recognizing that different constructs cannot be measured through identical procedures but instead require tailored approaches. This perspective shifts the focus toward the careful selection of methods and the critical understanding of tools that enable a more profound exploration of the phenomenon under study.

To address previous limitations, we need methodological approaches that prioritize the intended purpose of the instrument and help articulating it, foster reflection on its requirements, assessment context, and allow for the flexible integration of multiple sources of information. Therefore, it is necessary to move beyond traditional approaches that restrict the validation process to a limited set of questions and instead promote alternatives that consider the full range of possibilities for addressing potential threats. In this regard, pragmatism, one of the main foundations of Mixed Methods Research (MMR), becomes central. MMR has transformed the way research designs are conceptualized, as it enables researchers to approach phenomena from an open perspective in which specific procedures are not predetermined but are selected based on their ability to address the research question ([Creswell & Plano-Clark, 2011](#); [Tashakkori & Teddlie, 1998](#)). Grounded in pragmatism epistemology foundations, MMR emphasizes the centrality of the research objective, using it as the key element guiding the selection of design, techniques, and procedures within a study ([Johnson & Christensen, 2019](#)). The contributions of MMR, along with empirical evidence supporting its usefulness for addressing complex phenomena, suggest that this approach may also offer valuable insights in the context of validity. A closer examination of the principles underlying both “modern” validity theory and MMR reveals a clear conceptual alignment between them. From the earliest discussions of validity, scholars

have emphasized its dynamic nature, framing validation as an evolving and context-dependent process. [Messick \(1989\)](#) described validation as a hypothesis-testing endeavor that involves the use of multiple procedures to gather information, the consistency of which is then evaluated to determine the extent to which the intended interpretations can be supported. He also underscored the influence of context and group characteristics on test scores, as well as the limitations of purely local validation efforts, issues later expanded upon in frameworks such as the ecological model ([Chen & Zumbo, 2017](#)). Moreover, the integration of data from multiple sources is a recurrent theme across these perspectives, as the accumulation of validity evidence inherently involves combining information to strengthen the overall validity argument.

MMR provides two others important contributions to the discussion of validity. On one hand, the incorporation of qualitative procedures to offer relevant information about the same phenomena examined through quantitative approaches, but from a different perspective ([Dellinger & Leech, 2007](#)). This idea helps recover aspects of validity that have been largely overlooked, particularly those sources of evidence that are, by their nature, difficult to access through quantitative methods like response processes and intended and unintended testing consequences. MMR can help place these sources on an equal footing with other types of evidence traditionally included in validation studies. The reviews by [Cizek et al. \(2008\)](#) and [Zumbo \(2025\)](#) highlight how empirical studies on validity often fail to incorporate certain essential elements. According to their findings, only a small minority of studies up to the date of publication adopt a unified view of the concept of validity, with most treating validity as a property of the test itself. Consequently, the conceptualization of validity in terms of evidence is often missing. Among the studies that do adopt this perspective, references to evidence based on response processes and testing consequences are rare, reflecting, to some extent, the uncertainty researchers face when addressing these sources of information. Adopting approaches that incorporate qualitative data and sources of evidence not typically collected does not imply that these elements should be always mandatory. In fact, validation studies based on more traditional sources of validity evidence can, in many cases, provide sufficient support. However, in other situations, the characteristics of the target construct, the assessment instruments or context itself may introduce threats related to response processes and the consequences of evaluation. [Kane \(2013a\)](#) emphasized that “our inability to specify all potential consequences in precise quantitative terms does not eliminate the need to evaluate the consequences” (p. 61). Moreover, fairness in measurement has gained prominence with the rise of comparative studies across linguistic, cultural and sociodemographic groups that can involve decision-making processes capable of significantly affecting the lives of those being assessed differently. In fact, fairness cannot be separated from the concept of validity ([Lane & Marion, 2025](#)). Therefore, evaluating consequences can, in many instances, contribute to ensuring equity among individuals and groups.

On the other hand, MMR can also contribute to validation by helping to “summarize” and “report” validity arguments, an unresolved issue in the latest *Standards*, as [Lane and Marion \(2025\)](#) point out, and one they describe as a “challenging activity” when professionals and practitioners must integrate and synthesize diverse types of validity evidence and supporting arguments. MMR emphasizes the development of a “narrative” that synthesizes and summarizes both

quantitative and qualitative findings when evaluating intervention programs, while also being adaptable to diverse audiences, including stakeholders, practitioners, academics, and policymakers.

MMR is not conceived as an alternative to traditional validation studies but rather as an approach that enriches the information obtained through more commonly used quantitative methods. The relevance of MMR within the validation process becomes evident when considering some of its defining features, such as its nature as a hypothesis-testing process, its dynamic character, and the centrality of the study's purpose (Johnson & Christensen, 2019). Furthermore, from a modern perspective on validity, there is a clear call to incorporate multiple sources of evidence. In this context, MMR can be viewed as a response to this call, as it facilitates the integration of information from often overlooked sources of evidence, such as response processes or the consequences of test use.

Some counterarguments to these ideas have also emerged in literature. For example, Russell (2022) contended that it is not possible to synthesize or integrate evidence, that test developers are not responsible for the potential misuse of scores, and that certain arguments, such as those related to consequences, cannot be addressed until some time has passed after the evaluation. However, integration becomes meaningful when it involves gathering different perspectives on the same phenomenon from the "parts" involved in the assessment to inform decision-making; while we cannot be accountable for all possible uses, guidelines can be established to help prevent misuse; and although consequences cannot be evaluated *a priori*, they can indeed be anticipated.

The previous reflections on measurement highlight several challenges that we, as psychometricians, must address. As argued, the limited practical applicability of proposals for developing validity studies, combined with the inherent dynamism of psychological and educational constructs in evolving contexts, which pose ongoing challenges for validation, has led researchers and practitioners to rely on tools that offer simplified validation processes, make use of available data, or involve the design of readily feasible empirical studies. However, as emphasized in the *Standards*, validation should aim to achieve a comprehensive understanding of the phenomenon under study, enabling the confirmation of the appropriateness of the interpretations made based on the test scores.

Three key considerations should guide the development of an approach that facilitates both the evaluation of how test-takers' responses support the intended interpretations and uses of scores, and the understanding of the arguments that justify these interpretations and uses. The first concerns the conceptualization of validity in terms of sources of evidence that can inform all aspects of the phenomenon under study, thereby ensuring that potential threats to the proposed interpretations are identified. The second involves adopting a methodological framework capable of addressing a phenomenon as complex as validity while providing guidance for integrating data from multiple sources of evidence and incorporating contextual factors that may influence measurement. The third relates to selecting procedures and techniques that best address the validation projects, including qualitative methods that allow for the examination of aspects of the phenomenon not accessible through quantitative approaches, such as the causes or origins of statistically identified factors. These methods should also facilitate the collection of less frequently explored types of evidence, such as those based on response processes and testing consequences.

All of this underscores the need for a validation protocol that, while methodologically rich and rigorous, provides the necessary flexibility to accommodate validation processes for instruments with varying characteristics, assessing constructs of different natures in diverse contexts. This must be done in such a way that, as described in the *Standards*, it is possible to generate "a coherent argument regarding the degree to which evidence and theory support the interpretation of test scores for specific uses" (AERA et al., 2014, p. 21). Such a protocol should also encourage potential users to reflect on validity arguments, which should be the cornerstone of the process beyond the specific tools required.

For these reasons, the present study aims to propose PROVAL, a mixed-methods protocol for obtaining validity evidence, designed to guide the formulation of validation studies that support the intended uses and interpretations of instruments assessing psychological and educational variables across various contexts. PROVAL guides the design and development of validation studies by posing specific questions that facilitate the application of descriptions from previous frameworks to concrete situations. In other words, PROVAL not only outlines the steps for validation but also indicates how to construct validity arguments by articulating the intended purpose and tracing the full path to identifying potential threats. The proposed protocol is intended as a starting point for incorporating innovations related to validity theory, guidance on conducting validation studies, or new procedures for gathering validity evidence. This work also aims to illustrate the use of PROVAL by showing a specific application.

Description of PROVAL

PROVAL is a proposal to address validity studies based on both the theoretical and practical approaches available in literature and the demands of professionals and researchers. PROVAL was developed following the previously described guidelines and tools. Specifically, it draws on the phases proposed by Sireci (2013) and Sireci and Benítez (2023), as well as on the guidelines for formulating validity arguments proposed by Kane (1992). Within the argument-based approach to validity, Kane (1992, 2006, 2013a, 2013b) distinguishes between the interpretation/use argument (IUA), which specifies the intended interpretations and uses of test scores, and the validity argument, which evaluates the plausibility of these interpretations based on empirical evidence. Although this distinction has been highly influential in contemporary validity theory, for the purposes of the PROVAL framework, we adopt a more integrated and operational use of the term "validity argument," consistent with the interpretation proposed by Sireci (2013), in which both the specification of intended interpretations and the evaluation of supporting evidence are considered within a unified argumentative structure. This choice was motivated by the applied nature of PROVAL, which aims to translate validity theory into a stepwise protocol that can be readily implemented by researchers and practitioners. Importantly, this does not imply a departure from Kane's framework, but rather a pragmatic adaptation of terminology to facilitate the operationalization of argument-based validation within a structured methodological guide. In line with the definition of the sources of evidence outlined in the *Standards* and the recommendations for the most appropriate procedures to gather each type of evidence, a series of steps are established to guide and structure the design and development of a validation study that

focuses on the demands arising from the intended meaning of the scores and the use of the target instrument.

PROVAL is a protocol composed of nine steps, each representing a single activity or task related to the validation process. This means that traditional questions, such as “articulating the intended purposes,” can be divided into multiple parts to ensure that users engage in a thorough reflection on all relevant elements, rather than providing general statements that omit important details. Each step includes guidelines aimed at helping researchers reflect on the elements that, depending on the content of the inference and the kind of instrument, may be relevant for obtaining validity evidence. To this end, each step is divided into three sections: description, reflections, and results. The first section provides a brief description of what the step entails. The second section outlines the questions that researchers and practitioners must ask themselves. The answers to these questions lead to the results in the third section. PROVAL is designed to be used for the validation of inferences from measurement instruments that assess psychological and educational variables. The application of PROVAL will culminate in providing evidence of the test’s utility in achieving the intended objective and offering arguments to support the intended purpose and scores’ interpretations. Both the Spanish and the English version of PROVAL are available in appendixes 1 and 2 in the following link: <https://osf.io/zcxjq/overview>. Table 1 summarizes the steps of the PROVAL protocol, a framework designed to guide the design and development of validation studies focused on the specific requirements of measurement instruments.

Although PROVAL is an evidence-based proposal, one of its main contributions is its accessibility to users and its responsiveness to practitioners’ needs. Consequently, the protocol will be regularly updated as additional evidence is collected, incorporating clarifications and improvements to continue facilitating the validation process. Final versions will be made available in the “Current version of PROVAL” folder within the OSF project.

Illustrating the Application of PROVAL with the Spanish Version of the Inventory of Attitudes Toward Sexist/Nonsexist Language (IASNL-G-S)

To make easier the understanding of each step proposed in PROVAL, the following presents its application to the Spanish version of the *Inventory of Attitudes Toward Sexist/Nonsexist Language* (IASNL-G-S). The instrument represents a typical research tool in psychology, exhibiting many of the common characteristics researchers encounter when selecting a scale. Additionally, it presents several challenges related to cultural considerations and the potential consequences for the assessed participants. The IASNL-G is an instrument developed by Parks and Roberton (1998) to evaluate attitudes toward the use of sexist/nonsexist language. The instrument consists of 21 items scored on a 5-point Likert-type scale, where lower scores indicate a negative attitude, and higher scores reflect positive attitudes. The items are divided into three sections: 1) The first section asks participants to rate their agreement with 12 statements; 2) the second section requires participants to indicate the level of sexism of specific words and expressions in four items; and 3) the final section includes five items that ask participants about their habitual use of certain expressions.

Three researchers collaborated on the Spanish adaptation of the IASNL-G (IASNL-G-S), employing a committee approach (Harkness & Schoua-Glusberg, 1998). The committee approach to translation design follows International Test Commission Guideline TD-2 (5), which advocates for a double translation and reconciliation procedure (International Test Commission, 2017). Two researchers independently translated the original English version into Spanish, while a third researcher acted as the adjudicator. The adjudicator reviewed and compared both translations, resolving discrepancies to create a final integrated version that incorporated contributions from both translators. The adjudicator’s role involved two main tasks. First, when both translators produced identical or comparable translations, the adjudicator accepted the consensus version. In cases where one translation was clearer, this version was selected. Second, for elements where discrepancies arose, the adjudicator marked specific items for discussion. A joint meeting was held to address these discrepancies, where each researcher presented their rationale, and final decisions were made on the best translation choices. All three researchers were familiar with the content of the scale and

Table 1
Summary of the PROVAL Validation Protocol

Step	Title	Description	Main Result
1	Intended purpose and scores’ interpretations	Defines the test purpose, context of use, targeted population, and the decisions to be made based on scores.	Formulation of the intended purpose and score interpretations.
2	Formulation of validity arguments and assumptions	Identifies arguments based on the five sources of validity evidence from the <i>Standards</i> (AERA et al., 2014).	Formulation of validity assumptions and hypotheses to be investigated.
3	Prioritization of validation studies	Organizes the validity assumptions based on their relevance according to the test’s intended use (e.g., diagnosis, selection).	A prioritized list of assumptions to guide the study.
4	Nature of the evidence needed	Determines who can provide the information (participants) and whether data will be quantitative or qualitative.	Identification of necessary instruments, participants, and specific data requirements.
5	Selection of collection and analysis procedures	Identifies appropriate methods for data collection and the specific analytical techniques to be used.	Detailed description of the study’s phases and methodology.
6	Design of the validation study	Combines previous steps to establish the order, priority, and nature of the design (qualitative, quantitative, or mixed).	Final configuration and illustration of the validation study design.
7	Development of validation studies	Executes data collection according to the design, addressing recruitment and ethical considerations.	Preparation of all materials and successful data collection.
8	Data analysis and interpretation of results	Performs analyses and relates findings back to the initial validity assumptions.	Evidence supporting or refuting the validity assumptions.
9	Integration of evidence and conclusions	Synthesizes all findings to make a final decision on the relevance of the intended interpretations.	Final conclusions on the usefulness and validity of the test.

had experience with methodological aspects of translation and adaptation processes.

The application of PROVAL framework in the present study shows the first seven steps, which cover everything from outlining the intended use of the instrument to describing the complete design of the validation study and main aspects related to the development of the validation studies. In each step, responses to the reflections and the results derived from those reflections are provided. Appendixes 3 and 4 show the application of PROVAL to the IASNL-G-S in both Spanish and English (<https://osf.io/zcxjq/overview>). The illustrated steps cover the main contributions of PROVAL, although the documents will be updated once the validation studies are completed and the remaining steps are finalized. This will provide a complete example of a comprehensive validation study, including the interpretation and integration of results. Both the protocol and the illustration are conceived as dynamic tools that will be updated to incorporate innovations related to the concept of validity and derived from evidence obtained through the application of the tool in different contexts.

The first step presents information related to the intended purpose and characteristics of both the target population and the instrument. Since this is an adaptation, the information was derived from the original version, which included details about score interpretations. When developing new measures, the scores' interpretations should be described in that step. In the example, the results from step 1 include the description of the scale, the scores interpretations and the target population.

Step 2 evaluates whether available information in the scientific literature about the intended instrument can be interpreted as validity evidence. An affirmative response would apply if prior studies provided validity evidence. In that case, previous results will be summarized to show the current knowledge and identify the outstanding validation questions. In this case, some studies were found where different versions of IASNL-G were applied and provided data about the instrument's functioning. For instance, one of the validation studies concluded that the construct is unidimensional and include both beliefs about nonsexist language and recognition of sexist language, whereas the section measuring "use or willingness to use nonsexist language" seems to represent an independent construct or a linguistic artifact (Parks & Robertson, 2000). Therefore, that information is summarized as part of validity evidence based on internal structure, while other sources of evidence show information as "There are not studies investigating the relations between the scores from IASNL-G and other instruments measuring theoretically related constructs. However, the study presenting the original version of IASNL-G indicates that the scores could be related to those from the Gender-Specific Language Scale (GSLs)". In the results of step 2, we proposed validity assumptions according to the previous evidence, as well as additional assumptions reflecting both our goals and other claims requiring support from validity evidence. For instance, for validity evidence based on relations to other variables, the validity assumption was: "The total score on the IASNL-S will be positively related to the total score on the Gender Equality Attitudes Scale (Lorenzo Moledo et al., 2023)." That was the process for the five sources of validity evidence.

