UNIVERSIDAD ALBIZU

Ciencias de la Conducta
2022, Vol 37, No. 1, 4-22

ISSN: 2380-3002

Reliability and Validity of a Brief Structured Diagnostic Measure for Depression in Youths

Confiabilidad y Validez de un Instrumento Diagnóstico Estructurado Breve para la Depresión en Adolescentes

Eduardo Cumba-Avilés, Ph.D.[1]
Martha Quiles-Jiménez, Ph.D.[1]

[1]Institute for Psychological Research, University of Puerto Rico-Río Piedras Campus

Cite as: Cumba-Avilés, E., & Quiles-Jiménez, M. (2022). Reliability and Validity of a Brief Structured Diagnostic Measure for Depression in Youths. *Ciencias de la Conducta/Behavioral Sciences Journal, 37*(1), 4-22.

Author's Note. **Corresponding author:** Dr. Eduardo Cumba-Avilés, University of Puerto Rico, School of Social Sciences, Institute for Psychological Research, 9 Ave. Universidad #901, San Juan, PR 00925-2509; e-mail: eduardo.cumba1@upr.edu **Funding:** Work on this article was possible thanks to funds granted to the Institute for Psychological Research by the Presidential Office of the University of Puerto Rico, and funds granted to the first author by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health, under award number R03DK092547. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors have no conflict of interest to disclose.

ABSTRACT

Disorders characterized by depressive symptoms/signs (DCBDS) are frequent and impairing among adolescents. Psychometrically appropriate diagnostic tools ease clinical research and practices. We assessed the psychometrics of a Brief Structured Diagnostic Measure for Depression (BSDMD) among 621 adolescents from Puerto Rico recruited since November 2010 to December 2012. They completed the Children's Depression Inventory (CDI) and the Depressive Symptoms Spectrum Assessment Inventory (DSSAI). We examined the diagnostic inter-rater reliability (IRR; Cohen's κ) of BSDMD-generated diagnoses, assessed the internal consistency (α) of its symptoms and impairment scores, and documented their validity, considering Pearson correlations with external criteria (e.g., CDI, DSAAI, and self-efficacy for depression scores, and youths' lifetime number of depressive episodes of five or more symptoms) and significant associations (examined via Student's *t*-tests and odds ratios) with a history of suicidal attempt or depression treatment, and with CDI and DSSAI cut-off criteria. We found an average diagnostic IRR of .919, alpha coefficients from .74–.89 for continuous scores, and significant associations with external criteria described for continuous and diagnostic data, as applicable. Our findings portray the BSDMD as reliable and valid when assessing DCBDS among Hispanic adolescents for the described sample.
Keywords*:* adolescent depression, diagnostic measures, Hispanics, psychometrics

UNIVERSIDAD ALBIZU

Ciencias de la Conducta

CC 4.0 2022, Vol 37, No. 1, 4-22
ISSN: 2380-3002

RESUMEN

Los trastornos caracterizados por síntomas/signos depresivos (TCPSD) son frecuentes y perjudiciales entre adolescentes. Las herramientas diagnósticas psicométricamente apropiadas facilitan las investigaciones y prácticas clínicas. Evaluamos psicométricamente el Instrumento Diagnóstico Estructurado Breve para la Depresión (IDEBD) con 621 adolescentes de Puerto Rico, reclutadas(os) entre noviembre de 2010 y diciembre de 2012. Completaron el *Children's Depression Inventory* (CDI) y el Inventario para la Evaluación del Espectro de la Sintomatología Depresiva (INEESD). Examinamos la confiabilidad diagnóstica entre evaluadores (CDEE; κ de Cohen) partiendo del IDEBD, la consistencia interna (α) de puntuaciones de síntomas e impedimento y su validez, considerando correlaciones Pearson con criterios externos (p. ej., puntuaciones del CDI, INEESD, autoeficacia para la depresión y número de episodios depresivos de cinco síntomas o más presentados en la vida) y asociaciones significativas (examinadas vía pruebas *t* de Student y *odds ratios*) con el historial suicida o de tratamiento antidepresivo y los puntos de corte del CDI e INEESD. Encontramos una CDEE promedio de .919, coeficientes alfa de .74–.89 para puntajes continuos y asociaciones significativas de datos continuos y diagnósticos con los criterios externos descritos, según aplicable. Nuestros hallazgos proyectan un IDEBD válido y confiable al evaluar TCPSD entre adolescentes hispanos/as de esta muestra.

Palabras Claves: depresión en adolescentes, instrumentos diagnósticos, origen hispano, psicometría

RELIABILITY AND VALIDITY OF A BRIEF STRUCTURED DIAGNOSTIC MEASURE FOR DEPRESSION IN YOUTHS

Major Depressive Disorder (MDD) is an impairing mental health problem. In most cases, its first episode appears around ages 13–19 years, indicating a high lifetime prevalence among adolescents, which estimate ranges between 15%–20% (Birmaher et al., 1996). The past 12- month prevalence of MDD in Puerto Rican youths aged 11–17 years was 4.42%. Another 5.25% met criteria for minor depression, a sub-threshold group with similar correlates and comorbidity (González-Tejera et al., 2005), diagnosed as Depressive Disorder Not Otherwise Specified (DDNOS) in the Diagnostic and Statistical Manual for Mental Disorders (4th ed., text revised; American Psychiatric Association; 2000). Based on the same sample, when computing MDD rates for youths 13–17 years old, researchers found an estimate of 5.8% (Costello et al., 2006).

A recent study with Puerto Rican adolescents from a large community sample showed that 21.9% had significant depression symptoms, independently from diagnostic status or impairment (Moscoso-Álvarez et al., 2020). Some of these youths will meet criteria for MDD. Yet, a large proportion of youths with symptoms of depression (even with substantial distress or impairment) will not meet the full criteria for this disorder. Many of them will meet criteria for other psychiatric disorders, such as Dysthymia, DDNOS, Adjustment Disorder with Depressed Mood (ADDM), or will present no mental disorder. As rating scales usually provide scores on the frequency/severity of depressive symptoms only, we must also have reliable and valid clinical tools to assess the full diagnostic criteria for disorders characterized by depressive symptoms (DCBDS).

