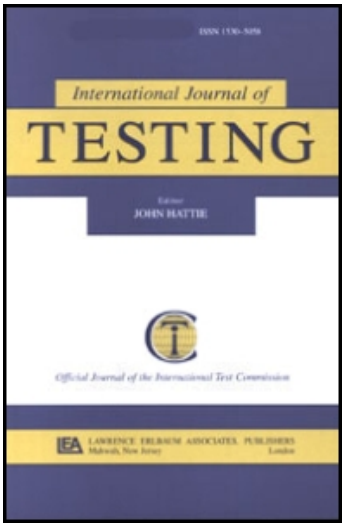

International Journal of Testing
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653658

Measurement Invariance of the Reynolds Depression Adolescent Scale across Gender and Age

Eduardo Fonseca-Pedrero[ab]; Craig Wells[c]; Mercedes Paino[ab]; Serafín Lemos-Giráldez[ab]; Úrsula Villazón-García[a]; Susana Sierra[a]; Ma Paz García-Portilla González[ab]; Julio Bobes[ab]; José Muñiz[ab]
[a] University of Oviedo, Spain [b] Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM, [c] School of Education, University of Massachusetts Amherst,

Online publication date: 19 August 2010

To cite this Article Fonseca-Pedrero, Eduardo , Wells, Craig , Paino, Mercedes , Lemos-Giráldez, Serafín , Villazón-García, Úrsula , Sierra, Susana , González, Ma Paz García-Portilla , Bobes, Julio and Muñiz, José(2010) 'Measurement Invariance of the Reynolds Depression Adolescent Scale across Gender and Age', International Journal of Testing, 10: 2, 133 — 148
To link to this Article: DOI: 10.1080/15305050903580822
URL: http://dx.doi.org/10.1080/15305050903580822

International Journal of Testing, 10: 133–148, 2010

ISSN: 1530-5058 print / 1532-7574 online
DOI: 10.1080/15305050903580822

Routledge
Taylor & Francis Group

Measurement Invariance of the Reynolds Depression Adolescent Scale across Gender and Age

Eduardo Fonseca-Pedrero
University of Oviedo, Spain; Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM

Craig Wells
School of Education, University of Massachusetts Amherst

Mercedes Paino and Serafín Lemos-Giráldez
University of Oviedo, Spain; Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM

Úrsula Villazón-García and Susana Sierra
University of Oviedo, Spain

Mª Paz García-Portilla González, Julio Bobes, and José Muñiz
University of Oviedo, Spain; Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM

The main objective of the present study was to examine measurement invariance of the Reynolds Depression Adolescent Scale (RADS) (Reynolds, 1987) across gender and age in a representative sample of nonclinical adolescents. The sample was composed of 1,659 participants, 801 males (48.3%), with a mean age of 15.9 years ($SD = 1.2$). Confirmatory factor analysis supported Reynolds' (2002) four-factor model, consisting of the Anhedonia, Somatic Complaints, Negative Self-Evaluation, and Dysphoric Mood dimensions. In addition, the results support the measurement

This research was funded by the Spanish Ministry of Science and Innovation (BES-2006-12797, PSI 2008-06220, PSI 2008-03934) and by the Instituto Carlos III, Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM).

Correspondence should be sent to Eduardo Fonseca-Pedrero, Facultad de Psicología, Plaza Feijoo, s/n, Oviedo 33003, Spain. E-mail: efonseca@cop.es

invariance of the RADS across gender and age. When latent means were compared as a function of gender and age, statistically significant differences were found. Females obtained higher scores than males in Somatic Complaints, Negative Self-Evaluation, and Dysphoric Mood. On the other hand, males obtained higher scores in Anhedonia. As a function of age, the 17–19 year olds obtained higher scores than the 14–16 year olds in Dysphoric Mood. The results suggest that the RADS could be used as an efficient self-report to test measurement invariance of depressive symptomatology across gender and age. Future studies should continue to explore measurement invariance across cultures and the incorporation of the latest advances in assessment methods in the clinical field.

Keywords: depression, RADS, self-reports, validation, measurement invariance, adolescents

Empirical evidence indicates that depressive symptomatology is a psychological phenomenon that can start during childhood and adolescence and is fairly common in this population subset (Birmaher et al., 1996; Cicchetti & Toth, 1998; Kessler, Berglund, Demler, Jin, Merikangas, & Walters, 2005). Overall, the childhood and adolescent prevalence rates fall within the 3%–8% range (Birmaher et al., 1996; Costello, Mustillo, Erkanli, Keeler, & Angold, 2003). In addition, it has been found that when using diverse measurement instruments for assessing depressive symptomatology, between 20% and 50% of adolescents exhibit subclinical levels of depression (Kessler, Avenevoli, & Merikangas, 2001; Petersen, Compas, Brooks-Gunn, & Stemmler, 1993). Subthreshold depressive symptoms in youths is associated with substantial functional impairment and is one of the best established risk factors for the onset of full-syndrome depressive disorders (Klein, Shankman, Lewinsohn, & Seeley, 2009; Lewinsohn, Solomon, Seeley, & Zeiss, 2000).

