Article

# Spanish Adaptation of the Mentalization Questionnaire (MZQ) in Community Adolescents and Adults

Jacqueline Nonweiler[1] ⬥, Eduardo Doval[1] ⬥, Neus Barrantes-Vidal[1,2] ⬥ and Sergi Ballespí[1] ⬥

1 Universitat Autònoma de Barcelona (Spain)
2 CIBER de Salud Mental, Instituto de Salud Carlos III (Spain)

## ABSTRACT

**Background:** Despite the role of mentalization in mental health outcomes and prevention, psychometrically-evaluated screening measures for mentalization remain sparse. One widely-used mentalization questionnaire is the Mentalization Questionnaire (MZQ; Hausberg et al., 2012), which we aimed to adapt and validate for use in Spanish. **Method:** We adapted the MZQ to European Spanish and evaluated its psychometric properties in both adolescent ($n$ = 389, ages 12-19, $M$ = 14.5) and adult community samples ($n$ = 382, $M$ = 48). **Results:** Confirmatory factor analysis resulted in a unidimensional structure including all items. This model had better goodness of fit than the original and other adaptations. Invariance analysis showed the same structure in adolescents compared by sex and age, and additionally in the adult versus adolescent samples. Evidence for convergent and discriminant validity was found. Internal consistency values in both adolescents and parents were fair and in the adolescent sample the MZQ scores remained moderately stable after re-test. **Conclusions:** The Spanish adaptation of the MZQ presents similar evidence of reliability and validity in the adolescent and adult samples. The results support this being a suitable version for evaluating mentalization in the general population.

## Adaptación Española del Cuestionario de Mentalización (Mentalization Questionnaire; MZQ) para Adolescentes y Adultos de Población General

### RESUMEN

**Antecedentes:** A pesar del creciente reconocimiento de la mentalización como factor en el desarrollo de la salud mental y la prevención, las medidas de cribado validadas psicométricamente siguen siendo escasas. El Cuestionario de Mentalización (MZQ; Hausberg et al., 2012) es uno de los más utilizados. El objetivo es adaptar y validar su uso al español. **Método:** Adaptamos el MZQ al español europeo en muestras comunitarias de adolescentes ($n$ = 389, edades 12-19, $M$ = 14,5) y de adultos ($n$ = 382, $M$ = 48). **Resultados:** Se realizó un análisis factorial confirmatorio que expuso una estructura unifactorial para ambas muestras. Este modelo presentó mejores índices de ajuste que los modelos presentados en la versión original y en las adaptaciones. El estudio de invariancia mostró la misma estructura en la muestra de adolescentes cuando se compararon por sexo y edad, y también en la muestra de adultos comparada con la muestra de adolescentes. Se encontraron evidencias de validez convergente y discriminante. **Conclusiones:** La adaptación del MZQ al español presenta evidencias de validez y fiabilidad similares en la muestra de adolescentes y en la de adultos. Los resultados apoyan que se trata de una versión apta para evaluar la mentalización en población general.

Mentalization has garnered notable interest for its role in mental health (Johnson et al., 2022). Comprised of four neuroscientifically-based polarities, this high-order cognition is the ability to notice and comprehend one's own and other peoples' internal mental states such as thoughts, feelings, emotions, motivations, and intents (Bateman & Fonagy, 2004; Fonagy, 1993). Increasing evidence points to mentalization as a global protective factor against psychopathology (Ballespí et al., 2018) and suffering in the presence of psychological disorders, and constitutes a common active ingredient of mental health treatments (Luyten et al., 2020). Mentalization is consistently, negatively associated with internalizing symptoms (Chevalier et al., 2023) and neuroticism (e.g. Dimitrijević et al., 2018), but has positive associations with resiliency (Fonagy & Campbell, 2017) and secure attachment (Fonagy & Target, 1997).

Despite the popularity and utility of mentalization as a construct for better understanding psychopathology, mentalization assessment remains an issue in research. Mentalizing abilities are classically assessed through interviews such as the Adult Attachment Interview (Main & Goldwyn, 1998) and the Child Attachment Interview (George et al., 1996), which require transcription and subsequent analysis using the Reflective Functioning Scale (RFS; Fonagy et al., 1998), which allows clinicians and researchers to classify mentalizing capacities by evaluating interview responses. Ensink and colleagues (Ensink et al., 2015) then adapted these measures for children, further including scales for self- and other- mentalizing polarities. While attachment interviews and analysis using RFS are the most thorough means of evaluating mentalization, this method requires extensive time investment, training, and accreditation both to administer and score the interview. These two factors result in a costly operation, which is not always available or reasonable, particularly in involved research studies or broad community samples where a screening instrument would be more time- and cost-efficient.

Mentalization is also commonly evaluated with measures that evaluate related concepts such as Theory of Mind (i.e., Strange Stories Task) (Happé, 1994), scales to assess emotional intelligence, including the Mayer-Salovey-Caruso Emotional Intelligence Test (Mayer et al., 2002), and the Movie for the Assessment of Social Cognition (Dziobek et al., 2006). More specific scales like the Trait Meta-Mood Scale also exist. The TMMS is broadly used as a measure of emotional self-awareness that can be used to approach dimensions of self-mentalizing (Vives et al., 2021; Yildirim et al., 2022a, 2022b, 2022c). While helpful for mentalization, the aforementioned measures do not specifically capture the construct nor do many of them provide in-depth information regarding the four polarities of mentalization as outlined and recently called for in state-of-the-art research (Luyten et al., 2020).

Few mentalization questionnaires that exist, and even fewer have been adapted to Spanish. The Reflective Function Questionnaire (Fonagy et al., 2016) is an 8-item questionnaire, which has been adapted for youth (Sharp et al., 2022). Additionally, the Parental Reflective Function Questionnaire (Luyten et al., 2017), whereby parents reflect on their internal experiences as well as those of their child(ren) has also been translated to Spanish (London, 2020). Apart from these three (similar) questionnaires, a paucity of mentalization questionnaires exist in Spanish, despite a) it

being the official language in many countries worldwide and b) questionnaires being the most efficient and cost-effective method for evaluating mentalization.

