

This is the **accepted version** of the journal article:

Pueyo, Natalia; Navarro Pastor, Jose Blas; de la Osa, Nuria; [et al.]. «Age and Sex-specific Cut-off Scores for the Teacher-report Inventory of Callous Unemotional Traits on Children». Psychological Assessment, 2022. DOI 10.1037/pas0001125

This version is available at <https://ddd.uab.cat/record/258407>

under the terms of the  **IN** COPYRIGHT license

Post-print de: Pueyo, N., Navarro, J.B., de la Osa, N., Penelo, E., & Ezpeleta, L. (2022). Age and Sex-specific Cut-off Scores for the Teacher-report Inventory of Callous Unemotional Traits on Children. *Psychological Assessment*. doi: <https://doi.org/10.1037/pas0001125>.

Age and Sex-specific Cut-off Scores for the Teacher-report Inventory of Callous Unemotional Traits on Children

Natalia Pueyo², José-Blas Navarro^{1,3}, Núria de la Osa^{1,2}, Eva Penelo^{1,3}, Lourdes Ezpeleta^{1,2}

¹Unitat d'Epidemiologia i de Diagnòstic en Psicopatologia del Desenvolupament

²Departament de Psicologia Clínica i de la Salut

Universitat Autònoma de Barcelona, Barcelona, Spain

³Departament de Psicobiologia i de Metodologia de les Ciències de la Salut

Universitat Autònoma de Barcelona, Barcelona, Spain

Author Note

Natalia Pueyo <https://orcid.org/0000-0003-3251-7859>

J. Blas Navarro <https://orcid.org/0000-0001-5929-4224>

Núria de la Osa <https://orcid.org/0000-0003-4499-0942>

Eva Penelo <https://orcid.org/0000-0001-6796-7660>

Lourdes Ezpeleta <https://orcid.org/0000-0002-8957-083X>

Natalia Pueyo: Conceptualization, Writing- Original draft. J. Blas Navarro: Data curation, Formal analysis, Methodology. Núria de la Osa: Writing- review and editing. Eva Penelo: Writing- review and editing. Lourdes Ezpeleta: Supervision.

The authors do not have any conflict of interest to disclose.

This work was supported by the Spanish Ministry of Science, Innovation and Universities [Grant PGC2018-095239-B-I00 (MICIU/FEDER)].

The project was approved by the Ethics Committee on Animal and Human Experimentation of the Universitat Autònoma de Barcelona. Families gave written consent for the assessment.

This study follows the Journal Article Reporting Standards for quantitative and longitudinal studies and the eight guidelines of the Transparency and Openness Promotion. Therefore, we inform that the data, materials, and analysis syntax of this study cannot be made publicly available due to ethical restrictions protecting the confidentiality of the families involved. Also, the study and the analysis plan were not preregistered.

Correspondence concerning this article should be addressed to Lourdes Ezpeleta, Departament de Psicologia Clínica i de la Salut. Edifici B, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona). Spain. Email: lourdes.ezpeleta@uab.cat

Abstract

This longitudinal study aims to provide meaningful cut-off scores for total score of the teacher-rated Inventory of Callous Unemotional Traits (ICU) from the preschool age until early adolescence, separately by age and sex. The ICU cut-off scores were determined by using low/high trajectories of oppositional defiant problems and conduct problems in a Spanish community sample of 620 children that were followed up between the ages of 3 and 13. Receiver operating characteristic (ROC) curves with the two trajectories as criteria and ICU total score at each age as a predictor were estimated by sex separately, and the area under the ROC curve (AUC) was obtained. Average ICU cut-off scores of 26 for boys and 22 for girls were found to be of moderate utility for the prediction of high trajectories of each of oppositional defiant problems and conduct problems. They identified cases with an average sensitivity of 66% and specificity of 70% for boys; and an average sensitivity of 69% and specificity of 72% for girls. The obtained cut-off scores might help clinical practitioners in their decision-making process when identifying low and high-risk groups of children.

Keywords: callous-unemotional traits, community samples, cut-off scores, early childhood, longitudinal.

Public Significance Statements

This study provides cut-off scores for the teacher-rated Inventory of Callous-Unemotional Traits in a community sample spanning early childhood until early adolescence, separately by age and sex. Cut-off scores of 26 for boys and 22 for girls are of moderate utility for the prediction of high trajectories of oppositional defiant problems, and conduct problems. These cut-off scores might help clinical practitioners when identifying low and high-risk groups of children.

Age and sex-specific Cut-off Scores for the Teacher-report Inventory of Callous Unemotional Traits on Children

Callous-unemotional traits (CU traits) (e.g., lack of guilt and empathy, shallow emotion display and irresponsibility) in children and adolescents are considered a risk factor for adult psychopathy (Frick et al., 2014). Even from the preschool age, CU traits can be identified and used in clinical settings to describe a subgroup of preschoolers who show psychosocial impairment, including conduct problems, difficulties in emotional recognition, and aggressive behavior (Frick & Myers, 2018; Kimonis et al., 2016). Among older children and adolescents, high CU traits are related to externalizing behavior and can predict greater antisocial behavior, criminality, violence, and impulsive behavior in adulthood (Frick et al., 2014).

To identify this subset of children and adolescents with CU traits, the Inventory of Callous-Unemotional Traits (ICU) (Frick, 2004) is among the most used instruments. The ICU was designed to improve certain limitations of the Antisocial Process Screening Device (APSD) (Frick & Hare, 2002) by providing a wider range of items to capture CU traits. Psychometric research on the ICU has widely focused on the self-reported version and has presented evidence for a 3-factor structure so that the 24 items of the ICU can be grouped into three subscales (Uncaring, Callousness, and Unemotional) (Essau et al., 2006). However, using the self-report ICU total score instead of the three subscale scores is more reliable to detect CU traits and shows better predictive and concurrent validity (Kimonis et al., 2008; Ray et al., 2016). A meta-analysis has shown high internal consistency and convergent validity, as well as validity evidence based on relations to external variables of the ICU total score across different samples (Cardinale & Marsh, 2020). This study also reveals that the total ICU scores of parent- or teacher-reports are able to predict externalizing behavior, low empathy, aggression, delinquency, hyperactivity, and higher psychopathic traits.

Although the ICU is widely used to assess CU traits, researchers have only recently begun to explore ICU cut-off scores. One explanation for this lack of thresholds might be that the ICU was not conceptualized as a categorical diagnostic tool (Kimonis & Goulter, 2017). It was rather designed to establish a dimensional description of CU traits (Herpers et al., 2017). By establishing an empirical cut-off score, the ICU could be used as a screening instrument for identifying children and adolescents who show high CU traits and, therefore, might be at higher risk of antisocial behavior (Pihet et al., 2015).

With the introduction of the “Limited Prosocial Emotions” specifier in the diagnosis of conduct disorder in the Diagnostic and Statistical Manual of Mental Disorders (5th Edition; DSM-5) (American Psychiatric Association, 2013), there is an unmet need for using a categorical perspective of CU traits and its corresponding clinical ICU cut-off scores. Different approaches to define these cut-off scores have been used. Providing normative data for the different informant versions of the ICU is one of them. There are European self-report ICU norms for adolescents aged 11 to 17 (Kemp et al., 2019) and for the parent-, teacher-, and self-report versions for children and adolescents between the ages of 6 to 18 (Ueno et al., 2021). There are also parent-report ICU norms in the United States for children aged 5 to 12 (Bansal et al., 2020).