The assessment of depressive symptoms by means of self-reports has advanced considerably in the past few decades (Beck, Steer, & Brown, 1996; Kovacs, 1992; Lang & Tisher, 1987; Reynolds, 1987; Reynolds, 2002). This type of assessment contributes to an interesting aspect within psychological assessment as it offers valuable information regarding the adolescents' subjective experiences, which could not be obtained using other resources (Reynolds, 1998). There is no doubt that an aspect of great relevance is having measurement instruments at our disposal with adequate psychometric properties that allow the efficient assessment and early detection of these types of participants. Within the field of depressive symptomatology assessment in children and adolescents, great efforts have been made. There are a number of instruments with adequate psychometric properties in the literature, such as the Children's Depression Scale (CDS) (Lang & Tisher, 1987), the Children's Depression Inventory (CDI) (Kovacs, 1992), or the Reynolds Adolescent Depression Scale (RADS) (Reynolds, 1987, 2002). Likewise, other

measurement tools such as the Beck Depression Inventory I/II (BDI-I/II) (Beck et al., 1996; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), initially developed for the assessment of depressive symptomatology in adult populations, have also been widely used in nonclinical adolescent populations (Byrne, Baron, Larsson, & Melin, 1996; Byrne & Stewart, 2006; Byrne, Stewart, Kennard, & Lee, 2007).

In particular, the RADS is a self-report measure composed of 30 items designed for the assessment of depressive symptom severity in adolescents aged 11–20 years, both in school and clinical settings (Reynolds, 1987, 2002). Different research studies conducted with the RADS have shown that it is a measurement instrument with adequate psychometric properties. Levels of internal consistency have ranged from .91 to .96 and test-retest reliability of .87 has been reported (Figueras-Masip, Amador-Campos, & Peró-Caballero, 2008; Krefetz, Steer, Gulab, & Beck, 2002; Maharajh, Ali, & Konings, 2006; Reynolds, 1994, 1998; Reynolds & Mazza, 1998). Regarding validity, studies have reported high correlations with different subscales that measure depressive symptomatology (e.g., BDI) as well as adequate levels of sensitivity and specificity (Campbell, Byrne, & Baron, 1994; d'Acremonta & Van der Linden, 2007; Krefetz et al., 2002; Maharajh et al., 2006; Reynolds, 1994, 1998; Reynolds & Mazza, 1998).

Furthermore, data from previous studies show that depressive symptomatology fluctuates as a function of gender and age (Costello et al., 2003; Hankin & Abramson, 1999). In children, there appears to be the same proportion of females and males with depressive symptoms, although during adolescence, the rate seems to double in females (Hankin & Abramson, 1999). Likewise, the prevalence increases with age and is more frequent in adolescents than in children (Costello et al., 2003), with an earlier onset in females than in males (Angold, Erkanli, Silberg, Eaves, & Costello, 2002). Age and gender differences have also been found in depressive symptomatology using the RADS (Figueras-Masip et al., 2008; Hyun, Nam, Kang, & Reynolds, 2009; Maharajh et al., 2006; Walker, Merry, Watson, Robinson, Crengle, & Schaaf, 2005), although there are some contradictory results (Reynolds & Mazza, 1998).

When mean scores are used to compare groups (e.g., male/female; youths/adults) it is important that the scores have the same meaning in each group; that is, the assessment is invariant across groups. In a classic study, Horn and McArdle (1992) defined measurement invariance as: "*whether or not, under different conditions of observing and studying phenomena, measurement operations yield measure of the same attribute*" (p. 117). When comparisons between groups are made, it is typically assumed that both the measurement instrument and the psychological construct underlying said instrument behave in the same manner and have the same significance across the groups being compared (Byrne, 2008; Byrne & Stewart, 2006). If measurement invariance does not hold, the validity of the inferences and interpretations extracted from the data may be erroneous or unfounded (Byrne, 2008; Rusticus, Hubley, & Zumbo, 2008). Therefore, it is

crucial to examine measurement invariance of the assessment tool so that findings based on comparisons of the groups can be valid. Thus, it would be inappropriate to make comparisons regarding depressive symptoms if, for example, males and females interpret the content of the items differently or if this construct does not behave in the same manner across groups.

The interest in the study of measurement invariance in the field of depressive symptomatology assessment is presently well-developed (Byrne, Baron, & Balev, 1996; Carle, Millsap, & Cole, 2008; Crockett, Randall, Shen, Russell, & Driscoll, 2005; García, Aluja, & Del Barrio, 2008; Motl, Dishman, Birnbaum, & Lytle, 2005; Nguyen, Kitner-Triolo, Evans, & Zonderman, 2004). Specifically, there are various studies that examined measurement invariance across gender or culture in the general adolescent population (Byrne & Baron, 1994; Byrne, Baron, & Balev, 1996; Byrne & Stewart, 2006; Byrne et al., 2007; Carle et al., 2008). For instance, Byrne, Baron, and Balev (1996) used the BDI (Beck et al., 1961) in a sample of 691 Bulgarian adolescents and found that the factorial structure was invariant across gender. In another study, Byrne and Stewart (2006) used a sample of Chinese adolescents ($n = 1460$) and a sample of American adolescents ($n = 451$) and found that the second-order factorial structure underlying the BDI-II (Beck et al., 1996) was invariant across cultures. However, measurement invariance of the factorial structure of the RADS across gender and age has not been exhaustively examined.

Within the framework of research aimed at understanding and analyzing the structure of depressive symptomatology in general adolescent populations, the main objective of the current work was to analyze the measurement invariance of the RADS across gender and age in a representative sample of nonclinical adolescents. This objective contributes to the advancement in the understanding of the structure and content of depressive symptomatology in adolescent populations. Likewise, we should not lose sight of the fact that adolescence is considered to be a developmental stage of special risk for the development of a variety of psychological problems among which depressive disorder is found. Consequently, it is of great interest to have measurement instruments with adequate psychometric properties that permit rapid and efficient assessment at our disposal.