In step 3, validity assumptions were prioritized based on their relevance. The prioritization helps to organize the process to ensure

that the most relevant validity assumptions are addressed. Although collecting the five sources of validity evidence will provide a wider perspective, sometimes it won't be possible and, in these cases, the order will guide how to proceed. For completing the step, PROVAL suggests using the indications on PETEYC, a tool based on previous classifications of test's uses and evidence collected through experts' evaluations. That tool is part of a research project and is publicly available (<https://osf.io/2x6mr/overview>). In the example, all the assumptions were included as the goal of the illustration is to present the most ideal situation. According to the reflections, assumptions were ordered by grouping them under the five sources of validity, and were prioritized as follows: test content, internal structure, relations to other variables, response processes, and testing consequences. That indicates that the main validation study should be focused on collecting validity evidence based on test content followed by those based on internal structure and relations with other variables.

Step 4 identified who can provide the intended data. In the illustration, we would need the participation of two groups of people: experts and participants from the target population. These two groups of informants were linked to specific sources of validity evidence, as questions to reflect require thinking on the intended evidence. For instance, in reflection about the test content, the answers show that: experts can confirm whether the items in the IASNL-G-S reflect the construct "attitudes towards sexist/non-sexist language" and identify causes when that does not occur. Therefore, experts will give quantitative and qualitative information.

Building on this, the reflections in step 5 helped with structuring the validation studies in different phases, aligning them with the groups providing the information and the procedures required to collect the intended data. Questions are designed to describe the phases around the groups of participants and the type of information needed. In the example, three different phases will allow extracting the intended information. First, a group of experts will participate in an expert's judgement task. Secondly, a representative sample from the target population will respond to the IASNL-G-S and other instruments in a pilot study. Finally, a smaller group from the target population will participate in cognitive interviews.

Taking into account these phases and their relevance, step 6 describes the study design as follows [(QUAN+qual) → QUAN → (quan+QUAL)], by using the standard nomenclature for mixed-methods designs (Johnson & Onwuegbuzie, 2004). The representation indicates that the design will have three phases, with two of them collecting both quantitative and qualitative data. The phases will be sequential, as each one will be based on results from the previous. Quantitative data will be the main source in the two initial phases, whereas qualitative data will have a most relevant role in the last one. Finally, some specifications were added in step 7 to facilitate the preparation of materials and the planning of tasks.

After step 7, the subsequent steps involve conducting the phases, analyzing data from each phase, and integrating the results to formulate comprehensive conclusions about the "support" for the intended use of the instrument. These conclusions will guide decisions on necessary modifications and lead to the creation of the final version of the instrument. The illustration will be updated to incorporate all the details related to the development of the study.

Discussion

The present study aimed to propose PROVAL, a mixed-methods protocol for obtaining validity evidence designed to guide the development of validation studies focused on supporting the intended purposes and scores' interpretations of instruments assessing psychological and educational variables in various contexts. The proposal, based on the definition of validity and sources of validity evidence from the *Standards* (AERA et al., 1999, 2014), provides a framework that integrates both the validation guide through validity arguments, as indicated by Kane (1992), and the general phases outlined by Sireci (2013). The result is a series of steps that help researchers reflect on the validity evidence required by the validity argument by responding to specific questions.

To facilitate the understanding of the process, a specific case illustrated the application of PROVAL. Specifically, PROVAL was used to design the validation studies for the Spanish version of the IASNL-G (IASNL-G-S), an instrument originally created to assess attitudes toward sexist/nonsexist language (Parks & Robertson, 2000). The result of the application was a mixed-methods sequential design validation study, where three different phases address the validity questions raised by the validity claims that support the intended interpretation and test use. The first phase focuses on validity evidence based on test content, incorporating expert judgment to assess the overlap between the items and the intended construct. The second phase proposes a pilot study to collect responses from participants within the target population, aiming to analyze the psychometric properties of the items and the scale, as well as to gather validity evidence based on the internal structure and relationships with other variables.

These two phases represent the most common tasks performed in validation studies and would likely be included even if PROVAL was not applied. However, the needs related to these phases become clearer when responding to the reflections, which will also guide the design of the experts' tasks, the identification of appropriate instruments for the pilot study, and the recognition of potential benefits that additional procedures, such as qualitative strategies, could offer.

In addition, the application of PROVAL provided two more relevant pieces of data: the response processes involved in answering the items and the consequences perceived by responders. In the third phase, interviews emerged as an appropriate procedure, as validity evidence based on response processes and testing consequences is needed to support the intended use of the instrument. Furthermore, the data to be collected during the application of this procedure is specified. Thus, PROVAL helps extend the validation process by incorporating additional sources of evidence that could provide a better understanding of the instrument's weaknesses.

The final design proposed by PROVAL would change for different purposes and scores' interpretations, but also if other gaps were detected in relation to the scale's functioning. That points to the flexibility of PROVAL to guide processes with diverse characteristics.

PROVAL helps fill some of the gaps in designing validation studies. First, it encourages approaching the validation process as inherently connected to the intended instrument and prompts reflection on the main limitations that should be addressed during the validation study, rejecting automatic processes where the same steps are conducted. Second, it emphasizes the role of validity evidence based on response processes and testing consequences, traditionally

“ignored” (Cizek et al., 2008), by making their contribution explicit and providing guidance on how to incorporate procedures to collect such evidence. Third, it highlights elements that could be better understood through qualitative procedures, proposing alternative designs such as those based on MMR. Finally, it presents the validation process as a comprehensive task that should not be automatically designed but instead enriched by a deep study of the target instrument, while remaining accessible to non-experts, as it consists of a step-by-step, ordered process.

Regarding the benefits of the proposal, the approach to validation of PROVAL promotes a comprehensive and reflective understanding of validity that goes beyond the automatic processes sometimes proposed. The reflections it encourages are linked to methodological issues that often generate uncertainty and “resistance” among professionals from other fields, who also need to deal with the measurement of psychological variables. Therefore, this project addresses a fundamental need: making validation procedures accessible to professionals and researchers, avoiding reliance on automatic and routine processes limited to the most familiar ones, and promoting reflective actions based on the identification of validity demands and the selection of the most suitable methodology for each situation. This practical and theoretical proposal also has an aspect related to the transfer of scientific knowledge to society. The dissemination of the tool will be facilitated, along with support for researchers and professionals who wish to apply it. Therefore, PROVAL promotes a closer connection between the concept of validity and users of measurement instruments. In addition, PROVAL is intended to serve as a permanent tool, incorporating new findings from both theoretical and empirical studies. As such, it is designed as a dynamic protocol that is open to users' experiences and responsive to innovations in the literature.

PROVAL complements other tools currently available in the field, such as the Test Review Model published by the European Federation of Psychologists' Associations (EFPA, 2025) or the COSMIN Taxonomy of Measurement Properties (Mokkink et al., 2010). The EFPA model identifies elements to consider when assessing test quality, including instrument description, the quality of test materials, reliability, interpretation norms, and validity, while primarily focusing on the utility of validity evidence provided by test authors and the three “classic” sources of validity evidence: content, internal structure, and relations with other variables. COSMIN, in turn, is widely recognized as one of the most comprehensive tools for evaluating the measurement properties of health-related instruments, particularly for evaluative purposes. Although PROVAL does not aim to assess measurement properties directly, it is designed to assist professionals and practitioners in conducting validation studies guided by the intended use of the measures across diverse contexts and fields, including education, social sciences, and health. Importantly, PROVAL includes the collection of validity evidence related to response processes and testing consequences, incorporating fairness within the latter, and adopts a broader perspective than EFPA or COSMIN, which focus primarily on specific sources of evidence or accessibility for particular populations. Therefore, PROVAL, conceived as a dynamic tool, could also be useful to address validity issues identified in both the EFPA and COSMIN frameworks.

Despite the contributions of this study, no evidence has been gathered yet regarding the utility of PROVAL beyond the specific

examples provided. Therefore, future studies will focus on gathering validity evidence to support the intended purpose of the tool. Specifically, we will conduct empirical studies focused on identifying potential limitations of PROVAL and implementing improvements that facilitate its application across different fields. We will also create a space for users of PROVAL to collect information about their experiences with the tool, and proposal to implement adjustments to improve its usability. This will promote the creation of a collaborative and dynamic product that will be periodically reviewed to incorporate any innovations and improvements detected by researchers.

Author Contributions

Isabel Benítez: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing. **Andrés González-Gómez:** Formal Analysis, Supervision, Validation, Visualization, Writing – Review & Editing. **José Luis Padilla:** Supervision, Validation, Writing – Review & Editing.

Funding

Grant funded by Consejería de Universidad, Investigación e Innovación and by ERDF Andalusia Program 2021-2027 (C-SEJ-245-UGR23); and by MICIU/AEI/ 10.13039/501100011033 and ERDF/EU (PID2023-147219NB-I00). 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 Interest

The authors declare that there are no conflicts of interest.

Data Availability Statement

Materials are available at the following link: <https://osf.io/2x6mr/overview>

Acknowledgments

The Mind, Brain and Behavior Research Center receives funding from grants CEX2023-001312- M by MICIU/AEI/10.13039/501100011033 and UCE-PP2023-11 by the University of Granada.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Chen, M. Y., & Zumbo, B. D. (2017). Ecological framework of item responding as validity evidence: An application of multilevel DIF modeling using PISA data. In B.D. Zumbo & A.M. Hubley (Eds.), *Understanding and investigating response processes in validation research* (pp. 53-68). Springer. https://doi.org/10.1007/978-3-319-56129-5_4
- Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement*, 68(3), 397-412. <https://doi.org/10.1177/0013164407310130>
- Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education*, 49(6), 560-575. <https://doi.org/10.1111/medu.12678>
- Creswell, J. W., & Plano-Clark, V. L. (2011). *Designing and conducting mixed methods research*. Sage Publications.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. ERIC.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Dellinger, A. B., & Leech, N. L. (2007). Toward a unified validation framework in mixed methods research. *Journal of Mixed Methods Research*, 1(4), 309-332. <https://doi.org/10.1177/1558689807306147>
- European Federation of Psychologists' Associations AISBL (EFPA). (2025). *Model for the review, description and evaluation of psychological and educational tests (Test review model)*. <https://www.efpa.eu/>
- Engelhardt, L., & Goldhammer, F. (2019). Validating test score interpretations using time information. *Frontiers in Psychology*, 10, Article 1131. <https://doi.org/10.3389/fpsyg.2019.01131>
- Harkness, J., & Schoua-Glusberg, A. (1998). Questionnaires in translation. In J. Harkness (Ed.), *Cross-cultural survey equivalence* (pp. 87-126). Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://www.ssoar.info/ssoar/handle/document/49733>
- Iliescu, D., & Greiff, S. (2021). On consequential validity. *European Journal of Psychological Assessment*, 37(3), 163-166. <https://doi.org/10.1027/1015-5759/a000664>
- International Test Commission. (2017). *The ITC guidelines for translating and adapting tests (Second edition)*. [www.InTestCom.org]
- Johnson, R. B., & Christensen, L. (2019). *Educational research: Quantitative, qualitative, and mixed approaches*. Sage publications.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7), 14-26. <https://doi.org/10.3102/0013189X033007014>
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535. <https://doi.org/10.1037/0033-2909.112.3.527>
- Kane, M. T. (2006). Validation In R.L. Breenan (Ed.), *Educational Measurement* (pp. 17-64). American Council on Education and Praeger.
- Kane, M. T. (2013a). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. <https://psycnet.apa.org/doi/10.1111/jedm.12000> <https://doi.org/10.1111/jedm.12000>
- Kane, M. (2013b). Validation as a pragmatic, scientific activity. *Journal of Educational Measurement*, 50, 1115-1122. <https://doi.org/10.2307/23353803>
- Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, 26(1), 127-135. <https://doi.org/10.7334/psicothema2013.258>
- Lane, S., & Marion, S. F. (2025). Validity and validation. In L. L. Cook & M. J. Pitoniak (Eds.), *Educational measurement* (5th ed.,

- pp. 191–275). Oxford University Press. <https://doi.org/10.1093/oso/9780197654965.003.0004>
- Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, 36(8), 437–448. <https://doi.org/10.3102/0013189X07311286>
- Lorenzo Moledo, M., Quiroga-Carrillo, A., & Ferraces Otero, M. J. (2023). Estudio psicométrico de una escala sobre actitudes hacia la igualdad de género con mujeres reclusas. *Revista Española de Investigación Criminológica*, 20(2), Article e657. <https://doi.org/10.46381/reic.v20i2.657>
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement* (pp. 13–103). American Council on Education and Macmillan.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter LM, & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- Neuert, C. E., & Lenzner, T. (2016). Incorporating eye tracking into cognitive interviewing to pretest survey questions. *International Journal of Social Research Methodology*, 19(5), 501–519. <https://doi.org/10.1080/13645579.2015.1049448>
- Oren, C., Kennet-Cohen, T., & Turval, E. (2014). Demonstrating the validity of three general scores of PET in predicting higher education achievement in Israel. *Psicothema*, 26(1), 117–126. <https://doi.org/10.7334/psicothema2013.257>
- Padilla, J.L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136–144. <https://doi.org/10.7334/psicothema2013.259>
- Parks, J. B., & Robertson, M. A. (1998). Influence of age, gender, and context on attitudes toward sexist/nonsexist language: Is sport a special case? *Sex Roles*, 38(5), 477–494. <https://doi.org/10.1023/A:1018766023667>
- Parks, J. B., & Robertson, M. A. (2000). Development and validation of an instrument to measure attitudes toward sexist/nonsexist language. *Sex Roles*, 42, 415–438. <https://doi.org/10.1023/A:1007002422225>
- Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26(1), 108–116. <https://doi.org/10.7334/psicothema2013.260>
- Rushton, P. W., Routhier, F., Miller, W. C., Auger, C., & Lavoie, M.-P. (2015). French-Canadian translation of the WheelCon-M (WheelCon-M-F) and evaluation of its validity evidence using telephone administration. *Disability and Rehabilitation*, 37(9), 812–819. <https://doi.org/10.3109/09638288.2014.941019>
- Russell, M. (2022). Clarifying the terminology of validity and the investigative stages of validation. *Educational Measurement: Issues and Practice*, 41(2), 25–35. <https://doi.org/10.1111/emip.12453>
- 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>
- Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107. <https://doi.org/10.7334/psicothema2013.256>
- Sireci, S. G. (1998). The construct of content validity. *Social indicators research*, 45(1), 83–117.
- Sireci, S.G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. In R. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 19–37). Information Age Publishing Inc.
- Sireci, S. G. (2013). Agreeing on validity arguments. *Journal of Educational Measurement*, 50(1), 99–104. <https://doi.org/10.2307/23353800>
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Vol. 46). Sage.
- Tsang, S., Royse, C. F., & Terkawi, A. S. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia*, 11(1), 80–89. https://doi.org/10.4103/sja.SJA_203_17
- Yasir, A. S. (2016). Cross cultural adaptation & psychometric validation of instruments: Step-wise description. *International Journal of Psychiatry*, 1(1), 1–4. <https://doi.org/10.33140/IJP/01/01/00001>
- Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.) *The concept of validity: Revisions, new directions and applications* (pp. 65–82). Information Age Publishing.
- Zumbo, B. D. (2025). Looking through the lens of DIF: Embracing the many ways of being human. *The Score*, 47(4), 2–5.
- Zumbo, B. D., & Hubley, A. M. (Eds.). (2017). *Understanding and investigating response processes in validation research* (Vol. 69). Springer International Publishing. <https://doi.org/10.1007/978-3-319-56129-5>

Article

Psychometric Properties of the Negative and Positive Affect Scales (NAPAS) in Patients with Breast Cancer

M. Victoria Cerezo , Lucía Sánchez-Espíldora, Rafael Alarcón , Lorena M. Soria-Reyes  and María J. Blanca 

University of Malaga (Spain)

ARTICLE INFO

Received: 16/01/2026

Accepted: 19/03/2026

Keywords:

Subjective well-being

Positive affect

Negative affect

Instrumental study

ABSTRACT

Background: Breast cancer (BC) has a substantial impact on women's emotional well-being. This study examined the psychometric properties of the Negative and Positive Affect Scales (NAPAS) to provide validity evidence for their use in women with BC. **Method:** We evaluated validity evidence based on internal structure and on relationships with other variables (general health, life satisfaction, self-esteem and resilience), as well as item analysis, reliability of scale scores and measurement invariance across age groups ($N = 202$). **Results:** Confirmatory factor analysis supported a two-correlated factor structure, negative affect (NA) and positive affect (PA), $\chi^2(53) = 137.807$, CFI = .994, NNFI = .993, RMSEA = 0.089, SRMR = .067. This structure was invariant between age groups (≤ 50 vs. > 50 years). The homogeneity indices and the reliability of the scores were adequate. PA was positively associated with life satisfaction, self-esteem and resilience, and negatively associated with general health and its dimensions. NA showed the opposite pattern of associations. **Conclusions:** The findings support the use of the NAPAS as a measure of affective well-being in women with BC. Given its strong associations with mental health indicators, the scale may be useful in clinical settings for assessment and intervention planning.

