Research in Developmental Disabilities 32 (2011) 1872–1883

Contents lists available at ScienceDirect

Research in Developmental Disabilities

The Center for Epidemiologic Studies Depression Scale: Factor validity and reliability in a French sample of adolescents with Intellectual Disability

Christophe Maïano [a,1,*], Alexandre J.S. Morin [b,c,1], Jérôme Bégarie [d]

[a] Institute of Movement Sciences Etienne-Jules Marey (UMR 6233), CNRS-University of Aix-Marseille II, Marseille, France
[b] Department of Psychology, University of Sherbrooke, Sherbrooke, Canada
[c] Educational Excellence and Equity (E3) Research Program, Center for Educational Research (University of Western Sydney), Australia
[d] Institut Médico-Educatif Valfleurs, Croix-Rouge Française, Grasse, France

ARTICLE INFO

Article history:
Received 4 March 2011
Received in revised form 17 March 2011
Accepted 17 March 2011
Available online 6 May 2011

Keywords:
CES-D
Depression
Intellectual Disability
Adolescents
Likert-graphical answer scale
Gender
Age
Invariance

ABSTRACT

The purpose of this study was to test the factor validity and reliability of the Center for Epidemiologic Studies Depression Scale (CES-D) within a sample of adolescents with mild to moderate Intellectual Disability (ID). A total sample of 189 adolescents (121 boys and 68 girls), aged between 12 and 18 years old, with mild to moderate ID were involved in two studies. In study 1, the content, phrasing and answering format of the CES-D were adapted for adolescents with ID. This instrument was renamed CES-D for ID (CES-D-ID) and two different versions based on two alternative answer scales (Likert and Likert-graphical) were developed and their psychometric properties were verified in study 2. The results provided support for the factor validity, reliability and invariance across gender and age of a 14-item version of the CES-D-ID based on a Likert-graphical answer scale.

1. Introduction

It has long been recognized among the scientific and clinical community that youths with Intellectual Disability (ID) are vulnerable to depression and may suffer from it as much as their "normal" peers (e.g. Gillberg, Presson, Grufman, & Themner, 1986; Manikam, Matson, Coe, & Hillman, 1995; Matson, Barett, & Helsel, 1988; Menolascino, 1965; Reiss & Rojahn, 1993; Reynolds & Miller, 1985; for recent reviews see Borthwick-Duffy, 1994; Dykens, 2000; Hurley, 2006; Janowky & Davis, 2005; Kiddle & Dagnan, 2011; Lunsky & Palucka, 2004; McBrien, 2003; Rola, 1996; Whitaker & Read, 2006). However, very few results are available regarding the prevalence of depressive symptoms in large sample of youths with ID (for an exception see Dekker, Koot, van der Ende, & Verhulst, 2002). As a consequence, the prevalence and symptomatology of depression among large sample of youths with ID may remain unidentified or undiagnosed. Indeed, most of the empirical studies that were conducted on depression in youths with ID (e.g. Dekker & Koot, 2003; Douma, Dekker, Verhulst, & Koot, 2006; Emerson, 2003; Gillberg et al., 1986) essentially relied on detailed psychiatric examinations or semi-structured interviews

* Corresponding author.
E-mail address: christophe.maiano@univmed.fr (C. Maïano).
[1] Since both contributed equally to the preparation of this paper, the order of appearance of the first and second authors (CM and AJSM) was determined at random: both should be considered first authors.

0891-4222/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ridd.2011.03.016

(administered either to parents, teachers, other informants or directly to the youth) and reported prevalence rate of depression, which is similar to the rates observed in "normal" adults or adolescents (e.g. Angold & Costello, 2001; Bebbington et al., 1998; Emerson, 2003; Fleming & Offord, 1990; Kessler, 2002; Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993; Newman et al., 1996; Offord et al., 1996; Weissman et al., 1996).

One possible reason for the paucity of large sample epidemiologic research on in youths with ID may be linked to the relative absence of specific epidemiologic standardized self-report assessment tool for quantifying depressive symptoms among youths with ID (e.g. Morin, Cobigo, Rivard, & Lépine, 2010; Perez-Achiaga, Nelson, & Hassiotis, 2009; Ross & Oliver, 2003). Indeed, previous studies relied on standardized interviews and protocols involving third person reports that, although they possess excellent psychometric properties, are not well suited to the conduct of large-scale epidemiologic studies. Indeed, these methods are associated with (Esbensen, Seltzer, Greenberg, & Benson, 2005; Finlay & Lyons, 2002): (i) a substantial consumption of time, attention and concentration during administration; (ii) more difficulties in the identification of non-observable symptoms that are generally harder to verbalize for youths with ID (feelings of worthlessness, suicidal ideation, and depressed mood); (iii) few controls for youths with ID natural predisposition to acquiescence bias; and (iv) elevated costs when used in large samples. The preeminence of standardized protocols is linked to the fact that it is well documented that third person reports of depression tend to be more accurate among populations with ID than self-report instruments (e.g. Benavidez & Matson, 1993; Kazdin, Matson, & Senatore, 1983; Masi, Brovedani, Mucci, & Favilla, 2002; Matson, 1982; Matson, 1983; Matson, 1986; Matson et al., 1999; Rojahn, Warren, & Ohringer, 1994). However, as mentioned by Hartley and MacLean (2006, p. 814) "*self-report measures are also important to the field of ID because they allow people with ID to have an active role in research. This follows from the general movement that people with ID need to have a greater say in their lives and should be encouraged to voice their concerns and perspectives*". In addition, self-reports are generally better suited to large scale studies, less time consuming, and have the advantage of providing access to important subjective information to which participants are often the only one to have access (Argus, Terrya, Bramstona, & Dinsdale, 2004).

For these reasons, various self-reported depression instruments initially developed for people without ID were adapted for use with in populations with ID [e.g. Birleson Depressive Short form Self-Rating Scale (Benson & Ivins, 1992); Beck Depression Inventory I and II (Lindsay & Skene, 2007; Powell, 2003); Zung Self-Report Depression scale (Powell, 2003); Hospital Anxiety and Depression Scale (Dagnan et al., 2008); Profile of Mood States (Argus et al., 2004)]. Some of these instruments have even been available for some time (e.g. Kazdin et al., 1983; Matson, 1982; Matson, 1986). New instruments were also specifically developed for individuals with ID [e.g. Glasgow depression scale for people with learning disabilities (Cuthill, Espie, & Cooper, 2003); Self-Report Depression Questionnaire (Esbensen et al., 2005; Reynolds & Baker, 1988)]. However, these self-report instruments present characteristics or limitations that make them sub-optimal in the context of large scale epidemiologic studies. Indeed, some of these instruments (for reviews see Perez-Achiaga et al., 2009; Ross & Oliver, 2003; Singh, Sood, Sonenklar, & Ellis, 1991): (i) present suboptimal psychometric properties (Powell, 2003) or were developed using small samples (e.g. $n \leq 100$ in Cuthill et al., 2003), rendering uncertain the generalizability of their psychometric characteristics; (ii) are too long to be integrated in large scope epidemiologic studies, either because they include too many items (i.e. often more than 20 items and up to 32 items in Esbensen et al., 2005; Reynolds & Baker, 1988), or because they rely on behaviorally anchored answer scales (Lindsay & Skene, 2007; Powell, 2003), which may also be suboptimal given the linguistic level of individuals with ID; (iii) do only cover a restricted range of depressive symptoms (Argus et al., 2004; Dagnan et al., 2008); (iv) were designed as screening or clinical follow-up tools rather than as epidemiologic instruments (Argus et al., 2004; Dagnan et al., 2008; Esbensen et al., 2005; Lindsay & Skene, 2007; Powell, 2003; Reynolds & Baker, 1988); (v) relied on purely verbal Likert-type answer scales (Cuthill et al., 2003; Esbensen et al., 2005; Reynolds & Baker, 1988), which may also be suboptimal for individuals with ID (we return to this aspect later). Regarding this last limitation, some studies even had to screen participants for their ability to answer affective questions on a Likert type scale (Esbensen et al., 2005).