METHOD

Participants

Stratified random cluster sampling was conducted at the classroom level in an approximate population of 37,000 students selected from the Principality of Asturias, a region situated in northern Spain. The sample was selected such that it would be representative of the population of adolescents in this geographical

region. The students belonged to different public and concerted Educational Centers of Compulsory Secondary Education and Vocational Training as well as to different socioeconomic levels. The layers were created as a function of the geographical zone (East, West, and Center) and the educational stage (compulsory and post-compulsory), where the extraction probability of the school depended on the number of students. There were 1,780 students in the initial sample, although some participants were excluded due to their high scores on the infrequency scale (more than three points) ($n = 69$), being older than 19 years of age ($n = 17$), not completing demographical data (e.g., gender and age) ($n = 9$), or not completing all the administered self-reports ($n = 26$). Thus, the final sample was composed of 1,659 students; 801 males (48.3%) and 858 (51.7%) females from 41 centers and 95 classrooms. The mean age was 15.93 years ($SD = 1.22$), with an age range of 14 to 19 years. The distribution by age was 14 year olds ($n = 209$), 15 year olds ($n = 439$), 16 year olds ($n = 480$), 17 year olds ($n = 351$), 18 year olds ($n = 156$), and 19 year olds ($n = 29$). In accordance with the Spanish educational system (compulsory education and post-compulsory education), two age groups were created: adolescents from 14 to 16 years of age ($n = 1123$) and from 17 to 19 years of age ($n = 536$).

Procedure

Contact with the principals of compulsory secondary education and vocational training centers was made by letter or telephone. The administration of the questionnaire was conducted in a collective manner in groups of 15–25 participants in a classroom within the school timetable. They were reminded repeatedly of the confidentiality of their responses and that their participation was voluntary. Written informed consent to participate in the study was obtained from the adolescents. For subjects younger than 18 years, parents were asked to provide written informed consent in order for their child to participate in the study. Participants did not receive any type of incentive for their participation in the study. The administration took place under the supervision of the researchers.

Measurement Instrument

The Reynolds Adolescent Depression Scale (RADS) (Reynolds, 1987, 2002) is used to assess the severity of depressive symptomatology in adolescents aged 11 to 20 years. It is composed of 30 items in a Likert response format with 4 options (1 = almost never, 2 = hardly ever, 3 = sometimes, 4 = most of the time). The total scores range from 30 to 120, with the cut-off score for determining the severity of the depressive symptomatology set at 77 points (Reynolds, 1987). Reynolds (2002) recently proposed four scales: Anhedonia/Negative Affect, Somatic Complaints, Negative Self-Evaluation, and Dysphoric Mood. The RADS has been extensively

employed in a variety of topics, samples, and nationalities presenting adequate psychometric properties (Campbell et al., 1994; d'Acremonta & Van der Linden, 2007; Hyun et al., 2009; Maharajh et al., 2006; Reynolds, 1998; Reynolds & Mazza, 1998; Walker et al., 2005). The Spanish version, which was validated in a sample of 1,384 nonclinical and 217 clinical adolescents, was used for the present study (Figueras-Masip et al., 2008). The internal consistency and test-retest reliability for the Spanish version has ranged from .82 to .90 (nonclinical sample) and .84 to .91 (clinical sample), respectively (Figueras-Masip et al., 2008). The correlation with the Children's Depression Inventory (CDI) (Kovacs, 1992) was .81, and with the depression and internalizing scales of the Youth Self-Report (YSR) (Achenbach & Edelbrock, 1987) the correlation was .50.

The Oviedo Infrequency Scale (INF-OV) (Fonseca-Pedrero, Lemos-Giráldez, Paino, Villazón-García, & Muñiz, 2009) was administered to the participants to detect those who responded in a random, pseudorandom, or dishonest manner. The INF-OV instrument is a self-report composed of 12 items in a 5-point Likert-scale format (1 = completely disagree; 5 = completely agree) that was developed following guidelines for test construction (Schmeiser & Welch, 2006). Students with more than three incorrect responses on the INF-OV scale were eliminated from the sample. For this study, a total of 69 participants were excluded based on their responses to the INF-OV scale.

Data Analyses

Descriptive statistics. The mean and standard deviations were calculated for each of the items both for the total sample and as a function of gender and two age groups. In addition, coefficient alpha was used to estimate the internal consistency of the scales.

Confirmatory factor analysis. Confirmatory factor analysis (CFA) was implemented to examine the four-factor model proposed by Reynolds (2002) using the entire sample. The four factors were based on the four scales proposed by Reynolds (2002): Anhedonia/ Negative Affect, Somatic Complaints, Negative Self-Evaluation, and Dysphoric Mood. Each item loaded on one factor, and one of the loadings for each factor was fixed to 1.0 to establish the scale for each latent variable (the item for which the factor loading is fixed to 1.0 is referred to as the referent item). Because the data were based on Likert-type responses, which are ordinal, the asymptotic covariance matrix was used with the Robust Maximum Likelihood estimation method (Jöreskorg & Sörbom, 1993). Although this method assumes multivariate normality, and our data had slight deviations from normality, it has been shown to be robust to this violation (Curran, West, & Finch, 1996). Following Brown's (2006) and Kline's (2005) guidelines, the goodness-of-fit indices employed to assess the model fit were Satorra-Bentler scaled statistic ($S - B\chi^2$),

the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR).