Such normative data help to define distribution-based cut-off scores that can be used to distinguish children and adolescents with low and high CU traits. To establish them, percentiles equal to or above 80 or 90 have been used in different datasets. For instance, in Kemp et al.’s (2019) European sample of self-report versions of the ICU, the 95th percentile is used as a cut-off, obtaining direct scores between 37 and 41 for boys and 32 and 35 for girls. Kumsta et al., (2012) used the 80th percentile of the parent - and self-reported ICU scores (direct scores of 46 and 30, respectively) in an adoption study on adolescents to identify high CU trait groups. This 80th percentile was also used among a clinical sample of male

adolescents using the teacher-report version (direct scores equal to or above 40) (Levy et al., 2015). Among community adolescents aged 11 to 13, the 94th percentile (direct scores above 39) for the self-report ICU has been applied as a cut-off to identify high CU traits (Viding et al., 2009).

These studies have mostly explored the parent and self-reported versions of the ICU, with resulting cut-off scores for the ICU total score that vary from 36 to 46 for the parent version and from 30 to 39 for the self-report version. It is of concern that these studies apply the 80th or 90th percentile arbitrarily on the three informant versions of the ICU, without taking into consideration the informant of the ICU, type of sample (community, clinical, at-risk samples), the sex of the children or their age. Likewise, the resulting low/high CU trait groups could be very heterogeneous and include a large number of false positives or false negatives, depending on the criterion variables that might have been used (Kimonis et al., 2014; Szabó et al., 2017).

Therefore, in recent years, there has been growing interest in determining ICU cut-off scores following associations with psychological and behavioral outcomes (Kemp et al., 2021). From this perspective, ICU cut-off scores are established by using an external criterion measure which allows identifying not only high CU trait groups but also concurrent psychosocial difficulties such as aggression, violent behavior, criminal activity, conduct problems, or bullying (Feilhauer et al., 2012; Kemp et al., 2021; Kimonis et al., 2014). For example, Kimonis et al. (2014) identified that direct ICU scores of 24 and 27 for mother-reports and father-reports, respectively, could detect co-occurring high CU traits and high conduct problems and predict bullying behavior among children aged 6 to 12 years.

Focusing on the ICU self-report, Feilhauer et al. (2012) found that a direct score of 26 could identify high CU traits and discriminate adolescent offenders from the control group. To date, only two studies have explored empirical teacher-reported ICU cut-off scores.

Docherty et al. (2017) explored the three versions of the ICU (parent, teacher and self-report) and found that a cut-off direct score of 33 for the teacher-report version of the ICU could predict aggression, violence, and detained status in a mixed sample of community and justice-involved adolescents. Similar cut-off scores were found by Kemp et al. (2019) for a community sample of youth, with sex-specific empirical thresholds for the teacher-report ICU where direct scores between 35-50 in adolescent boys and 35-38 in adolescent girls identified high CU traits, conduct problems, and peer-reported meanness.

These previous studies on ICU cut-off scores indicate that the age and sex of the children, type of informant, and even the settings of the study should be taken into consideration when establishing the cut-off scores (Carvalho et al., 2018; Kimonis, Fanti & Singh, 2014). Therefore, cut-off scores should be sex-specific to provide a valid diagnostic approach and guarantee a more efficient intervention strategy (Kemp et al., 2021), for two reasons; because the prevalence of CU traits is higher in boys than in girls (Ueno et al., 2021) and girls with elevated CU traits show more severe psychosocial difficulties than affected boys (Euler et al., 2015).

Moreover, age-specific cut-off scores are also needed because CU traits may undergo changes between childhood and adolescence (Essau et al., 2006). Even though CU traits remain rather stable over time on an individual level, there might be fluctuations of normative scores of CU traits at different ages on a cohort level. One stream of research has revealed that children show higher self-reported ICU scores than adolescents (Carvalho et al., 2018; Ueno et al., 2021) whereas, in another study, adolescents scored higher than their younger peers (Essau, 2006). As the effect of age needs to be further explored, establishing thresholds for different age groups might be crucial to reliably detect high CU traits over childhood and adolescence.

Gaining insight into the teacher-reported ICU cut-off scores for preschoolers (starting at age 3) from community samples is of special interest for different reasons. First, exploring the less-studied teacher-report ICU scores will offer further evidence on its validity, putting greater emphasis on teachers as informants for CU traits. Preliminary studies have found that teachers seem to be more reliable as informants for CU traits than the child or adolescent itself (Docherty et al., 2017; Ueno et al., 2021). They have also observed that the teacher-report ICU scores show better classification accuracy than other informant-reported version scores (Kemp et al., 2021). Certain CU traits, for example, indifference about performance or difficulties in prosocial behavior, may be identified easier in school settings than in the context of the family, implying that teachers may be more aware of these aspects than parents or the youth themselves (Ueno et al., 2021). Second, ICU cut-off scores for preschool samples have not yet been established, but early childhood is recognized as being a crucial developmental period for socio-affective maturity (Carter et al., 2004). Early callous-unemotional behavior can predict severe psychosocial impairment, aggression, and rule-breaking behavior in late childhood (Waller et al., 2016). Therefore, providing validated and empirical cut-off scores to detect high CU traits from early years could improve the diagnostic process and enable clinical practitioners to identify a high-risk group of children with difficulties in prosocial emotions. Third, focusing on a community sample can help provide valid cut-off scores for the general population, making the questionnaire scores useful not only in a clinical context but also in educational settings.

Finally, this longitudinal study aims to provide meaningful cut-off scores for the teacher-rated ICU total score from the preschool age until early adolescence (ages 3-13), separately by age and sex. As the literature indicates that early childhood CU traits are associated with severe conduct problems (Longman et al., 2016) and oppositional defiant disorder (ODD) (Willoughby et al., 2011), ICU cut-off scores will be determined by using

developmental trajectories empirically obtained for each of the previous aforementioned two measures as external categorial criteria.

Method

Participants

The sample comes from a longitudinal study on risk factors of psychopathology in children starting at age 3. The sampling design included two phases which are summarized in Figure 1. In the first phase, 1341 families of 2283 (58.7%) children, randomly selected from the census of early childhood schools in Barcelona (Spain), agreed to participate (33.6% high socioeconomic status (SES), 43.1% middle, and 23.3% low; 50.9% boys). In the second phase, a parent-rating of ODD symptoms (8 items) based on the conduct problems scale of the Strengths and Difficulties Questionnaire (SDQ) plus additional items from the Diagnostic and Statistical Manual of Mental Disorders (4th Edition; DSM-IV) based ODD was used to screen children with possible psychological problems. Two groups were finally considered: the screening positive group (+) included all children with scores above the 90th percentile on the screening measure or with a positive response for any of the eight DSM-IV ODD symptoms ($n = 417$; 49.0% boys) and the screening negative group (-) incorporated a random group of children who did not reach the positive criteria ($n = 205$; 51.2% boys). Both groups ($n = 622$) participated in the longitudinal study.

The sample used in the present study included 620 children (because there was no ICU data for two children) who were assessed yearly from the age of 3 to 13 years (11 assessment points). Table 1 shows a description of the available sample at the age of 3 and 13 years. SES was assessed according to the Hollingshead Four-Factor Index of Social Status (Hollingshead, 1975). This index includes weighted scaled scores of the occupation and educational attainment of the mother and father, which were categorized in 5 groups. The lowest sample size was found at age 13 when 318 children participated. There were no

differences in sex ($p = .630$) due to attrition, although the available sample at age 13 had a higher SES ($p < .001$).