The Center for Epidemiologic Studies Depression Scales (CES-D; see Table 1) is a self-report scale specifically designed for the measurement of depressive symptoms in the context of epidemiologic studies (Radloff, 1977). This instrument was initially validated in a large clinical and nonclinical sample of North American adults ($n > 5000$). The principal component analyses done by Radloff (1977) provided support for a 20-item first-order four-factor model: (i) depressive affect (DA: items 3, 6, 9, 10, 14 and 17); (ii) positive affect (PA: items 4, 8, 12, and 16); (iii) somatic complaints and retarded activity (SRA: items 1, 2, 5, 7, 11, 13 and 20); and (iv) disturbed interpersonal relationships (INT: items 15 and 19). Additional analyses among these various samples also demonstrated that the full scale presented: (i) acceptable internal consistency coefficients ranging from $\alpha = .85$ to .90 in the nonclinical and clinical samples; (ii) moderate test–retest reliability coefficients ranging from $r = .51$ to .32 for time intervals varying between 2 weeks and 12 months; (iii) moderate correlations with several convergent measures of depressive symptoms, general psychopathology, positive and negative affects, social desirability, medication, aggression, etc. Following this initial study, the CES-D has been widely cross-culturally adapted and validated in China (Cheung & Bagley, 1998), France (Fuhrer & Rouillon, 1989; Morin et al., in press), Germany (Hautzinger, 1988), Greece (Fountoulakis et al., 2001), Italy (Fava, 1983), Portugal (Gonçalves & Fagulha, 2004), Russia (Dershem, Patsiorkovski, & O'Brien, 1996) and Spain (Vázquez, Blanco, & López, 2007). Since the CES-D is one of the most widely used instrument in epidemiologic studies in the general population and given the lack of adequate self-report instruments available for the epidemiologic evaluation of ID youths, the main objective of the present series of studies was to investigate the factor validity and reliability of the CES-D within a population of adolescents with mild to moderate ID.

Table 1
Items from the CES-D-ID (Original French and English items from the CES-D in parentheses).

1	Je suis embêté(e) par des choses qui d'habitude ne m'embêtent pas – I am bothered by things that don't usually bother me. (J'ai été contrarié(e) par des choses qui d'habitude ne me dérangent pas - I was bothered by things that usually don't bother me)	SRA[†]
2	J'ai faim – I'm hungry. (Je n'ai pas eu envie de manger, j'ai manqué d'appétit - I did not feel like eating; my appetite was poor)	SRA[*]
3	Je pense à de mauvaises choses – I think about bad things. (J'ai eu l'impression que je ne pouvais pas sortir du cafard, même avec l'aide de ma famille et de mes ami(e)s - I felt that I could not shake off the blues even with help from my family and friends)	DA[†]
4	Je me sens aussi bon(ne) que les autres personnes – I feel that I am just as good as other people. (J'ai eu le sentiment d'être aussi bien que les autres - I felt that I was just as good as other people)	PA[*]
5	J'ai des difficultés à faire attention à ce que je fais – I have a hard time keeping my mind on what I do. (J'ai eu du mal à me concentrer sur ce que je faisais - I had trouble keeping my mind on what I was doing)	SRA[†]
6	Je me sens déprimé(e) – I feel depressed. (Je me suis senti(e) déprimée - I felt depressed)	DA[†]
7	Je sens que tout ce que je fais me fatigue – I feel that everything I do tires me. (J'ai eu l'impression que toute action me demandait un effort - I felt that everything I did was an effort)	SRA
8	J'ai plein d'espoir dans l'avenir – I feel hopeful about the future. (J'ai été confiant(e) en l'avenir - I felt hopeful about the future)	PA[†,*]
9	Je sens que ma vie est ratée – I feel that my life is a failure (J'ai pensé que ma vie était un échec - I thought my life had been a failure)	DA[†]
10	J'ai peur – I am afraid. (Je me suis senti(e) craintif(ve) - I felt fearful)	DA
11	Je dors mal – I sleep badly. (Mon sommeil n'a pas été bon - My sleep was restless)	SRA[†]
12	Je suis heureux(se) – I am happy. (J'ai été heureux(se) - I was happy)	PA[†,*]
13	J'arrive à faire plein de choses – I can do a lot of things. (J'ai manqué d'entrain - I could not get "going")	SRA[†,*]
14	Je me sens seul(e) – I feel lonely. (Je me suis senti(e) seul(e) - I felt lonely)	DA[†]
15	Les gens sont gentils – People are nice. (Les autres ont été hostiles envers moi - People were unfriendly)	INT[†,*]
16	Je profite de la vie – I enjoy life. (J'ai profité de la vie - I enjoyed life)	PA[†,*]
17	Je pleure beaucoup – I cry often. (J'ai eu des crises de larmes - I had crying spells)	DA
18	Je me sens triste – I feel sad. (Je me suis senti(e) triste - I felt sad)	DA[†]
19	Je sens que les gens m'aiment bien – I feel that people like me. (J'ai eu l'impression que les gens ne m'aimaient pas - I felt that people disliked me)	INT[†,*]
20	Je parle moins que d'habitude – I talk less than usual. (J'ai parlé moins que d'habitude - I talked less than usual)	SRA

DA: depressive affect; PA: positive affect; SRA: somatic and retarded activity; INT: interpersonal.
[†] Items retained in the final version.
[*] Reversed score.

In the regular CES-D, the answers to each item are given on a four-point Likert rating scale and the participant is asked to indicate the frequency with which he/she experienced the corresponding symptom during the last week: 0 = rarely or none of the time (less than days); 1 = some or little of the time (1–2 days); 2 = occasionally or a moderate amount of the time (3–4 days); 3 = most or all of the time (5–7 days). From the 20 items, 4 are worded in the positive direction to break possible responses tendencies and to assess positive affects. The score of the DA, PA, SRA and INT subscales range from 0 to 21, 0 to 12, 0 to 21 and 0 to 6, respectively. The overall scale score of depression can thus vary from 0 to 60, with higher scores indicating a greater severity of depressive symptoms. However, since this instrument was not originally designed for adolescents with ID, the items content and phrasing, and the response format of the regular CES-D may present several problems for youths with ID (Finlay & Lyons, 2001). In order to better adapt self-report questionnaires to populations with ID, Finlay and Lyons (2001) highlighted several potential solutions. First, the item content and phrasing can be improved by: (i) verifying participants' understanding of the items before proceeding with the answer, which may prove cumbersome in epidemiologic studies but highly useful in a preliminary study designed to verify the appropriateness of those items; (ii) keeping the item structure simple and avoiding long and complex sentences; (iii) using negative formulations of words but avoiding adding *no* or *not* to positive phrasings; (iv) using significant events markers for time related questions; and (iv) using concrete situations, events or vocabulary, while avoiding technical vocabulary, psychiatric symptoms and abstract or general concepts that may prove difficult to understand. Second, Finlay and Lyons (2001) suggested including pictorial representations of the response alternative provided by the Likert-type scale in order to render it easier to understand for youths with ID. This recommendation is not new and have been previously implemented by several researchers interested in the assessment of individuals with ID (e.g. March, 1992; Sigelman & Budd, 1986; Wadsworth & Harper, 1991; cited by Finlay

& Lyons, 2001). This proposal was recently tested in Hartley and MacLean's (2006) review of 51 published studies using Likert-type scales and/or pictorial representation among individuals with ID. Results from this study showed that the Likert-type scales with pictorial representations of response alternatives produced the best factor validity and reliability, and higher response rates among people with ID – especially with people presenting moderate to profound ID. According to Hartley and MacLean (2006), these results are due to the fact that pictorial representations help people with ID in distinguishing the subtle differences among responses provided by the Likert-type scale.

Considering these issues, the purpose of the first study was to verify the content and phrasing clarity of the CES-D in a sample of adolescents with ID and to develop two alternative response formats that would be tested in a second study: one using a Likert-type answer scale and one using a graphically illustrated Likert scale. The second study sought to examine the factor validity and reliability (internal consistency and temporal stability) of both forms of the CES-D-ID and to assess the measurement and latent mean invariance of the CES-D-ID across gender and age categories.