Measurement invariance. Measurement invariance is frequently tested via multigroup comparisons using structural equation modeling within the framework of a CFA. Basically, a hierarchical set of steps are followed when testing measurement invariance, typically starting with the determination of a well-fitting multigroup and baseline model and continuing with the establishment of successive equivalence constraints in the model parameters across groups (Byrne, 2008; Byrne & Stewart, 2006). The baseline model is called the configural model, which is the first and least restrictive model specified and is important because it represents the baseline model against which all subsequent specified invariance models are compared. The configural model is established by specifying and testing the CFA model for each group separately. Once the theoretical model has been validated in both groups, configural invariance is then examined, requiring that the same pattern of fixed and freely estimated model parameters is equivalent across groups; however, other than the referent item that is used to establish the scale of each latent variable, no equality constraints are imposed on the model parameters between groups. Configural invariance is tested by assessing the model fit. When configural invariance is met (i.e., the model fits the data), it suggests that at least the general factor structure is similar, but not necessarily equivalent, across groups. The next step is to impose equality constraints on the factor loadings across the groups to test metric or weak invariance. If the model fit with the constrained parameters is significantly and practically worse than the baseline or configural model, then weak invariance is not supported. When metric invariance is met, it suggests that the same unit of measurement is being used for the item across the groups and that the participants interpret and respond to the measure in a similar manner (Horn & McArdle, 1992). The final step is to impose constraints on the item intercepts and factor loadings to test strong or scalar invariance across groups. The confirmation of the invariance of the intercepts permits comparison of the latent means in both groups (Meredith, 1993).

The analyzed models are nested in that the imposed constraints are progressively added. The fit of nested models may be assessed by comparing the respective chi-square fit statistic or goodness-of-fit indices between the model with additional constraints to the less restricted model (Byrne & Stewart, 2006; Cheung & Rensvold, 2002). Both criteria have been extensively used in the literature; however, when they are used in conjunction they are often in disagreement, causing researchers to reach contradictory conclusions (Rusticus et al., 2008). Due to the limitations of the $\Delta\chi^2$ regarding its sensitivity to sample size, Cheung and Rensvold (2002) proposed a more practical criterion, the ΔCFI, to determine if nested models are practically equivalent. In this study, when ΔCFI is greater than .01 between two nested models, the more constrained model is rejected since the

additional constraints have produced practically worse fit. However, if the change in CFI is less than or equal to .01, it is considered that all specified equal constraints are tenable; therefore, we can continue with the next step in the analysis of measurement invariance. However, when this criterion is not met and some of the parameters (e.g., factorial loadings or intercepts) are not specified to be equal across groups, partial measurement invariance can be considered (Byrne, Shavelson, & Muthén, 1989). SPSS 15.0 and LISREL 8.73 (Jöreskorg & Sörbom, 1993) were used for all data analyses.

RESULTS

Descriptive Statistics

The mean of the total score on the RADS for the entire sample was 50.62 ($SD = 10.02$). Overall, the mean for females ($\bar{X}_F = 51.69$, $SD_F = 9.75$) was slightly larger than that of males ($\bar{X}_M = 49.47$, $SD_M = 10.19$). The mean of the total score for the 14–16 year age group ($\bar{X}_{14-16} = 50.37$, $SD_{14-16} = 10.09$) was much higher than the mean for the 17–19 year age group ($\bar{X}_{17-19} = 41.13$, $SD_{17-19} = 9.86$). Furthermore, the total scores ranged from 33 to 102 points, with 2.2% of the entire sample scoring above the cut-off point (77 points).

The item means and standard deviations for the entire sample and as a function of gender and age groups are shown in Table 1. It is apparent that for most items, the means were relatively low, indicating that most of the participants tended to respond to the lower categories (i.e., "almost never" and "hardly ever"). Furthermore, the analysis of the multivariate descriptive statistics showed that the multivariate Kurtosis coefficient was 1359.5 ($Z = 185.4$; $p < 0.001$), revealing that these data do not likely come from a population that follows a multivariate normal distribution, but, as mentioned above, the asymptotic covariance matrix with the Robust Maximum Likelihood estimation method was used in our analyses, which is robust to deviations from normality (Jöreskorg & Sörbom, 1993). The levels of internal consistency for the different subscales proposed by Reynolds (2002) for the total sample were Anhedonia (.61), Somatic Complaints (.61), Negative Self-Evaluation, (.77) and Dysphoric Mood (.75); for males they were (in the same order) .61, .57, .78, and .75 and for females .61, .57, .78, and .73, respectively. With respect to the age groups, levels of internal consistency were .60, .62, .77, and .74 for 14–16 years group and .64, .60, .78, and .76 for adolescents of 17–19 years group.