Propiedades Psicométricas de las Escalas de Afecto Negativo y Positivo (NAPAS) en Pacientes con Cáncer de Mama

RESUMEN

Palabras clave:

Bienestar subjetivo

Afecto positivo

Afecto negativo

Estudio instrumental

Antecedentes: El cáncer de mama (CM) impacta sustancialmente en el bienestar emocional de las pacientes. Este estudio examinó las propiedades psicométricas de las Escalas de Afecto Negativo y Positivo (NAPAS) con el fin de aportar evidencia de validez para su uso en mujeres con CM. **Método:** Se presentan evidencias de validez basadas en la estructura interna y en las relaciones con otras variables (salud general, satisfacción vital, autoestima y resiliencia), análisis de ítems, fiabilidad de las puntuaciones e invariancia de medida entre grupos de edad ($N = 202$). **Resultados:** El análisis factorial confirmatorio respaldó una estructura de dos factores correlacionados, afecto negativo (AN) y afecto positivo (AP), $\chi^2(53) = 137.807$, CFI = .994, NNFI = .993, RMSEA = 0.089, SRMR = 0.067, que fue invariante entre los grupos de edad (≤ 50 y > 50 años). Los índices de homogeneidad y la fiabilidad de las puntuaciones fueron adecuados. El AP se asoció positivamente con satisfacción vital, autoestima y resiliencia, y negativamente con salud general; el AN mostró el patrón contrario. **Conclusiones:** Los hallazgos respaldan el uso de NAPAS como medida de bienestar afectivo en mujeres con CM, pudiendo ser útil en contextos clínicos para la evaluación y planificación de intervenciones.

Psychological research has shown that subjective well-being may be affected by a variety of adverse health conditions, including cancer (Alarcón et al., 2020; Cerezo et al., 2020; Cerezo, Soria-Reyes et al., 2022). Cancer is one of the main causes of morbidity and mortality worldwide, with the International Agency for Research on Cancer predicting that the number of cases will surpass 44 million globally by 2040. Although survival rates have improved, breast cancer still accounts for 15.5% of all cancer deaths among women (International Agency for Research on Cancer, 2023). As expected, a diagnosis of breast cancer can have a significant negative impact on a woman's psychological adjustment (Cardenal et al., 2012; Cerezo et al., 2020), including her subjective well-being (Alarcón et al., 2020; Cerezo, Soria-Reyes et al., 2022, 2024).

According to Diener et al. (1999), subjective well-being refers to a person's evaluation of their own life, and it encompasses two closely related aspects: cognitive and affective. The cognitive component corresponds to life satisfaction, to how the person views their life as a whole, and thus it reflects the gap between prior expectations and current life circumstances. The affective component refers to the emotions experienced in response to life events and it includes states of pleasure, referred to as *positive affect*, and displeasure, referred to as *negative affect* (Diener et al., 1999; Pavot & Diener, 2008). This affective component plays a central role in the emotional and physical evolution of breast cancer patients, and its assessment provides important information that is not captured by measures focused solely on anxiety or depression. Research has highlighted that both types of affects significantly influence prognosis and quality of life after treatment, as high negative affect is associated with greater psychological distress, more pain and fatigue, and poorer recovery, while positive affect acts as a protective factor linked to better adaptation and daily functioning (Rosenblom et al., 2025; Zhao et al., 2024). Given that positive and negative affect can be experienced simultaneously, achieving a balance in the management of both affective states is more important for subjective well-being than simply increasing positive affect or decreasing negative affect. In line with this perspective, having an adequate instrument to assess affect in patients with breast cancer is essential. Such tools can provide valuable information about patients' emotional experiences throughout the course of the disease and help guide interventions aimed at promoting emotional balance, thereby enhancing their capacity to cope with illness.

When assessing subjective well-being, the cognitive and affective components are generally considered separately, typically using Likert-type scales to capture respondents' intensity and frequency of experiences (Ferrando et al., 2025). With regard to the measurement of life satisfaction, a variety of tools are available, including the Harmony in Life Scale (HILS; Kjell et al., 2016), the Cantril Ladder (Cantril, 1965), the Riverside Life Satisfaction Scale (RLSS; Margolis et al., 2019), and the Satisfaction with Life Scale (SWLS; Diener et al., 1985). Of these, the SWLS is the most widely used and it has recently been validated in the breast cancer population (Cerezo, Soria-Reyes et al., 2022).

A number of different instruments may likewise be used to assess affective well-being, for example, the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988), the 10-item International PANAS Short Form (I-PANAS-SF; Kercher, 1992; Thompson, 2007), the Scale of Positive and Negative Experience (SPANE; Diener et al., 2010) and the Negative and Positive Affect Scales

(NAPAS; Mroczek & Kolarz, 1998), for which a short form is also available (Joshanloo, 2017). The PANAS is one of the most widely used tools for assessing the affective well-being, and its psychometric properties have been analysed in several different populations worldwide (e.g., Davis et al., 2022; Joshanloo, 2016; López-Gómez et al., 2015; Medvedev et al., 2023). Despite its popularity, however, it has certain limitations. For instance, and as noted by Diener et al. (2010), some of its items use the words "active" and "strong", which do not necessarily refer to feelings or emotions. Similar ambiguities have been highlighted by López-Gómez et al. (2015) with respect to Spanish translations of the PANAS. Assessments across languages highlight that items developed in one language may not always function equivalently when translated or adapted to other languages (Badham et al., 2025), which is particularly relevant for instruments such as the PANAS, where some items may be interpreted differently across linguistic or cultural contexts. In this regard, the NAPAS appears to overcome these limitations, insofar as its items employ adjectives that refer unambiguously to emotional states. Cerezo, Soria-Reyes et al. (2022) found that the correlation between life satisfaction (measured using the SWLS) and affect (both positive and negative) was stronger when affect was assessed using the NAPAS rather than the PANAS, leading them to suggest that the descriptive language used in the former reflects more clearly the emotional state of women with breast cancer.

The NAPAS has also been widely used in both the general population (Gao et al., 2023; Haas et al., 2021; Hadžić & Kantar, 2021; Joshanloo, 2019; Sirgy et al., 2020) and specific groups, including older adults (González-Herero & Extremera, 2010), cancer survivors (Gallagher et al., 2019) and women with breast cancer (Alarcón et al., 2020; Cerezo, Álvarez-Olmo et al., 2022; Cerezo et al., 2014; Cerezo, Soria-Reyes et al., 2022, 2024). However, its psychometric properties have not been fully analysed, and only a few instrumental studies have been conducted with this instrument (Chan et al., 2020; Joshanloo, 2016, 2017; Joshanloo & Bakhshi, 2016).

The 12 items that comprise the NAPAS were first reported in a study conducted in the USA by Mroczek and Kolarz (1998), who followed Watson and Tellegen (1985) in conceptualising positive and negative affect separately. The items are distributed equally across the two dimensions (positive and negative affect), with each item being responded to using a 5-point Likert-type scale. The items consist of adjectives (or adjectival phrases) and the respondent is asked to indicate how often they have experienced each of these affective states during the past month. The results of subsequent studies involving confirmatory factor analysis support a structure of two related factors for the NAPAS (Chan et al., 2020; Joshanloo, 2017; Joshanloo & Bakhshi, 2016), and this structure has shown measurement invariance across gender (Joshanloo & Bakhshi, 2016) and age (Chan et al., 2020).

Regarding the reliability of scale scores, Mroczek and Kolarz (1998) reported Cronbach's alpha coefficients of .87 and .91 for the negative and positive affect scales, respectively. Similar results were obtained in later studies, with reliability coefficients above .80 for both scales (Chan et al., 2020; Joshanloo, 2016, 2017). In terms of validity evidence based on relationships with other variables, Mroczek and Kolarz (1998) analysed sociodemographic, personality and contextual variables using regression models. The results showed a different pattern of relationships for men and women. Among men, older age, being married and lower education were

associated with higher scores on positive affect. In the sample of women, being married was likewise associated with a higher level of positive affect, while the relationship with age followed a quadratic function. For both men and women, scores on positive affect were positively associated with extraversion, and negatively associated with neuroticism, occupational stress and relative stress. The opposite pattern of associations was observed for scores on negative affect. Very few studies have subsequently explored these relationships further. Joshanloo (2017) confirmed the association between scores on positive and negative affect and those on extraversion and neuroticism, and also reported direct relationships between positive affect and self-esteem and life satisfaction, and inverse relationships between negative affect and these two variables.

In addition to the fact that only a few studies have examined the properties of the NAPAS, those which have done present important limitations. For example, most studies only provide validity evidence based on internal structure, without exploring validity based on relationships with other variables (Chan et al., 2020; Joshanloo & Bakhshi, 2016). Furthermore, relationships are only established with life satisfaction (Joshanloo, 2016, 2017) or with neuroticism, extraversion and self-esteem (Joshanloo, 2017). Finally, there are no studies examining the psychometric properties of the NAPAS in the Spanish-speaking population, whether in general or specifically among cancer patients. The aim of the present study was therefore to conduct an in-depth analysis of the psychometric properties of the Spanish version of the NAPAS. More specifically, we sought to provide evidence supporting the use of NAPAS scores as a measure of the affective component of subjective well-being in a sample of Spanish women with breast cancer. This is important as the inferences that can be drawn from test scores are specific to a population and a context (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The information obtained through application of the instrument could contribute to the design of more tailored psychological interventions. To this end, and based on previous findings (Chan et al., 2020; Joshanloo, 2017; Joshanloo & Bakhshi, 2016), we first test the model of two related factors to obtain validity evidence based on internal structure. Next, we examine measurement invariance across age groups, the reliability of scale scores and item homogeneity indices. To provide validity evidence based on relationships with other variables, we calculate Pearson correlation coefficients between scores on the NAPAS and scores on measures of general health (somatic symptoms, anxiety and insomnia, social dysfunction and severe depression), life satisfaction, self-esteem and resilience. It is expected to find that scores on positive affect correlate positively with scores on life satisfaction, self-esteem and resilience, and negatively with scores on general health and its dimensions. Scores on negative affect are expected to show the opposite pattern of associations with these variables.

Method

Participants

The sample comprised 202 Spanish women diagnosed with breast cancer, aged between 31 and 74 years ($M = 51.56$, $SD = 9.66$). Of these, 72.3% resided in the autonomous community of Andalusia, while the remainder lived in other regions of Spain. They were

most commonly married or living with their partner (43.6%), had a university education (40.1%), and were currently employed outside the home (36.1%). Their age at the time of diagnosis ranged between 21 and 73 years ($M = 46.94$, $SD = 9.18$). In terms of their clinical status, approximately half of the sample (51.5%) was at stage II of the TNM (Tumour, Nodes, Metastasis) classification system.

The sample was recruited through various health clinics providing breast cancer care in Spain associated with the Spanish Federation of Breast Cancer Patients (FECMA). The inclusion criteria were: signing informed consent, having a confirmed diagnosis of breast cancer, and no diagnosis of any other cancer or of a mental health disorder. Table 1 shows the sample characteristics.

Table 1
Sample Characteristics

Variables	N	%
Age (years)		
≤ 50	97	48.0
> 50	105	52.0
Civil status		
Married/living with partner	88	43.6
Single	86	42.6
Divorced	20	9.9
Widowed	8	4.0
Educational level		
Primary	49	24.3
Secondary	72	35.6
University	81	40.1
Employment status		
Working at home	44	21.8
Working outside the home	73	36.1
Medical leave	34	16.8
Unemployed	17	8.4
Retired	34	16.8
Time since diagnosis (years)		
< 2	66	32.7
≥ 2	136	67.3
Cancer stage		
0	7	3.5
I	26	12.9
II	104	51.5
III	62	30.7
IV	3	1.5

Instruments

1) Sociodemographic questionnaire, created ad hoc for the present study and used to gather data on age, civil status, level of education, employment status and disease-related variables (time since diagnosis and cancer stage).

2) Negative and Positive Affect Scales (NAPAS; Mroczek & Kolarz, 1998), translated into Spanish by González-Herero and Extremera (2010). The NAPAS measures the affective component of subjective well-being by asking the respondent about emotions and feelings experienced during the past 30 days. It comprises 12 items, each rated using a 5-point Likert-type scale (1 = *none of the time*, 5 = *all of the time*) and distributed equally across the negative and positive affect scales. Higher scores indicate higher levels of the type of affect being measured.

3) General Health Questionnaire (GHQ-28; Goldberg & Hillier, 1979), in its Spanish version (Lobo et al., 1986). The GHQ-28 comprises 28 items that explore somatic symptoms, anxiety and insomnia, social dysfunction and severe depression (four subscales of 7 items each). All items are rated using a 4-point Likert-type scale (0 = *not at all*, 3 = *much more than usual*), and both a total score and subscale scores are calculated. Higher scores indicate poorer health status on the corresponding indicator. In the present sample, McDonald's ω were .87 for scores on the somatic symptoms subscale, .93 for anxiety and insomnia, .86 for social dysfunction, .92 for severe depression and .94 for the total score.

4) Satisfaction With Life Scale (SWLS; Diener et al., 1985), in its Spanish version (Vázquez et al., 2013), which has been validated in breast cancer patients (Cerezo, Soria-Reyes et al., 2022). The SWLS evaluates the cognitive component of subjective well-being (i.e. satisfaction with life as a whole) and comprises 5 items, each rated using a 7-point Likert-type scale (1 = *strongly disagree*, 7 = *strongly agree*). Higher scores reflect a higher level of life satisfaction. McDonald's ω of scale scores in the present sample was .84.

5) Self-Esteem Scale (SES; Rosenberg, 1965), in its Spanish version (Atienza et al., 2000). The SES comprises 10 items, each rated on a 4-point Likert-type scale (1 = *strongly disagree*, 4 = *strongly agree*). Higher scores indicate higher levels of self-esteem. McDonald's ω of scale scores in the present sample was .89.

6) 10-item version of the Connor-Davidson Resilience Scale (CD-RISC 10; Campbell-Sills & Stein, 2007), in its Spanish version (Notario-Pacheco et al., 2011), which has been validated in breast cancer patients (Alarcón et al., 2020). Each item of the CD-RISC-10 is rated using a 5-point Likert-type scale (0 = *not true at all*, 4 = *true nearly all the time*), and higher scores indicate higher levels of resilience. McDonald's ω of scale scores in the present sample was .85.

Procedure

The study was conducted in accordance with the ethical standards of the Declaration of Helsinki and it was approved by the Experimental Ethics Committee of the University of Malaga, Spain. Participants were recruited by convenience sampling through various health clinics providing care to women with breast cancer. Staff of these clinics contacted potential participants and invited them to take part in the research. Those who expressed an interest were then contacted via e-mail in order to inform them about the study objectives and procedures. It was made clear at this point that participation was entirely voluntary and that all data collected would remain anonymous and confidential and be used solely for research purposes. No incentives were offered for taking part. Subsequent to signing (electronically) an informed consent form, and with the aim of avoiding fatigue among respondents, the sample was divided into two groups, randomly split using the random function of the Excel program. Each group was then sent an e-mail with a link to the questionnaires they were required to complete. Both groups completed the sociodemographic questionnaire, and the following set of instruments: subsample 1 ($n = 102$) completed the NAPAS and the GHQ-28 (total of 40 items), while subsample 2 ($n = 100$) completed the NAPAS, SWLS, SES and CD-RISC 10 (total of 37 items). There were no missing data as the questionnaires could not be submitted unless all items had been responded to.

Data Analysis

We first calculated descriptive statistics (mean, standard deviation, skewness and kurtosis) for scores (both item and total) on the NAPAS and the other instruments administered. Then, in order to obtain validity evidence based on the internal structure of the NAPAS, we conducted a confirmatory factor analysis (CFA) using the *R* package *lavaan* (Rosseel, 2012), testing the structure with two related factors. The method used was diagonally weighted least squares (DWLS) estimation with the polychoric correlation matrix, which has been shown to provide accurate parameter estimates when dealing with categorical items (Li, 2016) and tends to perform better than unweighted least squares (ULS) when the sample size is small (Forero et al., 2009). Model fit was assessed by calculating the chi-square statistic (χ^2) and the following fit indices: comparative fit index (CFI), Bentler-Bonett non-normed fit index (NNFI), root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR). The criteria used in interpreting these indices were as follows: CFI and NNFI values of .95 or higher indicated a good fit (Hu & Bentler, 1999); RMSEA values below .08 indicated a reasonable fit (Browne & Cudeck, 1992) and those below .06 a satisfactory fit (Hu & Bentler, 1999); SRMR values close to .08 indicated adequate fit (Hu & Bentler, 1999). Next, we analysed configural, metric and scalar invariance across two age groups (≤ 50 and > 50 years) after first determining the baseline model for both groups. This cut-off point was chosen because 50 years of age is the age at which women in Spain undergo population screening for early detection of breast cancer. Measurement invariance implies that an instrument measures the same construct in the same way across different groups, at various time points or under different conditions (Cieciuch & Davidov, 2015). Configural invariance requires that each construct is measured by the same items. Metric invariance is tested by constraining the factor loadings between the observed items and the latent variable to be equal across the compared groups. Finally, scalar invariance is tested by constraining both the factor loadings and the thresholds to be equal across groups (Chen, 2007; Cieciuch & Davidov, 2015). Here, we assessed invariance by comparing the CFI, RMSEA, and SRMR values of the more constrained models with those obtained in the less constrained model. Equality constraints are tenable when the decrease in the CFI is no more than .010, and when the increase in the RMSEA is not greater than .015; the increase in the SRMR should be no greater than .030 for metric invariance and no more than .010 for scalar invariance (Chen, 2007; Cheung & Rensvold, 2002).