2. Materials and methods

2.1. Participants and procedures

Study 1. The participants were 20 adolescents (10 boys, M_{age} = 14.56 years, S.D.$_{age}$ = 2.13; 10 girls, M_{age} = 14.60 years, S.D.$_{age}$ = 2.12), aged between 12 and 18 years (M_{age} = 14.58 years, S.D.$_{age}$ = 2.06) and identified as having mild to moderate ID level by the Departmental Commission for the Right of Self-sufficiency of People with Disabilities (DCRSPD). All of these adolescents had an Intellectual Quotient (IQ) within the range of 70–35, were limited in their adaptive behavioral skills (Luckasson et al., 1992), attended full time one of two specialized school for ID adolescents. All parents and participants gave written informed consent (none of the youths either declined to participate or dropped out of the study). Items from the original French version of the CES-D were read aloud by the interviewer and the adolescents were then asked whether they understood the sentence, the format of delivery and the response alternatives. The suitability of the format, the content and the phrasing were then more directly probed with open-ended questions, such as: What is rarely or none of the time; some of a little of the time; occasionally or a moderate amount of the time; most or all of the time? What is being restless? What are crying spells?

Study 2. Participants were 169 adolescents (111 boys, M_{age} = 14.21 years, S.D.$_{age}$ = 1.48; 58 girls, M_{age} = 14.05 years, S.D.$_{age}$ = 1.38), aged between 12 and 18 years (M_{age} = 14.15 years, S.D.$_{age}$ = 1.44) and identified as having mild to moderate ID level by the DCRSPD (IQ between 70 and 35 and limited adaptive behavioral skills). On the basis of the adolescents' current educational placement, two separate groups were identified: (i) adolescents schooled full time in a regular school within a self-contained class with other adolescents with ID (n = 130); and (ii) adolescents enrolled full time in a specialized school with other adolescents with ID (n = 39). This sample was drawn from seven regular schools and two specialized establishments that agreed to participate. All parents and participants gave written informed consent and none declined to participate or dropped out of the study.

The two versions (i.e. Likert and Likert-graphical rating scales) of the CES-D-ID developed in study 1 were administered in the same day, with a 1-h interval, to all participants in quiet classroom conditions in classes of up to 12 adolescents. As in study 1, items were read aloud by the interviewer and the adolescents were then asked to answer directly on the questionnaire. In case of inability to directly respond to the questionnaire, the participants were asked to point to the interviewer his/her answer. Additionally, 13 adolescents (10 boys and 3 girls; M_{age} = 16.54 years, S.D.$_{age}$ = 1.56) with mild to moderate ID were retested after an 11-day period.

2.2. Measures

Three versions of the CES-D were used in the current series of studies. In the first study, the previously described original four-point Likert scale French version was used (Fuhrer & Rouillon, 1989). In the second study, two adaptations of the CES-D-ID (i.e. items wording and phrasing was simplified while retaining the original meaning) were used, one with a modified Likert-type answer scale and one using a graphically illustrated Likert scales. The Likert-graphical rating scale is illustrated in Fig. 1 and was developed by combining the Likert-type answer scale and the Wong-Baker facial pain rating scale (Wong & Baker, 1988).

2.3. Data analysis

In the second study, analyses were conducted in several stages. Because of the significant multivariate non-normality of the data (the normalized kurtosis coefficient was of 29.34), analyses were performed using Bootstrapped Maximum Likelihood estimation with AMOS 7.0 (Arbuckle, 2006). Thus, all fit values provided in this study were based upon the Bollen-Stine bootstrap p-value and bootstrap adjusted chi-square and goodness of fit indexes (Yuan & Hayashi, 2003).

In the first stage, six Confirmatory Factor Analytic (CFA) models commonly tested and reported in the literature (for reviews see: Perreira, Deeb-Sossa, Harris, & Bollen, 2005; Shafer, 2006; Sheehan, Fifield, Reisine, & Tennen, 1995) were tested on both versions of the CES-D-ID. The first model a priori hypothesized that answers to the CES-D-ID could be explained by a single factor of depression, and that the measurement error terms would be uncorrelated. The second model a priori

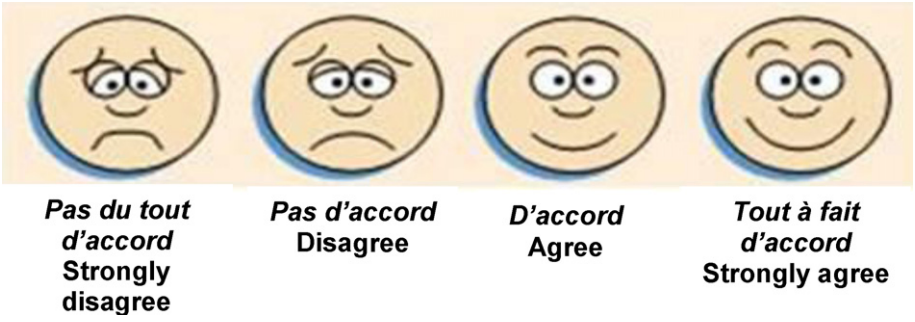

Fig. 1. Likert-graphical scale response.

hypothesized that the answers to the CES-D-ID could be explained by two factors (DA-SRA; PA-INT), that each item would have a non-zero loading on the CES-D-ID factor it was designed to measure and zero loadings on all other factors, that the two factors would be correlated, and that the measurement error terms would be uncorrelated. The third and fourth models both a priori hypothesized that the answers to the CES-D-ID could be explained by two distinct three factors models (3A: PA-DA, SRA, INT; 3B: DA-SRA, PA, INT), that each item would have a non-zero loading on the CES-D-ID factor it was designed to measure and zero loadings on all other factors, that the three factors would be correlated, and that the measurement error terms would be uncorrelated. The fifth model hypothesized that the answers to the CES-D-ID could be explained by the four a priori factors (DA, PA, SRA and INT), that each item would have a non-zero loading on the CES-D-ID factor it was designed to measure and zero loadings on all other factors, that the four factors would be correlated, and that the measurement error terms would be uncorrelated. Finally, the sixth model hypothesized that answers to the CES-D-ID could be explained by the four a priori first-order factors, that each item would have a non-zero loading on the lower-order CES-D-ID factor it was designed to measure and zero loadings on all other factors, that the four first-order factors would load on a single second-order factor representing depression, and that measurement errors would be uncorrelated.

The CES-D version (Likert or Likert-graphical) with the best fitting solution will be retained. In case of inadequate fit of all of the estimated models, the best model and version (Likert or Likert-graphical) of the CES-D-ID with 20 items from Radloff (1977) will be used as a starting point and modified on the basis of an examination of the model parameters (factor loadings, uniquenesses, standardized residuals covariances, etc.) and modification indices. The retained CFA model was then re-estimated to determine whether the modification resulted in an improved fit. This process was continued until a reasonable model can be generated as indicated by the absolute and incremental fit indices. Finally, the temporal stability of the best version of the CES-D-ID was also estimated using test–retest reliability correlations and intraclass correlations (Shrout & Fleiss, 1979; ICC) with their 95% confidence interval (95% CI) for scale scores uncorrected for measurement errors on the data from the 13 adolescents who were re-tested after an 11-day period.

In the two following stages, the invariance of the six-factor of the final retained CFA model was used to test the across age category (Stage 2) and gender (Stage 3). Regarding the reduced number of participants in several of the ages categories, this variable was dichotomized into early (12–14 years old, $n = 105$) and middle-late (15–18 years old, $n = 64$) adolescence (Simmons & Blyth, 1987; Steane & Heald, 1987). The measurement invariance tests were directly conducted across age categories and gender groups in the sequential order recommended by Meredith (1993). In these analyses, each model was compared to the preceding one that served as a reference model (Vandenberg & Lance, 2000).