Confirmatory Factor Analysis

The goodness-of-fit statistic and indices for the four-factor model proposed by Reynolds (2002) were: S-B $\chi^2 = 2005.4$; $df = 399$; $p < .001$; CFI = .956;

TABLE 1
Univariate Descriptive Statistics for the Items in the Reynolds Depression Adolescent Scale

Items	Total ($n = 1659$) Mean (SD)	Male ($n = 801$) Mean (SD)	Female ($n = 858$) Mean (SD)	14–16 years ($n = 1123$) Mean (SD)	17–19 years ($n = 536$) Mean (SD)
1	1.69 (.75)	1.70 (.74)	1.67 (.75)	1.68 (.75)	1.73 (.75)
2	2.82 (.91)	2.72 (.91)	2.90 (.90)	2.80 (.92)	2.87 (.90)
3	1.47 (.68)	1.43 (.68)	1.52 (.68)	1.46 (.68)	1.51 (.69)
4	1.34 (.70)	1.32 (.71)	1.35 (.69)	1.34 (.70)	1.33 (.70)
5	3.03 (.80)	2.90 (.85)	3.14 (.74)	3.02 (.82)	3.04 (.76)
6	1.44 (.68)	1.41 (.67)	1.47 (.69)	1.43 (.68)	1.46 (.67)
7	1.73 (.70)	1.60 (.69)	1.85 (.69)	1.72 (.72)	1.76 (.65)
8	1.66 (.71)	1.36 (.59)	1.93 (.69)	1.65 (.71)	1.68 (.69)
9	1.47 (.72)	1.44 (.76)	1.49 (.71)	1.46 (.72)	1.45 (.72)
10	1.48 (.77)	1.49 (.74)	1.47 (.79)	1.46 (.76)	1.53 (.78)
11	1.38 (.63)	1.32 (.60)	1.43 (.66)	1.36 (.62)	1.43 (.67)
12	1.70 (.84)	1.77 (.88)	1.63 (.80)	1.70 (.85)	1.70 (.83)
13	1.54 (.87)	1.45 (.83)	1.62 (.89)	1.53 (.88)	1.55 (.86)
14	1.11 (.45)	1.12 (.50)	1.10 (.39)	1.13 (.47)	1.09 (.39)
15	1.30 (.60)	1.27 (.58)	1.32 (.62)	1.31 (.61)	1.29 (.57)
16	1.54 (.69)	1.46 (.69)	1.63 (.69)	1.52 (.70)	1.61 (.67)
17	2.20 (1.0)	2.21 (1.0)	2.18 (.97)	2.20 (1.0)	2.20 (.99)
18	1.99 (.79)	1.93 (.83)	2.06 (.75)	1.98 (.81)	2.03 (.76)
19	1.31 (.59)	1.35 (.64)	1.27 (.55)	1.31 (.59)	1.31 (.61)
20	1.32 (.64)	1.28 (.62)	1.36 (.65)	1.33 (.65)	1.32 (.63)
21	1.24 (.58)	1.25 (.58)	1.24 (.57)	1.25 (.59)	1.24 (.56)
22	2.60 (.81)	1.61 (.83)	2.97 (.80)	2.61 (.84)	2.56 (.76)
23	1.33 (.65)	1.40 (.69)	1.26 (.61)	1.33(.66)	1.34 (.64)
24	1.62 (.86)	1.57 (.85)	1.67 (.88)	1.61 (.85)	1.65 (.90)
25	1.24 (.59)	1.28 (.63)	1.20 (.56)	1.23 (.61)	1.25 (.57)
26	2.03 (.76)	1.93 (.79)	2.12 (.74)	1.98 (.77)	2.14 (.75)
27	1.56 (.69)	1.44 (1.1)	1.68 (.73)	1.55 (.68)	1.60 (.71)
28	2.08 (.80)	2.12 (.86)	2.05 (.75)	2.09 (.81)	2.08 (.79)
29	1.81 (.86)	1.73 (.81)	1.89 (.91)	1.84 (.88)	1.76 (.83)
30	1.53 (.74)	1.50 (.75)	1.56 (.74)	1.51 (.74)	1.57 (.75)

Note: Items 1, 5, 10, 12, 23, 25, and 29 have been reverse-scored. Higher scores indicate more severe depressive symptomatology.

RMSEA = .050 [90% C.I = .047–.051]; SRMR = .051. The goodness-of-fit indices for this model supported adequate model fit in that the *CFI* value was higher than .90, the *RMSEA* value was less than .05, and the *SRMR* was less than .08. The estimated standardized factor loadings for the four-factor model are shown in Table 2. As can be observed, all standardized coefficients were statistically significant, oscillating between .17 and .52.

TABLE 2
Standardized Coefficients for the Four-Factor Model by Reynolds (2002)

	Factors			
Items	Dysphoric Mood	Anhedonia	Negative Self-Evaluation	Somatic Complaints
1		.51		
2	.17			
3	.46			
4			.38	
5		.23		
6	.36			
7	.51			
8	.41			
9			.44	
10		.28		
11				.22
12		.45		
13			.48	
14			.24	
15			.29	
16	.40			
17				.53
18				.40
19			.20	
20			.43	
21	.26			
22				.30
23		.27		
24				.36
25		.23		
26	.47			
27				.26
28				.38
29		.20		
30			.52	

Note: All standardized coefficients were statistically significant ($p < .01$).

Measurement Invariance

Gender. Measurement invariance for the four-factor model hypothesized by Reynolds (2002) for males and females was studied. The results are presented in Table 3. The goodness-of-fit indices obtained for both groups were satisfactory, indicating an adequate fit to the data. Likewise, the configural model in which no equality constraints (beyond the referent item factor loading) were imposed represented an excellent fit to the data. Next, metric invariance for both groups was tested. The CFI values for the configural model were identical to the metric

TABLE 3
Configural, Weak and Strong Measurement Invariance for Reynolds Depression Adolescent Scale: Goodness-of-Fit Statistics