Pediatric mental health diagnosis in research settings typically relies on one of two main approaches: computer-derived algorithms or clinician-based diagnoses. For decades, many researchers considered clinician-based diagnosis as a strategy that was more representative of what occurs in actual clinical settings. However, its use has been criticized for at least three reasons: (1) the diverse procedures and algorithms used during the interviews and diagnostic decision making; 2) the reduced level of agreement associated with decisions based on this disparity; and 3) the difficulties associated to achieve that all clinicians receive the same training to assure using the same procedures and diagnostic algorithms (Mellsop et al., 1982; Spitzer & Fleiss, 1974). Precisely, researchers introduced computer-derived mental health diagnoses in studies mainly to standardize the diagnostic interviews, reduce the burden and time consumption of training interviewers in psychopathology and diagnostic issues, and presumably increase diagnostic reliability (Andrews & Peters, 1998; Komiti et al., 2001; Wittchen, 1994). Yet, in some studies and most clinical settings, the use of structured diagnostic interviews may be too time-consuming, particularly when used with children or adolescents (Renou et al., 2004). Most structured diagnostic interviews for the pediatric population require an administration time that ranges from 30 to 120 minutes, depending on the population (e.g.., known patients or controls), responses to stem questions, and details assessed while ascertaining the presence of symptoms and other criteria (Giannakopoulos, 2017; Leffler et al., 2015; Neuschwander et al., 2017). This represent a considerable time burden in clinical settings and particularly in clinical studies that contain a variety of other outcome measures and assessments of inclusion and exclusion criteria. In addition,

the attention span in children and adolescents is usually shorter than in adults. This may contribute to reduce the precision of their reports when exposed to prolonged assessments (Edelbrock et al., 1985; Essau et al., 1999). Since these interviews contain many questions addressing different periods, duration of episodes, and frequencies of occurrence, some youth (particularly youngest children) may require more time to grasp time-related concepts and even some may provide answers without truly understanding most questions. This lack of understanding may be more frequent when assessing depressive disorders (Breton et al., 1995).

Most commonly used diagnostic instruments include modules to assess depressive disorders. Some instruments, like the Diagnostic Interview Schedule for Children-IV (DISC-IV; Shaffer et al., 2000) and the MINI International Neuropsychiatric Interview for Children and Adolescents (MINI-KID; Sheehan et al., 2010), have separate sections for Dysthymia (currently known as Persistent Depressive Disorder) and MDD. However, many modules designed for assessing depressive disorders via diagnostic instruments, and used in epidemiological or clinical studies (including clinical trials), do not fully consider exclusion criteria. This was the case with youths in the island-wide representative sample from the most recent pediatric epidemiological study conducted in Puerto Rico (Canino et al., 2004). In the sample of that study, "the diagnoses reported correspond to last year prevalence, obtained without implementing exclusionary criteria related to other DSM-IV diagnoses" (González-Tejera et al., 2005; p. 290). In addition, many of these instruments lacked specific questions to rule out mood disorders due to medical illness or substance-induced disorders (e.g., induced by medications, alcohol, or drugs).

Since computer-generated diagnoses were designed for epidemiologic studies, the assessment period used in these diagnostic interviews is typically oriented to identify symptoms of pediatric disorders presented at least once in a lifetime, at some point during the last year or at some time during the last month (e.g., Shaffer et al., 2000). When considering MDD, for example, its diagnostic criteria require assessing the presence of symptoms during 2 weeks. Because many available diagnostic tools for children and adolescents designed to generate MDD diagnoses by computer algorithms do not include questions to assess the presence of symptoms during the last 2 weeks, they do not provide accurate information about the presence of a current MDD in youths. This limitation could affect conclusions of epidemiological studies since researchers identify as correlates of depressive disorders some variables that they assessed using a time frame that met the definition of "current" while using for the disorders a time frame that did not. This issue has also repercussions for studies in which a current diagnosis of a depressive disorder is an inclusion criterion (i. e., clinical trials), especially if researchers considered the presence of the disorder an outcome measure but the disorder was not truly present at baseline.

Assessment of depressive symptoms and DCBDS in youths requires measures with appropriate psychometric properties. The most valid and reliable measures available for assessing depression among children and adolescents living in Puerto Rico are rating scales (Cumba-Avilés & Feliciano-López, 2013). Spanish versions of modules that assess DCBDS included in a couple of diagnostic interviews (e.g., the DISC-IV and the MINI-KID) were used for clinical trials on adolescent depression in Puerto Rico (Bernal et al., 2019; Cumba-Avilés & Sáez-Santiago, 2016;

Cumba-Avilés, 2017). However, when using these modules, clinical evaluators have to make additional probes either to assess diagnostic criteria in the past 2 weeks (i.e., if using the DISC-IV) or to evaluate exclusion criteria properly. These limitations in diagnostic interviews available in Spanish to assess DCBDS in Puerto Rico result in an extension of the assessment time. There is a need for diagnostic measures for assessing adolescents that collect all the information required to conduct a diagnosis of DCBDS (including exclusion criteria) while keeping a brief assessment time. Validation of such a brief diagnostic measure could facilitate the appropriate identification of cases in school, clinical, and research settings. Identification of these cases will ease their appropriate and early referral for treatment.

In this study, we aimed to assess the reliability and validity of a brief measure to assess criteria for DCBDS among adolescents from Puerto Rico. Specifically, we assessed the diagnostic reliability of the measure by examining agreement between two clinicians who classified cases in our sample based on diagnostic data collected with this measure. We also compared the prevalence of disorders obtained through decisions of the two clinicians and against rates published in the adolescent research literature. In addition, we assessed the internal consistency of symptoms scores and impairment scores obtained in the measure and provided initial evidence on their validity. Finally, we documented the validity of diagnoses generated through the new diagnostic measure by examining its association with several external criteria, including cut-off points from scales validated for adolescents from Puerto Rico.

Our first hypothesis was that raw agreement between clinicians would be of 95% or above and that agreement corrected by chance (Cohen's kappa; Cohen, 1960) would be .81 or above for all diagnostic categories, which represents almost perfect to perfect agreement (Landis & Koch, 1977). We also expected that symptoms scores and impairments scores would show an internal consistency (as measured by Cronbach's alpha) of .70 or above and significant correlations (mostly moderate, according to Champion, 1981) with validity criteria. Finally, we hypothesized that diagnoses generated through the brief measure will be significantly associated with external criteria such as a history of suicidal attempts and a history of any depression treatment, as well as with cut-off points of two validated rating scales.

METHOD

Participants were 621 youth (64.09% girls) aged 12–18 years (M = 15.07; SD = 1.56). About 92.27% (n = 573) were Puerto Ricans, 5.15% (n = 32) were Dominicans and the rest were from other Hispanic groups. They coursed junior-high (n = 326) and high-school grades in 23 public schools (n = 332) from San Juan districts and 19 private schools from four municipalities of the San Juan Metropolitan area. They must understand Spanish, and should not evidence any neurological, sensory, other cognitive or physical problems that could hinder participation.

Primary caregivers completed the Socio-Demographic Data Form (SDF). Most children lived in Metropolitan (89.53%) and urban (80.19%) areas. A 37.36% (232) lived in households with biological/foster parents who were married. In other 7.57% (47) of cases, they just lived together. Other youth lived in household with either divorced (n = 172), separated (n = 127), widowers (n = 27) or single parents (n = 16), who were never married nor lived together. A woman

was the primary caregiver in 601 of the homes. Most of them reported that their family belonged to an upper-middle (36.39%) or lower-middle (54.91%) socioeconomic status (SES). Most primary caregivers reported having a full-time job (59.74%) or a part-time one (27.54%). Their mean schooling was 14.75 years (*SD* = 2.98). About 49.11% had never achieved a bachelor's degree. Mean caregivers' age was 42.90 years (*SD* = 7.05; range was 28– 72). The mean household size was 3.96 members (*SD* = 1.21; range was 2–9).