Assessment of fit for the CFA models was based on multiple indicators (Byrne, 2005; Hu & Bentler, 1999; Vandenberg & Lance, 2000): the Chi-square statistic (χ^2), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Standardized Root Mean square Residual (SRMR), the Root Mean Square Error of Approximation (RMSEA), the 90% confidence interval of the RMSEA and the Expected Cross-Validation Index (ECVI). Values greater than .90 for CFI and TLI are considered to be indicative of adequate model fit (Byrne, 2005; Hu & Bentler, 1999; Vandenberg & Lance, 2000), although values approaching .95 are preferable. Values smaller than .08 or .06 for the RMSEA and smaller than .10 and .08 for the SRMR support respectively acceptable and good model fit (Hu & Bentler, 1999; Vandenberg & Lance, 2000). Concerning the RMSEA 90% CI, values of less than .05 for the lower bound (left side) and less than .08 for the upper bounds (right side) or containing 0 for the lower bound and less .05 for the upper bounds (right side) indicate respectively acceptable and good model fit (MacCallum, Browne, & Sugawara, 1996). Because χ^2 difference tests cannot be legitimately performed on non-nested models, the Expected cross-validation index (ECVI) was also used to compare non-nested models based on different answer scales (Browne & Cudeck, 1993). In comparing different models, the model with the lowest ECVI is deemed to provide the best representation of the data (Browne & Cudeck, 1993). The factor loadings, square multiple correlations, standard errors and t values were inspected for appropriate sign and/or magnitude. Critical values for the tests of multi-group invariance in CFAs models were evaluated by the examination of several criterions: χ^2 difference tests and changes in CFI, TLI and RMSEA (Chen, 2007; Cheung & Rensvold, 2002; Marsh, 2007; Vandenberg & Lance, 2000). A CFI difference of .01 or less and a RMSEA difference of .015 or less between a reference model and the following model indicate that the measurement invariance hypothesis should not be rejected. Changes in fit indices that incorporate a correction for parsimony (ΔRMSEA and ΔTLI) should also be inspected as it is possible for these indices to improve with the addition of invariance constraints (Marsh,

2007). Finally, the reliability was computed from the model's standardized parameters, using the formula by McDonald (1970): $\omega = (\sum|\lambda i|)^2/([\sum|\lambda i|]^2 + \sum\delta ii)$ where λi are the factor loadings and δii are the error variances.

3. Results

3.1. Study 1: format and content evaluation of the CES-D

Items from the French and English original versions of the CES-D are reported in parentheses in Table 1. Analyses of the items' content with the ID adolescents from the sample reveal that many words were not understood (e.g. item 1: *I was bothered*; item 5: *I have trouble keeping my mind*; item 8: *I felt hopeful*; item 10: *I feel fearful*, item 17: *I had crying spells*. . .). In all of these cases the wording of the problematic items was simplified while retaining their original meaning. This adapted French version is reproduced in Table 1 outside of the parentheses, along with their English equivalents in italics. Moreover, the interview with ID adolescents also revealed that the response format was problematic and required modification. Indeed, adolescents with ID tend to have trouble remembering past affective states or generalizing such states across a period (Finlay & Lyons, 2001; Finlay & Lyons, 2002). Thus, the decision was made to modify the CES-D instructions and answer scale to refer to the present state of the examinee rather than the frequency to his or her state "over the last week". Some of the preceding authors who developed depression questionnaires for youths with ID appear to have reach similar conclusions and choose to rely on a present state evaluation (e.g. Dagnan et al., 2008; Esbensen et al., 2005). All participants from this sample were then gathered again 1 week later and presented with the adapted version of the questionnaire (the CES-D-ID). This time, all adolescents clearly understood the items of the adapted version and both of the response formats (Likert or Likert-graphical) appeared adequate. We further did probe those whose affective state had changed and indeed, they showed difficulties in remembering the state they previously were in the past week.

3.2. Study 2: factor validity of the CES-D-ID

Stage 1. The goodness-of-fit statistics and factor loadings-uniquenesses of the six CFA measurement models of the two CES-D-ID versions are displayed in Tables 2 and 3. The results from these six CFA models (i.e. one-factor, two-factor, three-factor A and B, four-factor, four first-order factor and one second-order factor) showed for both versions (Table 2): (i) significant bootstrapped χ^2 values, (ii) CFI and TLI under .90, and (iii) RMSEA and SRMR exceeding .08 in most cases, with the exception of the RMSEAs in the correlated four factor model (.057 and .061).

Given the unsatisfactory fit of these various models, they were all deemed inadequate for French adolescents with ID. Nevertheless, analyses revealed that: (i) the Likert-graphical version provided the best results (i.e. this version presented the lowest scores on the ECVI, RMSEA and SRMR and the highest scores on the CFI and TLI) compared to the Likert-type scale version; and (ii) the best fit indices were obtained in both versions for the original four-factor model. Consequently, a reduced model was developed on the basis of an examination of the properties of the items from the best model (i.e. first-order four correlated factor model) of the Likert-graphical version. This examination revealed that some items: (i) presented potentially important correlated uniquenesses or cross loadings according to the modification indices; and/or (ii) had large standardized residuals covariances (i.e. $\geq|2.58|$). Thus, as illustrated in Table 3, the most problematic items (i.e. 2, 4, 7, 10, 17, 20) were removed. After deleting those six items, a new 14-item version of the Likert-graphical CES-D-ID was obtained. For this version, the results from the CFA showed significant bootstrapped χ^2 values (Table 2), but CFI and TLI values that exceeded .95 and SRMR and RMSEAs values that were lower than .06. All loadings in this CFA model were significant and substantial (Table 3).

In this 14-item version, the average scale score for the whole sample on the DA, PA, SRA and INT subscales were respectively of 6.02 (S.D. = 4.17; range = 0–15), 7.18 (S.D. = 1.90; range = 0–9), 5.59 (S.D. = 2.88; range = 0–12), 4.44 (S.D. = 1.61; range = 0–6). These scales also presented acceptable reliability (ω) coefficients of .83 for DA, .79 for PA, .79 for SRA, and .67 for INT. Latent variables correlations were in most cases statistically significant and ranged from .012 (for DA/PA) to .849 (for DA/SRA). Finally, the test–retest reliability correlation coefficients were modest to satisfactory: (i) $r^{tt} = .64$ ($p \leq .01$) for DA, .87 ($p \leq .01$) for PA, .39 ($p \geq .05$) for SRC and .91 ($p \leq .01$) for INT; (ii) ICC = .64 (95% CI: 17 to 87; $p \leq .01$) for DA, .87 (95% CI: .63 to .96; $p \leq .01$) for PA, .39 (95% CI: −19 to.76; $p \geq .05$) for SRC and .88 (95% CI: .64 to .96; $p \leq .01$) for INT.

Stage 2. The results from the measurement invariance tests are reported in Table 2. The results from the four first steps of the age-related measurement invariance tests (hypothesis A–D) showed that (i) none of the χ^2 difference tests were significant; (ii) the CFI, TLI, SRMR and RMSEA indicated adequate model fit; and (iii) the ΔCFI and ΔRMSEA never exceeded .01 and .015. The fifth level of measurement invariance (hypothesis E: equality constraints on the factors' variances–covariances) resulted in a significant bootstrap χ^2, modest goodness of fit-indices, and a ΔRMSEA that did not exceed .015. However, this model resulted in a significant $\Delta\chi^2$, and a ΔCFI exceeding .01 and an important decrease (.040) in TLI. Modification indices for this model were inspected and suggested that the age-group equality constraint should be relaxed for the covariance between the positive affect and interpersonal factors, and between the depressive affect and positive affect factors. This showed that the strength of association between these factors was stronger for the middle-late adolescent subgroup. The sixth model (hypothesis E′) freely estimated these parameters across age groups while keeping the other equality constraints and provided evidence of partial invariance of the factor variances/covariance matrix (i.e. non-

Table 2
Goodness-of-fit statistics of CES-D-ID models.[a]