Model	S-B χ^2	*df*	CFI	RMSEA	RMSEA 90% CI	SRMR	ΔCFI
Gender							
Male ($n = 801$)	1078.1	399	.961	.046	.043–.049	.054	
Female ($n = 858$)	1291.2	399	.945	.051	.048–.054	.056	
Multiple group							
Configural Invariance	2367.6	798	.953	.049	.047–.051	.056	–.010
Weak Invariance	2395.6	824	.953	.048	.046–.050	.059	–.010
Strong Invariance	2874.5	850	.940	.054	.051–.056	.060	.013
Partial Strong Invariance	2650.1	849	.946	.051	.048–.055	.060	–.010
Age							
14–16 years ($n = 1123$)	1509.2	399	.950	.050	.047–.053	.053	
17–19 years ($n = 536$)	936.9	399	.952	.050	.047–.053	.063	
Multiple group							
Configural Invariance	2442.4	798	.951	.050	.048–.052	.063	–.010
Weak Invariance	2439.1	824	.952	.049	.046–.051	.068	–.010
Strong Invariance	2499.9	850	.951	.048	.046–.051	.068	–.010

Note: *S-B*χ^2 = Satorra-Bentler scaled statistic; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation; 90% CI = 90% Confidence Interval; *SRMR* = Standardized Root Mean Square Residual.

invariant model, which indicated that the hypothesis of metric invariance was tenable. Subsequently, strong measurement invariance was tested where the item intercepts and factor loadings were constrained to be equal across groups. The ΔCFI between the constrained and the unconstrained models was .013, indicating that strong invariance was not supported. Inspection of the modification indices suggested the relaxed intercept of item 8 in the Dysphoric Mood scale. In this case, the ΔCFI was less than .01, therefore, according to the recommendations by Cheung and Rensvold (2002), partial strong invariance was accepted. Hence, the results support configural, metric, and partial strong invariance in depressive symptomatology measured by the RADS across gender.

Age. Subsequently, measurement invariance of the RADS across age was analyzed. The goodness-of-fit indices for the sample of adolescents from 14 to 16 years of age and from 17 to 19 years of age are shown in Table 3. The configural model in which no equality constraints were imposed provided adequate fit to the data. As can be observed, when the equivalence of the factorial loadings and intercept values were incorporated, the difference in the ΔCFI between the configural and the constrained models did not exceed .01; therefore, we concluded

that the factorial structure of the RADS was operating equivalently across the two adolescents age groups.

Tests for latent mean differences. Next, latent mean differences across groups were estimated fixing the latent mean values to zero first in males, and then in the 14–16 year group. For comparisons among groups in the latent means, statistical significance was based on the *z* statistic. The comparison of the gender groups on the latent means revealed statistically significant differences in the latent means of all four factors of the RADS: Anhedonia (–.063; $p < .05$), Somatic Complaints (.034; $p < .05$), Negative Self-Evaluation (.042; $p < .05$), and Dysphoric Mood (.130; $p < .001$). For example, in the case of the Dysphoric Mood factor, the .130 value indicated that, on average, females scored .130 units higher than males. The comparison of the age groups on the latent means revealed statistically significant differences in Dysphoric Mood (.055; $p < .05$), indicating that on average, the 17–19 year adolescents scored .055 units above the 14–16 year adolescents.

DISCUSSION AND CONCLUSIONS

The main objective of the current study was to examine measurement invariance of the Reynolds Adolescent Depression Scale (RADS) (Reynolds, 1987) across gender and age groups in a representative sample of nonclinical adolescents. The results showed that the RADS is a measurement instrument with adequate psychometric properties, which can be administered in a rapid, efficient, and noninvasive manner; therefore, it is useful for the assessment of depressive symptom severity in nonclinical adolescents. The levels of internal consistency for the RADS total score obtained in the current study are similar to those reported in the previous literature, although slightly lower (Figueras-Masip et al., 2008; Maharajh et al., 2006; Reynolds, 1998, 2002; Reynolds & Mazza, 1998; Walker et al., 2005).

The model proposed by Reynolds (2002), comprising the Anhedonia, Dysphoric Mood, Somatic Complaints, and Negative Self-Evaluation factors, exhibited reasonably good fit to the data. This finding indicates that the factorial structure of depression, measured by the RADS, can be specified as a structure composed of four correlated factors; however, although these findings are totally convergent with the work of Reynolds (2002), Walker el al. (2005) and Hyun et al. (2009), previous studies have also found the one-factor solution as the most parsimonious (Figueras-Masip et al., 2008). Likewise, the four-factor model hypothesized by Reynolds (2002) was shown to be invariant across gender and age of the adolescents in our study. Previous studies, which have used other self-reports such as the BDI (Beck et al., 1961) or the CDI (Kovacs, 1992) and were carried out in adult populations and in nonclinical adolescent populations, have obtained similar results (Byrne & Baron, 1994; Byrne, Baron, & Balev, 1996; Byrne & Stewart,

2006; Byrne et al., 2007; Carle et al., 2008; Crockett et al., 2005; García et al., 2008); notwithstanding, we must keep in mind that the strict comparison of studies is hindered by the type of sample and the measurement instrument employed as well as by the statistical analyses performed. Specifically, various studies have examined measurement invariance in the general adolescent population (Byrne & Baron, 1994; Byrne, Baron, & Balev, 1996; Byrne & Stewart, 2006; Byrne et al., 2007; Carle et al., 2008; Motl et al., 2005). For example, Byrne et al. (1996) used the BDI (Beck et al., 1961) in Bulgarian adolescents and found that its factorial structure was invariant across gender. In another study, Carle et al. (2008) used the CDI (Kovacs, 1992) in a sample of American adolescents and showed that the items in the CDI provided invariant measurement across gender. On the other hand, when the invariance of depressive symptomatology across the age of adolescents is examined, very few studies are found in the literature, for which future research should study the role that this variable plays more in depth.