Next, reliability of scale scores was assessed by calculating McDonald's ω coefficient, considering values of .70 or higher as satisfactory, in line with commonly used rule for acceptable reliability (Howitt & Cramer, 2020). Corrected item-total correlation coefficients (homogeneity index) for each of the NAPAS scales (negative and positive affect) were calculated, interpreting values above .30 as satisfactory (De Vaus, 2002). Finally, validity evidence based on relationships with other variables was obtained by calculating Pearson correlation coefficients between scores on the NAPAS and scores on the measures of general health, life satisfaction, self-esteem and resilience. In accordance with Cohen's (1988) criteria, coefficients were interpreted as follows: $|.10|$, low; $|.30|$, moderate; $|.50|$ or higher, strong correlation.

Results

Descriptive Analysis

Table 2 shows descriptive statistics for NAPAS items and its two scales, as well as for the other variables assessed.

Table 2
Descriptive Statistics for NAPAS Items, its two Scales and the Other Variables Assessed in the Study

Variables	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Negative Affect ^a	14.06	5.22	0.35	-0.37
1. So sad nothing could cheer you up	2.45	1.01	0.08	-0.83
2. Nervous	2.70	1.09	0.15	-0.73
3. Restless or fidgety	2.38	1.12	0.35	-0.80
4. Hopeless	2.13	1.07	0.50	-0.65
5. That everything was an effort	2.45	1.14	0.33	-0.68
6. Worthless	1.98	1.15	0.88	-0.31
Positive Affect ^a	19.05	4.91	-0.12	-0.48
1. Cheerful	3.45	0.84	-0.33	-0.17
2. In good spirits	3.48	0.96	-0.43	-0.39
3. Extremely happy	2.97	1.06	-0.07	-0.59
4. Calm and peaceful	2.94	0.97	-0.11	-0.54
5. Satisfied	3.23	0.99	-0.27	-0.24
6. Full of life	2.99	1.09	-0.17	-0.55
General Health (GHQ-28) ^b	27.91	14.86	0.60	-0.17
Somatic Symptoms ^b	8.18	4.85	0.39	-0.79
Anxiety and insomnia ^b	8.45	5.44	0.28	-0.72
Social dysfunction ^b	8.20	3.83	0.54	0.36
Severe Depression ^b	3.09	4.62	1.88	2.95
Life Satisfaction (SWLS) ^c	21.78	5.72	-0.44	-0.36
Self-esteem (SES) ^c	29.95	5.83	-0.25	-0.20
Resilience (CD-RISC 10) ^c	25.06	7.03	-0.61	0.46

Note. ^a *N* = 202; ^b *n*₁ = 102; ^c *n*₂ = 100.

Validity Evidence Based on Internal Structure

The results of the CFA (see Table 3) indicated that all fit indices were adequate according to the aforementioned cut-off criteria. This was the case for both the total sample and the two age-based subsamples, thus supporting the model with two related factors. The analysis also provided evidence of measurement invariance, insofar as comparison of the less constrained with the more constrained model showed a decrease of less than .010 in the CFI and increases of less than .015 in the RMSEA and less than .030 in the SRMR. Figure 1 shows the standardised parameter estimates for the total sample. All values are statistically significant.

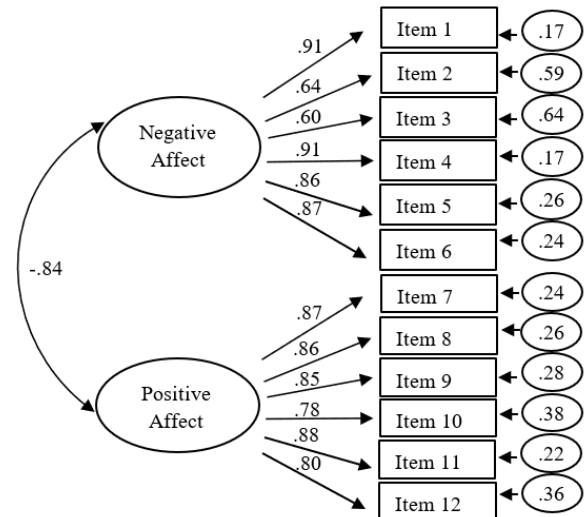
Table 3
Fit Indices for Confirmatory Factor Analysis and for Measurement Invariance

Model	χ^2	<i>df</i>	CFI	NNFI	SRMR	RMSEA	Δ CFI	Δ RMSEA	Δ SRMR
Total sample	137.807	53	.994	.993	.067	.089 [.071, .108]			
≤ 50 years	80.816	53	.996	.995	.076	.074 [.038, .105]			
> 50 years	77.571	53	.997	.996	.066	.067 [.030, .097]			
Configural invariance	158.387	106	.997	.996	.071	.070 [.046, .092]			
Metric invariance	195.109	116	.995	.994	.079	.083 [.062, .102]	<.010	<.015	<.030
Scalar invariance	202.563	150	.997	.997	.071	.059 [.036, .079]	<.010	<.015	<.010

Note. *N* = 202; ≤ 50 years, *n* = 97; > 50 years, *n* = 105; χ^2 = Chi-Square Statistic; *df* = Degrees of Freedom; CFI = Comparative Fit Index; NNFI = Non-Normed Fit Index; RMSEA = Root Mean Square Error of Approximation with 90% Confidence Interval; Δ CFI = CFI Less Constrained Model – CFI More Constrained Model; Δ RMSEA = RMSEA Less Constrained Model – RMSEA More Constrained Model; Δ SRMR = SRMR Less Constrained Model – SRMR More Constrained Model.

Figure 1

Factorial Loadings for the NAPAS Two-Factor Related Model With Standardized Factor Loadings for the Total Sample



Reliability of Scores and Item Analysis

Reliability of NAPAS scores was satisfactory, with ω values of .88 and .91 for the negative and positive affect scales, respectively. Table 4 shows corrected item-total correlations for each item of the NAPAS with respect to its corresponding scale. All correlations were above the established threshold of .30.

Table 4
Corrected Item-Total Correlations for Items of the NAPAS

Items	Item-Total Correlation
Factor 1: Negative Affect	
1. So sad nothing could cheer you up	.81
2. Nervous	.60
3. Restless or fidgety	.54
4. Hopeless	.79
5. That everything was an effort	.71
6. Worthless	.69
Factor 2: Positive Affect	
1. Cheerful	.79
2. In good spirits	.78
3. Extremely happy	.78
4. Calm and peaceful	.64
5. Satisfied	.80
6. Full of life	.70

Validity Evidence Based on Relationships With Other Variables

Table 5 shows correlations between scores on the NAPAS and scores on the criterion scales. Scores on the negative affect scale were strongly and negatively correlated with scores on life satisfaction, self-esteem and resilience, and strongly and positively correlated with scores on general health, anxiety and insomnia, social dysfunction and severe depression. A moderate positive correlation was observed between scores on the negative affect scale and the somatic symptoms dimension of general health.

For scores on the positive affect scale there was a strong positive correlation with scores on life satisfaction, self-esteem and resilience, and a strong negative correlation with scores on general health, anxiety and insomnia, social dysfunction and severe depression. A moderate negative correlation was observed between scores on positive affect and the somatic symptoms dimension of general health.

Table 5
Pearson Correlations Between NAPAS Scores and Scores on the Other Study Variables

Variables	Negative Affect	Positive Affect
General Health (GHQ-28) ^{1a}	.71***	-.62***
Somatic Symptoms	.35***	-.35**
Anxiety and Insomnia	.67***	-.52***
Social Dysfunction	.58***	-.55***
Severe Depression	.66***	-.55***
Life Satisfaction (SWLS) ²	-.57***	.66***
Self-Esteem (SES) ²	-.63***	.57***
Resilience (CD-RISC 10) ²	-.58***	.60***

Note. $n_1 = 102$; $n_2 = 100$; ** $p \leq .01$, *** $p \leq .001$

Discussion

The purpose of this study was to analyse the psychometric properties of the NAPAS and to provide evidence supporting its use as a tool for assessing the affective component of subjective well-being among women with breast cancer. To this end, we examined its internal structure, tested measurement invariance across two age groups, analysed the reliability of scale scores, conducted an item analysis and obtained validity evidence based on relationships with other variables (general health, life satisfaction, self-esteem and resilience).

Our findings supported the correlated two-factor model, consistent with prior research (Chan et al., 2020; Joshanloo, 2017; Joshanloo & Bakhshi, 2016), indicating that positive affect and negative affect are related but distinct dimensions of affective experience. The results align with the view that positive and negative affect represent distinct psychological constructs rather than opposite poles of a single bipolar continuum. This conceptualization is explicitly supported by Mroczek and Kolarz (1998), who noted that these dimensions must be treated separately and emphasized that the NAPAS was originally developed under this assumption, designed to measure positive and negative affect as separate, non-bipolar constructs. The correlated two-factor model allows for the derivation of separate positive and negative affect scores, facilitating their independent interpretation while accounting for their correlation. This structure yielded adequate fit indices for the two age groups considered (≤ 50 and > 50 years), and also showed measurement invariance across these groups. This means that the NAPAS items measure positive and negative affect in the same way in both these groups.

Item homogeneity indices were satisfactory for all items, indicating consistency among items when measuring the same construct. Reliability of scale scores was also satisfactory, with ω values of .88 and .91 for scores on the negative and positive affect scales, respectively, both above the established threshold of .70 and in line with coefficients reported previously (Chan et al., 2020; Joshanloo, 2016, 2017; Mroczek & Kolarz, 1998).

As for the relationship with other variables, the results obtained are again as expected. Scores on positive affect were strongly and positively correlated with scores on life satisfaction, self-esteem and resilience, and strongly and negatively correlated with scores on three of the four dimensions of general health; the negative correlation with the other general health factor (somatic symptoms) was of moderate magnitude. The opposite pattern of correlations was observed for scores on negative affect. Overall, these findings indicate that those women who experience more positive affective states (such as feeling cheerful, calm and full of life) also tend to be more satisfied with their life as a whole, to have a more positive view of themselves, are better able to deal with adversity and show better psychological adjustment in general; in turn, they are less likely to experience somatic symptoms, anxiety and insomnia, social dysfunction and severe depression. By contrast, those women who experience stronger negative emotions (such as feeling sad, anxious, restless, hopeless and worthless, or that everything is an effort) tend to show poorer psychological adjustment and general health.

The results obtained are consistent with those reported in the general population and add to existing validity evidence for the NAPAS. These studies likewise found that scores on positive affect were positively correlated with positive psychological variables such as life satisfaction, extraversion and self-esteem, while scores on negative affect were positively associated with indicators of emotional distress and instability, such as neuroticism, occupational stress and relative stress (Joshanloo, 2016, 2017; Mroczek & Kolarz, 1998). The relationship between positive and negative affect and the variables general health (including its different dimensions) and resilience has not previously been studied for the NAPAS, although the results obtained for these variables with the PANAS (Cerezo, Soria-Reyes et al., 2022; López-Gómez et al., 2015) are similar to those obtained here.

A noteworthy finding from the correlational analysis is that although all correlations were strong, scores on positive affect were more strongly correlated with life satisfaction than with the other variables measured. This suggests that it is these two constructs which make the most significant contribution to subjective well-being in the current sample. In a recent study that assessed affect using the SPANE, Jovanović and Joshanloo (2022) similarly found that positive affect had a strong effect on life satisfaction, leading them to conclude that people rely more on positive emotional experiences (rather than negative ones) when evaluating how satisfied they are with life. Here, we also found that the strongest correlation for scores on negative affect was with general health, which suggests that negative emotions are more closely linked to physical and psychological problems (Hu & Gruber, 2008). Future studies should aim to analyse in depth the relationship between life satisfaction and positive affect, as well as the possible causal relationship between negative affect and indicators of health problems.

Overall, our results suggest that the NAPAS has adequate psychometric properties for use in the breast cancer population.

Although this study does not provide evidence regarding its specific use in the clinical context, it is expected to be useful for the assessment of affective state, identifying states that should be improved with psychological intervention. Given its brevity, and insofar as scores on its two scales are associated with a set of variables that are indicative of general health and psychological adjustment, the NAPAS could be used as a tool to detect possible mental health problems, thus helping to identify those women for whom a more detailed assessment is indicated. One of the goals of intervention might be to help patients to achieve a balance between both emotional states, as achieving subjective well-being relies more on this emotional equilibrium than on simply increasing positive affect or reducing negative affect. Thus, interventions should aim to regulate emotions rather than eliminate negative ones, thereby enhancing patients' subjective well-being and strengthening their capacity to cope with the disease. In a meta-analysis, [van Agteren et al. \(2021\)](#) concluded that mindfulness-based and multi-component positive psychological interventions were the most effective in improving various indicators of subjective well-being in both clinical and non-clinical populations. They also found that singular positive psychological interventions, as well as interventions based on cognitive and behavioural therapy, acceptance and commitment therapy, and reminiscence were useful in this respect.

This study has a number of limitations that should be acknowledged. First, our use of a convenience sample may limit the generalizability of results. Second, the sample size does not allow the analysis of other more complex factor structures such as the bifactor model, which require large samples ([Kretzschmar & Gignac, 2019](#); [Zhang et al., 2021](#)). Future studies with larger sample sizes could evaluate whether the fit of the bifactor model is superior to that based on two related factors, and also use other criteria for the interpretation of fit indices, such as dynamic fit index cutoffs ([McNeish, 2023](#); [McNeish & Manapat, 2023](#)). Third, all the study variables were measured using self-report tools, and hence response bias might be an issue. Fourth, the cross-sectional nature of the study means that no causal relationships can be established between the different variables. Finally, although we report validity evidence based on relationships with several different variables, further research is needed to explore the potentially important role of other aspects of well-being (e.g., flourishing) and personality traits (e.g., introversion, neuroticism, etc.) in breast cancer.

Despite these limitations, the present study has two important strengths. First, it reports useful information about the use and interpretation of NAPAS scores among women with breast cancer, adding to existing validity and reliability evidence. In addition, it provides evidence regarding the relationship between the affective component of subjective well-being and a series of positive psychological variables, namely life satisfaction, self-esteem and resilience, an association that has barely been analysed in instrumental studies. Evidence is also provided about the relationship with variables inherent to psychological maladjustment and general health, namely somatic symptoms, anxiety and insomnia, social dysfunction and severe depression. The NAPAS can help professionals identify aspects of subjective well-being that need improvement to promote better adjustment and coping with the disease.

Author Contributions

M. Victoria Cerezo: Conceptualization, Methodology, Data Curation, Writing - Original Draft. **Lucía Sánchez-Espíldora:** Writing - Original Draft. **Rafael Alarcón:** Methodology, Writing - Original Draft, and Review. **Lorena M. Soria-Reyes:** Conceptualization, Data Curation, Writing - Original Draft and Review. **María J. Blanca:** Conceptualization, Methodology, Software, Data Curation, Writing - Original Draft, Supervision, Validation, Funding, Review and Editing.

Funding

University of Malaga with the Andalusia FEDER Program 2021-2027 (PPRO-CTS110-G-2023). 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 there is no conflict of interest.

Data Availability Statement

The data is available at <https://doi.org/10.24310/riuma.38142>

References

- Alarcón, R., Cerezo, M. V., Hevilla, S. & Blanca, M. J. (2020). Psychometric properties of the Connor-Davidson Resilience Scale in women with breast cancer. *International Journal of Clinical and Health Psychology*, 20(1), 81–89. <https://doi.org/10.1016/j.ijchp.2019.11.001>
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *The standards for educational and psychological testing*. AERA Publications.
- Atienza, F., Moreno, Y. & Balaguer, I. (2000). Análisis de la dimensionalidad de la Escala de Autoestima de Rosenberg en una muestra de adolescentes valencianos [Analysis of the dimensionality of the Rosenberg Self-Esteem Scale in a sample of Valencian adolescents]. *Revista de Psicología, Universitas Tarraconensis*, 22, 29–42.
- Badham, L., Oliveri, M. E., & Sireci, S. G. (2025). Navigating multi-language assessments: Best practices for test development, linking, and evaluation. *Psicothema*, 37(3), 1–11. <https://doi.org/10.70478/psicothema.2025.37.19>
- Browne, M. W. & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230–258. <https://doi.org/10.1177/0049124192021002005>
- Campbell-Sills, L. & Stein, M. B. (2007). Psychometric analysis and refinement of the Connor–Davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress*, 20(6), 1019–1028. <https://doi.org/10.1002/jts.20271>
- Cantril, H. (1965). Discovering people's aspirations: The method used. In H. Cantril (ed.) *The pattern of human concerns* (pp. 21–29). Rutgers University Press.