Perhaps the most widely-used and researched questionnaire evaluating mentalization beyond the various versions of the RFQ is the Mentalization Questionnaire (Hausberg et al., 2012), originally validated for use with inpatients with mental conditions. In the original version, 15 items are structured under four factors which evaluate pre-mentalizing modes and aberrant mentalization, 'refusing self-reflection' (4 items), 'emotional awareness' (4 items), 'psychic equivalence mode' (4 items), and 'regulation of affect' (3 items). Although the MZQ was initially validated in a clinical population, validations in Italian (Ponti et al., 2019), Korean (Song & Choi, 2017), and Finnish (Eloranta et al., 2020) show adequate psychometric properties in nonclinical adolescents, suggesting that a Spanish adaptation of this questionnaire could be useful for obtaining a fast, reliable, and comprehensive measure for mentalization in community samples. Another unpublished version exists in Dutch (Paridaens, n.d.). Like the original English version, the Italian and Korean versions contain four factors (with different items in each factor compared to the original and each other). The published Finnish and unpublished Dutch versions both contain one factor but omit some items.

As such, the aim of this study is to successfully adapt and validate the MZQ to European Spanish, providing evidence of reliability and evidence of validity in adolescent and adult samples. Specifically, we aim to provide evidence for validity based on the internal structure using confirmatory factor analysis and invariance analysis. We further intend to provide evidence of convergent and discriminant validity based on associations of MZQ scores (which reflect degree of mentalization deficit) with other questionnaires, such as those measuring internalizing symptoms (positively associated), and resiliency and secure attachment style (negatively associated). Finally, we aim to provide evidence for internal consistency.

## Method

### Participants

Of the 1735 participants invited to participate in the study, 389 adolescents (22.42%) and 382 (22.02%) of the parents participated. Since MZQ was completed in the context of a broader mental health study targeted toward nonclinical adolescents, the inclusion criterion was that students were not diagnosed with severe mental conditions such as intellectual disability, autism or psychosis. Participant samples were analyzed independently: one comprised of 389 adolescents (191 girls, 198 boys) between the ages of 12 and 19 ($M = 14.4$ $SD = 1.68$) and the second, their 382 parents, either fathers aged 37-67 ($M = 49.0$, $SD = 5.07$) or mothers aged 32-69 ($M = 46.6$, $SD = 4.48$). Of the adult participants that participated, 188 (49.2%) indicated their sex (22.34% men, 77.66% women). Socioeconomic status showed no statistically significant differences between adolescents of either sex ($\chi^2 = 8.56$, $df = 4$, $p = .07$). Breakdown of socioeconomic status was 7.5% low, 9.8% middle-low, 14.7% middle class, 37.8% middle-high, and 30.3% high. Participant sample size was determined based on power analyses for primary research objectives.

## Instruments

### *Mentalization Questionnaire (MZQ)*

The MZQ (Hausberg et al., 2012) was originally designed to assess mentalization deficits in the clinical population. This 15-item, self-report measure prompts respondents to rank their level of agreement with items such as "most of the time it is better not to feel anything". Higher scores indicate worse mentalization. The original MZQ reports four factors: 'refusing self-reflection' (items 5, 9, 13, 14), 'emotional awareness' (items 8, 10, 11, 15), 'psychic equivalence mode' (items 1, 4, 7, 12), and 'regulation of affect' (items 2, 3, 6), which yields adequate internal consistency ($\alpha$ = .81) and moderate test-retest reliability ($r$ = .67) (Hausberg et al., 2012). Subsequent adaptations of the MZQ have variations in factor structure and number of items. The Spanish adaptation is analyzed in the current paper.

### *Beck's Depression Inventory-2 (BDI)*

BDI (Beck et al., 1996; Sanz et al., 2003) is widely used to measure depression. The internal consistency of the current sample is excellent ($\alpha$ = .91), slightly higher than that obtained by Sanz et al. (2003) in the Spanish adaptation ($\alpha$ = .87). For the purposes of this study, the BDI scores were utilized to ascertain evidence of convergent validity of the MZQ scores for the adolescent sample.

### *Multidimensional Anxiety Scale for Children (MASC)*

Symptoms of anxiety were measured using the MASC (March et al., 1999), which includes four subscales: physical symptoms, harm and avoidance, social anxiety, and separation anxiety. Higher scores indicate more anxious symptoms. The Spanish version of the MASC boasts good reliability and validity evidence (García-Villamisar & Yenes, 2002), and for the current sample the internal consistency is $\alpha$ = .89. MASC scores were utilized to explore evidence of convergent validity of the MZQ scores in the adolescent sample.

### *General Health Questionnaire (GHQ)*

GHQ (Goldberg & Hillier, 1979; Lobo et al., 1986) was used to measure symptoms of mental health conditions in the adult sample. The Spanish version (Lobo et al., 1986) is widely used with good psychometric properties (Ames-Guerreroa et al., 2017). All four subscales consist of seven items. Items such as "have you recently lost much sleep over worry?" prompt four response options according to agreement or frequency. In the current sample, internal consistency values are: somatic symptoms, $\alpha$ = .78; anxiety, $\alpha$ = .88; social functioning, $\alpha$ = .76; depression, $\alpha$ = .77. GHQ scores were used to obtain evidence of convergent validity of the MZQ scores in the parent sample.

### *Connor-Davidson Resilience Scale 10 (CD-RISC 10)*

The CD-RISC 10 (Campbell-Sills & Stein, 2007; Connor & Davidson, 2003) focuses on an individual's ability to manage stressors. This research utilized a shortened version of the original scale, and reliably differentiates individuals with greater and lesser resilience. This 10-item version possesses excellent psychometric properties and prompts respondents to rank their frequency of prompts like 'can deal with whatever comes'. The Spanish version of the CD-RISC (Notario-Pacheco et al., 2011) has been validated and widely used in both adolescents and adults (e.g., Blanco et al., 2019; Notario-Pacheco et al., 2011). The CD-RISC scores were used for both samples to obtain evidence of convergent validity of MZQ scores. In the present study, the internal consistency of CD-RISC scores was $\alpha$ = .73 in adolescents and $\alpha$ = .91 in adults.

### *Relationship Scales Questionnaire (RSQ)*

The RSQ (Griffin & Bartholomew, 1994) is a 30-item self-report measure for attachment style. For this study, only 'secure attachment' subscale (5 items) was utilized. It includes items such as "I find it easy to get emotionally close to others". The RSQ has been widely used and adapted to several languages, and the Spanish version has been used in both adolescent (Bustamante et al., 2010; Magaz et al., 2011) and adult samples (Papalia & Widom, 2023) with similar psychometric properties to the English version. The current samples have internal consistency scores of $\alpha$ = .73 for adolescents, and $\alpha$ = .71 for adults. RSQ secure scores were utilized for evidence of convergent validity of MZQ scores in these samples.