The comparison in the latent means across gender and age yielded statistically significant differences. Females obtained higher scores than males in Somatic Complaints, Negative Self-Evaluation, and Dysphoric Mood, whereas males obtained higher scores in Anhedonia. As a function of age, adolescents between 17 and 19 years obtained higher scores in Dysphoric Mood as compared to the younger group. Consistent with the previous literature, the expression of depressive symptomatology varies as a function of age and gender (Angold et al., 2002; Carle et al., 2008; Costello et al., 2003; Figueras-Masip et al., 2008; Hankin & Abramson, 1999; Maharajh et al., 2006; Reynolds, 2002; Walker et al., 2005). In general terms, the prevalence of depressive symptoms has an earlier onset in females than in males (Angold et al., 2002) and increases with age, being more frequent in adolescents than in children (Costello et al., 2003). Using the raw scores of the RADS subscales or the RADS total score, female adolescents obtain higher scores than males in depressive dimensions, except in Anhedonia where males obtained higher scores and the older adolescents also obtained higher scores in comparison to the younger adolescents (Figueras-Masip et al., 2008; Hyun et al., 2009; Maharajh et al., 2006; Reynolds & Mazza, 1998; Walker et al., 2005). These data confirm the tendency to find significant differences between males and females in depressive symptomatology beginning in adolescence and continuing into adulthood.

The results found in this study should be interpreted in light of some possible limitations. First, the sample was composed exclusively of adolescents. It is well known that adolescence is a developmental period in which a great variety of neuromaturational, social, emotional, and self-identity changes occur and that may have an influence in the phenomenological expression of this construct. Second, in this study, information was gathered based solely on self-reports for which we consider that it would have been interesting to complete this information with a clinical interview or with a hetero-report administered to the participants' parents.

Third, it must be taken into account that no information about the presence of psychological problems in the adolescents or in their relatives was obtained in this study.

The assessment of depressive symptomatology in young populations permits the better understanding of this psychological construct as well as the possibility of having at our disposal instruments with psychometric guarantees that permit early evaluation and intervention in these types of participants. Future research should continue to advance in the study of measurement invariance of depressive symptomatology across cultures and incorporate the new advances in psychological and educational measurement through the use of computerized-adaptive testing.

REFERENCES

Achenbach, M. T., & Edelbrock, C. S. (1987). *Manual for the Youth Self Report and profile*. Burlington, VT: University of Vermont. Department Psychiatry.

Angold, A., Erkanli, A., Silberg, J., Eaves, L., & Costello, E. J. (2002). Depression scale scores in 8–17-year-olds: Effects of age and gender. *Journal of Child Psychology and Psychiatry*, *43*, 1052–1063.

Beck, A. T., Steer, R., & Brown, G. (1996). *Beck Depression Inventory Manual* (2nd ed). San Antonio, TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Birmaher, B., Ryan, N., Williamson, D., Brent, D., Kaufman, J., Dahl, R., et al. (1996). Childhood and adolescent depression: A review of the past 10 years: Part 1. *Journal of the American Academy of Child and Adolescent Psychiatry*, *35*, 1427–1439.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, *20*, 872–882.

Byrne, B. M., & Baron, P. (1994). Measuring adolescent depression: Tests of equivalent factorial structure for English and French versions of the Beck Depression Inventory. *Applied Psychology: An International Review*, *43*, 33–47.

Byrne, B. M., Baron, P., & Balev, J. (1996). The Beck Depression Inventory: Testing for its factorial validity and invariance across gender for Bulgarian non-clinical adolescents. *Personality and Individual Differences*, *21*, 641–651.

Byrne, B. M., Baron, P., Larsson, B., & Melin, L. (1996). Measuring depression for Swedish nonclinical adolescents: Factorial validity and equivalence of the Beck Depression Inventory across gender. *Scandinavian Journal of Psychology*, *37*, 37–45.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*, 287–321.

Byrne, B. M., Stewart, S. M., Kennard, B. D., & Lee, P. W. H. (2007). The Beck Depression Inventory-II: Testing for measurement equivalence and factor mean differences across Hong Kong and American adolescents. *International Journal of Testing*, *7*, 293–309.

Campbell, T. L., Byrne, B. M., & Baron, P. (1994). The Reynolds Adolescent Depression Scale: An exploratory factor analytic study. *Revue Européenne de Psychologie Appliquée*, *44*, 319–324.

Carle, A. C., Millsap, R. E., & Cole, D. A. (2008). Measurement bias across gender on the Children's Depression Inventory. Evidence for invariance from two latent variable models. *Educational and Psychological Measurement*, *68*, 281–303.

Cicchetti, D., & Toth, S. L. (1998). The development of depression in children and adolescents. *American Psychologist*, *53*, 221–241.

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, *60*, 837–844.

Crockett, L. J., Randall, B. A., Shen, Y. L., Russell, S. T., & Driscoll, A. K. (2005). Measurement equivalence of the Center for Epidemiological Studies Depression Scale for Latino and Anglo adolescents: A national study. *Journal Consultant Clinical Psychology*, *73*, 47–58.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.

d'Acremonta, M., & Van der Linden, M. (2007). How is impulsivity related to depression in adolescence? Evidence from a French validation of the cognitive emotion regulation questionnaire. *Journal of Adolescence*, *30*, 271–282.

