

Claus Barkmann
Michael Erhart
Michael Schulte-Markwort
and the BELLA Study Group

The German version of the Centre for Epidemiological Studies Depression Scale for Children: psychometric evaluation in a population-based survey of 7 to 17 years old children and adolescents – results of the BELLA study

Members of the BELLA Study Group:
Ulrike Ravens-Sieberer (Principal Investigator), Claus Barkmann, Susanne Bettge, Monika Bullinger, Manfred Döpfner, Michael Erhart, Beate Herpertz-Dahlmann, Heike Hölling, Franz Resch, Aribert Rothenberger, Michael Schulte-Markwort, Nora Wille, Hans-Ulrich Wittchen.

Prof. Dr. C. Barkmann (✉) · Dr. M. Erhart
Prof. Dr. M. Schulte-Markwort
Centre for Obstetrics and Paediatrics
Dept. of Psychosomatics in Children and Adolescents
University Clinic Hamburg-Eppendorf
Martinistraße 52, W29
20246 Hamburg, Germany
Tel.: +49-40/42803-3697
Fax: +49-40/42803-5105
E-Mail: barkmann@uke.uni-hamburg.de

■ **Abstract** *Objectives* To examine the psychometric properties and test the theoretical quality of the German version of the Centre for Epidemiological Studies depression scale for children (CES-DC), a 20-item screening instrument measuring the frequency of parent- and self-reported depressive symptoms in children and adolescents.

Methods Using a population-based, representative sample of $n = 2,863$ 7 to 17-year-old German children and adolescents, factorial validity were determined by means of linear structural equation modelling. Cross-sectional coefficients of reliability, inter-rater agreement as well as descriptive statistics of the scales were calculated. *Results* In a population-based German sample, the four-factor version of the CES-DC following Radloff (Appl Psychol Meas 1:385–401, 1977) is considered to have good factorial

validity and stability across age and informant versions. The main problems of the questionnaire are the high item difficulties, strong floor effects of the scales and low cross-sectional reliability, which are acceptable only for screening purposes. The low inter-rater agreement indicates that parental assessment can replace self-assessment only to a limited degree. *Conclusion* The strengths and weaknesses of the CES-DC are discussed taking previous data and comparable tests into consideration. Particular advantages are the existence of the parent-report form and the adult version, as well as its multifactorial structure. Parental assessment should be supplemented by self-report data whenever possible.

■ **Key words** depression – children – diagnostic test – psychometric properties – factor structure

Introduction

In general, without professional therapeutic intervention most depressive symptoms in children and adolescents are of minor severity and shorter duration, and are remittent [15]. On the other hand, multiple, severe, and long lasting depressive symptoms call for a diagnostic assessment in order to evaluate

the existence of a depressive disorder according to DSM-IV or ICD-10, and, if applicable, to introduce the necessary clinical measures [7]. There are numerous different psychometric tests that determine the type, extent and distribution of depressive symptoms in children and adolescents [9, 22, 30]. Two of the most popular ones are the Children's Depression Inventory (CDI) [18] and the Children's Depression Scale (CDS) [20]. In German-speaking

countries, the “Depressionsinventar für Kinder und Jugendliche” (Depression Test for Children and Adolescents, DIKJ, translation of the CDI, [31]) as well as the “Depressionstest für Kinder” (“Depression test for children,” DTK, [26]) are available. The Centre for Epidemiological Studies Depression Scale for Children (CES-DC) is a questionnaire that has been developed especially for the screening of depressive symptoms corresponding to different depressive disorders in population-based samples [32]. The CES-DC is currently the only internationally disseminated test that uses a multidimensional approach in measuring depressiveness in children and adolescents. It has only 20 items and is available in both the parent- as well as self-report version.

The one-sided paper-pencil test comprises a short introduction and 20 fully standardised items to evaluate depressive symptoms (see Table 2). The items consist of short and simple statements in the first person about the emotional, cognitive and behaviour-related components of depressiveness. All items are evaluated on a four-point Likert scale in relation to their incidence during the last week (0 = “not at all,” 1 = “a little,” 2 = “some,” 3 = “a lot”). Different forms, such as a parent-report form, parallel or short form are not officially on hand; the authors, however, report the collection of mother reports [32]. Up to now there was only a Spanish [2] in addition to the English version. According to its authors, the CES-DC is applicable to children and adolescents between 6 and 17 years of age. The instrument can be administered on an individual or group level and takes an average of 5 min to complete. In order to obtain a valid measurement, adequate cognitive understanding and sufficient capacity of introspection are required. No special knowledge or abilities are necessary to perform the test. A total score “Depressiveness” is calculated by summing up the unweighted 20-item scores (items 4, 8, 12 and 16 with reversed polarity). The higher the total score, the more depressive the tested child is. According to Radloff [24], the total score should not be calculated when more than four items are not answered. Evaluation by hand takes approximately 5 min.