### *Big Five Inventory (BFI)*

The BFI (Benet-Martínez & John, 1998; John & Srivastava, 1999) is a 44-item measure for the big five personality traits in adolescents. The Neuroticism scale (8 items) was used to explore convergent validity in the adolescent sample, and the Openness to Experience scale (10 items) was utilized to explore discriminant validity. The BFI has a validated five-factor structure consistent with the original version in Spanish, with good internal consistency ($\alpha$ = .72) (Reyes et al., 2014) and no evidence of cultural differences compared to the English version (Benet-Martínez & John, 1998). For the current sample, internal consistency for Neuroticism was $\alpha$ = .79 and for Openness to Experience was $\alpha$ = .71. In the adolescent sample, the BFI scores for Neuroticism were used to obtain evidence of convergent validity and the Openness to Experience scores to obtain evidence of discriminant validity of MZQ scores.

## Procedure

Following the adaptation process outlined by the International Test Commission (Hernández et al., 2020) and in alignment with best practice for digital technology-based assessment (Oliden et al., 2023), the MZQ was adapted to Spanish. After receiving ethical approval according to the Declaration of Helsinki (Ethics Committee of the Universitat Autònoma de Barcelona, CEEAH: 2603) for a broader project focused on adolescents, families were provided with written informed consent materials which were agreed to before proceeding to data collection.

Ten schools with similar characteristics of urbanicity, size, family socioeconomic status, educational approach and geographic location were invited to participate in the project according to their proximity to the research center. Five of the ten invited schools agreed to collaborate, with principal reasons that families refrained from participation being low project interest, being too busy, and

preferring not to provide data about mental states. Families were recruited through schools, and notified about the objectives, relevance, and implications of the study through a letter distributed by the school, then invited to a meeting to resolve any questions or concerns that participants have regarding their participation.

After informed consent was received, data were collected through the schools to simplify logistics. Participants included adolescents and their parents, who received copies of all aforementioned questionnaires in a closed, sealed envelope with alphanumeric identity encryption. Further contact with participants was only established to rectify cases with missing or out-of-value data. Retest was conducted after 30 days.

Participants were analyzed in two independent groups: parents and adolescents. This is sensible considering the different mentalization abilities and development level between the two age groups.

**Data Analysis**

Before completing analysis of the responses of the MZQ, a multiple imputation was conducted for missing values. Next, factor structure was examined in the adolescent and parent samples, mapping different models with Confirmatory Factor Analysis (CFA). The use of the diagonally weighted least squares (DWLS) estimator was appropriate given the categorical nature of the items and lack of normality (DiStefano & Morgan, 2014; Forero et al., 2009). Adjusted models were evaluated by chi square, comparative factor index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Residual (SRMR). According to the Goodness of Fit Index, CFI and TLI less than .90, RMSEA less than .08 and SRMR less than .10 are considered good, while excellent goodness of fit is indicated when CFI and TLI are greater than .95, RMSEA is less than .06 and SRMR is less than .08 (Brown, 2015; Schreiber et al., 2006).

Five factor models were analyzed in both samples. The (M1) original model (Hausberg et al., 2012) with four non-correlated factors is formed by items 5, 9, 13, and 14 (f1), 8, 10, 11, and 15 (f2), 1, 4, 7 and 12 (f3) and 2, 3, and 6 (f4). Further, the (M2) Finnish version model (Eloranta et al., 2020) was utilized, which comprises four independent factors formed by items 2, 4, 5, 7, 8, 13 and 14 (f1), 9, 10, and 11 (f2), 1 and 12 (f3) and 3, 6, and 15 (f4). The (M3) Korean version (Song & Choi, 2017) excludes items 7 and 9 and contains four correlated factors (2, 5, 8, 14 and 15 (f1); 10, 11 (f2); 1, 4, 12 and 13 (f3); 3 and 6 (f4)). Model 4 (M4) corresponds with the one-factor structure determined in the Italian adaptation (Ponti et al., 2019) which omitted items 2, 4, 10 and 13, with inter-item correlation between items 1 and 12. Finally, model 5 (M5) is consistent with the single-factor Dutch version (Paridaens, n.d.) which included all 15 original items. Goodness of fit of all analyzed models can be found in Table 1.

Once the model with best fit was determined, various evaluations for invariance were analyzed. Invariance was evaluated using two different possibilities in adolescents: by sex and age (below and above 15 years old, according to previous research which delineates early from late adolescence (e.g., van Lang et al., 2007). In each analysis, nested models were analyzed with progressively more stringent criteria (Vandenberg & Lance, 2000). Factor equivalence loadings for both groups (weak or metric invariance), intercept equivalence of the items (strong or scalar invariance) and

equivalence of the uniqueness of items (strict invariance) were verified. In both groups, metric invariance ensures equivalence of the meaning of the measure, scalar invariance signifies the equivalence of means of each item, and strict invariance demonstrates the equivalence of the variance of items that is not explained by the factor itself. The comparison between the different nested models was conducted using differences of fit comparative fit index, Tucker-Lewis Index, and root-mean-square error approximation (ΔCFI, ΔTLI and ΔRMSEA). According to Cheung & Rensvold (2002) and Marsh et al. (Marsh et al., 2013), decreases in CFI and TLI of less than .01 and increase in RMSEA greater than .015 are considered indicators of invariance. In the absence of complete invariance, partial invariance is evaluated by freely estimating some of the model parameters. The parameters to be released were identified with the modification indices, and a broad partial invariance criterion was used to determine the maximum number of parameters released (Byrne et al., 1989).

To evaluate internal consistency, both alpha and omega (Doval et al., 2023) were calculated. The omega coefficient was measured by considering the characteristics of each factorial model. Test-retest reliability was estimated with absolute agreement intraclass correlation coefficient for individual measures (Liljequist et al., 2019).

To provide evidence for validity related to other variables, correlations of MZQ scores were conducted with scores on secure attachment, depression, anxiety, neuroticism and resilience (evidence of convergent validity) and with the scores of openness to experience (evidence of discriminant validity).

All analyses were conducted using R. Several packages were utilized: *Psych* for descriptive analysis (Revelle, 2021), MVN (Korkmaz et al., 2014) for evaluating normality, *Amelia* (Honaker et al., 2011) for missing values, *Iavaan* (Rosseel, 2012) for AFC model fit, *SemTools* (Jorgensen et al., 2022) for invariance analysis and reliability coefficient calculation, *Hmisc* (Harrell, 2022) for Pearson and *irr* (Gamer et al., 2012) for intraclass correlation.