Figueras-Masip, A., Amador-Campos, J. A., & Peró-Caballero, M. (2008). Características psicométricas de la *Reynolds Adolescent Depression Scale* en población comunitaria y clínica [Psychometric characteristics of the RADS in community and clinical populations]. *International Journal of Clinical and Health Psychology*, *8*, 247–266.

Fonseca-Pedrero, E., Lemos-Giráldez, S., Paino, M., Villazón-García, U., & Muñiz, J. (2009). Validation of the Schizotypal Personality Questionnaire Brief form in adolescents. *Schizophrenia Research*, *111*, 53–60.

García, L. F., Aluja, A., & Del Barrio, V. (2008). Testing the hierarchical structure of the Children's Depression Inventory. A multigroup analysis. *Assessment*, *15*, 153–164.

Hankin, B. L., & Abramson, L. Y. (1999). Development of gender differences in depression: Description and possible explanations. *Psychological Medicine*, *28*, 372–379.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement equivalence in aging research. *Experimental Aging Research*, *18*, 117–144.

Hyun, M.-S., Nam, K.-A., Kang, H. S., & Reynolds, W. M. (2009). Reynolds Adolescent Depression Scale—second edition: initial validation of the Korean version. *Journal of Advanced Nursing*, *65*, 642–651.

Jöreskorg, K. G., & Sörbom, D. (1993). *LISREL 8 User's Reference Guide*. Chicago: Scientific Software International.

Kessler, R. C., Avenevoli, S., & Merikangas, R. K. (2001). Mood disorders in children and adolescents: An epidemiological perspective. *Biological Psychiatry*, *49*, 1002–1014.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distribution of DSM-IV disorders in the National Comorbidity Survey replication. *Archives of General Psychiatry*, *62*, 593–602.

Klein, D. N., Shankman, S. A., Lewinsohn, P. M., & Seeley, J. R. (2009). Subthreshold depressive disorder in adolescents: predictors of escalation to full-syndrome depressive disorders. *Journal of American Academy of Child and Adolescent Psychiatry*, *48*, 703–710.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). New York: The Guilford Press.

Kovacs, M. (1992). *Manual of the Children's Depression Inventory*. Toronto: Multi-Heath Systems, Inc.

Krefetz, D. G., Steer, A. S., Gulab, N. A., & Beck, A. T. (2002). Convergent validity of the Beck Depression Inventory–II with the Reynolds Adolescent Depression Scale in psychiatric inpatients. *Journal of Personality Assessment*, *78*, 451–460.

Lang, M., & Tisher, M. (1987). *Children's Depression Scale (CDS) Manual (North American Edition)*. Palo Alto, CA: Consulting Psychologists Press.

Lewinsohn, P. M., Solomon, A., Seeley, J. R., & Zeiss, A. (2000). Clinical implications of "subthreshold" depressive symptoms. *Journal of Abnormal Psychology*, *109*, 345–351.

Maharajh, H. D., Ali, A., & Konings, M. (2006). Adolescent depression in Trinidad and Tobago. *European Child & Adolescent Psychiatry*, *15*, 30–37.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Motl, R. W., Dishman, R. K., Birnbaum, A. S., & Lytle, L. A. (2005). Longitudinal invariance of the Center for Epidemiologic Studies-Depression scale among girls and boys in middle school. *Educational and Psychological Measurement*, *65*, 90–108.

Nguyen, H. T., Kitner-Triolo, M., Evans, M. K., & Zonderman, A. B. (2004). Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. *Psychiatry Research*, *126*, 177–187.

Petersen, A. C., Compas, B. E., Brooks-Gunn, J., & Stemmler, M. (1993). Depression in adolescence. *American Psychologist*, *48*, 155–168

Reynolds, W. M. (1987). *Reynolds Adolescent Depression Scale. Professional Manual*. Odessa: Psychological Assessment Resources, Inc.

Reynolds, W. M. (1994). Assessment of depression in children and adolescents by self-report questionnaires. In W. M. Reynolds & H. F. Johnston (Eds.), *Handbook of Depression in Children and Adolescents: Issues in Clinical Child Psychology* (pp. 209–234). New York: Plenum.

Reynolds, W. M. (1998). Depression in children and adolescents. In T. H. Ollendick (Ed.), *Comprehensive Clinical Psychology: Vol. 4. Children and Adolescents: Clinical Formulations and Treatment* (pp. 419–461). New York: Pergamon Press.

Reynolds, W. M. (2002). *Reynolds Adolescent Depression Scale—2nd Edition. Professional Manual*. Odessa: Psychological Assessment Resources, Inc.

Reynolds, W. M., & Mazza, J. J. (1998). Reliability and validity of the Reynolds Adolescent Depression Scale with young adolescents. *Journal of School Psychology*, *36*, 295–312.

Rusticus, S. A., Hubley, A. M., & Zumbo, B. D. (2008). Measurement invariance of the Appearance Schemas Inventory-Revised and the Body Image Quality of Life Inventory across age and gender. *Assessment*, *15*, 60–71.

Schmeiser, C. B., & Welch, C. (2006). Test development. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.) (pp. 307–353). Westport, CT: American Council on Education/Praeger.

Walker, L., Merry, S., Watson, P. D., Robinson, E., Crengle, S., & Schaaf, D. (2005). The Reynolds Adolescent Depression Scale in New Zealand adolescents. *Australian and New Zealand Journal of Psychiatry*, *39*, 136–140.