Measures

The Inventory of Callous-Unemotional Traits (ICU; Frick, 2004) includes 24 items that teachers responded to annually using a 4-point Likert-type scale from 0 (*not at all true*) to 3 (*definitely true*). The ICU total score is the sum of the ratings of all the items, reversed when necessary, and higher scores indicate greater CU traits. The total score demonstrated good internal consistency (Cronbach's α in the present sample ranged from .88 to .93 over time).

The *Strengths and Difficulties Questionnaire* (SDQ) (Goodman, 1997) is a brief screening questionnaire for the mental health of children based on five scales of five items each (0: *not true* to 2: *certainly true*). Teachers completed the questionnaire annually when children were between 3 and 13 years old. Different scale scores of the adapted Spanish version were used. For oppositional defiant problems (ODP), three items of the conduct problems scale of the SDQ were used (temper tantrums, disobedient and spiteful), together with five additional items that were included based on the DSM-IV ODD symptomatology criteria (annoys, blames, touchy, angry, argumentative). For conduct problems, the specific SDQ scale was used. Ordinal alpha values ranged from .91 to .96 for ODP scores and from .83 to .91 for conduct problem scores.

The *Diagnostic Interview for Children and Adolescents for Parents of Preschool and Young Children* (DICA-PPYC; Ezpeleta et al., 2011) is a semi-structured diagnostic interview for assessing a set of common psychological disorders according to the DSM-5 criteria. ODD and Conduct disorder (CD) diagnoses were used to describe the level of psychopathology in the sample at ages 3 and 13.

Procedure

This project was approved by the Ethics Committee on Animal and Human Experimentation of the author's institution and follows the ethical standards of the 1964 Declaration of Helsinki and its later amendments. Families were recruited from schools and those who met screening criteria and gave their consent to participate in the study were assessed annually by teachers.

Data Analysis

The statistical analysis was conducted with MPlus 8.5 and Stata 16. As the sampling design was two-stage, all the analyses were weighted by the inverse probability of selection in the second phase of sampling. To estimate groups of individual trajectories for direct scores of ODP and SDQ conduct problems, Growth Mixture Models (GMM) with one process at each time were estimated. The Robust Maximum Likelihood (MLR) estimation method was used. The growth models included intercept (I), linear (S), and quadratic trend (Q) fitted over the 11 available annual assessments from ages 3 to 13. Time was rescaled to 0-10, so the first-year assessment (at age 3) represented the intercept (i.e., the basal direct score). Models with two growth patterns (trajectories) were obtained for each of the two processes. The accuracy of the classification in two trajectories was assessed through adequate average posterior probabilities, entropy values equal to or greater than .70, and a minimum of 20 participants in each trajectory.

Receiver operating characteristic (ROC) curves with the two trajectories as criteria and the ICU total score at each age as a predictor were estimated separately by sex, and the area under the ROC curve (AUC) was obtained. Following Hosmer et al., (2013) a value of $AUC < .70$ is considered non-predictive; AUC from .70 to .80 is considered acceptable; AUC from .80 to .90 is considered excellent and $AUC > .90$ is considered outstanding. The optimal ICU total score cut-off was selected as the score that maximizes sensitivity (Se) and specificity (Sp) by minimizing the square of the distance between the point (0, 1) on the

upper left-hand corner of ROC space and any point on ROC curve, with equals costs for false positive and false negative misclassifications (Hajian-Tilaki, 2013).

Transparency and Openness

This study follows the Journal Article Reporting Standards for quantitative and longitudinal studies and the eight guidelines of the Transparency and Openness Promotion. Therefore, we inform that the data, materials, and analysis syntax of this study cannot be made publicly available due to ethical restrictions protecting the confidentiality of the families involved. Also, the study and the analysis plan were not preregistered.

Results

Descriptive Data

Table 2 shows the mean scores for the ICU total score at each follow-up along childhood from 3 to 13 years separately by sex. The mean ICU scores showed fluctuations over age and across sex. Focusing on age, the highest mean ICU scores were found at ages 10 to 13 among boys and at ages 3 and 4 among girls. Focusing on sex, boys' mean scores ranged between 21.5 (age 11) and 24.3 (age 13), whereas girls' mean scores were significantly lower than the boys' scores, ranging between 17.5 (age 11) and 19.3 (age 3).

When applying the 80th percentile to establish normative cut-off scores for identifying high CU traits (Kumsta et al., 2012), ICU cut-off scores showed age and sex-specific variations, with scores ranging between 29 and 36 for boys and between 23 and 27 for girls (Table 2).

Growth Mixture Models

For each analyzed score (ODP and SDQ-conduct), the 2-trajectory solution (low and high) from GMM was selected to act as criterion for estimating ROC curves with the ICU total score as predictor. Average posterior probabilities were in the range .90-.97, entropy values were in the range .80-.86, and the minimum *N* was 77 for the high ODP trajectory, all

of them indicating an adequate goodness of fit. Figure 2 shows the 2-trajectory solutions for the two measures. A reference line with the percentile 75 value has been added to each figure to help delimitate low versus high scores.

Average Sensitivity, Specificity and Predictive Value of ICU Scores for ODP and Conduct Problems

The ICU total score was able to distinguish boys on low vs. high ODP trajectories (with average values of AUC = 73.7%; Se = 67.8%; Sp = 69.3%) and conduct problem trajectories (with average values of AUC = 73.7%; Se = 67.7%; SP = 69.5%). The optimal average cutoff scores fell around a raw score of 26 for both measures (Tables 3-4).

For girls, the ICU total score was also able to discriminate low vs. high ODP trajectories (with average values of AUC = 74.4%; Se = 69.5%; Sp = 72.6%) and conduct problem trajectories (with average values of AUC = 74.2%; Se = 69.6%; Sp = 72.1%). The optimal average cutoff score fell around a raw score of 22 for both measures (Tables 3-4).

Age and Sex-Specific Sensitivity, Specificity and Predictive Value of ICU Scores for ODP and Conduct Problems

For ODP trajectories (Table 3), the ICU total score showed acceptable-excellent predictability, especially between the ages of 6 to 9 where the AUCs reached their highest value, ranging between 75.2% and 81.8% for boys; and between 80.7% and 84.7% for girls. The ICU raw total cut-off score at which there was optimal discrimination between belonging to a low or high ODP trajectory ranged between 24 and 29 for boys; and between 21 and 27 for girls (excluding the cut-off of 17 at age 13). The best discrimination was found at age 6, when a cut-off score of 28 could identify boys on a high ODP trajectory with an AUC = 81.8%, providing the best trade-off between Se (72.1%) and Sp (75.4%). For girls, the best discrimination was achieved at age 8, when a cut-off score of 23 could discriminate belonging to a high ODP trajectory with an AUC = 85.3%, with Se = 78.9% and Se = 77.0%.