**Results**

**Descriptive Statistics**

In the adolescent sample, missing items were present for 28 participants, were low in frequency (0.41%), and affected 8 items (between 0.26% and 2.83%). In all cases with missing responses, only one item of the entire questionnaire was not answered. Table 1 includes descriptive statistics of each item. Means and standard deviations ranged between 2.26 and 3.48 (*SD* 1.12 - 1.39), with skewness between -0.50 and 0.69, while values for kurtosis were between -1.17 and -0.71. The Shapiro-Wilk test confirmed that all items were non-normally distributed (*p* < .001).

In the parent sample, there was also a small percentage of missing items (1.82% of all data), though in this case 13 of 15 MZQ items (86.7%) had a missing value, with frequencies between 0.26% and 14.14% (for item 2). Regarding participants, 72 individuals (18.5%) continued without responding to between 1 and 3 items of the questionnaire. Means and standard deviations for the parent sample ranged between 1.63 and 2.77 (*SD* 0.92 - 1.28), with skewness between -0.06 and 1.90 and kurtosis between -1.34 and 2.61. Non-normal distribution was further confirmed with the Shapiro-Wilk test (*p* < .001).

**Table 1**
*Descriptive Statistics of MZQ Items in Adolescent and Parent Samples*

| Item | Adolescents (*n* = 389) | | | | Parents (*n* = 382) | | | |
|------|------|------|----------|----------|------|------|----------|----------|
| | $M^a$ | SD | Skewness | Kurtosis | $M^a$ | SD | Skewness | Kurtosis |
| 1 | 3.18 | 1.30 | -0.25 | -1.15 | 2.66 | 1.26 | 0.17 | -1.34 |
| 2 | 2.87 | 1.12 | 0.05 | -0.78 | 2.77 | 1.13 | 0.14 | -1.26 |
| 3 | 2.69 | 1.30 | 0.13 | -1.18 | 1.92 | 1.13 | 0.99 | -0.27 |
| 4 | 3.15 | 1.37 | -0.20 | -1.24 | 2.27 | 1.28 | 0.61 | -0.98 |
| 5 | 2.45 | 1.31 | 0.54 | -0.85 | 1.57 | 1.04 | 1.90 | 2.61 |
| 6 | 3.05 | 1.30 | -0.11 | -1.18 | 2.16 | 1.19 | 0.79 | -0.61 |
| 7 | 2.81 | 1.26 | 0.03 | -1.08 | 2.38 | 1.22 | 0.46 | -1.00 |
| 8 | 2.85 | 1.21 | 0.08 | -0.91 | 2.30 | 1.28 | 0.58 | -0.99 |
| 9 | 3.48 | 1.22 | -0.50 | -0.73 | 2.82 | 1.27 | 0.06 | -1.32 |
| 10 | 2.98 | 1.25 | -0.11 | -1.08 | 2.24 | 1.21 | 0.62 | -0.88 |
| 11 | 2.87 | 1.21 | 0.00 | -1.00 | 2.26 | 1.18 | -0.60 | -0.84 |
| 12 | 3.11 | 1.39 | -0.17 | -1.30 | 2.32 | 1.22 | 0.55 | -0.95 |
| 13 | 2.26 | 1.28 | 0.69 | -0.71 | 2.05 | 1.18 | 0.95 | -0.18 |
| 14 | 3.18 | 1.30 | -0.20 | -1.15 | 2.74 | 1.28 | 0.18 | -1.27 |
| 15 | 2.77 | 1.38 | 0.16 | -1.24 | 1.63 | 0.92 | 1.52 | 1.62 |

*Note. M* = mean score on the corresponding item.

## Factor Structure of the Adolescent Sample

Table 2 (available as a supplementary table; all supplementary materials are accessible at: https://figshare.com/s/e5fee1fd7479d7dc204f) shows goodness of fit for the analyzed models corresponding to adolescents and adults. The models with multifactor structure (M1, M2 and M3) clearly show poor goodness of fit, however, the fitness of models with one factor (M4 and M5) are acceptable. When M5 is adjusted for covariance of error in items 1 and 12, fit of the model is improved substantially. The results of this new, updated model (M5b) are also provided in Table 2. This new model (M5b) is a single factor model which avoids eliminating any of the 15 original items but accounts for error covariance of items 1 and 12. This model provides the best goodness of fit of all items analyzed ($\chi^2$ = 95.484, *df* = 89, *p* = .30, CFI = .995, TLI = .994, RMSEA = .014 with limits between .00 and .031; SRMR = .047).

The fact that model M5b includes all 15 items of the original questionnaire and that its goodness of fit is much better than the other analyzed models led us to conclude that this model is most representative of the internal structure of the adolescent sample for the MZQ Spanish adaptation. The values of the standardized factor loadings of model M5b can be seen in Table 3 (supplementary material). Of note is low factor loading of item 2 (.145 in the present sample). In all analyzed models, item 2 had a low factor loading. Further, high factor loading for item 15 must be highlighted.

Alpha coefficient for the total questionnaire comprised of 15 items and was α = .756 [IC95%: .719, .790], while those for omega corrected for inter-item error correlations (Raykov, 2004) was ω = .742. Both coefficients are provided because in the presence of inter-item error correlations, Cronbach's alpha tends to be an overestimation of internal consistency (Bentler, 2021; Raykov, 2001).

## Factor Structure of the Parent Sample

Table 2 shows goodness of fit for the models analyzed using the parent sample. Like the adolescent sample, all multifactorial models (M1, M2 and M3) showed poor goodness of fit. Nonetheless, M4 (single factor excluding items 2, 4, 10 and 13 with associated error between items 1 and 12) showed excellent goodness of fit ($\chi^2$ = 27.81. *df* = 43. *p* = .965. CFI = 1.00, TLI = 1.022, RMSA = .000 with upper and lower limit equal to .000, and SRMR = .036). Like for the adolescent sample, M5 (single-factor model including all items) was a poor fit, but modification analysis indices suggested modeling the error covariance between items 1 and 12 (M5b), which resulted in good fit ($\chi^2$ = 112.46, *df* = 89, *p* = .047, CFI = .983, TLI = .979. RMSA = .026 with lower limit .003 and upper limit .040, and SRMR = .053). Similar to the adolescent sample, the fact that this model included all original MZQ items, compared to M4, led us to consider it the optimal model for the factor structure of this questionnaire. Factor loadings of the selected model can be seen in Table 3 (Supplementary Material). Reliability coefficients alpha and omega corrected for correlated errors of this model were α = .762 [IC95%: .705, .780] and ω = .744.