- Cardenal, V., Cerezo, M. V., Martínez, J., Ortiz-Tallo, M. & Blanca, M. J. (2012). Personality, emotions and coping styles: Predictive value for the evolution of cancer patients. *The Spanish Journal of Psychology*, 15(2), 756–767. https://doi.org/10.5209/rev_SJOP.2012.v15.n2.38887
- Cerezo, M. V., Álvarez-Olmo, A. & Rueda, P. (2022). General health and resilience of breast cancer patients: The mediator role of affective well-being. *International Journal of Environmental Research and Public Health*, 19(9). Article 5398 <https://doi.org/10.3390/ijerph19095398>
- Cerezo, M. V., Blanca, M. J. & Ferragut, M. (2020). Personality profiles and psychological adjustment in breast cancer patients. *International Journal of Environmental Research and Public Health*, 17(24), 9452. <https://doi.org/10.3390/ijerph17249452>
- Cerezo, M. V., Ortiz-Tallo, M., Cardenal, V. & De La Torre-Luque, A. (2014). Positive psychology group intervention for breast cancer patients: A randomised trial. *Psychological Reports*, 115(1), 44–64. <https://doi.org/10.2466/15.20.PR0.115c17z7>
- Cerezo, M. V., Soria-Reyes, L. M., Alarcón, R. & Blanca, M. J. (2022). The Satisfaction with Life Scale in breast cancer patients: Psychometric properties. *International Journal of Clinical and Health Psychology*, 22(1). Article 100274. <https://doi.org/10.1016/j.ijchp.2021.100274>
- Cerezo, M. V., Soria-Reyes, L. M., Alarcón, R., & Blanca, M. J. (2024). The Flourishing Scale: Psychometric properties in breast cancer patients. *Integrative cancer therapies*, 23, 15347354241249935. <https://doi.org/10.1177/15347354241249935>
- Chan, M. H. M., Gerhardt, M. & Feng, X. (2020). Measurement invariance across age groups and over 20 years' time of the Negative and Positive Affect Scale (NAPAS). *European Journal of Psychological Assessment*, 36(4), 537–544. <https://doi.org/10.1027/1015-5759/a000529>
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling*, 14, 464–504. <https://doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Cieciuch, J. & Davidov, E. (2015). Establishing measurement invariance across online and offline samples. A tutorial with the software packages Amos and Mplus. *Studia Psychologica*, 15, 83–99. <https://doi.org/10.21697/sp.2015.14.2.06>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Davis, R. C., Arce, M. A., Tobin, K. E., et al. (2022). Testing measurement invariance of the Positive and Negative Affect Schedule (PANAS) in American and Arab university students. *International Journal of Mental Health and Addiction*, 20(2), 874–887. <https://doi.org/10.1007/s11469-020-00411-z>
- De Vaus, D. (2002). *Analyzing Social Science Data: 50 Key Problems in Data Analysis*. Sage.
- Diener, E., Emmons, R. A., Larsen, R. J. & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13
- Diener, E., Suh, E. M., Lucas, R. E. & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125(2), 276–302. <http://doi.org/10.1037/0033-2909.125.2.276>
- Diener, E., Wirtz, D., Tov, W. et al. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, 97(2), 143–156. <https://doi.org/10.1007/s11205-009-9493-y>
- Ferrando, P. J., Morales-Vives, F., Casas, J. M., & Muñoz, J. (2025). Likert scales: A practical guide to design, construction and use. *Psicothema*, 37(4), 1–15. <https://doi.org/10.70478/psicothema.2025.37.24>
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16(4), 625–641. <https://doi.org/10.1080/10705510903203573>
- Gallagher, M. W., Long, L. J., Richardson, A. & D'Souza, J. M. (2019). Resilience and coping in cancer survivors: The unique effects of optimism and mastery. *Cognitive Therapy and Research*, 43(1), 32–44. <https://doi.org/10.1007/s10608-018-9975-9>
- Gao, Q., Tang, W., Yang, Y. & Fu, E. (2023). Children's emotional intelligence and aggressive behavior: The mediating roles of positive affect and negative affect. *Heliyon*, 9(10). Article e20366. <https://doi.org/10.1016/j.heliyon.2023.e20366>
- Goldberg, D. P. & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine*, 9(1), 139–145. <https://doi.org/10.1017/S0033291700021644>
- González-Herero, V. & Extremera, N. (2010). Daily life activities as mediators of the relationship between personality variables and subjective well-being among older people. *Personality and Individual Differences*, 49, 124–129. <https://doi.org/10.1016/j.paid.2010.03.019>
- Haas, B. W., Hoefft, F. & Omura, K. (2021). The role of culture on the link between worldviews on nature and psychological health during the COVID-19 pandemic. *Personality and Individual Differences*, 170. Article 110336. <https://doi.org/10.1016/j.paid.2020.110336>
- Hadžić, A. & Kantar, D. (2021). The relationships between attachment dimensions and affect in adulthood: The mediating effects of psychological flexibility. *Primenjena Psihologija*, 14(2), 173–188. <https://doi.org/10.19090/PP.2021.2.173-188>
- Howitt, D., & Cramer, D. (2020). *Research methods in psychology* (6th ed.). Pearson.
- Hu, L. T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hu, J. & Gruber, K. J. (2008). Positive and negative affect and health functioning indicators among older adults with chronic illnesses. *Issues in Mental Health Nursing*, 29(8), 895–911. <https://doi.org/10.1080/01612840802182938>
- International Agency for Research on Cancer. (2023). *Number of new cases in 2020, both sexes, all ages*. <https://geo.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf>
- Joshanloo, M. (2016). Factor structure of subjective well-being in Iran. *Journal of Personality Assessment*, 98(4), 435–443. <https://doi.org/10.1080/00223891.2015.1117473>
- Joshanloo, M. (2017). Factor structure and criterion validity of original and short versions of the Negative and Positive Affect Scale (NAPAS). *Personality and Individual Differences*, 105, 233–237. <https://doi.org/10.1016/j.paid.2016.09.060>

- Joshanloo, M. (2019). Lay conceptions of happiness: Associations with reported well-being, personality traits, and materialism. *Frontiers in Psychology*, 10, Article: 2377 <https://doi.org/10.3389/fpsyg.2019.02377>
- Joshanloo, M. & Bakhshi, A. (2016). The factor structure and measurement invariance of positive and negative affect a study in Iran and the USA. *European Journal of Psychological Assessment*, 32(4), 265–272. <https://doi.org/10.1027/1015-5759/a000252>
- Jovanović, V. & Joshanloo, M. (2022). The contribution of positive and negative affect to life satisfaction across age. *Applied Research in Quality of Life*, 17(2), 511–524. <https://doi.org/10.1007/s11482-020-09903-5>
- Kercher, K. (1992). Assessing subjective well-being in the old-old: The PANAS as a measure of orthogonal dimensions of positive and negative affect. *Research on Aging*, 14(2), 131–168. <https://doi.org/10.1177/0164027592142001>
- Kjell, O. N. E., Daukantaitė, D., Hefferon, K. & Sikström, S. (2016). The Harmony in Life Scale complements the Satisfaction with Life Scale: Expanding the conceptualization of the cognitive component of subjective well-being. *Social Indicators Research*, 126(2), 893–919. <https://doi.org/10.1007/s11205-015-0903-z>
- Kretschmar, A., & Gignac, G. E. (2019). At what sample size do latent variable correlations stabilize? *Journal of Research in Personality*, 80, 17–22. <https://doi.org/10.1016/j.jrp.2019.03.007>
- Li, C. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, 21(3), 369–387. <https://doi.org/10.1037/met0000093>
- Lobo, A., Pérez-Echeverría, M. J. & Artañ, J. (1986). Validity of the scaled version of the General Health Questionnaire (GHQ-28) in a Spanish population. *Psychological Medicine*, 16(1), 135–140. <https://doi.org/10.1017/S0033291700002579>
- López-Gómez, I., Hervás, G. & Vázquez, C. (2015). Adaptación de las “Escala de Afecto Positivo y Negativo” (PANAS) en una muestra general Española [Adaptation of the Positive and Negative Affect Schedule (PANAS) in a general Spanish population]. *Behavioral Psychology / Psicología Conductual*, 23(3), 529–548.
- Margolis, S., Schwitzgebel, E., Ozer, D. J. & Lyubomirsky, S. (2019). A new measure of life satisfaction: the Riverside Life Satisfaction Scale. *Journal of Personality Assessment*, 101, 621–630. <https://doi.org/10.1080/00223891.2018.1464457>
- McNeish, D. (2023). Dynamic fit index cutoffs for factor analysis with Likert, ordinal, or binary responses. *American Psychologist*, 78(9), 1061–1075. <https://doi.org/10.1037/amp0001213>
- McNeish, D., & Manapat, P. D. (2023). Dynamic fit index cutoffs for hierarchical and second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, 31(1), 27–47. <https://doi.org/10.1080/10705511.2023.2225132>
- Medvedev, O. N., Roemer, A., Krägeloh, C. U., Sandham, M. H. & Siegert, R. J. (2023). Enhancing the precision of the Positive and Negative Affect Schedule (PANAS) using Rasch analysis. *Current Psychology*, 42(2), 1554–1563. <https://doi.org/10.1007/s12144-021-01556-3>
- Mroczek, D. K. & Kolarz, C. M. (1998). The effect of age on positive and negative affect: A developmental perspective on happiness. *Journal of Personality and Social Psychology*, 75(5), 1333–1349. <https://doi.org/10.1037/0022-3514.75.5.1333>
- Notario-Pacheco, B., Solera-Martínez, M., Serrano-Parra, M. D., Bartolomé-Gutiérrez, R., García-Campayo, J. & Martínez-Vizcaino, V. (2011). Reliability and validity of the Spanish version of the 10-item Connor-Davidson Resilience Scale (10-item CD-RISC) in young adults. *Health and Quality of Life Outcomes*, 9, 1–6. <https://doi.org/10.1186/1477-7525-9-63>
- Pavot, W. & Diener, E. (2008). The Satisfaction With Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, 3(2), 137–152. <https://doi.org/10.1080/17439760701756946>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. University Press
- Rosenblom, M. F., Reme, S. E., & Jacobsen, H. B. (2025). The role of negative affect and experiential avoidance in postsurgical pain and fatigue among Norwegian women with breast cancer. *European Journal of Pain*, 29(2), e4788. <https://doi.org/10.1002/ejp.4788>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sirgy, M. J., Lee, D. J., Park, S., Joshanloo, M. & Kim, M. (2020). Work–family spillover and subjective well-being: The moderating role of coping strategies. *Journal of Happiness Studies*, 21(8), 2909–2929. <https://doi.org/10.1007/s10902-019-00205-8>
- Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, 38(2), 227–242. <https://doi.org/10.1177/0022022106297301>
- van Agteren, J., Iasiello, M., Lo, L., et al. (2021). A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. *Nature Human Behaviour*, 5(5), 631–652. <https://doi.org/10.1038/s41562-021-01093-w>
- Vázquez, C., Duque, A. & Hervás, G. (2013). Satisfaction with Life Scale in a representative sample of Spanish adults: Validation and normative data. *The Spanish Journal of Psychology*, 16, Article E82. <https://doi.org/10.1017/sjp.2013.82>
- Watson, D. & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98(2), 219–235. <https://doi.org/10.1037/0033-2909.98.2.219>
- Watson, D., Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Zhang, B., Sun, T., Cao, M., & Drasgow, F. (2021). Using bifactor models to examine the predictive validity of hierarchical constructs: Pros, cons, and solutions. *Organizational Research Methods*, 24(3), 530–571. <https://doi.org/10.1177/1094428120915522>
- Zhao, J., Yang, D. S., Liu, Y. Q., Wu, Y. K., Chen, C., Li, J. T., & Wu, R. G. (2024). Characteristics of positive and negative effects on the quality of life of breast cancer patients. *BMC psychiatry*, 24(1), 926. <https://doi.org/10.1186/s12888-024-06311-z>

Article

Psychometric Evaluation and Normative Framework for the Movie for the Assessment of Social Cognition in Spanish-Speaking Adults

Aina Llompart-Casasnovas¹, Adoración Castro^{1,2}, Margalida Gili^{1,2}, Guillem Navarra-Ventura^{1,2}, Roberto Rodríguez-Jiménez^{3,4}, Miguel A. Ortega⁵, Leticia Boada^{4,6}, Mercè Jodar^{4,7}, Narcís Cardoner^{4,8}, Albert Sesé^{1,2}, Miquel Roca^{1,2} and Guillermo Lahera^{4,5}

¹ University of the Balearic Islands (Spain)

² Health Research Institute of the Balearic Islands (Spain)

³ Universidad Complutense de Madrid (Spain)

⁴ CIBER de Salud Mental (Spain)

⁵ Universidad de Alcalá (Spain)

⁶ Hospital Gregorio Marañón (Spain)

⁷ Hospital Universitari Parc Taulí (Spain)

⁸ Hospital de la Sant Creu i Sant Pau (Spain)

ARTICLE INFO

Received: 03/10/2025

Accepted: 24/03/2026

Keywords:

Theory of mind
Mentalization
Factor analysis
Internal consistency
Percentiles

ABSTRACT

Background: The Movie for the Assessment of Social Cognition-Spanish (MASC-SP) assesses general mentalization performance and associated error patterns: over-mentalization, under-mentalization, and non-mentalization. This study aims to provide the first large-scale psychometric evaluation and normative framework for MASC-SP scores in Spanish-speaking adults. **Method:** 2,582 healthy adults completed the MASC-SP via a secure online platform. The psychometric evaluation included factor analysis, measurement invariance, internal consistency, and floor/ceiling effects. Normative data were stratified by demographic variables and are presented as empirical percentiles and z-scores. **Results:** MASC-SP scores showed a predominantly unifactorial structure, with internal consistency ranging from questionable to good. A floor effect was observed in the non-mentalization error score. Female participants and those with a university education scored higher and made fewer errors than men and those without a university education. Age-related results were heterogeneous. **Conclusions:** The results support interpreting the MASC-SP total score as a robust quantitative indicator of general mentalization performance and the error scores as secondary qualitative indicators of different error patterns. Normative data stratified by sex and educational level provide a useful reference for research and clinical settings. However, norms stratified by age groups should be applied with caution due to their questionable clinical and developmental relevance.

Evaluación Psicométrica y Marco Normativo para la *Movie for the Assessment of Social Cognition* en Adultos Castellanohablantes

RESUMEN

Antecedentes: La *Movie for the Assessment of Social Cognition-Spanish* (MASC-SP) evalúa rendimiento mentalizador general y patrones de error asociados: sobre-mentalización, infra-mentalización y no-mentalización. El objetivo es proporcionar la primera evaluación psicométrica y marco normativo a gran escala para las puntuaciones de la MASC-SP en adultos castellanohablantes. **Método:** 2.582 adultos sanos completaron la MASC-SP mediante una plataforma en línea. La evaluación psicométrica incluyó análisis factorial, invarianza de medida, consistencia interna y efectos suelo/techo. Los datos normativos se estratificaron según variables demográficas y se presentan como percentiles y puntuaciones *z* empíricas. **Resultados:** Las puntuaciones mostraron una estructura predominantemente unifactorial, con una consistencia interna entre cuestionable y buena. Se observó un efecto suelo en la puntuación de error de no-mentalización. Las mujeres y los participantes con estudios universitarios obtuvieron puntuaciones más altas y cometieron menos errores que sus contrapartes. Los resultados relacionados con la edad fueron heterogéneos. **Conclusiones:** Los resultados respaldan la interpretación de la puntuación total de la MASC-SP como un indicador cuantitativo sólido del rendimiento mentalizador general y de las puntuaciones de error como indicadores cualitativos secundarios. Los datos normativos estratificados por sexo y educación proporcionan una referencia útil para contextos de investigación y clínicos. Las normas estratificadas por grupos de edad deben aplicarse con cautela debido a su cuestionable relevancia clínica y de desarrollo.

Palabras clave:

Teoría de la mente
Mentalización
Análisis factorial
Consistencia interna
Percentiles

Cite as: Llompart-Casasnovas, A., Castro, A., Gili, M., Navarra-Ventura, G., Rodríguez-Jiménez, R., Ortega, M. A., Boada, L., Jodar, M., Cardoner, Sesé, A., N., Roca, M., & Lahera, G. (2026). Psychometric evaluation and normative framework for the Movie for the Assessment of Social Cognition in Spanish-speaking adults. *Psicothema*, 38(4), 213-223. <https://doi.org/10.70478/psicothema.2026.38.20>

Corresponding author: Guillem Navarra-Ventura, g.navarra@uib.cat

Social cognition (SC) enables the perception, processing, and interpretation of social information, facilitates understanding of others' behaviors and intentions, and promotes effective functioning in social interactions (Arioli et al., 2018). It is a multidimensional and hierarchically organized construct that encompasses several interrelated domains, of which the most studied are emotional processing and theory of mind (ToM) (Schurz et al., 2021). To date, research has focused on key subcomponents, such as emotion recognition, mental state reasoning (inferring thoughts/beliefs), and mental state decoding (inferring feelings) (Mitchell & Phillips, 2015).