Measures

Children´s Depression Inventory (CDI)

This scale measures depressive symptoms in youth aged 7–17 years for the past 2 weeks. Its 27 items provide three options scored as 0, 1, or 2. We used the Spanish CDI distributed by Multi-Health Systems (Kovacs, 2001). Its internal reliability in this sample was .86.

Depressive Symptoms Spectrum Assessment Inventory (DSSAI)

The DSSAI (INEESD by its Spanish acronym) is a self-report measure of depression in youth aged 12 years and older. It provides Total scores for the last 2 weeks (L2W) and the last 6 months (L6M). The measure contains 120 items in a Likert-type format with options from 0 (*Never or almost never*) to 3 (*Very frequently*). These items classify into 10 clinical subthemes within the spectrum of depressive symptoms. A preliminary report (*N* = 201) of the psychometric properties of the DSSAI Total scales and subscales is available in Feliciano-López and Cumba-Avilés (2014). Psychometric properties for the final sample (*N* = 621) were very similar to those obtained in the initial report (Cumba-Avilés & Feliciano-López, 2015). The internal consistency of both DSSAI Total scores was .98. Test-retest values were above .85 for both Total scales. Its correlation with CDI scores and other external criteria documented its validity.

Brief Structured Diagnostic Measure for Depression (BSDMD)

The BSDMD assesses diagnostic criteria for MDD, Dysthymia, or DDNOS ever in a lifetime and during the most recent episode. Besides specific DSM symptoms, this instrument assesses exclusion criteria, impairment, frequency of episodes, age of onset and age at the latest symptom occurrence, the duration of the first and the most recent episode, lifetime history of suicide attempt, and history of depression treatment, among other relevant data. Using data collected through the BSDMD, clinicians can derive clinical diagnoses for DCBDS in the past 2 weeks, the past 6 months, the past 12 months, and anytime in a lifetime. Questions in this measure also allow for a distinction between Adjustment Disorder (current, last-year, or lifetime) and depressive disorders or between the latter and no disorder. In addition to questions aimed to assess criteria for DCBDS, the BSDMD includes two sets of questions to assess perceived self-efficacy for depression in the past 2 weeks and the past 6 months. For each time frame, and using a scale from 0 to 10, adolescents rated their confidence on their ability to manage situations and symptoms faced when depressed by their own and their confidence in their ability to ask others for help to deal with them.

Suicide Risk Interview for Adolescents (FERSA by its Spanish acronym)

We used the FERSA at interviews to assess the lethality of suicidal ideation/behavior. The development of the FERSA and details regarding the protocol for suicide risk assessment used in

this study are described elsewhere (Cumba-Avilés & Feliciano-López, 2013).

Procedure

After approval by the university IRB (#0910-111), we met with school personnel to explain study procedures. We obtained the authorization of the Puerto Rico Department of Education for collecting data in public schools. Directors/executive directors of private schools provided authorization for their particular institutions. At informative meetings with students, we explained study procedures and gave them an envelope with the SDF, Consent/Assent Forms, and informational sheets. We asked youths to deliver the envelope to their guardians to authorize their participation and complete the SDF. Teens signed the forms if assented to participate and handed over documents in the envelope to school staff, who called research staff to pick them up at school. We scheduled informative meetings and assessment dates in coordination with school staff. Assessment sessions were in a self-report format and lasted about 60 minutes. As in our pilot study (Cumba-Avilés & Feliciano-López, 2013), we conducted in-depth interviews for risk assessment with youths who reported suicidal ideation. We instructed those with depressive symptoms and no suicidal ideation to ask for specific help from a mental health professional. We provided a document with examples of mental health care providers' contact information.

Data Analyses

We conducted all statistical analyses with SPSS 27.0. Two licensed clinicians used data collected through the BSDMD to achieve separate diagnostic decisions per case on several diagnostic categories. These were current MDD, past 6 months MDD, last-year MDD, lifetime MDD, current DDNOS, past 6 months DDNOS, last-year DDNOS, lifetime DDNOS, current Dysthymia, lifetime Dysthymia, last 6 months ADDM, last-year ADDM, lifetime ADDM, and none of these disorders. Clinicians also rated their decisions (Yes / No) in summary categories applicable to MDD, DDNOS, and Dysthymia, as follows: Any Current Depressive Disorder, Any Past 6 Months Depressive Disorder, Any Last-Year Depressive Disorder, and Any Lifetime Depressive Disorder. We computed frequencies and percentages based on diagnostic decisions made by each clinician to estimate their reported prevalence for each category of disorders. Using Cohen's kappa (κ) coefficient, we assessed the inter-rater reliability (IRR) of diagnostic decisions made by these clinicians. Based on the asymptotic standard error, we also provided a 95% Confidence Interval (CI) for each κ coefficient. After identifying cases and diagnostic categories in which disagreements occurred, we discussed and resolved each of them. Resolution of disagreements could result in support for the decision of any individual rater or a decision different from the originally proposed by any rater. We estimated a key code for diagnostic decisions (final decisions) after resolving disagreements. Next, we estimated another set of κ coefficients, to report then the level of agreement between each rater and the key code.

We also analyzed the reliability, inter-correlations, and validity of other elements of the BSDMD. For example, we estimated Cronbach's alpha (α) coefficients for scores in these areas: number of symptoms presented in the first depressive episode (FE; 0 to 9), number of symptoms presented in the most recent depressive episode (MRE; 0 to 9), and number of impaired areas attributable to depressive symptoms (0 to 6). We used Pearson product-moment coefficient to

assess the inter-correlation between the impairment score and symptoms score of the FE and MRE. We examined the concurrent validity of impairment scores and symptoms scores from the FE and the MRE by estimating their correlations with the following criteria: CDI scores, DSSAI scores, and self-efficacy for depression scores. In addition, we further examined the validity of symptoms scores and impairment scores by comparing the means obtained in these variables by groups defined by a history of suicidal attempt or history of depression treatment, using independent sample *t*-tests and Hedges' *g* for estimating effect sizes.

To assess the validity of the final diagnostic decisions against external criteria, we used two strategies. First, we used odds ratio (*OR*) to examine the association of diagnostic categories with a lifetime history of suicidal attempts or history of depression treatment. Second, we estimated *OR*s to test the association of diagnostic categories with meeting cut-off criteria from the CDI, the DSSAI-Total score for the L2W, or at least one of these criteria. For all *OR*s, we estimated bias-corrected 95% CIs with a bootstrapping procedure of 1000 samples. In Table 1, we summarized all the statistical analyses conducted and provided a brief description of them, as well as about our hypotheses for specific analyses.

RESULTS

Inter-Rater Agreement in Prevalence and Classification of Cases by Diagnostic Category

Based on reports from Raters 1 and 2, about 25% of the sample had a lifetime history of Any Depressive Disorder (Table 2). The rate of Any Last-Year Depressive Disorder was around 20% and the prevalence of Any Last 6 Months Depressive Disorder was 17.71%. According to Rater 1, about 14.65% of the adolescents met criteria of Any Current Depressive Disorder, while Rater 2 estimated that this occurred in 13.69% of the cases. Inter-rater raw agreement in these categories ranged from 97.75% (Any Current Depressive Disorder) to 99.84% (Any Lifetime Depressive Disorder). Kappa coefficients for these categories ranged from .907 to .996 ($p \leq .001$).