Stages	Version	Description	χ^2(B-S)	df	CFI	TLI	SRMR	RMSEA	RMSEA 90% CI	ECVI	$\Delta\chi^2$(df)	ΔCFI	ΔTLI	ΔRMSEA
1	Likert	One factor	199.99*	170	.588	.540	.128	.099	.088–.110	3.15				
		Two factors	196.82*	169	.819	.797	.096	.066	.053–.078	2.22				
		Three factors A	195.37*	167	.653	.605	.123	.091	.080–.103	2.90				
		Three factors B	194.63*	167	.816	.816	.094	.063	.049–.075	2.16				
		Four factors	190.69*	164	.850	.826	.092	.061	.047–.074	2.13				
		Four 1st-order factors and one 2nd-order factor	190.65*	166	.695	.650	.144	.086	.074–.098	2.74				
1	Likert-graphic	One factor	207.66*	170	.701	.666	.101	.079	.067–.091	2.55				
		Two factors	204.23*	169	.812	.789	.090	.063	.050–.076	2.16				
		Three factors A	204.40*	167	.752	.718	.097	.073	.060–.085	2.39				
		Three factors B	201.60*	167	.816	.791	.088	.063	.049–.075	2.16				
		Four factors	198.50*	164	.851	.827	.083	.057	.043–.070	2.06				
		Four 1st-order factors and one 2nd-order factor	201.24*	166	.827	.802	.144	.089	.047–.074	2.13				
		Alternative 4-factor	87.05	71	.974	.966	.059	.029	.000–.055	.89				
2	Likert-graphic	A - No invariance (age)	169.62*	142	.935	.916	.082	.034	.003–.052	1.830				
		B - λs invariant (age)	182.03*	152	.929	.915	.090	.034	.003–.052	1.785	12.41 (10)	−.006	−.001	.000
		C - λs, τs invariant (age)	192.29	162	.929	.920	.090	.033	.000–.051	2.062	10.26 (10)	.000	+.005	+.001
		D - λs, τs, δs invariant (age)	203.48	176	.935	.933	.090	.031	.000–.048	1.961	11.19 (14)	+.006	+.013	+.002
		E - λs, τs, δs, ξs/φs invariant (age)	232.28*	186	.891	.893	.089	.039	.019–.054	2.014	28.80 (10)*	−.044	−.040	+.008
		E′ - λs, τs, δs, partial ξs/φs invariant (age)	214.64	184	.928	.928	.094	.032	.000–.048	1.932	11.16 (8)	−.007	−.005	+.001
		F - λs, τs, δs, ξs, ηs invariant (age)	220.14	188	.924	.927	.095	.032	.000–.048	1.917	5.50 (4)	−.004	−.001	.000
3	Likert-graphic	A - No invariance (gender)	173.88*	142	.922	.900	.087	.037	.010–.054	1.856				
		B - λs invariant (gender)	182.39*	152	.926	.911	.093	.035	.000–.047	1.787	8.50 (10)	+.004	+.011	+.002
		C - λs, τs invariant (gender)	193.64*	162	.922	.913	.092	.034	.005–.051	2.070	11.25 (10)	−.004	+.002	+.001
		D - λs, τs, δs invariant (gender)	206.15	176	.926	.924	.095	.032	.000–.049	1.977	12.51 (14)	+.004	+.011	+.002
		E - λs, τs, δs, ξs/φs invariant (gender)	220.60*	186	.915	.917	.110	.033	.007–.049	1.944	14.45 (10)	−.011	−.007	+.001
		E′ - λs, τs, δs, partial ξs/φs invariant (gender)	212.48	185	.933	.934	.110	.030	.000–.047	1.907	6.33 (9)	+.007	+.010	−.002
		F - λs, τs, δs, ξs, ηs invariant (gender)	224.93*	189	.912	.915	.112	.034	.009–.050	1.934	12.45 (4)*	−.021	−.019	−.004

CFA: confirmatory factor analytic model; χ^2(B-S): Bollen-Stine chi-square; df: degrees of freedom; CFI: comparative fit index; TLI: Tucker-Lewis index; SRMR: standardized root mean square residual; RMSEA: root mean square error of approximation; ECVI: expected cross-validation index; RMSEA 90% CI = 90% confidence interval for the RMSEA point estimate; λ: factor loading; τ: intercept; δ: uniquenesses; ξ: factor variance; η: factor mean; $\Delta\chi^2$: change in goodness-of-fit χ^2 relative to the preceding model; Δdf: change in degrees of freedom relative to the preceding model; ΔCFI: change in comparative fit index relative to the preceding model; ΔTLI: change in Tucker-Lewis index relative to the preceding model; ΔRMSEA: change in root mean square error of approximation relative to the preceding model.

[a] Bootstrapped goodness of fit indexes are reported in this table because of the significant multivariate non-normality within these data.

* $p < .05$.

Table 3
CFA's factor loadings and uniquenesses for the total sample.

Factor	Item no.	Likert version	Likert-graphical version	
		Four factors	Four factors	Alt. four factors
		$\lambda(\delta)$	$\lambda(\delta)$	$\lambda(\delta)$
DA	3	.556(.309)†	.575(.331)†	.604(.364)†
	6	.680(.463)	.623(.388)	.647(.419)
	9	.640(.409)	.696(.485)	.709(.502)
	10	.467(.218)	.440(.194)	
	14	.472(.223)	.567(.321)	.545(.297)
	17	.654(.428)	.687(.472)	
	18	.638(.407)	.739(.546)	.695(.483)
PA	4	.379(.143)†	.371(.138)†	
	8	.508(.258)	.505(.255)	.433(.187)†
	12	.748(.560)	.523(.274)	.603(.364)
	16	.643(.413)	.421(.177)	.415(.172)
SRA	1	.567(.321)†	.519(.270)†	.495(.245)†
	2	.339(.115)	.263(.069)	
	5	.556(.309)	.592(.351)	.557(.310)
	7	.129(.017)☆	.194(.038)☆	
	11	.619(.383)	.579(.336)	.619(.383)
	13	.367(.135)	.384(.147)	.366(.134)
	20	.001(.000)☆	.164(.027)☆	
INT	15	.537(.289)†	.566(.320)†	.568(.322)†
	19	.823(.677)	.597(.356)	.595(.354)

λ: factor loading; δ: uniquenesses; DA: depressive affect; PA: positive affect; SRA: somatic and retarded activity; INT: interpersonal.
☆ Loadings that were non-significant ($p > .05$).
† Item that was set to be 1.0.

significant $\Delta\chi^2$; ΔCFI $< .01$; ΔRMSEA $< .015$). Finally, the last model (hypothesis F) supported the invariance of the latent means across age groups.

Stage 3. The four first steps of gender-based invariance testing (hypothesis A–D) resulted in significant bootstrap χ^2 but non-significant $\Delta\chi^2$, acceptable goodness of fit-indices, ΔCFIs that did not exceed .01 and ΔRMSEAs that did not exceed.015. The fifth level of measurement invariance (hypothesis E: equality constraints on the factors' variances–covariances) resulted in a significant bootstrap χ^2, acceptable goodness of fit-indices, non-significant $\Delta\chi^2$ and a ΔRMSEA that did not exceed .015. However, this model resulted in a ΔCFI exceeding .01. Modification indices for this model were inspected and suggested that the gender-group equality constraint should be relaxed for the variance of the PA factor. This showed that the PA factor had a greater level of variability in the girls' subgroup. The sixth model (hypothesis E′) freely estimated this parameter across gender while keeping the other factor variance/covariance constraints and provided evidence of partial invariance of the factor variances (i.e. non-significant $\Delta\chi^2$; ΔCFI $< .01$; ΔRMSEA $< .015$). Finally, the seventh model (hypothesis F) did not support the invariance of the latent means across gender (i.e. significant $\Delta\chi^2$, ΔCFI $> .01$ and ΔTLI $> .015$). Post hoc probing of this difference revealed that boys (Latent mean = .35, $t = 2.62$, $p < .05$, $d = .87$) presented significantly higher scores of PA than girls (Latent mean fixed to zero to act as reference). As these items were inverted before the analysis, this means that girls have more positive affects than boys.

4. Discussion

As recommended by Finlay and Lyons (2001), the objectives of the first study were to verify the clarity of the original CES-D items, instructions and response format within a sample of adolescents with ID and to use these results to develop two alternative response format versions of this questionnaire that were easy to understand for this population. The results led to the development of a Likert-type answer scale and of a Likert-graphical answer scale of the CES-D-ID (showing smiling and unsmiling faces). The resulting instrument (with both forms of answer scales) proved to be far easier to understand for ID adolescents. However, before the results from the second study can be considered, it should be noted that the results from this first study confirm, as suggested by Finlay and Lyons (2001), that the direct transposition of the CES-D to adolescents with ID without the preliminary verification of its applicability might be highly problematic.

The objectives of the second study were to directly evaluate the psychometric properties of both versions of the CES-D-ID in a sample of adolescents with ID. First, the present findings indicated that none of the various models tested (one-factor, two-factor, three-factor A and B, four-factor, four first-order factor and one second-order factor) in both of the versions (Likert and Likert-graphical) provided a satisfactory fit to the data in the present sample, due to the presence of a few problematic items. Nevertheless, the results revealed that the Likert-graphical version and the original four-factor model provided a slightly better fit to the data than the other alternatives that were considered. Consequently this model was retained for subsequent analysis and, after excluding the items deemed to be most responsible for the model

misspecifications (2, 4, 7, 10, 17, 20), a final CFA model was estimated with the remaining items. This 14-item four-factor model fitted the data well and presented: modest to elevated latent factor correlations, satisfactory internal consistency coefficients and test–retest reliability correlation coefficients that were of similar magnitude to those found by Radloff (1977). These results again confirmed the conclusions from Finlay and Lyons (2001) review in showing that it is important to adapt questionnaires before using them in populations with ID. Overall, the results provided evidence of the reliability and construct validity of an alternative 14-item CES-D-ID. Researchers can thus be confident in the use of this instrument among French adolescents with ID.