The CES-DC was not developed directly for children and adolescents, but was derived from the CES-D for adults developed by Radloff [24, 32]. Weissman et al. modified the CES-D items to make them more child-friendly (for example, CES-D item 7: “I felt like everything I did was an effort.” CES-DC modification: “I felt like I was too tired to do things” [32, see p. 738, Table 2]). Additionally, the verbal attributes of the response scale were converted from absolute to relative categories. Apart from these changes, no other modification was done. Further developmental stages with corresponding modifications of the CES-DC do not exist.

Weissman et al. [32] evaluated only the total score and did not make out a certain factorial structure for the CES-DC. Radloff [24] found a four-factor structure (PCA, Kaiser criterium) for the adult version in three field samples and in a group of psychiatric patients: “Somatic and retarded activity,” “Depressed affect,” “Positive affect,” and “Interpersonal” (for item-scale allocation see Table 2). Since then, a number of independent studies have been able to reproduce this solution on field samples of adults, as well as in adolescents. Today it is considered the standard for the CES-D [see meta-analysis, 29]. Meanwhile, two studies exist that have identified this four-factor model for the CES-DC. In the first study, Fendrich et al. [13] analysed the data of $n = 166$ 6 to 23-year-old participants of a risk study by means of an explorative factor analysis (PCA, Varimax). The resulting four-factor solution complied with Radloff’s model [24] according to the authors, but was not described in further detail.

A second study, headed by Olsson and von Knorring [23], also identified the Radloff model in $n = 2,272$ 16 and 17-year-old Swedish students. The varimax-rotated solution of the PCA explained 58.9% of the variance (38.6, 8.0, 6.8 and 5.5% per factor). Nonetheless, only 15 out of the 20 items were considered for the scales (items 2, 9, 11, 13 and 16 were omitted). All included items loaded positively and high on their factor ($a = 0.61$ – 0.88 ; [23, p. 84]).

The CES-DC showed high correlations with other depression tests ($r = 0.44$ and $r = 0.58$ for the CDI [11, 12], respectively; and $r = 0.81$ for the Beck Depression Inventory (BDI) [5, 23]). The correlations with depression-like constructs and total scores for emotional and behavioural problems showed the expected high effects as well ($r = 0.52$ for anxiety according to [11]); $r = 0.43$ for self-esteem according to Fendrich et al. [13]; $r = 0.59$ for the total score of the Child Behaviour Check-list (CBCL) [1] according to Weissman et al. [32]. Existing studies aiming to determine the diagnostic validity of the CES-DC are difficult to interpret due to methodological deficiencies. Overall, the CES-DC seems to be able to differentiate between children and adolescents with or without psychiatric diagnoses according to DSM-III [13, 32], but the discrimination of different depressive disorders as well as between depressive and other disorders is not convincing [12, 13]. The content validity is indirectly hypothesised because of the derivation from the CES-D [32].

Internal consistencies of the total score vary between Cronbach’s $\alpha = 0.77$ and $\alpha = 0.91$, depending on the study [12, 13, 23, 28]. This can be explained by the differences in the chosen samples, and especially by a repeatedly confirmed age correlation with lower consistencies for younger children (e.g. $\alpha = 0.78$ for

6- to 11-year-olds vs. $\alpha = 0.89$ for the total sample according to Fendrich et al. [13]). The test-retest reliability is $r = 0.79$ for a 1-week interval in 15-year-old Guatemalan students [3], and $r = 0.51$ for a 2-week interval in child and adolescent psychiatric patients after discharge [12]. The only study to date that has analysed the inter-rater agreement of parent- and self-reports found a correlation of $r = 0.04$ for 28 mother and child reports [32]. Effects of age and socioeconomic status have not been consistently proven yet; however, gender effects showed a clear pattern of small to medium effect sizes with higher scores for girls [6, 13, 23, 32].

Overall, the CES-DC has not been empirically evaluated as much as the CES-D and there exists only little and rather uncertain knowledge about its psychometric properties. This applies especially to its factorial structure and subscale characteristics. The adoption of the four-scale solution from the adult version in the CES-DC is reasonable, but has not yet been pertinently evaluated. Furthermore, there is almost no psychometric information about the parent-report version. While testing the parent- and self-report form on an identical factor structure, data about rater correspondence is particularly relevant from a practical point of view. Finally, normative scores for parent- and self-reports using representative population-based samples are not available at this point. Thus, the questions analysed here were as follows:

1. Can the four-factor structure, established for the CES-D by Radloff [23], also be confirmed for the CES-DC?
2. How reliable are the scales?
3. What are the scale properties, and which sociodemographic correlates do they show?
4. What are the differences between the parent- and the self-report?
5. Which normative scores could be assigned to the raw scores of the parent- and self-reports?