Patients with mental disorders often present SC deficits (Cotter et al., 2018), which tend to be more pronounced in the more complex, higher-order subcomponents than in the more basic subcomponents (Navarra-Ventura et al., 2022; Romero-Ferreiro et al., 2022). These deficits have been documented in Asperger syndrome (Dziobek et al., 2006; Lahera et al., 2014), schizophrenia (Andrade-González et al., 2021; Catalan et al., 2018; Engelstad et al., 2019; Fretland et al., 2015; Montag et al., 2011; Sahl et al., 2023; Vaskinn et al., 2018), bipolar disorder (Aidelbaum & Goghari, 2022; Montag et al., 2010; Santos et al., 2017), depression (Seitz et al., 2022; Wolkenstein et al., 2011; Zwick & Wolkenstein, 2017), and personality disorder (Andreou et al., 2015; Newbury-Helps et al., 2017; Preißler et al., 2010; Quek et al., 2018; Somma et al., 2019).

Demographic factors also influence SC performance. Older age and lower educational level have been associated with reduced mental state reasoning and decoding, while emotion recognition appears to be less affected by these factors (Fittipaldi et al., 2024). Other studies have revealed sex-related differences in temporo-limbic and frontal brain regions (Gur, 2002), and in large-scale connectivity patterns (Ingallhalikar et al., 2014), which are often accompanied by behavioral differences. Women tend to outperform men in emotion recognition, while sex differences in mental state reasoning and decoding are generally small or non-existent (Christov-Moore et al., 2014; Navarra-Ventura et al., 2018, 2021, 2022; Vaskinn et al., 2024).

These findings underscore the multifactorial and heterogeneous nature of SC and highlight the need for assessment instruments that are psychometrically sound and supported by normative data (Pinkham et al., 2018). Traditional instruments include the Pictures of Facial Affect (POFA) for emotion recognition (Ekman & Friesen, 1976), the Strange Stories Test (SST) for mental state reasoning (Happé, 1994), and the Reading the Mind in the Eyes Test (RMET) for mental state decoding (Baron-Cohen et al., 2001). Despite their widespread use, the scores of these instruments are limited by the presence of ceiling effects, their low ecological validity, and their exclusive focus on a single subcomponent of SC (Dziobek et al., 2006; Lahera et al., 2014). In contrast, real-life social interactions require the simultaneous integration of multiple dynamic signals and cues (facial expression, gaze, gestures, body language, speech prosody, verbal content) to understand others' behaviors and intentions (Cavieres et al., 2025).

Among the instruments designed to capture this complexity, the Movie for the Assessment of Social Cognition (MASC) (Dziobek et al., 2006) stands out. The MASC integrates verbal and non-verbal signals and cues related to emotional processing and ToM in video format, and its scores provide a more ecologically valid and dynamic assessment of SC than traditional instruments. Its multiple-choice question structure allows for distinguishing between typical

and impaired ToM performance (primary outcome) and identifying specific error patterns, including over-attribution, under-attribution, and non-attribution of mental states (secondary outcomes). This scoring system allows for a more nuanced assessment of mentalization abilities and associated error patterns, thereby improving the interpretability of results (Msika et al., 2024).

To date, the MASC has been adapted into several languages, including Spanish (MASC-SP) (Lahera et al., 2014), and there is growing evidence supporting the clinical sensitivity and cross-cultural applicability of its scores (Chang et al., 2024; Fossati et al., 2018; Huang et al., 2023; Martinez et al., 2017; Müller et al., 2016; Sarig-Shmueli et al., 2023). However, these studies are limited by small sample sizes, lack of psychometric evaluation of secondary scores, and lack of normative data, which has restricted their use to research settings and their applicability in clinical practice.

This study aims to provide the first large-scale psychometric evaluation and normative framework for MASC-SP scores in Spanish-speaking adults.

Method

Participants

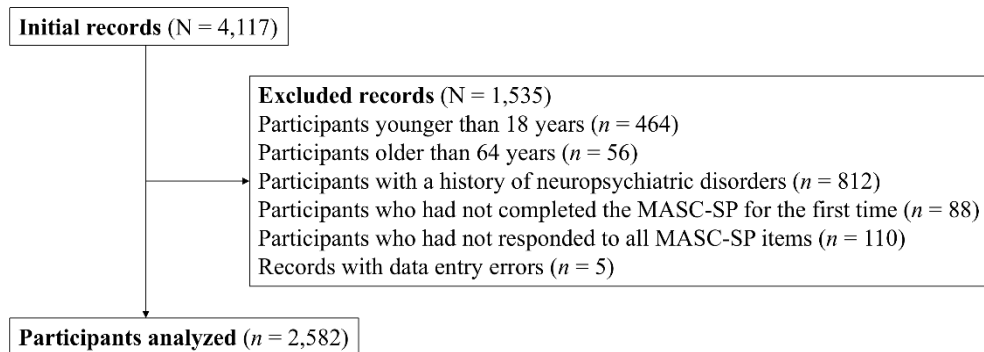
All native Spanish speakers residing in Spain were eligible to participate. Participants were recruited between June 2016 and May 2024 using non-probability methods, primarily convenience and snowball sampling, in community organizations, universities, and clinical settings through direct contact and study invitations.

The initial dataset comprised 4,117 records, which were filtered according to the following exclusion criteria: the participant was under 18 or over 64 years of age, had a history of neuropsychiatric disorders, had previously completed the MASC-SP, or had not completed all items on the MASC-SP. Five additional records were excluded due to data entry errors. After applying these criteria, 1,535 records were excluded, resulting in a final sample of 2,582 participants (Figure 1).

Instruments

The MASC-SP (Lahera et al., 2014) assesses mentalization abilities through real-life social situations presented in video format. The video shows four people (two men, two women) interacting at a dinner party, and participants are asked to infer the characters' beliefs/thoughts (mental state reasoning) and feelings (mental state decoding). The video pauses several times to present up to 45 multiple-choice questions about the characters' behaviors and intentions. Each question includes four answer options: one correct and three incorrect, allowing for the identification of different types of mentalization errors: over-mentalization (attributing a mental state when no mental explanation is required), under-mentalization (misattributing an existing mental state), and non-mentalization (total absence of mental inference, such as attributing physical causality to social situations and mental states). An example of the items and their response options can be found elsewhere (Dziobek et al., 2006; Lahera et al., 2014). A primary total score for correct answers, ranging from 0 (worst mentalization) to 45 (best mentalization), and three secondary scores for errors, ranging from 0 (no errors) to 45 (maximum errors), are calculated. The video lasts 15 minutes and

Figure 1
Flow Diagram



Note. MASC-SP, Movie for the Assessment of Social Cognition-Spanish.

participants have up to 30 seconds to answer each question, giving a total administration time of 45 minutes.

Studies examining the psychometric properties of the MASC and MASC-SP have primarily focused on the total score. In the original study (Dziobek et al., 2006), the total score effectively discriminated adults with Asperger syndrome from matched control group participants, achieving an area under the curve of .98. This performance exceeded that of the total scores of the POFA (.79), SST (.65), and RMET (.86). In the Spanish adaptation study (Lahera et al., 2014), moderate-to-strong Pearson correlations were found between the total score and the total scores of the POFA (.57), SST (.68), and RMET (.77), supporting its convergent validity. In contrast, correlations with general cognition, as measured by the Screen for Cognitive Impairment in Psychiatry, were not significant, providing evidence of discriminant validity. The total score also showed optimal internal consistency ($\alpha = .86$) and good test-retest reliability.

Procedure

The study was approved by the Ethics Committee of the University of the Balearic Islands in Mallorca (Balearic Islands), Spain (#011CER25), and was conducted in accordance with the principles of the Declaration of Helsinki (1964) and its subsequent amendments. The study was not preregistered.

All participants completed the MASC-SP via a password-protected online platform (<http://masc-cognicionsocial.es/>), supported by the Instrument Library of the Center for Biomedical Research in Mental Health Network (CIBERSAM) at the Carlos III Health Institute (ISCIII) in Madrid, Spain. Data were also collected on sex (male/female), age (in whole years), educational level (non-university/university), history of neuropsychiatric disorders (yes/no), and whether it was the first time the participant had completed the MASC-SP (yes/no). Participants who did not complete all items on the MASC-SP were excluded from the analysis due to missing data (Figure 1).

Data Analysis

Data normality was verified using histograms, skewness, kurtosis, and the Kolmogorov-Smirnov test. When score distributions deviated from normality, non-parametric analyses were used. Descriptive statistics are presented as frequencies and percentages, means (M) and standard deviations (SD), or medians (Mdn) and ranges (minimum-maximum). The effect size r is also reported. Analyses were performed using IBM SPSS Statistics 29, with statistical significance set at $p < .05$.

The psychometric properties of the MASC-SP scores were examined using Spearman correlations, exploratory factor analysis (EFA) with principal axis factoring and Oblimin rotation, and Cronbach's alpha for internal consistency. For correlations, scores were treated as continuous variables. For EFA and internal consistency, item-level responses were dichotomized, resulting in four binary indicators per item corresponding to total and error scores. We further examined the dimensional structure of the total score using EFA on the tetrachoric correlation matrix, as implemented in the R packages psych (Revelle, 2025) and GPArotation (Bernaards et al., 2025). Consistent with the original conceptualization of the instrument, we primarily evaluated a unidimensional solution, but as a precautionary measure, we also inspected alternative solutions with more than one factor. Decisions about factor retention were guided by a parallel analysis based on factor eigenvalues, in which the eigenvalues obtained from the empirical tetrachoric matrix were compared with those derived from randomly generated data matrices of the same size. Floor/ceiling effects were also evaluated, with values $> 15\%$ indicating possible clustering at the extremes.

In addition, we examined measurement invariance of the general mentalization factor across sex and educational level using multigroup confirmatory factor analysis (CFA) implemented in the lavaan 0.6-21 package (Rosseel, 2012). Building on this CFA model, we re-estimated it in a multigroup framework, treating all items as dichotomous and using probit CFA under the theta parameterization

with the robust Weighted Least Squares with Mean and Variance estimator (WLSMV). The multigroup model syntax was generated using the *semTools* package (Jorgensen et al., 2026), applying the identification strategy for categorical indicators proposed by Wu & Estabrook (2016). For each grouping variable, we then fitted a two-step hierarchy of invariance models. First, a configural model specified the same unidimensional factor structure across groups, while allowing all factor loadings and item thresholds to be freely estimated in each group. Second, a strong invariance model constrained both factor loadings and thresholds to equality across groups for the corresponding grouping variable. Because each item has only one threshold and residual variances are fixed for identification in this probit CFA setting, the joint equality of loadings and thresholds constitutes the practically relevant level of measurement invariance for comparing latent means with binary indicators (Muthén, 2013; Wu & Estabrook, 2016). Model fit was evaluated using chi-square test (χ^2), comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA) with 90% confidence intervals and associated close-fit tests, and standardized root mean square residual (SRMR), following conventional guidelines (Hu & Bentler, 1999). Measurement invariance for sex and educational level was assessed by comparing the configural and strong models using the Satorra-Bentler scaled χ^2 difference test for WLSMV, together with changes in CFI and RMSEA relative to recommended cutoffs for approximate invariance (Chen, 2007; Cheung & Rensvold, 2002).

In accordance with the Standards for Educational and Psychological Testing (American Educational Research Association et al., 2014), these procedures provide reliability evidence for the MASC-SP scores and validity evidence based on its internal structure and interrelationships. Validity claims refer to the interpretation and use of the MASC-SP scores, rather than to the properties of the test itself (Sireci & Benítez, 2023).

To generate normative data, participants were initially grouped into 5-year age intervals, with exceptions at both ends. A narrower interval was applied at the lower end (18-19 years) to capture possible differences related to the development of SC abilities in young adults, while a wider interval was used at the upper end (50-64 years) to address sample size limitations and improve the stability of normative estimates. To examine whether age-related differences in MASC-SP scores were meaningful for the derivation of age-stratified normative data, Kruskal-Wallis H tests were performed, followed by post hoc comparisons between adjacent groups using Mann-Whitney U tests. To define the final age groups, adjacent age groups that did not show statistically significant differences were progressively combined, resulting in aggregated comparison groups (Figures S1-S4). Mann-Whitney U tests were also performed to examine whether differences in MASC-SP scores by sex and educational level were meaningful for the derivation of normative data stratified by these variables. Where appropriate, Bonferroni corrections were applied. Normative data were stratified according to demographic variables that showed statistically significant differences in Mann-Whitney U tests and are presented as empirical percentiles and z -scores.

Results

Table 1 shows the sample characteristics. Descriptive statistics, Kolmogorov-Smirnov tests, and histograms of MASC-SP scores for all participants and stratified by demographic variables are presented in Tables S1 and S2 and Figures S5-S37.

Table 1
Sample Characteristics

	All participants (N = 2,582)
Female sex	1,804 (69.90%)
Age	29 (18-64)
University education (> 12 years)	1,896 (73.40%)
MASC-SP	
Total score	33 (0-45)
Over-mentalization errors	6 (0-45)
Under-mentalization errors	4 (0-29)
Non-mentalization errors	2 (0-45)

Note. MASC-SP, Movie for the Assessment of Social Cognition-Spanish. Results are presented as n (%) or median (range).

Psychometric Evaluation of MASC-SP Scores

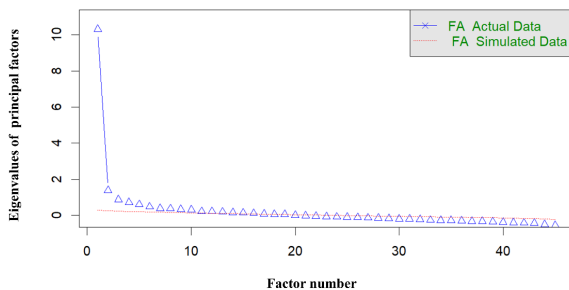
Spearman correlations showed moderate negative associations between the total score and all error scores, a weak positive association between under-mentalization and non-mentalization error scores, a negligible positive association between over-mentalization and non-mentalization error scores, and a non-significant association between over-mentalization and under-mentalization error scores (Table S3).

An EFA was performed for each score. Sampling adequacy was supported by Kaiser-Meyer-Olkin indices ranging from good (.72) to excellent (.91), and data suitability for factor analysis was confirmed by significant Bartlett's tests of sphericity (Table S4). Initial factor extraction suggested multiple factors with eigenvalues > 1 for each score. However, inspection of scree plots consistently indicated a clearly dominant first factor across all scores, characterized by a sharp drop from the first to the second factor, followed by a gradual, nearly linear decline in subsequent eigenvalues (Figures S38-S41). The first factor explained 13.32% of the variance in the total score, 6.83% in the over-mentalization error score, 8.83% in the under-mentalization error score, and 11.79% in the non-mentalization error score. No subsequent factor accounted for > 4.66% of the variance in any score. The item loadings for the first factor of each score are shown in Table S5. Loadings on subsequent factors were generally weak (< .30), often cross-loaded, and with some negative values. Therefore, given the primarily exploratory nature of the study and its applied objectives, retaining a predominantly unifactorial solution for each score was considered the most parsimonious, theoretically consistent, and clearly interpretable approach for normative and clinical purposes.

An additional EFA was performed on the tetrachoric correlation matrix of the total score. The first factor showed an eigenvalue of 10.32, representing 22.90% of the total variance of the items. The remaining factors had substantially smaller eigenvalues (all ≤ 1.38), confirming a marked dominance of the first dimension. To

complement these results, we performed a parallel analysis based on factor eigenvalues, which revealed that although a relatively large number of factors had eigenvalues slightly above those expected under random data, only the first factor clearly exceeded the simulated eigenvalues by a substantial margin. Subsequent factors showed much smaller increments over chance and a gradual, almost linear decline (Figure 2). In line with current recommendations for item-level factor analysis, we therefore retained a predominantly unidimensional solution for the total score, interpreting the first factor as a general mentalizing performance dimension and acknowledging that a considerable proportion of variance remains attributable to item-specific and residual components.

Figure 2
Parallel Analysis of the MASC-SP Total Score (45 Dichotomized Items)



Note. MASC-SP, Movie for the Assessment of Social Cognition-Spanish; FA, Factor analysis.

As a sensitivity check, we also explored a two-factor solution. The first factor accounted for 12.30% of the variance and showed moderate to high loadings for most items. The second factor accounted for an additional 10.60% of the variance, consisting mainly of a small subset of items with substantial cross-loadings and some negative values. This pattern did not yield a theoretically coherent or clearly interpretable multidimensional structure that was also consistent with the original conceptualization of the instrument, and it largely overlapped with the dominant general factor identified in the one-factor solution. Together with the results of the parallel analysis, these findings support retaining a predominantly unidimensional structure for the total score.

Cronbach's alpha values were .80 (total score), .59 (over-mentalization error score), .62 (under-mentalization error score), and .74 (non-mentalization error score).

No floor/ceiling effects were observed in the total score or error scores, except in the case of non-mentalization, which showed a floor effect (20.10%) (Table S6).