**Table 8**
*Correlations Between MZQ and Other Administered Questionnaires in Both Samples*

| Measures | MZQ Adolescents (*n*) | MZQ Parents (*n*) |
|----------|------------------------|--------------------|
| BDI Depression | .46*** (191) | --b |
| MASC Anxiety | .45*** (189) | --b |
| BFI Neuroticism | .38*** (183) | --b |
| BFI Openness | -.08*** (183) | --b |
| CD-RISC Resilience | -.23*** (389) | -.22* (110) |
| RSQ Secure Attachment | -.36*** (181) | -.19 (71) |
| GHQ Somatization | --a | .16* (184) |
| GHQ Depression | --a | .23** (186) |
| GHQ Anxiety | --a | -.22** (186) |
| GHQ Social Functioning | --a | .23** (186) |

*Note.* The abbreviations for measurement scales are in accordance with abbreviations for measures that are consistent throughout the paper. BDI = Beck's Depression Inventory, MASC = Multidimesional Anxiety Scale for Children, BFI = Big Five Inventory, CD-RISC = Connor Davidson Resilience Scale 10, RSQ = Relationship Scales Questionnaire, GHQ = General Health Questionnaire. Sample size is indicated between parentheses.
aThese sections are left blank as adolescents were not administered the GHQ, as it is a scale for adults.
bThese sections are left blank as parents were not administered the corresponding measures.
* *p* < .05
** *p* < .005
*** *p* < .001

## Measurement Invariance in Adolescents by Sex

The one-factor model was the model including error covariance between items 1 and 12. Table 4 (Supplementary Material) shows goodness of fit for the invariance analysis by sex in the adolescent sample. The factor structure remains consistent independent of sex. Metric invariance is partially met, though the factor loading of item 13 varies (higher in boys and irrelevant in girls). Further, the comparison shows partial scalar invariance, with intercept

differences for items 8 (2.89 in boys, 2.47 in girls) and 10 (3.06 in boys, 2.67 in girls). The uniqueness was equal in all items. These results indicate an acceptable partial invariance based on sex.

## Measurement Invariance in Adolescent Sample by Age

Further evaluating the adolescent sample, Table 5 (Supplementary Material) shows goodness of fit for invariance analysis by age. The analysis of metric invariance revealed differences in factor loadings for items 2 and 4. Further, scalar invariance showed intercept differences for items 12 and 13. This comparison allows for the identification of differences by age-group for error in item 13. The number of items affected at each level of analysis for variance is relatively low, indicating correct partial invariance between age groups.

## Adolescents vs. Parents: Measurement Invariance

Table 6 (Supplementary Material) shows goodness of fit for the analysis of invariance by respondent group (adolescents vs. parents). The factor structure is invariable in both samples. Nonetheless, metric invariance was only partially reached, due to differences in factor loadings of the items 7, 9 and 14, because, as can be seen in Table 1, they are higher in the parent sample, while for item 15 the factor loading is higher in the adolescent sample. Further, the scalar variance was partial, with intercept values that were higher in adolescents for item 4 (3.15 vs. 2.85), item 5 (2.45 vs. 2.10) and item 15 (2.77 vs. 2.55), and higher for parents for item 7 (2.82 vs. 3.05) and item 11 (2.87 vs. 3.05). There was no evidence for strict invariance.

## Questionnaire Validity: Correlations with Other Measures

Table 7 (Supplementary Material) shows the descriptive statistics of the MZQ score in the sample of adolescents and adults represented as whole groups, with adolescents also presented split into groups by sex and age. Table 8 shows the correlation of MZQ scores with scores of other questionnaires. In the adolescent sample, the correlations of MZQ scores with those of the BDI, MASC, and the neuroticism scale from the BFI are positive, while the scores for CD-RISC and RSQ correlate negatively with MZQ scores. These correlations can be considered moderate (between .23 and .46). Both the sign and robustness of correlations are evidence of convergent validity for the MZQ scores in the adolescent sample. The almost-zero correlation between the MZQ and the Openness scale of the BFI constitutes evidence for discriminant validity.

In the parent sample, moderately low correlations with the GHQ scales, positive correlations with the depression and social functioning subscales, and negative correlations with the anxiety subscale provide evidence for convergent validity of the MZQ scores. Analysis also revealed an unexpected near-zero correlation between MZQ score and secure attachment in parents.

## Test-retest Reliability

In the adolescent sample 180 participants completed the MZQ twice. The intra-class correlation of absolute agreement for mean measures between scores is .636 [IC95%: .540, .715], *p* < .001.

## Discussion

Modeling previous factor structures of the original MZQ along with its subsequent adaptation to Spanish resulted in the conclusion that M5b was the best-fit factor structure in both adolescents and adults. This model includes all 15 original items, and accounts for error covariance between items 1 and 12. When compared by sex, the same M5b factor structure was the best-fit model for boys and girls, with some small variations in factor loadings for individual items. Age-group analysis (below and above 15 years old) in the adolescent sample demonstrated partial invariance at the metric (variation between item 2 and 4) scalar (intercept differences between item 12 and 13) and strict levels (residual variance for item 13). Finally, invariance analysis between the adolescent and adult samples revealed the factor structure remains the same for the two groups.

The original MZQ revealed a four-factor structure. The original authors noted that "it remains to be seen whether the revealed four-factor structure of the questionnaire can be replicated…" (Hausberg et al., 2012). Each subsequent adaptation has concluded to a different factor structure to the model, suggesting that the original four-factor model may not be optimal for (other adaptations of) this questionnaire. Indeed, adaptations of the MZQ which do use four factor models (Korean, Finish) use *different* items to form the four factors. By contrast, the Italian adaptation utilizes one single factor but omits several items. Perhaps such variation from the original four-factor structure is because of significant overlap in 'distinct' constructs underneath the umbrella term 'mentalization', or more likely, the variation of use of a clinical sample in the original version vs. non-clinical adaptations, along with adult vs. adolescent validations.

Despite low factor loading for item 2, rather than electing to eliminate this item from the Spanish adaptation as in previous adaptations, we elected to maintain this item. First, maintaining the item did not result in significant changes to the psychometrics of the questionnaire. Second, and perhaps more importantly, by maintaining all 15 items of the original MZQ, the Spanish adaptation is eligible for cross-cultural comparisons which may not be possible for other versions.

More detailed inspection of the psychometric properties revealed that covariance of item error between items 1 and 12 was not a novel finding, as the Italian adaptation of the MZQ revealed the same covariance (Ponti et al., 2019). This covariance seems logical. Item 1 reads "If I expect to be criticized or offended, my fear increases more and more" while item 12 reads "Often I feel threatened by the idea that someone could criticize or offend me". While subtle differences of 'feeling threatened' versus 'expecting' exist, both items clearly prompt respondents to consider their level of discomfort of criticism or offense from other people.