In general, the level of inter-rater agreement for specific diagnostic categories, and the degree of similarity in prevalence rates by category, based on data collected in the BSDMD, was excellent. In fact, except for the Last-Year MDD category, in all other cases, the prevalence rates based on their separate diagnostic decisions differed less than 1% one from another. Unadjusted (raw) agreement for ADDM categories was over 99% with κ coefficients from .874 (last year) to .932 (last 6 months). Unadjusted agreement for specific categories of primary depressive disorders reflected values from 98.07% (current MDD) to 99.84% (current Dysthymia). Adjusted agreement for specific categories of these disorders was excellent, as reflected by κ coefficients that ranged from .854 (lifetime Dysthymia) to .960 (last 6 months MDD). Considering both specific and global diagnostic categories, the average adjusted IRR was .919. We did not find in any case within this sample a lower bound in the 95% CI of the κ value that was below .70.

Agreement Between Raters and Key Codes

As expected, when comparing each rater's decisions against key codes, we found that both raters kept a high level of similarity with final diagnostic decisions (Table 3). Specifically, the raw agreement between Rater 1 and key codes ranged from 99.03% to 100%, while the agreement

Table 1

Description, Interpretation, and Expected Findings of Statistics or Analytical Strategies Used

Statistic or Strategy	Definition or Purpose	Interpretation	Hypothesis
Raw agreement	It refers to the percent of cases in which raters agreed, ignoring any adjustments for agreement expected by chance.	An agreement of 90% or higher is strong; 80% is acceptable; in most cases < 80% is unacceptable.	At least 95%
Kappa coefficient (κ)	An indicator of the level of inter-rater agreement that corrects for the agreement expected by chance. It can assume values ranging from -1.00 to 1.00.	< .00 (Poor) .00 – .20 (Slight) .21 – .40 (Fair) .41 – .60 (Moderate) .61 – .80 (Substantial) .81 – 1.0 (Almost perfect)	Coefficients of .81 or above
Cronbach's alpha (α)	It is a measure of the closeness of the relationship among items in a scale. Alpha describes the extent to which all the items in a test measure the same concept or construct.	< .50 (Unacceptable) .50 – .59 (Poor) .60 – .69 (Questionable) .70 – .79 (Acceptable) .80 – .89 (Good) .90 – 1.0 (Excellent)	Values ≥ .70 for all summative scores (either scales or subscales)
Student *t*-test (*t*)	Used to compare the mean scores of two independent samples or groups.	A significant test suggests differences between the means beyond those expected by chance.	Significantly higher means for HSA and HDT groups
Hedges *g*	A standardized measure of the size of the difference between means, corrected for the total sample size.	0.20 = Small effect size 0.50 = Medium effect size 0.80 = Large effect size	Effect sizes will be large or medium.
Pearson's Correlation (*r*)	An indicator of how strong is the association between two continuous variables. It can assume values ranging from -1.00 to 1.00.	Absolute values of: .00 – .25 (Low) .26 – .50 (Moderate-low) .51 – .75 (Moderate-high) .76 – 1.0 (High)	Mostly moderate correlations (in absolute values)
Odds Ratio (*OR*)	A measure of association between two dichotomous variables; it is the quotient of the odds of occurrence of an event when a criterion is present vs. when the criterion is absent.	When *OR* = 1.0, an event is equally likely to occur if the criterion is absent or present. A significant OR has upper and lower limits in its 95% CI that does not include the 1.0 value.	Associations examined will result in significant ORs in most cases, except for ADDM.
Bootstrapping correction	A resampling technique used to estimate statistics on a population sampling a dataset with replacement. It produces many random samples based on data.	Provides a bias-corrected 95% CI for the statistics estimated, based on calculations from 1000 random samples.	N/A

Note. CI = Confidence interval; ADDM = Adjustment disorder with depressed mood; HSA = History of suicide attempt; HDT = History of depression treatment.

Table 2

Inter-Rater Reliability and Prevalence per Diagnostic Category Based on DSM-IV-TR Criteria

Diagnostic Category	Prevalence According to Rater 1 % (*f*)	Prevalence According to Rater 2 % (*f*)	Rater 1 vs. Rater 2 Unadjusted Agreement % (*f*)	Rater 1 vs. Rater 2 κ (95% IC)
Lifetime MDD	16.43 (102)	15.78 (98)	99.36 (617)	.976 [.952 – 1.00]
Last-year MDD	13.37 (83)	12.08 (75)	98.39 (611)	.928 [.882 – .974]
Last-6M MDD	11.59 (72)	11.11(69)	99.19 (616)	.960 [.924 – .996]
Current MDD	9.98 (62)	9.34 (58)	98.07 (609)	.889 [.825 – .953]
Lifetime Dysthymia	2.58 (16)	1.93 (12)	99.36 (617)	.854 [.710 – .998]
Current Dysthymia	0.97 (6)	0.81 (5)	99.84 (620)	.908 [.726 – 1.00]
Lifetime DDNOS	7.89 (49)	8.53 (53)	98.71 (613)	.915 [.855 – .975]
Last-year DDNOS	6.12 (38)	6.76 (42)	98.71 (613)	.893 [.819 – .967]
Last 6M DDNOS	5.15 (32)	5.80 (36)	99.03 (615)	.907 [.831 – .983]
Current DDNOS	3.70 (23)	3.54 (22)	99.19 (616)	.885 [.783 – .987]
Lifetime ADDM	5.31 (33)	4.67 (29)	99.03 (615)	.898 [.816 – .980]
Last-year ADDM	3.70 (23)	2.90 (18)	99.19 (616)	.874 [.762 – .986]
Current ADDM	2.58 (16)	2.25 (14)	99.68 (619)	.932 [.836 – 1.00]
Any Lifetime DD	25.76 (160)	25.60 (159)	99.84 (620)	.996 [.988 – 1.00]
Any Last-year DD	20.45 (127)	19.65 (122)	97.91 (608)	.935 [.899 – .971]
Any Last 6M DD	17.71 (110)	17.71 (110)	99.03 (615)	.967 [.941 – .993]
Any Current DD	14.65 (91)	13.69 (85)	97.75 (607)	.907 [.859 – .955]

Note. Global categories for any depressive disorder do not include adjustment disorders. An equal amount of agreements could result in different kappa coefficients depending on the diagnosis base rate and the distribution of disagreements. All coefficients are significant at $p \leq .001$. DSM-IV-TR = Diagnostic and Statistical Manual for Mental Disorders, 4th ed., text revised; Rater 1 = First author; Rater 2 = Second author; CI = Confidence interval; κ = Cohen's kappa; MDD = Major Depressive Disorder; DD = Depressive Disorder; DDNOS = Depressive Disorder Not Otherwise Specified; ADDM; Adjustment Disorder with Depressed Mood; 6M = 6 months.