For trajectories related to conduct problems (Table 4), the ICU total score showed acceptable predictability from the ages of 6 to 13 for boys (AUCs between 72.2% and 79.9%); and excellent predictability between the ages of 6 and 10 for girls (AUCs between 81.3% and 84.3%). The ICU total score cut-off at which there was optimal discrimination between belonging or not belonging to the high conduct problem trajectory ranged between 22 and 27 for boys (excluding the cut-off of 33 at age 13); and between 20 and 25 for girls (also excluding the cut-off of 18 at age 13). The best discrimination was found at age 8, when a cut-off score of 27 could identify the boys belonging to a high conduct problem trajectory with an AUC = 79.9%, providing the best trade-off between Se (75.9%) and Sp (75.6%). For girls, the best discrimination was identified at age 7, when a cut-off score of 24 could discriminate females on a high conduct problem trajectory with an AUC = 84.3%, with Se = 76.9% and Sp = 80.4%.

Overall, the empirical approach through ROC analysis revealed that average ICU cut-off scores of 26 for boys and 22 for girls can be of moderate utility for the prediction of high trajectories of each of ODP and conduct problems, identifying cases with an average Se of 66% and Sp of 70% for boys; and an average Se of 69% and Sp of 72% for girls.

Discussion

The aim of this study was to provide cut-off scores from preschool age to late childhood for the teacher-reported ICU version, separately by age and sex. Our results indicate that the discriminative capacity of the ICU total scores between low and high trajectories of each of ODP and conduct problem measures is more accurate in mid-childhood (6-9 years) than in the developmental extremes of early childhood (3-5 years) or early adolescence (10-13 years). Even if the ICU score has shown to be a valid tool to assess CU traits in preschoolers (Kimonis et al., 2016), this is the first study that has explored early childhood ICU cut-off scores. Further research in this early developmental stage is needed to understand how these

cut-off scores, the Se, and the Sp of the ICU may vary. More data are available among community and delinquency-involved adolescent samples, suggesting that the teacher-report ICU cut-off score is allocated around 33 (Docherty et al., 2017; Kemp et al., 2021), which is between 7 and 9 points higher than our obtained scores. Also, normative data shows that for the teacher-report ICU version, adolescents between 11 and 14 years score higher on the ICU than children aged 6 to 10 years (Ueno et al., 2021). These results may suggest that when children reach puberty, certain adolescent behaviors (such as rule-breaking, shallow emotional display, etc) may be compatible with CU traits. Thus, more CU behaviors are observed during adolescence, resulting in higher ICU cut-off scores for youth than for children. Consequently, AUC, Se, and Sp of ICU scores may change in adolescent samples, as our results indicate.

The type of informant of the ICU might also be a factor to consider in understanding the classification accuracy across ages. Our study focused on the teacher-reported ICU version, which is one of the less-studied ICU informant versions. Even though the role of teacher-student relationships in children with CU traits is only scarcely studied, teachers experience children in a school context where punishment insensitivity, impaired reward processing, low intrinsic motivation, and low socioemotional abilities can specifically be observed (Allen et al., 2018). Therefore, teachers, in comparison to parents, might have a more objective view when rating the ICU items as they are in touch with a great variety of children and adolescents. For example, in Ueno et al.'s (2021) study on ICU norms, teachers reported higher mean ICU scores across age and sex than parents. The same tendency was observed in ICU cut-off scores (Kemp et al., 2021) where the empirical teacher-reported ICU cut-off score (35) was 12 points higher than the parent-reported score (23). This study also found higher classification accuracy in the teacher-report version of the ICU score for conduct problems and peer-reported meanness in comparison to the parent- and self-report ICU

versions. These findings, together with our preliminary results, indicate that teachers are valid informants for CU traits in childhood and adolescence and that the teacher-report ICU version should be more widely used in clinical and research settings. This would help to gain further evidence on possible differences between parent and teacher-reported ICU versions and the different resulting ICU cut-off scores.

Our study also found sex-specific effects on ICU cut-off scores across age. The fact that lower cut-off scores were found among girls is in line with previous research (Kemp et al., 2021; Ueno et al., 2021). As girls generally score lower on CU traits than boys (Pihet et al., 2015), the current stream of research on CU traits has vastly focused on boys. Knowledge of how girls express CU traits is still limited, but preliminary findings show that girls with high CU traits engage in more internalizing behavior (depression and anxiety) (Euler et al., 2015) and show more chronic and severe ODD/conduct disorder trajectories (Kroneman et al., 2011) than boys. Interestingly, girls with CU traits might not robustly show a lack of guilt or empathy (Pardini et al., 2012), which might be explained by the biological and environmental differences between sexes, such as sex-specific brain structure differences in areas involving empathy or gender differences in the socialization process, where girls are more encouraged than boys to be empathic, caring, and sensitive towards others (Hipwell et al., 2007). In line with this hypothesis, Essau et al., (2006) found that the “uncaring” scale score of the ICU was not a predictor of problematic behavior in girls, but it was for boys. Applied to our results, it is of no surprise that the ICU scores showed lower classification accuracy in discriminating each of ODP and conduct problem trajectories when comparing girls to boys. Thus, it would be of interest to analyze ICU cut-off scores as predictors for internalizing behaviors, separately by sex, to better understand sex differences in CU traits together with other socioemotional difficulties such as anxiety or depression.

Finally, in our study, the proposed data-driven ICU cut-off scores to define high-risk groups of boys and girls are less stringent than when applying normative cut-off scores as, for example, the 80th percentile. Using such normative cut-off scores might result in identifying a heterogeneous group of children that share similar high CU traits but might differ in other characteristics. Therefore, this approach might be more useful in clinical samples to identify specifically high CU traits among children who might already have been identified as belonging to a high-risk group. But according to our results, when such percentiles are used as an ICU cut-off score, we are possibly excluding a group of children that are also susceptible to high risk, even though their ICU scores do not reach a “clinical significance.” For this reason, using the data-driven approach is especially important for early detection and prevention purposes when working with community samples. Because our data-driven approach shows lower ICU cut-off scores than using, for example, the 80th percentile cut-off, more children with high-risk trajectories could be identified. This larger group of children might show greater heterogeneity with probably more comorbidities and other demographic risk factors. That would be in line with previous research suggesting that the developmental pathways to CU traits are diverse (Frick et al., 2013). Therefore, the empirical ICU cut-off scores might be less stringent than the normative thresholds and might also help to detect a group of children who are at high risk and might need preventive and therapeutic intervention programs.

Constraints on Generality

The participants of our study were children from a community sample in Spain. We expect our results to generalize to other community sample children with similar socioeconomic and cultural backgrounds. A direct replication would follow a community sample of children from early childhood to early adolescence using the adapted versions of

the used instruments. We have no reason to believe that the results depend on other characteristics of the participants, materials, or contexts.

This study has notable strengths, such as a large sample size, community environment, empirical trajectories used as outcomes, and longitudinal approach of 11 years during which multiple teachers reported the children's development that was used to obtain ICU cut-off scores. Also, the study focused on the less-studied teacher-report ICU version, and both sexes were analyzed separately to provide more accurate ICU cut-off scores. The ROC analysis that was applied to establish empirical ICU cut-off scores is also a strength as only a few studies have used this approach (Docherty et al., 2017; Kemp et al., 2021; Kimonis et al., 2014), making it a promising field of research. In addition, normative data based on the 80th percentile was also presented. Finally, the external criterion was based not only on an individual score on certain tests but also on the whole childhood trajectory of two different, but somewhat overlapping, constructs of ODP and conduct problems.