Overall, comparisons by sex displayed consistency between the boys and girls. Nonetheless, girls had lower intercepts for two items–item 8, which reads "I tend to ignore feelings of physical tension or of discomfort until they compel my whole attention" and item 10, which reads, "Sometimes I only become aware of my feelings in retrospect". These two items could be reflective of higher emotional maturity in girls, who mature faster than boys for what are understood to be a variety of reasons including neurobiological development, and parental expectations (Frere et al., 2020; Rose &

Rudolph, 2006). Nonetheless, these differences were only discovered in two of 15 items, and thus no practical or relevant implications for maturational differences between boys and girls can be made.

The invariance analysis between the adolescent and adult samples indicates that the factor structure is consistent for the two groups. Nonetheless, at the metric level the invariance is partial since the factor loading of items 7, 9, 14 and 15 are different between samples. Scalar invariance was also partially different with differing intercepts for five items. Some variation in intercepts between adolescents, with partially developed mentalizing capacity, and adults with developed mentalizing capacity, are to be expected, however. For example, in the case of item 4, "I can only believe that someone likes me if I have enough realistic proof for it (e.g. a date, a gift, or a hug)" is reflective of the pre-mentalizing mode of psychic equivalence, whereby internal and external realities are assumed to be equivalent. Psychic equivalence is more common in adolescence (Harenski et al., 2018; Keulers et al., 2010), a developmental period when cognitive and emotional neurological features are undergoing rapid and consistent change and further development.

Finally, with all parameters released, strict invariance is observed. The factor invariance indicates that there are no relevant differences between adolescent and adult samples in their reliability of MZQ scores. Nonetheless, no variance at the level of factor means demonstrates different means for mentalization abilities, which are higher in the adolescent sample, indicating that the adolescent sample exhibited poorer mentalization than the adult sample.

Regarding validity analysis comparing the MZQ associations to other measures, most results were in the expected direction. Higher scores on the MZQ indicate poorer mentalization, and good mentalizing is considered a protective factor for a variety of mental conditions (Ballespí et al., 2018), thus positive correlations of MZQ scores with depression and anxiety symptoms were anticipated. Further, expected negative correlations with the protective factors of secure attachment and resilience were also found. Both sign and robustness of the remaining correlations are evidence of convergent validity for the MZQ in the adolescent sample. The almost-zero correlation between the MZQ and the Openness scale of the BFI, also expected, constitutes evidence of discriminant validity. Evidence of convergent validity of MZQ scores was also found in the parent sample, with moderately-low correlations with GHQ scales in the expected direction, which are positive for depression and social dysfunction, and negative for anxiety. One unexpected lack of relationship was found in the parent sample between secure attachment and the MZQ. This is surprising, as the association between good mentalizing and secure attachment is well-established, although particularly in adolescence, when expressions of socio-emotional experiences and affect are often pronounced (Bailen et al., 2019). Perhaps the intensity of emotional experiences in adolescence but not adulthood, paired with the fact that the current study evaluated this association in the general population (which routinely shows attenuated relationships compared to clinical samples), could be responsible for the lack of association.

To our knowledge, a previous adaptation of the MZQ to Spanish does not exist. Evidence suggests that this questionnaire effectively measures mentalization in non-clinical adolescents and adults. The present study benefits from its evaluation of the MZQ in two samples; adolescents were the primary sample as this study formed part of a larger project focused on adolescents, but adult data provides further confirmation for the utility of the MZQ in adult populations globally and is consistent with its utility in adults in the Finnish, Korean and Italian versions. One primary limitation of all mentalization self-report measures is that it is only possible to examine the mentalizing that someone is capable of regarding their own capacity to mentalize, and thus, with individuals for limited mentalizing abilities, there may be bias. Unfortunately, this is a drawback of all mentalization questionnaires.

Future research should utilize this validated, Spanish version of the MZQ to evaluate mentalization in larger adult samples and replicate the findings of this study, particularly in the clinical population. This adaptation offers research and future clinical potential for use of MZQ in preventative, community mental health studies or to explore the role of mentalization in the community. Considering the transdiagnostic, protective role of good mentalization, this scale can be applied for a wide range of mental health constructs including internalizing disorders and psychosis, which could help advance the understanding mentalization's contribution to salutogenesis, or to improve follow-up measures in thorough longitudinal designs to test treatment efficacy and efficiency. With one of the most widely used mentalization questionnaires now adapted to the Spanish language and validated for both adolescents and adults in non-clinical populations, horizons are broadened for researchers to conduct cost-efficient, large-sample mentalization research in the myriad countries with Spanish as a first language.

## Author Contributions

## Acknowledgements

## Funding

## Declaration of Interests

The authors declare no conflicts of interest.

## Data Availability Statement

The authors will provide the dataset without undue reservation to interested parties upon request. Those interested may contact the corresponding author.

## References

Ames-Guerreroa, R., Barreda-Parrab, V., & Huamani-Cahuac, J. (2017). Psychometric properties and factor invariance for the General Health Questionnaire (GHQ-28): study in Peruvian population exposed to the COVID-19 pandemic. *ASEAN Journal of Psychiatry, 21*(2), 1–12. https://doi.org/10.1101/2020.11.10.20229435

Bailen, N. H., Green, L. M., & Thompson, R. J. (2019). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review, 11*(1), 63–73. https://doi.org/10.1177/1754073918768878

Ballespí, S., Vives, J., Debbané, M., Sharp, C., & Barrantes-Vidal, N. (2018). Beyond diagnosis: Mentalization and mental health from a transdiagnostic point of view in adolescents from non-clinical population. *Psychiatry Research, 270*, 755–763. https://doi.org/10.1016/j.psychres.2018.10.048

Bateman, A. W., & Fonagy, P. (2004). Mentalization-based treatment of BPD. *Journal of Personality Disorders, 18*(1), 36–51. https://doi.org/10.1521/pedi.18.1.36.32772

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory (BDI-II)* (Vol. 10). Pearson.