Table 3

Level of Agreement between Individual Raters and Key Codes per Diagnostic Category

Diagnostic Category	Rater 1 vs. Key Codes UA % (*f*)	Rater 2 vs. Key Codes UA % (*f*)	Rater 1 vs. Key Codes κ (95% CI)	Rater 2 vs. Key Codes κ (95% CI)
Lifetime MDD	99.84 (620)	99.52 (618)	.994 [.982 – 1.00]	.982 [.962 – 1.00]
Last-year MDD	99.68 (619)	98.71 (613)	.986 [.966 – 1.00]	.941 [.899 – .983]
Last-6M MDD	99.68 (619)	99.68 (618)	.984 [.966 – 1.00]	.976 [.948 – 1.00]
Current MDD	99.36 (617)	98.71 (613)	.964 [.928 – 1.00]	.926 [.874 – .978]
Lifetime Dysthymia	99.68 (618)	99.84 (620)	.894 [.772 – 1.00]	.959 [.877 – 1.00]
Current Dysthymia	99.84 (620)	100.00 (621)	.908 [.726 – 1.00]	1.00 [1.00 – 1.00]
Lifetime DDNOS	99.36 (617)	99.36 (617)	.957 [.915 – .999]	.959 [.917 – 1.00]
Last-year DDNOS	99.03 (615)	99.68 (619)	.918 [.852 – .984]	.974 [.938 – 1.00]
Last 6M DDNOS	99.68 (618)	99.68 (618)	.951 [.895 – 1.00]	.954 [.902 – 1.00]

Current DDNOS	99.19 (616)	100.00 (621)	.885 [.783 – .987]	1.00 [1.00 – 1.00]
Lifetime ADDM	99.84 (620)	99.36 (617)	.984 [.952 – 1.00]	.932 [.864 – 1.00]
Last-year ADDM	99.36 (617)	99.84 (620)	.901 [.803 – .999]	.972 [.916 – 1.00]
Current ADDM	99.84 (620)	99.84 (620)	.967 [.901 – .1.00]	.965 [.895 – 1.00]
Any Lifetime DD	100.00 (621)	99.84 (620)	1.00 [1.00 – 1.00]	.996 [.988 – 1.00]
Any Last-year DD	99.19 (616)	98.71 (613)	.975 [.953 – .997]	.960 [.932 – .998]
Any Last 6M DD	99.68 (619)	99.36 (617)	.989 [.973 – 1.00]	.978 [.956 – 1.00]
Any Current DD	99.03 (615)	98.71 (613)	.961 [.929 – .993]	.947 [.909 – .985]

Note. Global categories for any depressive disorder do not include adjustment disorders. An equal amount of agreements could result in different kappa coefficients depending on the diagnosis base rate and the distribution of disagreements. All coefficients are significant at $p \leq .001$. Rater 1 = First author; Rater 2 = Second author; UA = Unadjusted agreement; CI = Confidence interval; κ = Cohen's kappa; MDD = Major Depressive Disorder; DD = Depressive Disorder; DDNOS = Depressive Disorder Not Otherwise Specified; ADDM; Adjustment Disorder with Depressed Mood. 6M = 6 months.

between Rater 2 and key codes reflected values from 98.71% to 100%. Kappa coefficients for Rater 1 against key codes ranged from .885 (current DDNOS) to 1.00 (Any Lifetime Depressive Disorder), with an average of .954. These coefficients ranged from .926 (Current MDD) to 1.00 (Current Dysthymia and Current DDNOS) if comparing Rater 2 and final decisions, with an average of .966. In no case, the lower bound in the 95% CI was below .70.

Internal Consistency, Inter-Correlations, and Validity of Continuous Scores

Cronbach's alpha for the symptoms score in the first depressive episode (FE) was .85, a good one according to George and Mallery (2003). This coefficient was .89 for the symptoms score in the most recent episode (MRE). The internal consistency of the score obtained by adding the number of impaired areas (NIA) was .74. According to Champion (1981), we found high (.76) and moderate-high (.58) correlations between the NIA score and symptoms score for the FE and MRE, respectively ($p \leq .001$). Supporting their validity, NIA scores correlated positively in a moderate-low magnitude with Total scores in the CDI (.47), positively and in a moderate-low or moderate-high magnitude with DSSAI-L2W (.49) and DSSAI-L6M (.51), respectively, and negatively and in a low magnitude (-.23) with scores of self-efficacy for depression (SED) in the past 2 weeks. All these correlations, however, were highly significant ($p \leq .001$). The same applied to symptoms scores in the FE and MRE, with correlations of .53 and .51, respectively, with the CDI scores (which were moderate-high, according to Champion, 1981), and associations of -.24 (low) and -.27 (moderate-low), respectively, with ratings of SED ($p \leq .001$). FE and MRE symptoms scores also correlated positively ($p \leq .001$) with Total scores of the DSSAI-L2W (.53 and .51, respectively) and the DSSAI-L6M (.56 and .52, respectively), and with the lifetime number of depressive episodes of five or more symptoms reported by youth (.53 and .63, respectively). All these coefficients were moderate-high in magnitude according to Champion (1981). The association of NIA scores with the latter variable was .46 ($p \leq .001$), which is a moderate-low correlation.

As shown by *t*-tests results (equal variances not assumed) in Table 4, youth with a history of suicide attempt (HSA) and those with a history of depression treatment (HDT) obtained significantly higher symptoms scores for the FE and the MRE, as well as higher NIA scores than their counterparts. Effects sizes (Hedges *g*; see Hedge & Olkin, 1985) for those differences were large for the group comparison based on HSA and medium for the comparison based on HDT.

Relationship Between Diagnostic Decisions Based on the BSDMD and External Criteria

We also examined the association of diagnostic categories with a lifetime HSA and HDT (Table 4). Except for the category that combined cases of Dysthymia and DDNOS, the odds for meeting criteria for diagnoses assessed with the BSDMD were significantly higher among youths with HSA than among their counterparts, with ORs ranging from 4.19 to 6.32 ($p \leq .001$). This means that adolescents with HSA were between just over four and more than six times more likely to meet criteria for MDD or any depressive disorder, considering four different timeframes. On the other hand, adolescents with HDT showed significantly higher odds of meeting criteria for all diagnostic categories in Table 4, with ORs ranging from 2.16 ($p \leq .05$) to 3.90 ($p \leq .001$). This means that adolescents with HDT were between just over two to almost four times more likely to meet criteria for MDD or any depressive disorder, considering the four different timeframes examined, as well as DDNOS or Dysthymia ever in a lifetime.

Finally, we tested if there was any association between diagnostic categories and cut-off criteria from the CDI (Criterion A), the DSSAI-Total score for the L2W (Criterion B), or meeting any of these criteria (Criterion C). As shown in Table 5, all diagnostic categories allusive to primary depressive disorders were significantly associated with higher odds of meeting any of the cut-off criteria assessed, but this was not the case for ADDM. *OR*s range from 2.16 ($p \leq .05$) to 22.13 ($p \leq .001$) for Criterion A, from 2.32 ($p \leq .01$) to 21.96 ($p \leq .001$) for Criterion B, and from 2.51 ($p \leq .01$) to 33.39 ($p \leq .001$) for Criterion C.