Despite these strengths, the results of this study need to be considered in light of some limitations. First, the study only focused on the teacher-report ICU version. As multi-informant approaches provide a more consistent CU trait assessment and are considered best-practice (Kimonis & Goulter, 2017), including other-report versions in future studies would be of interest. Second, the results might be valid for Spanish children only, and replication studies in other countries are needed to confirm the provided ICU cut-off scores. Finally, the predictive validity was analyzed through ODP and conduct problem trajectories. Both trajectories share three items of the SDQ, therefore, this overlap might explain the similar results on AUCs, sensitivity, and specificity. Including criterion variables such as clinical diagnoses of conduct disorder, ODD, or hyperactivity could therefore provide specific clinical ICU cut-off scores. As our study focused on a community sample, the prevalence of

such diagnoses was too low to include them as criterion variables. Further studies on more clinical samples might explore this empirical approach in establishing cut-off scores.

Clinical Implications

Our research could be a useful aid in CU trait assessments as researchers and clinical practitioners could apply the provided ICU cut-off scores to identify a subgroup of children with high CU traits and high risk of sustained ODP and conduct problem trajectories, especially if the provided cut-off scores have been tested and replicated in other studies. Our data-driven approach to establish ICU cut-off scores has resulted in less stringent thresholds that might identify a larger group of high-risk children than applying an arbitrary percentile. This is important because clinicians should take into consideration that children who do not necessarily show clinically high CU traits (e.g., over the 80th percentile) could benefit from early detection and prevention of other associated difficulties such as ODP or conduct problems. In this sense, working with such a categorical CU trait approach together with the dimensional model would be useful to personalize the treatment approach through a deeper understanding of specific problematic domains. Moreover, our results suggest that early childhood CU trait scores can already detect a subgroup of children in high-risk trajectories with moderate accuracy and average Se and Sp between 66% and 72%. These values of Se and Sp are important to prevent false negatives or false positives during the diagnostic process. In this sense, the developmental window of 6 to 10 years seems to be the most accurate for correctly identifying CU traits and ODP or conduct problems, implying clinical practitioners should pay more explicit attention when children show CU traits in this phase. Finally, CU trait assessments should be based on multi-informant reports, counting on teachers as valid informants for identifying CU trait behaviors.

In conclusion, this is the first study that provides empirical cut-off scores for the teacher-report ICU in a longitudinal sample of 3- to 13-year-old children. The provided ICU

scores demonstrated the capacity to identify children with high CU traits and ODP or conduct problem trajectories with average cut-off scores of 26 for boys and 22 for girls. These thresholds might help clinical practitioners in their decision-making process when identifying low and high-risk groups of children. However, it should be noted that a final diagnosis should never rely on a single test score but rather explore different sources of information. Future research should continue to evaluate empirical ICU cut-off scores according to the different informant-versions, age, sex, context, and other external criteria.

References

- Allen, J.L., Bird, E., & Chhoa, C.Y. (2018). Bad Boys and Mean Girls: Callous-Unemotional Traits, Management of Disruptive Behavior in School, the Teacher-Student Relationship and Academic Motivation. *Frontiers in Education*, 3,108. [https://doi:10.3389/feduc.2018.00108](https://doi.org/10.3389/feduc.2018.00108)
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA. American Psychiatric Association.
- Bansal, P. S., Babinski, D. E., Waxmonsky, J. G., & Waschbusch, D. A. (2020). Psychometric properties of parent ratings on the Inventory of Callous–Unemotional traits in a nationally representative sample of 5- to 12-year-olds. *Assessment*, 1–15. <https://doi.org/10.1177/1073191120964562>
- Cardinale, E. M., & Marsh, A. A. (2020). The reliability and validity of the Inventory of Callous Unemotional traits: A meta-analytic review. *Assessment*, 27(1), 57–71. <https://doi.org/10.1177/1073191117747392>
- Carter, A. S., Briggs-Gowan, M. J., Davis, N. O. (2004). Assessment of young children's social-emotional development and psychopathology: Recent advances and recommendations for practice. *Journal of Child Psychology and Psychiatry*, 45(1), 109–134. <https://doi.org/10.1046/j.0021-9630.2003.00316.x>
- Carvalho, M., Faria, M., Conceição, A., Matos, M. G. De, & Essau, C. A. (2018). Callous-Unemotional Traits in children and adolescents. Psychometric properties of the Portuguese version of the Inventory of Callous-Unemotional traits. *European Journal of Psychological Assessment*, 34(2), 101–110. <https://doi.org/10.1027/1015-5759/a000449>
- Docherty, M., Boxer, P., Huesmann, L. R., Brien, M. O., & Bushman, B. (2017). Assessing Callous-Unemotional traits in adolescents: Determining cutoff scores for the Inventory of Callous and Unemotional traits. *Journal of Clinical Psychology*, 73(3), 257–278.

<https://doi.org/10.1002/jclp.22313>

Essau, C. A., Sasagawa, S., Frick, P. J. (2006). Callous-Unemotional Traits in a community sample of adolescents. *Assessment*, 13(4), 454–469.

<https://doi.org/10.1177/1073191106287354>

Euler, F., Jenkel, N., Stadler, C., Schmeck, K., Fegert, J. M., Kölch, M., & Schmid, M.

(2015). Variants of girls and boys with conduct disorder: Anxiety symptoms and callous-unemotional traits. *Journal of Abnormal Child Psychology*, 43(4), 773–785.

<https://doi.org/10.1007/s10802-014-9946-x>

Ezpeleta, L., de la Osa, N., Granero, R., Domènech, J. M., & Reich, W. (2011). The

Diagnostic Interview of Children and Adolescents for Parents of Preschool and Young Children: psychometric properties in the general population. *Psychiatry Research*, 190, 137-144. doi: 10.1016/j.psychres.2011.04.034

Feilhauer, J., Cima, M., & Arntz, A. (2012). Assessing callous-unemotional traits across different groups of youths: Further cross-cultural validation of the Inventory of Callous-Unemotional Traits. *International Journal of Law and Psychiatry*, 35(4), 251–262.

<https://doi.org/10.1016/j.ijlp.2012.04.002>

Frick, P. J. (2004). The Inventory of Callous-Unemotional traits. In *Unpublished rating scale*.

Frick, P. J., Blair, R., & Castellanos, F. (2013). Callous-Unemotional Traits and

Developmental Pathways to the Disruptive Behavior Disorders. In: Tolan P.,

Leventhal B. (eds) *Disruptive Behavior Disorders. Advances in Development and*

Psychopathology: Brain Research Foundation Symposium Series, Vol 1. Springer,

New York, NY. https://doi.org/10.1007/978-1-4614-7557-6_4

Frick, P. J., & Hare, R. (2002). *Antisocial Process Screening Device*. Multi-Health Systems.

Frick, P. J., & Myers, T. D. W. (2018). Conduct Disorder and Callous- Unemotional Traits.