Subsequent CFAs analyses were performed with the objective of assessing the measurement and latent mean invariance of the 14-item French version of the CES-D-ID associated with a Likert-graphical scale, across gender and age categories. They showed that the measurement model and latent mean structure was fully invariant (configuration, loadings, intercepts and errors; Meredith, 1993) across age categories and gender. Researchers can thus be quite confident that this instrument would measure the same construct with the same precision in these groups – and thus that meaningful group comparisons can be conducted with this instrument. In fact, such between group comparisons were directly conducted as a follow up to these more classic tests of between group invariance (Vandenberg & Lance, 2000) and revealed a lack of age-based differences in depression. This finding is consistent with previous studies among students with mild ID (e.g. Heiman, 2001), but also in contrasts with the results from previous studies based on youths without ID, which show that average levels of depression tend to increase with age (for a review see: Tram & Cole, 2006). Tram and Cole (2006) explain this observation by the fact that this developmental period is characterized by an accumulation of important biopsychosocial developmental changes that exert a significant impact on youth, In addition, previous studies show that these age-related differences generally emerge in the context of a two-way interaction involving gender, such as a significant increase in depression symptoms is generally observed among girls, but not boys (e.g. Angold & Costello, 2001; Angold, Erkanli, Silberg, Eaves, & Costello, 2002; Cole et al., 2002; Hankin et al., 1998). Unfortunately, we did not have enough participants to test this possibility across four gender-by-age groups. However, whether these multiple transitions exert the same type of impact through the same interactive mechanisms in youths with ID remains an open question and may explain the diverging results we found in this study. This issue should thus be more precisely examined in future studies. Finally, this study relies on a cross-sectional sample that precludes the verification of the developmental change of depression symptoms during the adolescent years. It is thus unknown whether depression symptoms of adolescents with ID increase or not during adolescence. Consequently, future research should investigate more specifically this issue in the context of a longitudinal study involving a large sample of youths with ID.

Additionally, these results also revealed that girls, when compared to boys, presented a higher level of variability on the PA factor, as well as lower mean on this factor (illustrating more positive affects than boys). These results thus fail to support the very few studies of adolescents with ID which have found higher depressive scores in girls than in boys (e.g. Heiman, 2001; Heiman & Margalit, 1998; Reynolds & Miller, 1985). One potential explanation of those results would be that those mean-level differences noted by these studies may in fact have been an artifact of their reliance on manifest scale scores rather than latent variables, precluding the control for measurement errors in gender group comparisons. We do not believe that such a claim can be made on the basis of this single study, especially since we found evidence of strict measurement invariance across gender groups. However, these results clearly underline the need for future studies to devote a great deal more attention to potential measurement biases in self-report instruments designed to measure depression among ID adolescents. Moreover, it should be noted that this is not the first time that the absence of gender based-differences in self-reported instrument assessing the severity of depressive symptoms (rather than the prevalence of a depression diagnosis). Indeed, additional studies examining gender-based differences in depressive symptoms in individuals without ID with various methodologies (e.g., observed score methods, latent variable models, etc.) and depression instruments (CES-D, Goldberg Anxiety and Depression, Depressive Experience Questionnaire, Beck Depression Inventory, etc.) also reported both a lack of gender-based latent mean differences and measurement invariance (Christensen et al., 1999; Morin, Janosz, & Larivée, 2009; Moullec et al., 2011; Salokangas, Vaahtera, Pacriev, Sohlman, & Lehtinen, 2002; Steer, Beck, & Brown, 1989; Wenzel, Steer, & Beck, 2005). However surprising these results must be given the fact that gender-based differences in depression diagnosis are well documented (for reviews see: Angold & Costello, 2001; Hyde, Mezulis, & Abramson, 2008; Piccinelli & Wilkinson, 2000), they point to the fact that it is thus unknown whether girls really do present higher levels of depressive symptoms than boys. In addition, is also remains unknown whether these results more or less well documented in the general population will really fully generalize to populations with ID (Hodapp & Dykens, 2009). Consequently, as suggested by Tsakanikos, Bouras, Sturmey, and Holt (2006) and Lunsky (2003), respectively, the research on depression and gender differences in adolescents with ID "*have received a moderate amount of attention*" (p. 583), they are thus "*clearly at an early stage*" (p. 418).

In addition to what was previously noted, three limitations of the present studies must be taken into account in the interpretation of the results. First, the factor structure and measurement invariance analyses of the French CES-D-ID were based on a mixed (boys and girls) sample of a priori nonclinical adolescents with ID, which might not be considered representative of the French adolescent population, with or without ID. In addition, the present study rely on only two samples of adolescents with ID, only one of which was used to investigate the psychometric properties of this instrument. Consequently, these results must be interpreted with caution, especially given the fact that data-driven modifications were conducted on the instrument. Replication is thus needed to ensure that the fit of the final model does not simply reflect capitalization on chance. More work, thus, needs to be done to explore whether the French 14-item version of the CES-D-ID provides similar results with another sample of French adolescents with ID.

Second, it was impossible to obtain an "*objective*" and/or validated assessment of depression for the participants with which to compare the obtained CES-D-ID scores since to our knowledge no valid French version of semi-structured interviews, self report instruments, or third person protocols were available to assess the depression of adolescents with ID. This is an important part of the criterion-related validity of the CES-D-ID that remains to be examined in later studies and given the documented superiority of third person reports of depressions in youths with ID. For the moment, the use of the CES-D-ID as a diagnostic tool would thus be highly premature. In a related way, we did not evaluate the measurement invariance of this instrument across clinical and non-clinical ID populations. This was in part based by the initial objective of validating an epidemiological tool. However, this also means that the psychometric properties of this instrument in a clinical population remain unknown. Examining this question should be a future research priority, especially since the CES-D was initially designed as a screening tool designed to identify clinical depression in community samples.

Third, the present series of studies were based on a French sample of adolescents with ID, and the use of this instrument should thus be limited to samples with the same characteristics. Clearly, before the generalizability of this instrument to other cultural background (e.g. French speaking Maghreb adolescents) or linguistic groups (e.g. Spanish or English speaking adolescents) can be systematically investigated in other studies, its cross-cultural or linguistic use cannot be recommended. Therefore, examining the measurement and latent mean invariance of the 14-item version of the CES-D-ID associated with a Likert-graphical response scale across diverse cultural or linguistic groups should also represent a future research priority for the conduct of cross-cultural epidemiologic studies.

5. Conclusion

In conclusion, it is important that the present study be viewed as a preliminary step in the validation of an epidemiologic self-report depression assessment instrument available for youths with ID. Moreover, despite that the psychometric properties of the French 14-item version of the CES-D-ID associated with a Likert-graphical response scale found to be adequate and acceptable, they must be cross-validated in a second independent sample of French adolescents with ID. Nevertheless, based on the present results and on their limitations, it would be premature at this time to recommend the use of this instrument with clinical ID adolescent samples or with other cultural or linguistic groups. Therefore, this instrument may be used in research, but it may be premature to recommend it yet as a clinical instrument before additional studies are conducted.

Acknowledgements

The authors would like to thank the specialized classes and specialized establishment for their invaluable assistance in data collection, and second author (AJSM) also wishes to thank Pablo Sillon for his invaluable advices at previous stages of this research. This manuscript was designed and prepared in part while AJSM was a visiting scholar at the University of the Mediterranean (Aix-Marseille II).

References

Angold, A., & Costello, E. J. (2001). The epidemiology of depression in children and adolescents. In I. M. Goodyer (Ed.), *The depressed child and adolescent* (2nd ed., pp. 143–178). Cambridge, UK: Cambridge University Press.

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

Arbuckle, J. (2006). *AMOS 7.0 user's guide*. Chicago, IL: SPSS.

Argus, G. R., Terrya, P. C., Bramstona, P., & Dinsdale, S. L. (2004). Measurement of mood in adolescents with intellectual disability. *Research in Developmental Disabilities, 25*, 493–507.