DISCUSSION

We examined the reliability and validity of the BSDMD to assess criteria for DCBDS in adolescents. First, we assessed its diagnostic reliability by examining agreement between two clinicians who classified cases based on BSDMD data. Consistent with our first hypothesis, the raw agreement was over 95%, specifically, from 97.75% to 99.84%, across diagnostic categories. As expected, IRR corrected by chance was excellent, with κ values of .854–.996. These values compared favorably with IRR coefficients for depressive disorders (0.79 to 1.00) as assessed with the MINI-KID (Sheehan et al., 2010) and other Spanish versions of recognized diagnostic measures (Ulloa et al., 2006). The prevalence of disorders based on clinicians' decisions was consistent with published rates for adolescents. For instance, this was the case for lifetime Any Depressive Disorder (Kessler et al., 2001), lifetime (Birmaher et al., 1996) and last-year MDD (Center for Behavioral Health Statistics and Quality, 2016; Costello et al., 2006), but also for lifetime Dysthymia (Waslick et al., 2003), and last-year DDNOS (González-Tejera et al. 2005).

Second, we assessed the internal consistency of symptoms scores and impairment scores obtained in the BSDMD and provided initial evidence on their validity. As expected, we found a Cronbach's alpha of .70 or above (specifically, values from .74 to .89) for these scores. We also

Table 4

Association of External Criteria with Diagnostic, Impairment and Number of Symptoms Information Provided in the BSDMD

Variable	Hx of Suicide Attempt[a] (*n* = 56)	No Hx of Suicide Attempt (*n* = 565)	*t* (*ES*)/ *OR* [95% CI]	Hx of Depression Treatment[b] (*n* = 87)	No Hx of Depression Treatment (*n* = 534)	*t* (*ES*)/ *OR* [95% CI]
Number of impaired areas	1.86 (1.66)	0.70 (1.24)	5.08** (0.93)	1.68 (1.62)	0.66 (1.24)	5.58*** (0.79)
Number of symptoms in FE	4.02 (3.04)	1.66 (2.26)	5.64*** (1.00)	3.40 (2.78)	1.63 (2.28)	5.65*** (0.75)
Number of symptoms in MRE	2.59 (3.14)	0.77 (1.74)	4.27*** (0.95)	2.01 (2.75)	0.76 (1.76)	5.60*** (0.65)
Any Depressive-Lifetime	55.36% (31)	22.83% (129)	4.19*** [2.43 – 7.61]	51.72% (45)	21.54% (115)	3.90*** [2.48 – 6.56]
Any Depressive-Past Year	50.00% (28)	17.35% (98)	4.77*** [2.54 – 8.58]	37.9% (33)	17.4% (93)	2.90*** [1.72 – 4.76]
Any Depressive-L6M	46.43% (26)	14.51% (82)	5.11*** [2.89 – 8.76]	32.2% (28)	15.0% (80)	2.69*** [1.60 – 4.39]
Any Depressive-Current	41.07% (23)	11.68% (66)	5.27*** [2.89 – 9.23]	24.1% (21)	12.7% (68)	2.18** [1.20 – 3.67]
MDD-Lifetime	42.86% (24)	13.63% (77)	4.75*** [2.50 – 8.76]	32.2% (28)	13.7% (73)	3.00*** [1.78 – 5.11]
MDD-Past Year	39.29% (22)	10.44% (59)	5.55*** [2.96 – 10.16]	26.44% (23)	10.86% (58)	2.95*** [1.68 – 5.09]
MDD-L6M	37.50% (21)	8.67% (49)	6.32*** [3.31 – 11.47]	21.84% (19)	9.55% (51)	2.65** [1.46 – 4.73]
MDD-Current	32.14% (18)	7.79% (44)	5.61*** [2.82 – 10.45]	17.24% (15)	8.80% (47)	2.16* [1.01 – 3.97]
DDNOS/Dysthymia-Lifetime	12.50% (7)	9.91% (56)	1.30 [0.44 – 2.58]	20.69% (18)	8.43% (45)	2.83*** [1.47 – 5.16]

Note. We estimated bias-corrected CIs using a bootstrapping with 1000 samples. Hx = History; FE = First depressive episode; MRE = Most recent depressive episode; MDD = Major Depressive Disorder; DDNOS = Depressive Disorder Not Otherwise Specified; BSDMD = Brief Structured Diagnostic Measure for Depression; CI = Confidence interval; L6M = Last 6 months; ES = Hedges *g*.

[a] Degrees of freedom for *t*-tests (equal variances not assumed) were 61.29, 61.18, and 58.41, respectively, for variables in the first three rows. [b] Degrees of freedom for *t*-tests were 102.22, 105.74, and 97.81, respectively, for variables in the first three rows.

*$p \leq .05$; **$p \leq .01$; ***$p \leq .001$.

Table 5

Patterns of Association (Odds Ratio) between Depressive Symptoms Cut-off Points and Diagnoses Generated with the BSDMD

Variable	**Criterion A** CDI ≥ 13 (*n* = 123) vs. < 13 (*n* = 498) *OR* [95% CI]	**Criterion B** DSSAI-L2W ≥ P84 (*n* = 100) vs. < P84 (*n* = 521) *OR* [95% CI]	**Criterion C** A or B Present (*n* =148) vs. Both Absent (*n* = 473) *OR* [95% CI]
Any Depressive-Lifetime	7.23*** [4.65 – 11.30]	8.82*** [5.44 – 14.58]	7.48*** [4.98 – 11.92]
Any Depressive-Past Year	8.45*** [5.69 – 13.83]	9.75*** [6.06 – 15.24]	8.93*** [5.78 – 14.25]
Any Depressive-L6M	11.19*** [7.14 – 18.50]	13.01*** [8.14 – 22.16]	12.91*** [8.29 – 21.63]
Any Depressive-Current	15.40*** [9.03 – 27.17]	20.26*** [12.53 – 36.52]	21.44*** [12.32 – 42.72]
MDD-Lifetime	8.41*** [5.04 – 13.81]	9.80*** [5.88 – 15.72]	8.98*** [5.73 – 15.51]
MDD-Past Year	11.05*** [6.56 – 19.80]	11.02*** [6.53 – 18.47]	11.72*** [7.41 – 20.91]
MDD-L6M	15.04*** [8.81 – 29.36]	14.96*** [8.36 – 20.10]	18.06*** [10.29 – 36.96]
MDD-Current	22.13*** [12.61 – 44.75]	21.96*** [11.77 – 44.47]	33.39*** [17.01 – 100.49]
DDNOS/Dysthymia-Lifetime	2.06* [1.04 – 3.62]	2.32** [1.17 – 4.27]	2.51** [1.42 – 4.40]
DDNOS/Dysthymia-Past Year	2.17* [1.09 – 4.10]	2.57** [1.14 – 4.84]	2.53** [1.33 – 4.66]
DDNOS/Dysthymia-L6M	2.87** [1.35 – 6.17]	3.37*** [1.54 – 6.90]	3.52*** [1.69 – 7.40]
DDNOS/Dysthymia-Current	3.48** [1.51 – 8.41]	5.41*** [2.21 – 12.34]	5.09*** [2.34 – 13.66]
ADDM-Lifetime	0.56 [0.12 – 1.38]	0.33 [0.08 – 1.42]	0.58 [0.13 – 1.34]
ADDM-Past Year	0.75 [0.19 – 2.24]	0.28 [0.18 – 1.11]	0.59 [0.15 – 1.65]
ADDM-L6M (Current)	1.01 [0.26 – 3.40]	0.37 [0.25 – 1.69]	0.79 [0.22 – 2.86]

Note. We estimated bias-corrected CIs using a bootstrapping with 1000 samples. BSDMD = Brief Structured Diagnostic Measure for Depression; CDI = Children´s Depression Inventory; DSSAI; Depressive Symptoms Spectrum Assessment Inventory; CI = Confidence interval; MDD = Major Depressive Disorder; DDNOS = Depressive Disorder Not Otherwise Specified; ADDM = Adjustment Disorder with Depressed Mood; L2W = Last 2 weeks; L6M = Last 6 months; P84 = 84th percentile.