In J. W. & S. Ltd (Ed.), *The Wiley Handbook of Disruptive and Impulse-Control*

- Disorders* (pp. 37–54). John Wiley & Sons, Hoboken, NJ.
- Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychological Bulletin*, *140*(1), 1–57. <https://doi.org/10.1037/a0033076>
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology & Psychiatry*, *38*(5), 581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Herpers, P. C. M., Klip, H., Rommelse, N. N. J., Taylor, M. J., Greven, C. U., & Buitelaar, J. K. (2017). Taxometric analyses and predictive accuracy of callous-unemotional traits regarding quality of life and behavior problems in non-conduct disorder diagnoses. *Psychiatry Research*, *253*, 351–359. <https://doi.org/10.1016/j.psychres.2017.04.004>
- Hipwell, A. E., Pardini, D. A., Loeber, R., Semboer, M., Keenan, K., & Stouthamer-Loeber, M. (2007). Callous-unemotional behaviors in young girls: Shared and unique effects relative to conduct problems. *Journal of Clinical Child and Adolescent Psychology*, *36*(3), 293–304. <https://doi.org/10.1080/15374410701444165>
- Hollingshead, A. B. (1975). *Four factor index of social status*. New Haven: Unpublished manuscript, Yale University, Department of Sociology.
- Hosmer, D., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. 3rd ed. John Wiley & Sons, Hoboken, NJ.
- Hajian-Tilaki K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*, *4*(2), 627–635. <http://www.ncbi.nlm.nih.gov/pmc/articles/pmc3755824/>
- Kemp, E., Frick, P. J., & Robertson, E. (2019). *Inventory of Callous-Unemotional traits (ICU) - self-report youth version: T-scores & percentiles [Normative dataset]*.

<https://sites01.lsu.edu/faculty/pfricklab/icu>

Kemp, E., Frick, P.J., Matlasz, T., Clark, J., Robertson, E., Ray, J., Thornton, L., Wall Myers, Steinberg, L., & Cauffman, E. (2021). Developing Cutoff Scores for the Inventory of Callous-Unemotional Traits (ICU) in Justice-Involved and Community Samples. *Journal of Clinical Child & Adolescent Psychology*.

<https://doi.org/10.1080/15374416.2021.1955371>

Kimonis, E. R., Fanti, K. A., & Singh, J. P. (2014). Establishing cut-off scores for the parent-reported Inventory of Callous-Unemotional traits. *Archives of Forensic Psychology*, 1(1), 27–48.

Kimonis, E. R., Fanti, K. A., Anastassiou-Hadjicharalambous, X., Mertan, B., Goulter, N., & Katsimicha, E. (2016). Can callous-unemotional traits be reliably measured in preschoolers? *Journal of Abnormal Child Psychology*, 44(4), 625–638.

<https://doi.org/10.1007/s10802-015-0075-y>

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., Aucoin, K. J., & Morris, A. S. (2008). Assessing callous-unemotional traits in adolescent offenders: Validation of the inventory of callous-unemotional traits. *International Journal of Law and Psychiatry*, 31(3), 241–252.

<https://doi.org/10.1016/j.ijlp.2008.04.002>

Kimonis, E. R., & Goulter, N. (2017). Assessment of Callous and Related Traits. In Sturme, P (Ed.), *The Wiley Handbook of Violence and Aggression* (pp. 1–12). John Wiley & Sons, Hoboken, NJ.

Kroneman, L., Hipwell, A., Loeber, R., Koot, H., & Pardini, D. (2011). Contextual risk factors as predictors of disruptive behavior disorder trajectories in girls: Moderating effects of callous unemotional. *Journal of Child Psychology & Psychiatry*, 52(2), 167–175. <https://doi.org/10.1111/j.1469-7610.2010.02300.x> Contextual

- Kumsta, R., Sonuga-barke, E., & Rutter, M. (2012). Adolescent callous – unemotional traits and conduct disorder in adoptees exposed to severe early deprivation. *British Journal of Psychiatry*, 200(3), 197-201. doi:10.1192/bjp.bp.110.089441
- Leno, V. C., Bedford, R., Chandler, S., White, P., Yorke, I., & Charman, T. (2020). Callous-unemotional traits in youth with autism spectrum disorder (ASD): Replication of prevalence estimates and associations with gaze patterns when viewing fearful faces. *Development and Psychopathology*, 1-9. <https://doi.org/10.1017/s0954579420000449>
- Levy, T., Bloch, Y., Bar-Maisels, M., Gat-Yablonski, G., Djalovski, A., Borodkin, K., & Apter, A. (2015). Salivary oxytocin in adolescents with conduct problems and callous-unemotional traits. *European Child & Adolescent Psychiatry*, 24(12), 1543–1551. <http://www.ncbi.nlm.nih.gov/pubmed/26433370>
- Longman, T., Hawes, D. J., & Kohlhoff, J. (2016). Callous–Unemotional traits as markers for conduct problem severity in early childhood: A meta-analysis. *Child Psychiatry and Human Development*, 47(2), 326–334. <https://doi.org/10.1007/s10578-015-0564-9>
- Pardini, D., Stepp, S., Hipwell, A., Stouthamer-Loeber, M., & Loeber, R. (2012). The clinical utility of the proposed DSM-5 callous-unemotional subtype of conduct disorder in young girls. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(1), 62–73. <https://doi.org/10.1016/j.jaac.2011.10.005>
- Pihet, S., Etter, S., Schmid, M., & Kimonis, E. R. (2015). Assessing callous-unemotional traits in adolescents: Validity of the Inventory of Callous-Unemotional traits across gender, age, and community/institutionalized status. *Journal of Psychopathology and Behavioral Assessment*, 37(3), 407–421. <https://doi.org/10.1007/s10862-014-9472-8>
- Ray, J. V., Pechorro, P., & Goncalves, R. A. (2016). A comparison of self-report measures of Callous-Unemotional traits among incarcerated youth: Associations with aggression, conduct disorder, and offending behavior. *Criminal Justice and Behavior*, 43(10), 1293-

1309. <https://doi.org/10.1177/0093854815628027>

Szabó, E., Kocsel, N., Édes, A., Pap, D., Galambos, A., & Zsombók, T. (2017). Callous-unemotional traits and neural responses to emotional faces in a community sample of young adults. *Personality and Individual Differences*, *111*, 312–317.

<https://doi.org/10.1016/j.paid.2017.02.026>

Ueno, K., Ackermann, K., Freitag, C. M., & Schwenck, C. (2021). Assessing Callous–Unemotional traits in 6- to 18-year-olds: Reliability, validity, factor structure, and norms of the German version of the Inventory of Callous–Unemotional traits. *Assessment*, *28*(2), 567–584. <https://doi.org/10.1177/1073191119847766>

Viding, E., Simmonds, E., Petrides, K. V., & Frederickson, N. (2009). The contribution of callous-unemotional traits and conduct problems to bullying in early adolescence. *The Journal of Child Psychology and Psychiatry*, *4*, 471–481. <https://doi.org/10.1111/j.1469-7610.2008.02012.x>

Waller, R., Dishion, T. J., Shaw, D. S., Gardner, F., Wilson, M. N., & Hyde, L. W. (2016). Does early childhood callous-unemotional behavior uniquely predict behavior problems or callous-unemotional behavior in late childhood? *Development and Psychopathology*, *52*(11), 1805–1819. <https://doi.org/10.1002/cncr.27633>. Percutaneous

Willoughby, M. T., Waschbusch, D. A., Moore, G. A., & Propper, C. B. (2011). Using the ASEBA to screen for callous unemotional traits in early childhood: Factor structure, temporal stability, and utility. *Journal of Psychopathology and Behavioral Assessment*, *33*(1), 19–30. <https://doi.org/10.1007/s10862-010-9195-4>

Table 1*Descriptives of the Sample at Ages 3 and 13 Years*

	Age 3	Age 13
<i>N</i>	600	318
Sex (% boys)	50.4	49.8
SES (%)		
High	35.4	33.9
Middel-high/Middel	46.2	57.6
Middel-low/Low	18.4	8.5
Ethnia (%)		
Caucasian	91.1	95.2
Latino	4.7	1.5
Other	4.2	3.3
DSM-5 ODD diagnose (%)		
Boys	4.9	7.9
Girls	9.0	11.6
DSM-5 CP diagnose (%)		
Boys	1.2	0.0
Girls	0.6	0.0
ODP score – <i>M (SD)</i>		
Boys	3.3 (3.1)	3.1 (3.6)
Girls	3.2 (3.2)	1.9 (2.8)
SDQ-Conduct score – <i>M (SD)</i>		
Boys	2.0 (2.1)	1.5 (2.0)
Girls	1.7 (2.0)	0.9 (1.5)

Note. SES: Socioeconomic status. DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. ODD: Oppositional defiant disorder. CP: Conduct problems disorder. ODP: Oppositional defiant problems. SDQ: Strengths and Difficulties Questionnaire-Conduct problems.