Bebbington, P. E., Dunn, G., Jenkins, R., Lewis, G., Brugha, T., Farrell, M., et al. (1998). The influence of age and sex on the prevalence of depressive conditions: Report from the National Survey of Psychiatric Morbidity. *Psychological Medicine, 28*, 9–19.

Benavidez, D. A., & Matson, J. L. (1993). Assessment of depression in mentally retarded adolescents. *Research in Developmental Disabilities, 14*, 179–188.

Benson, B. A., & Ivins, J. (1992). Anger, depression and self-concept in adults with mental retardation. *Journal of Intellectual Disability Research, 36*, 169–175.

Borthwick-Duffy, S.A. (1994). Epidemiology and prevalence of psychopathology in people with mental retardation. *Journal of Consulting and Clinical Psychology, 62*, 17–27.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment, 85*, 17–32.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement. *Structural Equation Modeling, 14*, 464–504.

Cheung, C., & Bagley, C. (1998). Validating an American scale in Hong-Kong: The Center for Epidemiological Studies Depression scale (CES-D). *Journal of Psychology, 132*, 169–186.

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

Christensen, H., Jorm, A. F., Mackinnon, A. J., Korten, A. E., Jacomb, P. A., Henderson, A. S., et al. (1999). Age differences in depression and anxiety symptoms: A structural equation modeling of data from a general population sample. *Psychological Medicine, 29*, 325–339.

Cole, D. A., Tram, J. M., Martin, J. M., Hoffman, K. B., Ruiz, M. D., Jacquez, F. M., et al. (2002). Individual differences in the emergence of depressive symptoms in children and adolescents: A longitudinal investigation of parent and child reports. *Journal of Abnormal Psychology, 111*, 156–165.

Cuthill, F. M., Espie, C. A., & Cooper, S. A. (2003). Development and psychometric properties of the Glasgow Depression Scale for people with a Learning Disability: Individual and carer supplement versions. *British Journal of Psychiatry, 182*, 347–353.

Dagnan, D., Jahoda, A., McDowell, K., Masson, J., Banks, P., & Hare, D. (2008). The psychometric properties of the hospital anxiety and depressions scale adapted for use with people with intellectual disabilities. *Journal of Intellectual Disability Research, 52*, 942–949.

Dekker, M. C., & Koot, H. M. (2003). DSM-IV disorders in children with borderline to moderate intellectual disability. I. Prevalence and impact. *Journal of the Academy of Child and Adolescent Psychiatry, 42*, 915–922.
Dekker, M. C., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2002). Emotional and behavioral problems in children and adolescents with and without intellectual disability. *Journal of Child Psychology and Psychiatry, 43*, 1087–1098.
Dershem, L. D., Patsiorkovski, V. V., & O'Brien, D. J. (1996). The use of the CES-D for measuring symptoms of depression in three rural Russian villages. *Social Indicators Research, 39*, 89–108.
Douma, J. C. H., Dekker, M. C., Verhulst, F. C., & Koot, H. M. (2006). Self-reports on mental health problems of youth with moderate to borderline intellectual disabilities. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*, 1224–1231.
Dykens, E. M. (2000). Annotation: Psychopathology in children with intellectual disability. *Journal of Child Psychology and Psychiatry, 41*, 407–417.
Emerson, E. (2003). Prevalence of psychiatric disorders in children and adolescents with and without intellectual disability. *Journal of Intellectual Disability Research, 47*, 51–58.
Esbensen, A. J., Seltzer, M. M., Greenberg, J. S., & Benson, B. B. (2005). Psychometric evaluation of a self-report measure of depression for individuals with mental retardation. *American Journal of Mental Retardation, 110*, 469–481.
Fava, G. A. (1983). Assessing depressive symptoms across cultures: Italian validation of the CES-D self-rating scale. *Journal of Clinical Psychology, 39*, 249–251.
Finlay, W. M. L., & Lyons, E. (2001). Methodological issues in interviewing and using self-report questionnaires with people with mental retardation. *Psychological Assessment, 13*, 319–335.
Finlay, W. M. L., & Lyons, E. (2002). Acquiescence in interviews with people with mental retardation. *Mental Retardation, 40*, 14–29.
Fleming, J. E., & Offord, D. R. (1990). Epidemiology of childhood depressive disorders: A critical review. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*, 571–580.
Fountoulakis, K., Iacovides, A., Kleanthous, S., Samolis, S., Kaprinis, S. G., Sitzoglou, K., et al. (2001). Reliability, validity and psychometric properties of the Greek translation of the Center for Epidemiological Studies-Depression (CES-D) Scale. *BMC Psychiatry, 1*, 3.
Fuhrer, R., & Rouillon, F. (1989). La version française de l'échelle CES-D: Description et traduction de l'échelle d'autoévaluation [The French version of CES-D: Description and translation of the self-report scale]. *Psychiatrie and Psychobiologie, 4*, 163–166.
Gillberg, C., Presson, E., Grufman, M., & Themner, U. (1986). Psychiatric disorders in mildly and severely mentally retarded urban children and adolescents: Epidemiological aspects. *British Journal of Psychiatry, 149*, 68–74.
Gonçalves, B., & Fagulha, T. (2004). The Portuguese Version of the Center for Epidemiologic Studies Depression Scale (CES-D). *European Journal of Psychological Assessment, 20*, 339–348.
Hankin, B. L., Abramson, L. Y., Moffit, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology, 107*, 128–140.
Hartley, S. L., & MacLean, W. E., Jr. (2006). A review of the reliability and validity of Likert scales for people with intellectual disability. *Journal of Intellectual Disability Research, 50*, 813–827.
Hautzinger, M. (1988). Die CES-D Skala: Ein Depressionsmess instrument für Untersuchungen in der Allgemeinbevölkerung [The CES-D scale: A depression-rating scale for research in the general population]. *Diagnostica, 34*, 167–173.
Heiman, T. (2001). Depressive mood in students with mild intellectual disability: Students' reports and teachers' evaluations. *Journal of Intellectual Disability Research, 45*, 526–534.
Heiman, T., & Margalit, M. (1998). Loneliness, depression, and social skills among students with mild mental retardation in different educational settings. *Journal of Special Education, 32*, 154–163.
Hodapp, R. M., & Dykens, E. M. (2009). Intellectual disabilities and child psychiatry: Looking to the future. *Journal of Child Psychology and Psychiatry, 50*, 99–107.
Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
Hurley, A. D. (2006). Mood disorders in intellectual disability. *Current Opinion in Psychiatry, 19*, 465–469.
Hyde, J. S., Mezulis, A. H., & Abramson, L. Y. (2008). The ABCs of depression: Integrating affective, biological, and cognitive models to explain the emergence of the gender difference in depression. *Psychological Review, 115*, 291–313.
Janowky, D. S., & Davis, J. M. (2005). Diagnosis and treatment of depression in patients with mental retardation. *Current Psychiatry Reports, 7*, 421–428.
Kazdin, A. E., Matson, J. L., & Senatore, V. (1983). Assessment of depression in mentally retarded adults. *American Journal of Psychiatry, 140*, 1040–1043.
Kessler, R. C. (2002). Epidemiology of depression. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (pp. 23–42). London, UK: Guilford.
Kiddle, H., & Dagnan, D. (2011). Vulnerability to depression in adolescents with intellectual disabilities. *Advances in Mental Health and Intellectual Disabilities, 5*, 3–8.
Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology. I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *Journal of Abnormal Psychology, 102*, 133–144.
Lindsay, W. R., & Skene, D. D. (2007). The Beck Depression Inventory II and the Beck Anxiety Inventory in People with Intellectual Disabilities: Factor analyses and group data. *Journal of Applied Research in Intellectual Disabilities, 20*, 401–408.
Luckasson, R., Coulter, D. L., Polloway, E. A., Reiss, S., Schalock, R. L., Snell, M. E., et al. (1992). *Mental retardation: Definition, classification and systems of support*. Washington, DC: American Association of Mental Retardation.
Lunsky, Y. (2003). Depressive symptoms in intellectual disability: Does gender play a role? *Journal of Intellectual Disability Research, 47*, 417–427.
Lunsky, Y., & Palucka, A. (2004). Depression in intellectual disability. *Current Opinion in Psychiatry, 17*, 359–363.
MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.
Manikam, R., Matson, J. L., Coe, D. A., & Hillman, N. (1995). Adolescent depression: Relationships of self-report to intellectual and adaptive functioning. *Research in Developmental Disabilities, 16*, 349–364.
March, P. (1992). Do photographs help adults with severe mental handicaps to make choices. *British Journal of Mental Subnormality, XXXVII*, 122–128.
Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 774–798). New York, NY: Wiley.
Masi, G., Brovedani, P., Mucci, M., & Favilla, L. (2002). Assessment of anxiety and depression in adolescents with mental retardation. *Child Psychiatry and Human Development, 32*, 227–237.
Matson, J. L. (1982). The treatment of behavioral characteristics of depression in the mentally retarded. *Behavior Therapy, 13*, 209–218.
Matson, J. L. (1983). Depression in mentally retarded persons: Research findings and future directions. *Journal of Intellectual and Developmental Disability, 9*, 185–190.
Matson, J. L. (1986). Treatment outcome research for depression in mentally retarded children and youth: Methodological issues. *Psychopharmacology Bulletin, 22*, 1081–1085.
Matson, J. L., Barett, R., & Helsel, W. J. (1988). Depression in mentally retarded children. *Research in Developmental Disabilities, 9*, 39–46.
Matson, J. L., Rush, K. S., Hamilton, M., Anderson, S. J., Bamburg, J. W., Baglio, C. S., et al. (1999). Characteristics of depression as assessed by the diagnostic assessment for the severely handicapped-II (DASH-II). *Research in Developmental Disabilities, 20*, 305–313.
McBrien, A. (2003). Assessment and diagnosis of depression in people with intellectual disabilities. *Journal of Intellectual Disability Research, 47*, 1–13.
McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Statistical and Mathematical Psychology, 23*, 1–21.
Menolascino, F. J. (1965). Emotional disturbance and mental retardation. *American Journal of Mental Deficiency, 70*, 248–256.
Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543.
Morin, D., Cobigo, V., Rivard, M., & Lépine, M. (2010). Intellectual disabilities and depression: How to adapt psychological assessment and intervention. *Canadian Psychology, 51*, 185–193.