Measurement invariance across sex and educational level was evaluated with a two-step sequence of configural and strong models. For sex, the configural model, which specified the same unidimensional factor structure for women and men with freely estimated loadings and thresholds, showed good fit ($\chi^2(1,890) = 3497.80$; CFI = .955; TLI = .953; RMSEA = .026; SRMR = .068), indicating comparable latent structure across groups. Constraining factor loadings and thresholds

to equality (strong invariance) produced some deterioration in fit, although indices still indicated good overall fit ($\chi^2(1,934) = 3973.17$; CFI = .943; TLI = .942; RMSEA = .029; SRMR = .073), and the Satorra-Bentler scaled chi-square difference test was significant, $\Delta\chi^2(44) = 90.71, p < .001$. Even so, changes in approximate fit indices were small ($\Delta\text{CFI} = -.012$; $\Delta\text{RMSEA} = +.003$; $\Delta\text{SRMR} = +.005$) and within recommended cutoffs for approximate invariance, supporting the view that strong invariance across sex is tenable in practical terms and that latent mean comparisons between women and men are substantively interpretable. For educational level, the configural model also showed good fit, suggesting that the unidimensional mentalizing factor operated similarly among university-educated versus non-university-educated participants ($\chi^2(1,890) = 3378.77$; CFI = .953; TLI = .951; RMSEA = .025; SRMR = .066). Imposing equality constraints on factor loadings and thresholds across these groups (strong invariance) again led to some loss of fit, although the global indices remained acceptable ($\chi^2(1,934) = 3481.42$; CFI = .943; TLI = .941; RMSEA = .027; SRMR = .070), and the Satorra-Bentler scaled chi-square difference test indicated significantly worse fit than the configural model, $\Delta\chi^2(44) = 102.65, p < .001$. As in the case of sex, the shifts in approximate fit indices were modest ($\Delta\text{CFI} = -.010$; $\Delta\text{RMSEA} = +.002$; $\Delta\text{SRMR} = +.004$) and stayed within commonly used criteria for approximate invariance, indicating that strong invariance across educational level is also met to a practically acceptable extent and that latent mean differences between education-level groups can be interpreted as reflecting substantive rather than measurement artifacts.

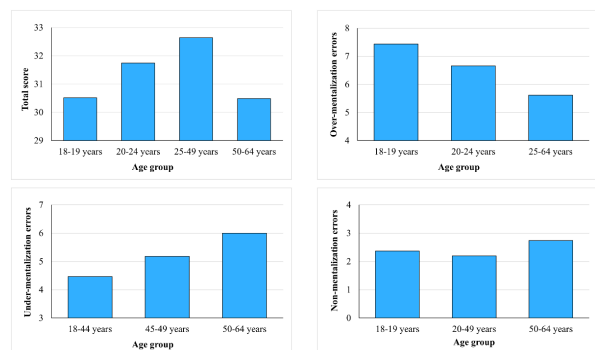
Normative Data for the MASC-SP Total Score

Women scored higher than men (Table S7), and participants with a university education also scored higher than those without (Table S8); this effect was observed in women and men (Table S9).

Total scores differed across initial age groups, $H(7) = 72.91, p < .001$ (Figure S1). In the final age groups, participants aged 18-19 scored lower than those aged 20-24, who in turn scored lower than those aged 25-49. Participants aged 25-49 scored higher than those aged 50-64 (Figure 3, Table S10).

Normative data are presented in Tables 2 and 3.

Figure 3
MASC-SP Scores by Age Group. Means are Reported



Note. MASC-SP, Movie for the Assessment of Social Cognition-Spanish.

Table 2Normative Data for the MASC-SP Total Score in Male Participants by Age Group and Educational Level ($N = 778$)

MASC-SP total score in men											
z-score	PC	Non-university education (≤ 12 years)					University education (> 12 years)				
		18-19	20-24	25-49	50-64	Total	18-19	20-24	25-49	50-64	Total
-2.326	1	10	0	0	-	0	-	1	11	11	11
-1.645	5	13	13	13	15	14	17	22	21	14	20
-1.282	10	19	18	17	18	18	19	27	27	25	27
-.842	20	23	24	-	22	23	27	30	30	28	30
-.524	30	25	28	23	24	25	30	32	31	30	31
-.253	40	27	30	26	25	27	31	-	33	32	33
0	50	29	31	28	27	29	-	33	34	33	-
.253	60	31	33	30	28	31	34	34	35	-	34
.524	70	32	34	33	32	33	35	35	36	34	35
.842	80	33	35	34	34	34	36	36	37	36	37
1.282	90	34	38	36	35	36	37	37	38	38	38
1.645	95	35	39	40	37	38	38	39	39	39	39
2.326	99	-	-	44	-	41	-	-	41	-	41
<i>n</i>		55	90	126	24	295	27	98	308	50	483
<i>M</i>		27.564	29.356	27.587	27.000	28.075	31.148	32.204	32.834	31.420	32.466
<i>SD</i>		6.058	8.102	7.928	6.269	7.560	6.163	5.936	5.532	6.276	5.740
<i>Mdn</i>		29	31	28	26.500	29	34	33	34	33	34
Range		10-37	0-41	0-45	15-38	0-45	17-38	1-42	3-42	11-39	1-42
Skewness		-1.030	-1.480	-.793	-.094	-.986	-1.231	-2.997	-1.806	-1.691	-2.002
Kurtosis		.848	2.755	1.321	-.687	1.596	.601	13.403	5.185	3.131	6.277

Note. MASC-SP, Movie for the Assessment of Social Cognition-Spanish; PC, Percentile; *M*, Mean; *SD*, Standard Deviation; *Mdn*, Median.**Table 3**Normative Data for the MASC-SP Total Score in Female Participants by Age Group and Educational Level ($N = 1,804$)

MASC-SP total score in women											
z-score	PC	Non-university education (≤ 12 years)					University education (> 12 years)				
		18-19	20-24	25-49	50-64	Total	18-19	20-24	25-49	50-64	Total
-2.326	1	13	11	7	9	11	20	13	14	14	14
-1.645	5	18	20	18	13	19	23	25	25	23	24
-1.282	10	22	24	22	16	21	26	27	29	26	27
-.842	20	26	26	27	20	26	28	29	31	29	30
-.524	30	28	29	29	22	28	30	31	32	30	32
-.253	40	30	31	30	28	30	31	32	33	32	33
0	50	31	32	32	29	31	33	33	34	33	34
.253	60	32	33	33	30	33	35	34	35	34	35
.524	70	33	34	34	-	34	36	35	36	35	36
.842	80	36	35	35	33	35	37	36	37	37	37
1.282	90	37	37	37	34	37	-	38	39	38	38
1.645	95	39	38	38	35	38	38	39	40	40	39
2.326	99	-	42	42	-	41	40	41	41	41	41
<i>n</i>		78	105	165	43	391	103	317	890	103	1,413
<i>M</i>		30.026	31.010	30.509	26.884	30.148	32.311	32.511	33.685	32.359	33.225
<i>SD</i>		5.930	5.402	6.152	6.877	6.097	4.627	5.193	4.877	4.974	4.971
<i>Mdn</i>		30.500	32	32	29	31	33	33	34	33	34
Range		13-40	11-42	7-43	9-36	7-43	20-40	9-45	6-45	14-41	6-45
Skewness		-.754	-.753	-1.300	-.777	-1.028	-.552	-1.471	-1.750	-.778	-1.511
Kurtosis		.661	.878	2.267	-.303	1.283	-.434	3.785	5.244	1.049	3.986

Note. MASC-SP, Movie for the Assessment of Social Cognition-Spanish; PC, Percentile; *M*, Mean; *SD*, Standard Deviation; *Mdn*, Median.

Normative Data for the MASC-SP Error Scores

Women made fewer over-mentalization and non-mentalization errors than men, with no significant sex differences in under-mentalization errors (Table S7). Participants with a university education made fewer over-mentalization, under-mentalization, and non-mentalization errors than those without (Table S8). The effect of educational level on over-mentalization and non-mentalization errors was observed in women and men (Table S9).

Over-mentalization errors differed across initial age groups, $H(7) = 106.50, p < .001$ (Figure S2). In the final age groups, participants aged 18-19 made more over-mentalization errors than those aged 20-24, who in turn made more than those aged 25-64 (Figure 3, Table S10).

Under-mentalization errors differed across initial age groups, $H(7) = 70.67, p < .001$ (Figure S3). In the final age groups, participants aged 18-44 made fewer under-mentalization errors than those aged 45-49, who in turn made fewer than those aged 50-64 (Figure 3, Table S10).

Non-mentalization errors differed across initial age groups, $H(7) = 28.50, p < .001$ (Figure S4). In the final age groups, participants aged 18-19 made more non-mentalization errors than those aged 20-49, who in turn made fewer than those aged 50-64 (Figure 3, Table S10).

Normative data are presented in Tables S11-S15.

Discussion

This study presents the first large-scale psychometric evaluation and normative framework for MASC-SP scores in Spanish-speaking adults. For the first time, the psychometric properties of total and error scores have been evaluated. The results support a predominantly unifactorial structure for each score, with good internal consistency for the total score, acceptable internal consistency for the non-mentalization error score, and lower but still interpretable internal consistency for the over-mentalization and under-mentalization error scores. Based on this psychometric foundation, normative data are also presented stratified by sex and educational level, which were shown to function under conditions of measurement invariance.

Spearman correlations showed negative associations between the total score and all error scores, along with positive and selective correlations between error scores, thus capturing partially overlapping but not redundant aspects of general mentalization performance and associated error patterns, in line with the original conceptualization of the instrument's scores (Dziobek et al., 2006; Lahera et al., 2014).

Although the proportion of variance explained by the one-factor solution was modest, subsequent factors accounted for much smaller amounts of variance and did not form coherent alternative structures. In this regard, the tetrachoric correlation matrix of the total score, together with the parallel analysis, provided converging evidence that only the first factor clearly exceeded random eigenvalues, supporting the retention of a predominantly unidimensional solution.

Consistent with this structure, the total score showed the highest reliability, supporting its use as a robust primary quantitative indicator of general mentalization performance in Spanish-speaking adults. In contrast, error scores showed more heterogeneous reliability, supporting their interpretation as complementary indicators of specific response patterns, rather than as scores reflecting independent latent

constructs, in line with their original conceptualization as secondary outcomes (Dziobek et al., 2006; Lahera et al., 2014).

The ceiling effect commonly observed in scores on traditional SC instruments (Pinkham et al., 2018) was not observed in our results, but a floor effect was detected in the non-mentalization error score, as one-fifth of participants, particularly those with university education, made few or no non-mentalization errors. This is consistent with results obtained in non-clinical samples characterized by intact mentalization abilities (Cotter et al., 2018) and supports that MASC-SP scores provide an adequate range of measurements and are sensitive to individual differences in healthy populations, especially beyond the lower end of the error distribution and within the mid-to-high range of error frequency. Importantly, greater discriminatory power of the non-mentalization error score is expected in clinical samples, whose error rates are typically higher than those of healthy populations (Dziobek et al., 2006; Lahera et al., 2014).

Overall, these results are consistent with those of previous psychometric studies, which have reported a comparable unifactorial solution (Chang et al., 2024; Fossati et al., 2018; Müller et al., 2016) and similar internal consistency for the total score (Chang et al., 2024; Dziobek et al., 2006; Fossati et al., 2018; Huang et al., 2023; Lahera et al., 2014; Martinez et al., 2017; Sarig-Shmueli et al., 2023). However, none of these studies examined the factor structure, reliability, and distributional properties of error scores, highlighting the novelty of the present work. These analyses were exploratory and did not aim to establish formal latent trait models, but rather to evaluate whether each error response category was characterized by a predominant source of shared variance, so that each error score could be considered a distinguishable and interpretable aspect of instrument performance beyond the total score.

This study also presents the first normative framework for MASC-SP scores using empirical percentiles and *z*-scores. The data were derived from a large-scale community sample and are applicable to research and clinical settings. Sex differences showed small-magnitude effect sizes, with women scoring higher overall and making fewer over-mentalization and non-mentalization errors than men. Educational level showed small-to-medium effect sizes, with university-educated participants scoring higher overall and making fewer mentalization errors of any kind than non-university-educated participants. Importantly, measurement invariance across sex and educational level was confirmed, indicating that differences between groups likely reflect substantive variations in general mentalization performance rather than measurement artifacts, justifying the development of group-stratified norms.

In contrast, the influence of age was heterogeneous. While total scores varied across age groups, with middle-aged participants scoring higher than younger and older groups, non-mentalization errors showed an inverse pattern, mirroring differences in total scores. Furthermore, over-mentalization errors tended to decrease with age, and under-mentalization errors tended to increase. Despite these results, differences between adjacent age groups were characterized by small-to-negligible effect sizes, calling into question their relevance as clinically meaningful or developmentally related changes. Consequently, age-stratified norms are presented for descriptive purposes only and should be used with caution. Therefore, it is recommended that non-age-stratified scores be prioritized over age-stratified scores when applying these norms in

research and clinical settings. Further studies are needed to provide continuous norms using age-conditioned metrics to reduce potential artifacts introduced by age binning.

Overall, these results are consistent with those of the Spanish adaptation of the MASC, in which healthy participants showed a similar scoring profile to that of the participants in the present study (Lahera et al., 2014), and with those of other studies that have examined the influence of sex and educational level on SC performance using traditional instruments (Christov-Moore et al., 2014; Navarra-Ventura et al., 2018, 2021, 2022; Vaskinn et al., 2024). The results are also consistent with recent evidence that challenges the traditional view of age-related cognitive change as a linear process and instead emphasizes that mentalizing performance across the lifespan is determined by a heterogeneous set of factors, including individual differences in cognitive/executive functions, brain reserve, and brain pathology (Arioli et al., 2018; Fittipaldi et al., 2024). In summary, when applying these norms, it is essential to bear in mind that educational level had a consistent influence on general mentalizing performance and associated error patterns, while sex only influenced specific scores. Age, on the other hand, exerted influences whose clinical or developmental significance is questionable.

When evaluating the strengths and weaknesses of normative data, several criteria have been proposed (Mitrushina et al., 2005): (1) ≥ 50 participants per cell; detailed reports on (2) sample composition and (3) test administration; presentation of data by (4) sex, (5) age, and (6) education/intelligence quotient (IQ); and (7) presentation of descriptive statistics for each normative cell. The present study meets five of these criteria. The absence of IQ data (criterion 6) is unlikely to be a major limitation, given the stratification of data by educational level and the well-established positive correlation between education and IQ (Deary & Johnson, 2010). A second limitation is that not all normative cells included ≥ 50 participants (criterion 1), particularly those corresponding to non-university-educated male participants in the younger and older age groups. This shortcoming is mitigated by the total sample size, which is larger than that of previous studies on the MASC and MASC-SP (Chang et al., 2024; Dziobek et al., 2006; Fossati et al., 2018; Huang et al., 2023; Lahera et al., 2014; Martínez et al., 2017; Müller et al., 2016; Sarig-Shmueli et al., 2023). Furthermore, when normative cells are considered independently of age groups, all of them meet the threshold of ≥ 50 participants, reinforcing the recommendation to prioritize non-age-stratified scores over age-stratified scores when using the norms. A third limitation concerns the use of non-probability sampling and voluntary online participation, which may have generated a certain degree of self-selection bias, resulting in a sample with a sociodemographic profile skewed toward university-educated young women. This limits the generalizability of the results and suggests that normative data should be considered preliminary and more representative of populations with sociodemographic characteristics similar to those in the present study. Finally, no information was collected on the region of residence of participants within Spain, which prevented reliable alignment of the norms with population benchmarks in key geographic strata. Future research using probabilistic and stratified sampling methods is needed to refine these norms and establish region-specific validations of the instrument scores. This is of great importance, especially for Latin American countries, as cultural, contextual, and linguistic factors may influence mentalization performance (Fittipaldi et al., 2024), so the direct application of these Spain-based norms to other Spanish-speaking populations outside Spain should be approached with caution.

In conclusion, this study presents the first large-scale psychometric evaluation and normative framework for MASC-SP scores in Spanish-speaking adults. The results provide reliability evidence for total and error scores, as well as validity evidence based on internal structure and interrelationships, supporting the interpretation of these scores as indicators of general mentalization performance and associated error patterns. Within this framework, the total score can be used as a robust primary quantitative indicator of general mentalization performance, while error scores are best understood as secondary qualitative indicators of specific response patterns. Normative data stratified by sex and educational level provide a useful reference for describing an individual's position relative to demographically comparable reference groups and for making group-level comparisons in research and clinical settings. However, the validity evidence presented is mainly based on internal structure and does not yet support higher-level uses such as individual diagnostic classification or treatment allocation. Extending the interpretation of these scores to such uses requires accumulating additional sources of validity evidence, particularly evidence based on relationships with other variables (Dziobek et al., 2006; Lahera et al., 2014), on response processes, and on testing consequences (American Educational Research Association et al., 2014; Sireci & Benítez, 2023).