Table 2*Descriptives of ICU Total Score at Each Follow-up Separately by Sex*

Age	Boys			Girls			Mean comparison
	<i>N</i>	<i>M (SD)</i>	<i>P80</i>	<i>N</i>	<i>M (SD)</i>	<i>P80</i>	<i>p</i>
3	313	23.0 (10.3)	32	307	19.3 (9.4)	26	< .001
4	308	22.6 (10.7)	29	301	19.0 (9.4)	27	< .001
5	281	21.8 (9.4)	30	283	17.7 (9.3)	25	< .001
6	234	23.2 (10.3)	31	230	17.7 (8.9)	24	< .001
7	232	22.4 (9.7)	31	237	17.5 (9.3)	26	< .001
8	218	22.7 (11.9)	32	211	17.6 (10.3)	25	< .001
9	226	22.0 (10.0)	30	219	18.5 (10.7)	27	< .001
10	216	23.6 (11.4)	33	213	17.2 (8.4)	23	< .001
11	224	21.5 (11.2)	30	232	17.5 (9.8)	25	< .001
12	177	22.9 (10.9)	31	188	18.6 (9.9)	27	< .001
13	158	24.3 (13.1)	36	160	18.3 (10.1)	27	< .001

Note. ICU: Inventory of Callous Unemotional Traits. *P80*: Cut-off scores for the 80th Percentile.

Table 3

AUC, Cut-off, Sensitivity and Specificity of ICU Total Score to Discriminate Trajectories of ODP Separately by Sex

Age	Boys					Girls				
	<i>N</i> High trajectory	<i>AUC</i> (%)	Cut-off	<i>Se</i> (%)	<i>Sp</i> (%)	<i>N</i> High trajectory	<i>AUC</i> (%)	Cut-off	<i>Se</i> (%)	<i>Sp</i> (%)
3	50	67.7 (62.2;72.9)	27	60.0	67.2	27	65.0 (59.4;70.3)	21	60.0	61.6
4	50	74.8 (69.5;79.5)	27	63.6	76.4	27	73.7 (68.3;78.5)	21	76.7	64.2
5	46	69.6 (63.9;74.9)	24	71.2	64.9	24	75.7 (70.3;80.6)	25	61.5	82.6
6	38	81.8 (76.2;86.5)	28	72.1	75.4	18	81.5 (75.9;86.3)	23	75.0	76.7
7	37	75.2 (69.1;80.7)	24	76.2	61.5	18	80.7 (75.1;85.5)	24	75.0	78.6
8	39	79.7 (73.7;84.9)	29	73.8	77.3	18	85.3 (79.9;89.8)	23	78.9	77.0
9	37	78.8 (72.9;84.0)	26	72.5	69.8	16	84.7 (79.3;89.2)	27	75.0	83.1
10	35	73.8 (67.3;79.6)	27	72.2	66.3	17	79.8 (73.9;84.9)	22	77.8	76.5
11	33	72.6 (66.2;78.4)	27	65.7	73.7	23	73.9 (67.8;79.4)	22	66.7	79.1
12	31	72.1 (64.8;78.7)	25	64.5	64.1	19	58.5 (51.2;65.5)	22	47.4	69.9
13	27	64.6 (56.5;72.1)	29	53.8	66.2	16	59.1 (51.1;66.8)	17	70.6	49.7
<i>M</i>		73.7	26.6	67.8	69.3		74.4	22.5	69.5	72.6

Note. *AUC*: Area under the ROC curve. *ICU*: Inventory of Callous Unemotional Traits. *ODP*: Oppositional Defiant Problems. *Se*: Sensitivity. *Sp*: Specificity.

Table 4

AUC, Cut-off, Sensitivity and Specificity of ICU Total Score to Discriminate Trajectories of SDQ-Conduct Separately by Sex

Age	Boys					Girls				
	<i>N</i> High trajectory	AUC (%)	Cut-off	<i>Se</i> (%)	<i>Sp</i> (%)	<i>N</i> High trajectory	AUC (%)	Cut-off	<i>Se</i> (%)	<i>Sp</i> (%)
3	67	68.7 (63.2;73.8)	27	60.3	69.3	31	59.7 (54.0;65.2)	21	55.3	61.6
4	66	74.8 (69.5;79.6)	26	63.4	74.8	31	68.8 (63.3;74.0)	21	65.8	63.9
5	61	67.8 (62.0;73.2)	24	65.7	65.7	29	68.1 (62.4;73.5)	25	52.9	82.8
6	48	77.7 (71.8;82.9)	27	66.1	74.2	21	83.8 (78.4;88.3)	23	80.0	78.5
7	53	72.2 (65.9;77.9)	24	68.4	62.2	22	84.3 (79.1;88.7)	24	76.9	80.4
8	49	79.9 (73.9;85.0)	27	75.9	75.6	20	83.6 (77.9;88.3)	22	78.3	76.0
9	51	74.0 (67.7;79.6)	25	65.5	67.1	19	84.0 (78.5;88.5)	23	81.8	73.1
10	46	75.3 (68.9;81.0)	27	70.8	68.7	20	81.3 (75.5;86.2)	22	78.3	77.9
11	45	73.5 (67.2;79.2)	25	69.4	69.8	25	78.8 (73.0;83.8)	20	75.0	74.4
12	37	75.2 (68.0;81.4)	22	84.6	57.5	21	64.6 (57.4;71.3)	22	54.5	71.2
13	30	72.4 (64.7;79.2)	33	54.8	80.0	16	59.0 (51.0;66.7)	18	66.7	53.5
<i>M</i>		73.7	26.1	67.7	69.5		74.2	21.9	69.6	72.1

Note. AUC: Area under the ROC curve. ICU: Inventory of Callous Unemotional Traits. SDQ: Strengths and Difficulties Questionnaire. *Se*: Sensitivity. *Sp*: Specificity.