Benet-Martínez, V., & John, O. P. (1998). Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*(3), 729-750. https://psycnet.apa.org/doi/10.1037/0022-3514.75.3.729

Bentler, P. M. (2021). Alpha, FACTT, and beyond. *Psychometrika, 86*(4), 861–868. https://www.doi.org/10.1007/s11336-021-09797-8

Blanco, V., Guisande, M. A., Sánchez, M. T., Otero, P., & Vázquez, F. L. (2019). Spanish validation of the 10-item Connor–Davidson Resilience Scale (CD-RISC 10) with non-professional caregivers. *Aging & Mental Health, 23*(2), 183–188. https://doi.org/10.1080/13607863.2017.1399340

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications. https://books.google.es/books?hl=en&lr=&id=tTL2BQAAQBAJ&oi=fnd&pg=PP1&dq=confirmatory+factor+analysis+for+applied+research&ots=alWxoRZL3B&sig=E0ShsQIBCG_nLJR990lvh3 FNfas&redir_esc=y#v=onepage &q=confirmatory%20factor%20analysis %20for%20ap plied%20research&f=false

Bustamante, J. G., Barona, E. J. G., & del Barco, B. L. (2010). Relation between attachment and emotional intelligence in teenagers. *REME, 13*(34), Article 8.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466. https://psycnet.apa.org/doi/10.1037/0033-2909.105.3.456

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: Official Publication of The International Society for Traumatic Stress Studies, 20*(6), 1019–1028. https://doi.org/10.1002/jts.20271

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chevalier, V., Simard, V., & Achim, J. (2023). Meta-analyses of the associations of mentalization and proxy variables with anxiety and internalizing problems. *Journal of Anxiety Disorders, 95*, Article 102694. https://doi.org/10.1016/j.janxdis.2023.102694

Connor, K. M., & Davidson, J. R. T. (2003). Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and Anxiety, 18*(2), 76–82. https://doi.org/10.1002/da.10113

Dimitrijević, A., Hanak, N., Dimitrijević, A., & Marjanović, Z. (2018). The Mentalization Scale (MentS): A self-report measure for the assessment of mentalizing capacity. *Journal of Personality Assessment, 100*(3), 268–280. https://doi.org/10.1080/00223891.2017.1310730

DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 425–438. https://doi.org/10.1080/10705511.2014.915373

Doval, E., Viladrich, C., & Angulo-Brunet, A. (2023). Coefficient alpha: the resistance of a classic. *Psicothema, 35*(1), 5–20. https://doi.org/10.7334/psicothema2022.321

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/10.1080/19012276.2020.1863852

Eloranta, S. J., Kaltiala, R., Lindberg, N., Kaivosoja, M., & Peltonen, K. (2020). Validating measurement tools for mentalization, emotion regulation difficulties and identity diffusion among Finnish adolescents. *Nordic Psychology*, 1–23. https://doi.org/10.1080/19012276.2020.1863852

Elosua Oliden, P., Aguado García, D., Fonseca Pedrero, E., Abad García, F. J., & Santamaría Fernández, P. (2023). New Trends in Digital Technology-Based Psychological and Educational Assessment. *Psicothema, 35*(1), 50–57. https://www.doi.org/10.7334/psicothema2022.241

Ensink, K., Normandin, L., Target, M., Fonagy, P., Sabourin, S., & Berthelot, N. (2015). Mentalization in children and mothers in the context of trauma: An initial study of the validity of the Child Reflective Functioning Scale. *British Journal of Developmental Psychology, 33*(2), 203–217. https://doi.org/10.1111/bjdp.12074

Fonagy, P. (1993). *Reflective Self Function in Early Attachment and Borderline States* (1st ed.). ERIC Clearinghouse.

Fonagy, P., & Campbell, C. (2017). Mentalizing, attachment and epistemic trust: how psychotherapy can promote resilience. *Psychiatria Hungarica, 32*(3), 283–287.

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y. W., Warren, F., Howard, S., Ghinai, R., Fearon, P., & Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. *PLoS ONE, 11*(7), Article e0158678. https://doi.org/10.1371/journal.pone.0158678

Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology, 9*(4), 679–700. https://doi.org/10.1017/S0954579497001399

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-functioning manual version 5.0, for application to adult attachment interviews*. University College London. https://discovery.ucl.ac.uk/id/eprint/1461016/

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.

Frere, P. B., Vetter, N. C., Artiges, E., Filippi, I., Miranda, R., Vulser, H., Paillère-Martinot, M.-L., Ziesch, V., Conrod, P., & Cattrell, A. (2020). Sex effects on structural maturation of the limbic system and outcomes on emotional regulation during adolescence. *NeuroImage, 210*, Article 116441.

Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). Package 'irr.' *Various Coefficients of Interrater Reliability and Agreement, 22*, 1–32.

García-Villamisar, D., & Yenes, A. V. E. (2002). Structure of anxiety symptoms in infancy: An exploratory study with the Multidimensional Anxiety Scale for children-Spanish version (MASC-SV). *Psiquis, 23*(2), 45–58.

George, C., Kaplan, N., & Main, M. (1996). Adult attachment interview. *Unpublished Manuscript, Department of Psychology, University of California Berkeley (Third Edition)*.

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

Griffin, D. W., & Bartholomew, K. (1994). *Relationship scales questionnaire. [Database record]*. APA PsycTests.

Happé, F. G. E. (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.

Harenski, C. L., Calhoun, V. D., Bustillo, J. R., Haas, B. W., Decety, J., Harenski, K. A., Caldwell, M. F., Van Rybroek, G. J., Koenigs, M., & Thornton, D. M. (2018). Functional connectivity during affective mentalizing in criminal offenders with psychotic disorders: Associations with clinical symptoms. *Psychiatry Research: Neuroimaging, 271*, 91–99.

Harrell, F. E. (2022). *Package "Hmisc": Harrell Miscellaneous.*

Hausberg, M. C., Schulz, H., Piegler, T., Happach, C. G., Klöpper, M., Brütt, A. L., Sammet, I., & Andreas, S. (2012). Is a self-rated instrument appropriate to assess mentalization in patients with mental disorders? Development and first validation of the Mentalization Questionnaire (MZQ). *Psychotherapy Research, 22*(6), 699–709. https://doi.org/10.1080/10503307.2012.709325

Hernández, A., Hidalgo, M. D., Hambleton, R. K., & Gómez Benito, J. (2020). International test commission guidelines for test adaptation: A criterion checklist. *Psicothema, 32*(3), 390–398.

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software, 45*, 1–47.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research, 2*(1999), 102–138.

Johnson, B. N., Kivity, Y., Rosenstein, L. K., LeBreton, J. M., & Levy, K. N. (2022). The association between mentalizing and psychopathology: A meta-analysis of the reading the mind in the eyes task across psychiatric disorders. *Clinical Psychology: Science and Practice, 29*(4), 423–439. https://doi.org/10.1037/cps0000105

Jorgensen, T. D., Pornprasertmanit, S., Shoemann, A. M. , Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling. R package version 0.5-6.* https://cran.r-project.org/package=semTools

Keulers, E., Evers, E., Stiers, P., & Jolles, J. (2010). Age, sex, and pubertal phase influence mentalizing about emotions and actions in adolescents. *Developmental Neuropsychology, 35*(5), 555–569. https://doi.org/10.1080/87565641.2010.494920

Korkmaz, S., Göksülük, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *R JOURNAL, 6*(2), 151-162.