Author Contributions

Aina Llompарт-Casasnovas: Data curation, Formal analysis, Visualization, Writing-Review & Editing. **Adoración Castro:** Writing-Review & Editing. **Margalida Gili:** Supervision, Writing-Review & Editing. **Guillem Navarra-Ventura:** Conceptualization, Data curation, Formal analysis, Project administration, Visualization, Writing-Original Draft, Writing-Review & Editing. **Roberto Rodríguez-Jimenez:** Funding acquisition, Investigation, Writing-Review & Editing. **Miguel A. Ortega:** Investigation, Writing-Review & Editing. **Leticia Boada:** Investigation, Writing-Review & Editing. **Mercè Jodar:** Writing-Review & Editing. **Narcís Cardoner:** Writing-Review & Editing. **Albert Sesé:** Formal analysis, Methodology, Writing-Review & Editing. **Miquel Roca:** Supervision, Writing-Review & Editing. **Guillermo Lahera:** Conceptualization, Funding acquisition, Investigation, Resources, Writing-Review & Editing.

Funding

This work was funded by the Instituto de Salud Carlos III (ISCIII) and co-funded by the European Regional Development Fund (ERDF) (grant number PI21/01252). Additional support was provided by the Regional Government of Madrid (R&D Program in Biomedicine, S2022/BMD-7216, AGES 3-CM) and by the Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM, ISCIII).

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.

Declaration of Interest

Guillem Navarra-Ventura, Roberto Rodríguez-Jimenez, Miquel Roca and Guillermo Lahera have served as consultants for,

participated in sponsored activities of, or received research grants from JanssenCilag, Lundbeck, Otsuka, Pfizer, Ferrer, Juste, Takeda, Esteve, Exeltis, Casen Recordati, Angelini, Adamed, Boehringer, or Rovi. The remaining authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Supplementary materials are accessible at https://osf.io/46gij/?view_only=ea167e207bcf4c319f41e2b19bdbcb23

References

- Aidelbaum, R., & Goghari, V. (2022). A visual task-based assessment of theory of mind and social perception within bipolar disorder. *Journal of Nervous & Mental Disease*, 210(1), 37–44. <https://doi.org/10.1097/NMD.0000000000001408>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Andrade-González, N., Sarasa, M., García-López, A., Leonés, I., Halverson, T. F., & Lahera, G. (2021). Mentalizing errors in patients with schizophrenia who received psychosocial rehabilitation: A case-control study. *Psychiatric Quarterly*, 92(3), 947–959. <https://doi.org/10.1007/s11126-020-09863-x>
- Andreou, C., Kelm, L., Bierbrodt, J., Braun, V., Lipp, M., Yassari, A. H., & Moritz, S. (2015). Factors contributing to social cognition impairment in borderline personality disorder and schizophrenia. *Psychiatry Research*, 229(3), 872–879. <https://doi.org/10.1016/j.psychres.2015.07.057>
- Arioli, M., Crespi, C., & Canessa, N. (2018). Social cognition through the lens of cognitive and clinical neuroscience. *BioMed Research International*, 2018, Article 4283427. <https://doi.org/10.1155/2018/4283427>
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42(2), 241–251.
- Bernaards, C. A., Gilbert, P., & Jennrich, R. I. (2025). *GPArotation: Gradient projection factor rotation (Version 2025.3-1)* [R package]. R Foundation for Statistical Computing. <https://CRAN.R-project.org/package=GPArotation>
- Catalan, A., Angosto, V., Díaz, A., Martínez, N., Guede, D., Pereda, M., Madrazo, A., Bustamante, S., Bilbao, A., Osa, L., Inchausti, L., & Gonzalez-Torres, M. A. (2018). The relationship between theory of mind deficits and neurocognition in first episode-psychosis. *Psychiatry Research*, 268, 361–367. <https://doi.org/10.1016/j.psychres.2018.06.066>
- Cavieres, A., Acuña, V., Arancibia, M., & Escobar, C. (2025). Advances in the ecological validity of research on social cognition in schizophrenia: A systematic review of the literature. *Schizophrenia Research: Cognition*, 39, Article 100333. <https://doi.org/10.1016/j.scog.2024.100333>
- Chang, Y., Yu, C., Huang, C., Wang, T., Dziobek, I., & Lane, H. (2024). Discrepancy of social cognition between bipolar disorders and major depressive disorders. *Brain and Behavior*, 14(1), Article e3365. <https://doi.org/10.1002/brb3.3365>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews*, 46, 604–627. <https://doi.org/10.1016/j.neubiorev.2014.09.001>
- Cotter, J., Granger, K., Backx, R., Hobbs, M., Looi, C. Y., & Barnett, J. H. (2018). Social cognitive dysfunction as a clinical marker: A systematic review of meta-analyses across 30 clinical conditions. *Neuroscience & Biobehavioral Reviews*, 84, 92–99. <https://doi.org/10.1016/j.neubiorev.2017.11.014>
- Deary, I. J., & Johnson, W. (2010). Intelligence and education: Causal perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology*, 39(5), 1362–1369. <https://doi.org/10.1093/ije/dyq072>
- Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Convit, A. (2006). Introducing MASC: A Movie for the Assessment of Social Cognition. *Journal of Autism and Developmental Disorders*, 36(5), 623–636. <https://doi.org/10.1007/s10803-006-0107-0>
- Ekman, P., & Friesen, W. C. (1976). *Pictures of facial affect*. Consulting Psychologists Press.
- Engelstad, K. N., Rund, B. R., Torgalsbøen, A.-K., Lau, B., Ueland, T., & Vaskinn, A. (2019). Large social cognitive impairments characterize homicide offenders with schizophrenia. *Psychiatry Research*, 272, 209–215. <https://doi.org/10.1016/j.psychres.2018.12.087>
- Fittipaldi, S., Legaz, A., Maito, M., Hernandez, H., Altschuler, F., Canziani, V., Moguilner, S., Gillan, C. M., Castillo, J., Lillo, P., Custodio, N., Avila-Funes, J. A., Cardona, J. F., Slachevsky, A., Henriquez, F., Fraile-Vazquez, M., Cruz de Souza, L., Borroni, B., Hornberger, M., ... Ibanez, A. (2024). Heterogeneous factors influence social cognition across diverse settings in brain health and age-related diseases. *Nature Mental Health*, 2(1), 63–75. <https://doi.org/10.1038/s44220-023-00164-3>
- Fossati, A., Borroni, S., Dziobek, I., Fonagy, P., & Somma, A. (2018). Thinking about assessment: Further evidence of the validity of the Movie for the Assessment of Social Cognition as a measure of mentalistic abilities. *Psychoanalytic Psychology*, 35(1), 127–141. <https://doi.org/10.1037/pap0000130>
- Fretland, R. A., Andersson, S., Sundet, K., Andreassen, O. A., Melle, I., & Vaskinn, A. (2015). Theory of mind in schizophrenia: Error types and associations with symptoms. *Schizophrenia Research*, 162(1–3), 42–46. <https://doi.org/10.1016/j.schres.2015.01.024>
- Gur, R. C. (2002). Sex differences in temporo-limbic and frontal brain volumes of healthy adults. *Cerebral Cortex*, 12(9), 998–1003. <https://doi.org/10.1093/cercor/12.9.998>
- Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters’ thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24(2), 129–154. <https://doi.org/10.1007/BF02172093>

- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Huang, Y.-L., Chen, T.-T., Dziobek, I., & Tseng, H.-H. (2023). Mentalizing in a Movie for the Assessment of Social Cognition (MASC): The validation in a taiwanese sample. *BMC Psychology*, 11(1), Article 287. <https://doi.org/10.1186/s40359-023-01321-0>
- Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Hakonarson, H., Gur, R. E., Gur, R. C., & Verma, R. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences*, 111(2), 823-828. <https://doi.org/10.1073/pnas.1316909110>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2026). *semTools: Useful tools for structural equation modeling (Versión 0.5-8)*. [R package]. R Foundation for Statistical Computing. <https://CRAN.R-project.org/package=semTools>
- Lahera, G., Boada, L., Pousa, E., Mirapeix, I., Morón-Nozaleda, G., Marinas, L., Gisbert, L., Pamiàs, M., & Parellada, M. (2014). Movie for the Assessment of Social Cognition (MASC): Spanish validation. *Journal of Autism and Developmental Disorders*, 44(8), 1886-1896. <https://doi.org/10.1007/s10803-014-2061-6>
- Martinez, G., Alexandre, C., Mam-Lam-Fook, C., Bendjema, N., Gaillard, R., Garel, P., Dziobek, I., Amado, I., & Krebs, M. O. (2017). Phenotypic continuum between autism and schizophrenia: Evidence from the Movie for the Assessment of Social Cognition (MASC). *Schizophrenia Research*, 185, 161-166. <https://doi.org/10.1016/j.schres.2017.01.012>
- Mitchell, R. L. C., & Phillips, L. H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, 70, 1-10. <https://doi.org/10.1016/j.neuropsychologia.2015.02.018>
- Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment*. Oxford University Press.
- Montag, C., Dziobek, I., Richter, I. S., Neuhaus, K., Lehmann, A., Sylla, R., Heekeren, H. R., Heinz, A., & Gallinat, J. (2011). Different aspects of theory of mind in paranoid schizophrenia: Evidence from a video-based assessment. *Psychiatry Research*, 186(2-3), 203-209. <https://doi.org/10.1016/j.psychres.2010.09.006>
- Montag, C., Ehrlich, A., Neuhaus, K., Dziobek, I., Heekeren, H. R., Heinz, A., & Gallinat, J. (2010). Theory of mind impairments in euthymic bipolar patients. *Journal of Affective Disorders*, 123(1-3), 264-269. <https://doi.org/10.1016/j.jad.2009.08.017>
- Msika, E.-F., Despres, M., Piolino, P., & Narme, P. (2024). Dynamic and/or multimodal assessments for social cognition in neuropsychology: Results from a systematic literature review. *The Clinical Neuropsychologist*, 38(4), 922-962. <https://doi.org/10.1080/13854046.2023.2266172>
- Müller, N., Baumeister, S., Dziobek, I., Banaschewski, T., & Poustka, L. (2016). Validation of the Movie for the Assessment of Social Cognition in adolescents with ASD: Fixation duration and pupil dilation as predictors of performance. *Journal of Autism and Developmental Disorders*, 46(9), 2831-2844. <https://doi.org/10.1007/s10803-016-2828-z>
- Muthén, B. (2013). *Binary item measurement invariance testing with categorical factor analysis* [Unpublished technical report]. Retrieved from <https://www.statmodel.com/download/Binary%20item%20measurement%20invariance%20testing.pdf>
- Navarra-Ventura, G., Fernandez-Gonzalo, S., Turon, M., Pousa, E., Palao, D., Cardoner, N., & Jodar, M. (2018). Gender differences in social cognition: A cross-sectional pilot study of recently diagnosed patients with schizophrenia and healthy subjects. *Canadian Journal of Psychiatry*, 63(8), 538-546. <https://doi.org/10.1177/0706743717746661>
- Navarra-Ventura, G., Vicent-Gil, M., Serra-Blasco, M., Cobo, J., Fernández-Gonzalo, S., Goldberg, X., Jodar, M., Crosas, J. M., Palao, D., Lahera, G., Vieta, E., & Cardoner, N. (2022). Higher order theory of mind in patients with bipolar disorder and schizophrenia/schizoaffective disorder. *European Archives of Psychiatry and Clinical Neuroscience*, 272(3), 497-507. <https://doi.org/10.1007/s00406-021-01265-9>
- Navarra-Ventura, G., Vicent-Gil, M., Serra-Blasco, M., Massons, C., Crosas, J. M., Cobo, J., Jubert, A., Jodar, M., Fernández-Gonzalo, S., Goldberg, X., Palao, D., Lahera, G., Vieta, E., & Cardoner, N. (2021). Group and sex differences in social cognition in bipolar disorder, schizophrenia/schizoaffective disorder and healthy people. *Comprehensive Psychiatry*, 109, Article 152258. <https://doi.org/10.1016/j.comppsy.2021.152258>
- Newbury-Helps, J., Feigenbaum, J., & Fonagy, P. (2017). Offenders with antisocial personality disorder display more impairments in mentalizing. *Journal of Personality Disorders*, 31(2), 232-255. <https://doi.org/10.1521/pedi.2016.30.246>
- Pinkham, A. E., Harvey, P. D., & Penn, D. L. (2018). Social cognition psychometric evaluation: Results of the final validation study. *Schizophrenia Bulletin*, 44(4), 737-748. <https://doi.org/10.1093/schbul/sbx117>
- Preißler, S., Dziobek, I., Ritter, K., Heekeren, H. R., & Roepke, S. (2010). Social cognition in borderline personality disorder: Evidence for disturbed recognition of the emotions, thoughts, and intentions of others. *Frontiers in Behavioral Neuroscience*, 4, Article 182. <https://doi.org/10.3389/fnbeh.2010.00182>
- Quek, J., Bennett, C., Melvin, G. A., Saeedi, N., Gordon, M. S., & Newman, L. K. (2018). An investigation of the mentalization-based model of borderline pathology in adolescents. *Comprehensive Psychiatry*, 84, 87-94. <https://doi.org/10.1016/j.comppsy.2018.04.005>
- Revelle, W. (2025). *psych: Procedures for psychological, psychometric, and personality research (Version 2.5.3)* [R package]. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Romero-Ferreiro, V., García-Fernández, L., Aparicio, A. I., Martínez-Gras, I., Dompablo, M., Sánchez-Pastor, L., Rentero, D., Alvarez-Mon, M. Á., Espejo-Saavedra, J. M., Lahera, G., Mari-Beffa, P., Santos, J. L., & Rodríguez-Jimenez, R. (2022). Emotional processing profile in patients with first episode schizophrenia: The influence of neurocognition. *Journal of Clinical Medicine*, 11(7), Article 2044. <https://doi.org/10.3390/jcm11072044>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Sahl, A. C., Rognlien, H. F., Andreassen, O. A., Melle, I., Ueland, T., & Vaskinn, A. (2023). Theory of mind in schizophrenia: A comparison of subgroups with low and high IQ. *Nordic Journal of Psychiatry*, 77(4), 329-335. <https://doi.org/10.1080/08039488.2022.2106512>
- Santos, J. M., Pousa, E., Soto, E., Comes, A., Roura, P., Arrufat, F. X., & Obiols, J. E. (2017). Theory of mind in euthymic bipolar patients and first-degree relatives. *The Journal of Nervous and Mental Disease*, 205(3), 207-212. <https://doi.org/10.1097/NMD.0000000000000595>

- Sarig-Shmueli, Y., Kurman, J., Talmon, S., Yoeli, N., Gur, E., Stein, D., & Rothschild-Yakar, L. (2023). Mentalizing ability, mentalizing impairments, and anorexia nervosa: Validation of the hebrew version of the Movie for the Assessment of Social Cognition (MASC). *Clinical Psychology & Psychotherapy*, 30(4), 873–881. <https://doi.org/10.1002/cpp.2848>
- Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*, 147(3), 293–327. <https://doi.org/10.1037/bul0000303>
- Seitz, K. I., Ehler, N., Schmitz, M., Schmitz, S. E., Dziobek, I., Herpertz, S. C., & Bertsch, K. (2022). Affective and cognitive theory of mind in posttraumatic stress, major depressive, and somatic symptom disorders: Association with childhood trauma. *British Journal of Clinical Psychology*, 61(3), 680–700. <https://doi.org/10.1111/bjc.12357>
- Sireci, S. G., & Benítez, I. (2023). Evidence for test validation: A guide for practitioners. *Psicothema*, 35(3), 217–226. <https://doi.org/10.7334/psicothema2022.477>
- Somma, A., Ferrara, M., Terronni, A., Frau, C., Ardizzone, I., Sharp, C., & Fossati, A. (2019). Hypermentalizing as a marker of borderline personality disorder in italian adolescents: A cross-cultural replication of Sharp and colleagues' (2011) findings. *Borderline Personality Disorder and Emotion Dysregulation*, 6(1), Article 5. <https://doi.org/10.1186/s40479-019-0104-5>
- Vaskinn, A., Andersson, S., Østefjells, T., Andreassen, O. A., & Sundet, K. (2018). Emotion perception, non-social cognition and symptoms as predictors of theory of mind in schizophrenia. *Comprehensive Psychiatry*, 85, 1–7. <https://doi.org/10.1016/j.comppsy.2018.05.002>
- Vaskinn, A., Ueland, T., Melle, I., & Sundet, K. (2024). Sex differences in social cognition among individuals with schizophrenia and in healthy control participants: A secondary analysis of published data. *Archives of Women's Mental Health*, 27(5), 661–667. <https://doi.org/10.1007/s00737-024-01422-8>
- Wolkenstein, L., Schönenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of Affective Disorders*, 132(1-2), 104–111. <https://doi.org/10.1016/j.jad.2011.02.010>
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014–1045. <https://doi.org/10.1007/s11336-016-9506-0>
- Zwack, J. C., & Wolkenstein, L. (2017). Facial emotion recognition, theory of mind and the role of facial mimicry in depression. *Journal of Affective Disorders*, 210, 90–99. <https://doi.org/10.1016/j.jad.2016.12.022>