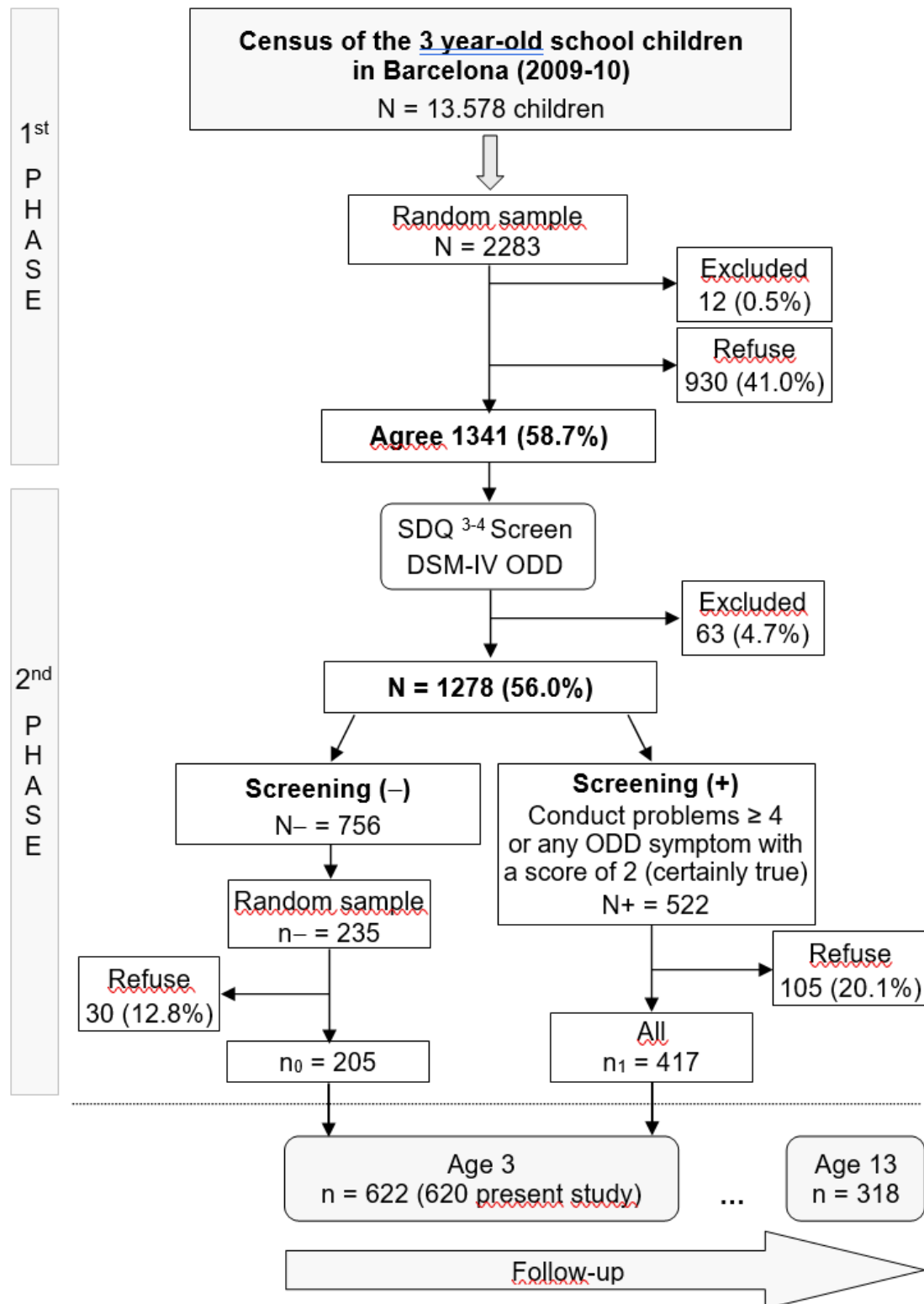
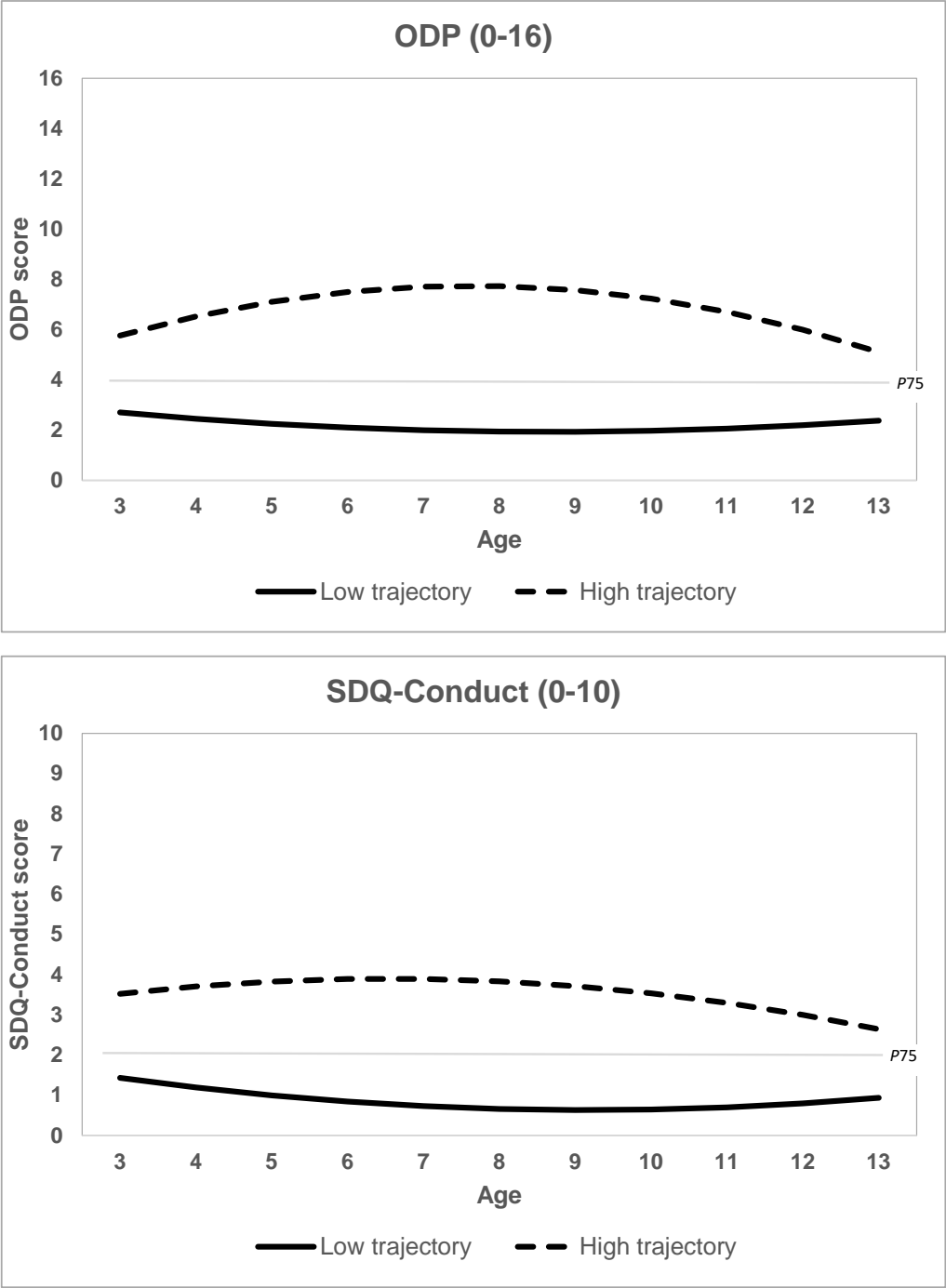
Figure 1*Study Sampling Design and Follow-ups*

Figure 2
Trajectories of ODP and SDQ-Conduct Problems Scores



Note. Each panel shows the 2 trajectories separately for each measure.
ODP: Oppositional defiant problems. SDQ: Strengths and Difficulties Questionnaire. P75: 75th Percentile