*$p \leq .05$; **$p \leq .01$; ***$p \leq .001$.

found significant correlations (mostly moderate and in the expected direction) with concurrent validity criteria such as Total scores in the rating scales (e. g., CDI, DSSAI-L2W, and DSSAI-L6M), scores of self-efficacy for depression in the past 2 weeks, and the lifetime number of depressive episodes of five or more symptoms. In addition, BSDMD continuous scores were significantly associated with external criteria such as HSA and HDT (Table 4). Thus, the combined evidence on the internal reliability, concurrent validity, and criterion-related validity of continuous scores yielded by the BSDMD was more than adequate.

In third place, we documented the validity of diagnoses generated through the BSDMD against criteria from youth clinical history. Consistent with our hypothesis, diagnostic categories were significantly associated with external criteria such as HSA and HDT (Table 4). Only the category that combined cases of Dysthymia and DDNOS showed a non-significant association with HSA, which is consistent with having (in many cases) milder forms of depression and with the absence of suicidality among diagnostic criteria for Dysthymia. The association of diagnostic categories with HSA followed closely the expected pattern of higher *OR*s for MDD diagnostic categories over Any Depressive Disorder categories. On the other hand, the associations of diagnostic categories with HDT followed the expected descending pattern as the period of the category came closer to a "current" disorder. The fact that some cases of current depression did not have a referral for treatment yet, as compared to cases whose depression was present since earlier, might help to explain the latter.

Finally, we also documented the validity of diagnoses generated through the BSDMD against cut-off points of validated rating scales. Specifically, diagnostic categories allusive to primary depressive disorders were significantly associated with higher odds of meeting any of the cut-off criteria assessed in the CDI, the DSSAI, or any of them, but (as expected) this was not the case for ADDM categories (Table 5). Consistent with the nature of the rating scale scores, as the time used to define the diagnostic category approximated the "current" period, the values of the *OR*s increased. The latter was true for MDD, Dysthymia/DDNOS, and Any Depressive Disorder. Taken together, the associations documented for BSDMD-generated diagnoses with criteria from youth clinical history and cutoff criteria from rating scales, provided substantial evidence supporting the criterion-related validity of the brief measure.

Several limitations deserve consideration when interpreting our results. We recruited a convenience sample, which may have introduced bias in the type of youth participating. In addition, our sample only included adolescents enrolled in schools. Findings may or not be different if we had selected a random sample or if adolescents who drop out from school had been included. We recommend further examination of the reliability and validity of the BSDMD among samples that include clinical cases (both outpatients and inpatients), school dropouts, and adolescents in correctional institutions. In addition, the validity and reliability of this measure should be examined in samples of adolescents from other Hispanic groups, as well as from other ethno-cultural and racial groups. Our study did not include an examination of the test-retest reliability of diagnoses, nor another diagnostic measure as concurrent validity criterion. Future

studies on the psychometric properties of the BSDMD should include a subsample who receives a second administration of this measure within a short time interval (e.g., 1 to 5 days) and a subsample who also completes another diagnostic measure such as the MDD, Dysthymia, and Adjustment Disorder modules of the standard MINI-KID.

Overall, the BSDMD is a promising tool for use in clinical research. It has a short-time administration (about 15 min), is suitable for self-report or interview format, and produces an appropriate assessment of the full criteria for DCBDS. Although designed at the term of DSM-IV-TR, its questions allow for a diagnosis based on the current DSM version. Still, more research to support its psychometric properties and its potential use in clinical settings is necessary.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.).

Andrews, G., & Peters, L. (1998). The psychometric properties of the Composite International Diagnostic Interview. *Social Psychiatry and Psychiatric Epidemiology, 33*(2), 80–88. https://doi.org/10.1007/s001270050026

Bernal, G., Rivera-Medina, C. L., Cumba-Avilés, E., Reyes-Rodríguez, M. L., Sáez-Santiago, E., Duarté-Vélez, Y., Nazario, L., Rodríguez-Quintana, N., & Rosselló, J. (2019). Can Cognitive-Behavioral Therapy be optimized with parent psychoeducation? A randomized effectiveness trial of adolescents with major depression in Puerto Rico. *Family Process, 58*(4), 832-854. https://doi.org/10.1111/famp.12455

Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., Kaufman, J., Dahl, R. E., Perel, J., & Nelson, B. (1996). Childhood and adolescent depression: A review of the past 10 years. Part I. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*(11), 1427–1439. https://doi.org/10.1097/00004583-199611000-00011

Breton, J. J., Bergeron, L., Valla, J. P., Lépine, S., Houde, L., & Gaudet, N. (1995). Do children aged 9 through 11 years understand the DISC Version 2.25 questions?. *Journal of the American Academy of Child and Adolescent Psychiatry, 34*(7), 946–956. https://doi.org/10.1097/00004583-199507000-00019

Canino, G., Shrout, P. E., Rubio-Stipec, M., Bird, H. R., Bravo, M., Ramirez, R., Chavez, L., Alegria, M., Bauermeister, J. J., Hohmann, A., Ribera, J., Garcia, P., & Martinez-Taboas, A. (2004). The DSM-IV rates of child and adolescent disorders in Puerto Rico: Prevalence, correlates, service use, and the effects of impairment. *Archives of General Psychiatry, 61*(1), 85–93. https://doi.org/10.1001/archpsyc.61.1.85

Center for Behavioral Health Statistics and Quality. (2016). *Key substance use and mental health indicators in the United States: Results from the 2015 National Survey on Drug Use and Health* (HHS Publication No. SMA 16-4984, NSDUH Series H-51). https://www.samhsa.gov/data/sites/default/files/NSDUH-FFR1-2015/NSDUH-FFR1-2015/NSDUH-FFR1-2015.pdf

Champion, D. J. (1981). *Basic statistics for social research* (2nd ed.). Prentice Hall.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Costello, E. J., Erkanli, A., & Angold, A. (2006). Is there an epidemic of child or adolescent depression? *Journal of Child Psychology and Psychiatry, 47*(12), 1263-1271. https://doi.org/10.1111/j.1469-7610.2006.01682.x

Cumba-Avilés, E. (2017). Cognitive-Behavioral Group Therapy for Latino youth with type 1 diabetes and depression: A case study. *Clinical Case Studies, 16*(1), 58-75. https://doi.org/10./1534650116668270

Cumba-Avilés, E., & Feliciano-López, V. (2013). Estudio piloto de validación del Inventario para la Evaluación del Espectro de la Sintomatología Depresiva (INEESD) [Pilot validation study of the Depressive Symptoms Spectrum Assessment Inventory (DSSAI)]. *Ciencias de la Conducta, 28*(1), 1-33.