Liljequist, D., Elfving, B., & Skavberg Roaldsen, K. (2019). Intraclass correlation–A discussion and demonstration of basic features. *PloS One, 14*(7), Article e0219854.

Lobo, A., Pérez-Echeverría, M. J., & Artal, 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

London, U. C. (2020). *The Reflective Functioning Questionnaire* (RFQ). https://www.ucl.ac.uk/psychoanalysis/research/reflective-functioning-questionnaire-rfq

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). The Mentalizing Approach to Psychopathology: State of the Art and Future Directions. *Annual Review of Clinical Psychology, 16*(9.1), 1–29. https://doi.org/10.1146/annurev-clinpsy-071919-015355

Luyten, P., Mayes, L. C., Nijssens, L., & Fonagy, P. (2017). The parental reflective functioning questionnaire: Development and preliminary validation. *PLoS ONE, 12*(5), Article e0176218. https://doi.org/10.1371/journal.pone.0176218

Magaz, A. M., Chorot, P., Sandin, B., Santed, M. A., & Valiente, R. M. (2011). Estilos de apego y acoso entre iguales (bullying) en adolescentes [Attachment patterns and peer bullying in adolescents]. *Revista de Psicopatología y Psicología Clínica, 16*(3), 207–221.

Main, M., & Goldwyn, R. (1998). Adult attachment rating and classification systems. *Unpublished Manuscript, Department of Psychology, University of California at Berkeley.*

March, J. S., Sullivan, K., & Parker, J. (1999). Test-retest reliability of the Multidimensional Anxiety Scale for Children. *Journal of Anxiety Disorders, 13*(4), 349–358. https://doi.org/10.1016/S0887-6185(99)00009-2

Marsh, H. W., Vallerand, R. J., Lafrenière, M.-A. K., Parker, P., Morin, A. J. S., Carbonneau, N., Jowett, S., Bureau, J. S., Fernet, C., & Guay, F. (2013). Passion: Does one scale fit all? Construct validity of two-factor passion scale and psychometric invariance over different activities and languages. *Psychological Assessment, 25*(3), 796–809. https://doi.org/10.1037/a0032573

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). Mayer-Salovey-Caruso emotional intelligence test (MSCEIT) item booklet. *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) [Database Record]. APA Psyctests.* https://doi.org/10.1037/t05047-000

Notario-Pacheco, B., Solera-Martínez, M., Serrano-Parra, M. D., Bartolomé-Gutiérrez, R., García-Campayo, J., & Martínez-Vizcaíno, 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), Article 63. https://doi.org/10.1186/1477-7525-9-63

Papalia, N., & Widom, C. S. (2023). Do insecure adult attachment styles mediate the relationship between childhood maltreatment and violent behavior? *Development and Psychopathology*, 1–12. https://doi.org/10.1017/S0954579422001468

Paridaens, P. (n.d.). *Reliability and Validity of the Menatlization Questionnaire (MZQ) in Forensic Care* [Undergraduate thesis]. http://arno.uvt.nl/show.cgi?fid=142737

Ponti, L., Stefanini, M. C., Gori, S., & Smorti, M. (2019). The assessment of mentalizing ability in adolescents: The Italian adaptation of the Mentalization Questionnaire (MZQ). TPM: Testing, Psychometrics, *Methodology in Applied Psychology, 26*(1), 29–38.

Raykov, T. (2001). Bias of coefficient afor fixed congeneric measures with correlated errors. *Applied Psychological Measurement, 25*(1), 69–76. https://doi.org/10.1177/0146621601025100

Raykov, T. (2004). Point and interval estimation of reliability for multiple-component measuring instruments via linear constraint covariance structure modeling. *Structural Equation Modeling, 11*(3), 342–356. https://doi.org/10.1207/s15328007sem1103_3

Revelle, W. (2021). *Package "psych": Procedures for Psychological, Psychometric, and Personality.*

Reyes, E. Z., Álvarez, C., Peredo, A. S., Miranda, A., & Rebolledo, I. M. (2014). Psychometric properties of the Big Five Inventory in a Mexican sample. *Salud Mental, 37*(6), 491–497.

Rose, A. J., & Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological Bulletin, 132*(1), 98–131. https://doi.org/10.1037/0033-2909.132.1.98

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

Sanz, J., Perdigón, A. L., & Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 2. Propiedades psicométricas en población general [Spanish Adaptation of Beck's Depression Inventory-II (BDI-II): Psychometric Properties in the General Population]. *Clínica y Salud, 14*(3), 249–280.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Sharp, C., Steinberg, L., McLaren, V., Weir, S., Ha, C., & Fonagy, P. (2022). Refinement of the Reflective Function Questionnaire for Youth (RFQY)

Scale B using item response theory. *Assessment, 29*(6), 1204–1215. https://doi.org/10.1177/10731911211003971

Song, H., & Choi, H.-A. (2017). Exploration of the Factor Structure of the Mentalization Questionnaire (MZQ) in 16–17-year-old Korean Adolescents. *Korean Journal of Clinical Psychology, 36*(3), 391–401.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002

van Lang, N. D. J., Ferdinand, R. F., & Verhulst, F. C. (2007). Predictors of future depression in early and late adolescence. *Journal of Affective Disorders, 97*(1–3), 137–144. https://doi.org/10.1016/j.jad.2006.06.007

Vives, J., Morales, C., Barrantes-Vidal, N., & Ballespí, S. (2021). Emotional comprehension is not related to duration of distress from daily life events. *International Journal of Environmental Research and Public Health, 18*(2), 459. https://doi.org/10.3390/ijerph18020459

Yildirim, D., Vives, J., & Ballespí, S. (2022a). Anxiety and depression: the moderating effects of attention to emotion and emotional clarity. *Psychological Reports, 126*(3), 1221–1234. https://doi.org/10.1177/00332941211070764

Yildirim, D., Vives, J., & Ballespí, S. (2022b). Meta-mood knowledge moderates the relationship between neuroticism and depression but not between neuroticism and anxiety in a sample of nonclinical adolescents. *Current Psychology, 187*, 11407. https://doi.org/10.1007/s12144-022-02864-y

Yildirim, D., Vives, J., & Ballespí, S. (2022c). Why do I feel what I feel? Examining individual differences in meta-mood knowledge as a moderator of the relationship between anxiety and depression in adolescents. *Personality and Individual Differences, 187*, Article 111407.