Morin, A. J. S., Janosz, M., & Larivée, S. (2009). The Montreal Adolescent Depression Development Project (MADDP): School life and depression following high school transition. *Psychiatry Research Journal, 1*, 183–231.
Morin, A. J. S., Moullec, G., Maïano, C., Layet, L., Just, J. F., & Ninot, G. in press. Psychometric properties of the center for epidemiologic studies depression scale (CES-D) in French clinical and nonclinical adults, *Revue d'Epidémiologie et de Santé Publique*.
Moullec, G., Maïano, C., Morin, A. J. S., Monthuy-Blanc, J., Rosello, L., & Ninot, G. (2011). A very short visual analog form of the Center for Epidemiologic Studies Depression Scale (CES-D) for the idiographic measurement of depression. *Journal of Affective Disorders, 128*, 220–234.
Newman, D. L., Moffitt, T. E., Caspi, A., Magdol, L., Silva, P. A., & Stanton, W. R. (1996). Psychiatric disorder in a birth cohort of young adults: Prevalence, comorbidity, clinical significance, and new cases incidence from ages 11 to 21. *Journal of Consulting and Clinical Psychology, 64*, 552–562.
Offord, D. R, Boyle, M. H., Campbell, D., Goering, P., Lin, E., Wong, M., et al. (1996). One-year prevalence of psychiatric disorder in Ontarians 15 to 64 years of age. *Canadian Journal of Psychiatry, 41*, 559–563.
Perez-Achiaga, N., Nelson, S., & Hassiotis, A. (2009). Instruments for the detection of depressive symptoms in people with intellectual disabilities: A systematic review. *Journal of Intellectual Disabilities, 13*, 1–21.
Perreira, K., Deeb-Sossa, N., Harris, K., & Bollen, K. (2005). What are we measuring? An evaluation of the CES-D across race/ethnicity and immigrant generation. *Social Forces, 83*, 1567–1602.
Piccinelli, M., & Wilkinson, G. (2000). Gender differences in depression. Critical review. *British Journal of Psychiatry, 177*, 486–492.
Powell, R. (2003). Psychometric properties of the Beck Depression Inventory and the Self-rating Depression Scale in adults with mental retardation. *Mental Retardation, 41*, 88–95.
Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.
Reiss, S., & Rojahn, J. (1993). Joint occurrence of depression and aggression in children and adults with mental retardation. *Journal of Intellectual Disability Research, 37*, 287–294.
Reynolds, W. M., & Baker, J. A. (1988). Assessment of depression in persons with mental retardation. *American Journal on Mental Retardation, 93*, 93–103.
Reynolds, W. M., & Miller, K. L. (1985). Depression and learned helplessness in mentally retarded and non mentally retarded adolescents: An initial investigation. *Applied Research in Mental Retardation, 6*, 295–306.
Rojahn, J., Warren, V. J., & Ohringer, S. (1994). A comparison of assessment methods for depression in mental retardation. *Journal of Autism & Developmental Disorders, 24*, 305–313.
Rola, J. (1996). Depressive disorders in mentally retarded children. A review of selected literature. Part II. *Psychiatria Polska, 30*, 211–219.
Ross, E., & Oliver, C. (2003). The assessment of mood in adults who have severe or profound mental retardation. *Clinical Psychology Review, 23*, 225–245.
Salokangas, R. K., Vaahtera, K., Pacriev, S., Sohlman, B., & Lehtinen, V. (2002). Gender differences in depressive symptoms. An artefact caused by measurement instruments? *Journal of Affective Disorders, 68*, 215–220.
Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology, 62*, 123–146.
Sheehan, T. J, Fifield, J., Reisine, S., & Tennen, H. (1995). The measurement structure of the Center for Epidemiologic Studies Depression Scale. *Journal of Personality Assessment, 64*, 507–521.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 2*, 420–428.
Sigelman, C. K., & Budd, E. C. (1986). Pictures as an aid in questioning mentally retarded persons. *Rehabilitation Counselling Bulletin, 29*, 173–181.
Simmons, R. G., & Blyth, D. A. (1987). *Moving into adolescence: The impact of pubertal change and school context*. New York, NY: Aldine DeGruyter.
Singh, N. N., Sood, A., Sonenklar, N., & Ellis, C. R. (1991). Assessment and diagnosis of mental illness in persons with mental retardation: Methods and measures. *Behavior Modification, 15*, 419–443.
Steane, J. E., & Heald, F. P. (1987). Adolescent growth and development. *Maryland Medical Journal, 36*, 923–926.
Steer, R. A., Beck, A. T., & Brown, G. (1989). Sex differences on the revised Beck Depression Inventory in outpatients with affective disorders. *Journal of Personality Assessment, 53*, 693–702.
Tram, J. M., & Cole, D. A. (2006). A multimethod examination of the stability of depressive symptoms in childhood and adolescence. *Journal of Abnormal Psychology, 115*, 674–686.
Tsakanikos, E., Bouras, N., Sturmey, P., & Holt, G. (2006). Psychiatric co-morbidity and gender differences in intellectual disability. *Journal of Intellectual Disability, 50*, 582–587.
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.
Vázquez, F. L., Blanco, V., & López, M. (2007). An adaptation of the Center for Epidemiologic Studies Depression Scale for use in non-psychiatric Spanish populations. *Psychiatry Research, 149*, 247–252.
Wadsworth, J. S., & Harper, D. C. (1991). Increasing the reliability of self-report by adults with moderate mental retardation. *Journal of the Association for Persons with Severe Handicaps, 16*, 228–232.
Weissman, M. M., Bland, R., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H.-G., et al. (1996). Cross-national epidemiology of major depression and bipolar disorder. *Journal of the American Medical Association, 276*, 293–299.
Wenzel, A., Steer, R. A., & Beck, A. T. (2005). Are there any gender differences in frequency of self-reported somatic symptoms of depression? *Journal of Affective Disorders, 89*, 177–181.
Whitaker, S., & Read, S. (2006). The prevalence of psychiatric disorders among people with intellectual disabilities: An analysis of the literature. *Journal of Applied Research in Intellectual Disabilities, 19*, 330–345.
Wong, D. L., & Baker, C. M. (1988). Pain in children: Comparison of assessment scales. *Pediatric Nursing, 14*, 9–17.
Yuan, K.-H., & Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three statistics for covariance structure models. *British Journal of Mathematical and Statistical Psychology, 56*, 93–110.