Cumba-Avilés, E., & Feliciano-López, V. (2015). Development and validation of an indigenous measure of depression for Puerto Rican adolescents. *American Psychological Association 2015 Convention Presentations and Abstracts, Volume 8* [Conference session abstract], 904-905. 123rd American Psychological Association Annual Convention, Toronto, Ontario. https://doi.org/10.1037/e506802016-001

Cumba-Avilés, E., & Sáez-Santiago, E. (2016). Research program on type 1 diabetes and youth depression in Puerto Rico. *Puerto Rican Journal of Psychology, 27*(1), 44-60. https://www.redalyc.org/pdf/2332/233245623004.pdf

Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R., & Conover, N. C. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child Development, 56*(1), 265–275. https://doi.org/10.2307/1130193

Essau, C. A., Conradt, J., & Petermann, F. (1999). Progress and unresolved issues in depressive disorders among children and adolescents. In C. A. Essau, & F. Petermann, (Eds.), *Depressive disorders in children and adolescents: Epidemiology, risk factors, and treatment* (pp. 461-486). Jason Aroson.

Feliciano-López, V., & Cumba-Avilés, E. (2014). Propiedades psicométricas del Inventario para la Evaluación del Espectro de la Sintomatología Depresiva en adolescentes [Psychometric properties of the Depressive Symptoms Spectrum Assessment Inventory in youth]. *Puerto Rican Journal of Psychology, 25*(2), 260-278. https://www.redalyc.org/pdf/2332/233245622007.pdf

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Giannakopoulos, G. (2017). Concurrent validity of the Computer-assisted Structured Diagnostic Interview (CASDI) for children and adolescents aged 8-18 years old. *The Open Psychology Journal*, 10, 1-10. 1 https://doi.org/10.2174/1874350101710010001

González-Tejera, G., Canino,G., Ramírez, R., Chávez, L., Shrout, P., Bird, H., Bravo, M. Martinez-Taboas, A., Ribera, J. & Bauermeister, J. (2005). Examining minor and major

depression in adolescents. *Journal of Child Psychology & Psychiatry*, *46*(8), 888-899. https://doi.org/10.1111/j.1469-7610.2005.00370.x

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Kessler, R. C., Avenevoli, S., & Ries Merikangas, K. (2001). Mood disorders in children and adolescents: An epidemiologic perspective. *Biological Psychiatry, 49*(12), 1002-1014. https://doi.org/10.1016/S0006-3223(01)01129-5

Komiti, A. A., Jackson, H. J., Judd, F. K., Cockram, A. M., Kyrios, M., Yeatman, R., Murray, G., Hordern, C., Wainwright, K., Allen, N., & Singh, B. (2001). A comparison of the Composite International Diagnostic Interview (CIDI-Auto) with clinical assessment in diagnosing mood and anxiety disorders. *The Australian and New Zealand Journal of Psychiatry, 35*(2), 224–230. https://doi.org/10.1046/j.1440-1614.2001.00868.x

Kovacs, M. (2001). *Children's Depression Inventory (CDI): Technical manual.* Multi-Health Systems.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Leffler, J. M., Riebel, J., & Hughes, H. M. (2015). A review of child and adolescent diagnostic interviews for clinical practitioners. *Assessment, 22*(6), 690–703. https://doi.org/10.1177/1073191114561253

Mellsop, G., Varghese, F., Joshua, S., & Hicks, A. (1982). The reliability of axis II of DSM-III. *The American Journal of Psychiatry, 139*(10), 1360–1361. https://doi.org/10.1176/ajp.139.10.1360

Moscoso-Álvarez, M. R., Colón-Jordán, H. M., Reyes-Pulliza, J. C., Rodríguez-Figueroa, L. (2020). *El uso de substancias en los escolares puertorriqueños: Consulta Juvenil 2018-2020* [Substance use among Puerto Rican students: Youth Survey 2018-2020]. Administración de Servicios de Salud Mental y Contra la Adicción. https://assmca.pr. gov/BibliotecaVirtual/Consultas/Consulta%20Juvenil%20X%20-%202018-2020.pdf

Neuschwander, M, In-Albon, T, Meyer, A. H., Schneider, S. (2017). Acceptance of a structured diagnostic interview in children, parents, and interviewers. *International Journal of Methods in Psychiatric Research, 26*, e1573. https://doi.org/10.1002/mpr.1573

Renou, S., Hergueta, T., Flament, M., Mouren-Simeoni, M. C., & Lecrubier, Y. (2004). Entretiens diagnostiques structurés en psychiatrie de l'enfant et de l'adolescent [Diagnostic structured interviews in child and adolescent's psychiatry]. *L'Encephale, 30*(2), 122–134. https://doi.org/10.1016/s0013-7006(04)95422-x

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(1), 28–38. https://doi.org/10.1097/00004583- 200001000-00014

Sheehan, D. V., Sheehan, K. H., Shytle, R. D., Janavs, J., Bannon, Y., Rogers, J. E., Milo, K. M.,

Stock, S. L., & Wilkinson, B. (2010). Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *Journal of Clinical Psychiatry, 71*(3), 313-326. https://doi.org/10.4088/JCP.09m05305whi

Spitzer, R., & Fleiss, J. (1974). A re-analysis of the reliability of psychiatric diagnosis. *British Journal of Psychiatry*, 125, 341–347. https://doi.org/10.1192/bjp.125.4.341

Ulloa, R. E., Ortiz, S., Higuera, F., Nogales, I., Fresán, A., Apiquian, R., Cortés, J., Arechavaleta, B., Foulliux, C., Martínez, P., Hernández, L., Domínguez, E., & de la Peña, F. (2006). Estudio de fiabilidad interevaluador de la versión en español de la entrevista Schedule for Affective Disorders and Schizophrenia for School-Age Children--Present and Lifetime version (K-SADS-PL) [Interrater reliability of the Spanish version of Schedule for Affective Disorders and Schizophrenia for School-Age Children--Present and Lifetime version (K-SADS-PL)]. *Actas Españolas de Psiquiatria, 34*(1), 36–40. https://www.researchgate.net/publication/7252226_Inter-rater_reliability_of_the_Spanish_Version_of_Schedule_for_Affective_Disorders_and_Schizophrenia_for_School-Age_Children-Present_and_Life_time_Version_K-SADS-PL

Waslick, B., Schoenholz, D., & Pizarro, R. (2003). Diagnosis and treatment of chronic depression in children and adolescents. *Journal of Psychiatric Practice, 9*(5), 354-366. https://doi.org/10.1097/00131746-200309000-00004

Wittchen, H. (1994). Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review. *Psychiatry Research*, 28, 57–84. https://doi.org/10.1016/0022-3956(94)90036-1