Methods

■ Design

The data are from the German Health Interview and Examination Survey for Children and Adolescents (KiGGS), prepared and conducted by the Robert Koch-Institute (Berlin) by order of the Federal Ministry of Health, as well as the Federal Ministry of Education and Research [19]. In this comprehensive, multi-methodological survey, nationwide representative data involving the physical and mental health of children and adolescents up to the age of 17 years were obtained. The BELLA module was specifically concerned

with questions about the type and distribution of mental health problems, health-related quality of life, as well as corresponding risk and protective factors. The methods applied within BELLA are described in detail in another paper of this supplement issue [25]. The compilation of the data used for the following analyses was performed by trained interviewers, aided by standardised, computer-assisted telephone interviews (CATI) of approximately 30 min. The German translation of the CES-DC was accomplished according to current standards [4]. For children 7–10 years of age, only the parent-report was used, because a full understanding of all items could not be expected in this age group. In youths between 11 and 17 years of age, the self-report was collected additionally.

■ Sample

Children and adolescents up to 17 years of age residing in the Federal Republic of Germany were defined as the population sample. The representativeness of the sample was established by a stratified two-stage probability approach. Thus, 167 sampling units in cities and communities distributed nationwide were selected by means of a stratified random process. The subjects were then chosen again using a random method from the address files of the registration offices. The field work was conducted between May 2003 and May 2006 in up to four sample points simultaneously. The BELLA participants were children between 7 and 17 years of age randomly selected from the KiGGS sample. After signing a declaration of consent, the participants were contacted via telephone on a date of their choice.

From the 28,299 families that had been invited to participate in the KiGGS study, 17,641 were included. From those, 4,199 were asked to participate in the BELLA module. Out of these 4,199 potential participants, 70.1% ($n = 2,942$) agreed to participate and 97.3% ($n = 2,863$) actually participated in the survey. Data from the parent interview was completely missing in 66 cases, whereas data was missing from the children interviews in 58 cases. Furthermore, three parent-reports and two self-reports had to be excluded because four or more item responses were missing in the CES-DC. Thus, the resulting sample consisted of 1,137 parent-reports for 7–10 years old, 1,657 parent-reports for 11–17 years old, and 1,681 self-reports for 11–17 years old.

■ Analyses

Items 4, 8 and 15 showed elevated, non-random missing data in the parent-report ($n = 40, 72$ and 41), which were replaced by multiple imputation (full

information maximum likelihood; [14]). All analyses were subdivided into groups according to age and informant—7–10 years old and 11–17 years old for parent-report, and 11–17 years old for self-report—in order to enable direct comparisons between parent- and self-reports. An interval scale level was assumed on account of the polytome, as well as a graphically and verbally equidistantly constructed response scale. Several cross-sectionally estimable psychometric statistics were calculated according to classical test theory. To ensure representativeness, a case specific weighting factor was used, which corrects the deviation of the sample from the population according to age, gender, region and nationality (reference data 31.12.2004). Hence, case numbers reported in the results section refer to weighted cases. Data were processed with SPSS 12.0, LISREL 8.54 and PRELIS 2.5.

Results

■ Confirmation of the four-factor structure according to Radloff [24]

The factorial structure was tested by means of confirmatory factor analyses based on a linear structural equation model using the polychoric correlation matrix. The indicators were interpreted as reflexive and the model was defined as recursive and congenerous. To ensure proper identification of the model parameters, one indicator path per latent variable was fixed at one and all other parameters were set free. Because in the three groups of the analysis 53% of the items showed a skewness >2, and 30% a kurtosis >7, a discrepancy function was chosen that did not presuppose a multinomial distribution [33]. Due to the large sample size, the asymptotically distribution-free estimation (ADF) could be used. The ADF method enables an efficient parameter estimation and inferential statistics, even in not normally distributed variables [16]. The goodness of fit was evaluated according to the rules compiled by Schermelleh-Engel and Moosbrugger [27] using the chi-square test, root mean square error of approximation (RMSEA), standardised root mean residual (SRMR), comparative fit index (CFI) and goodness of fit index (GFI).

The model fit can be classified as acceptable to good for all three groups, according to the chosen indices presented in Table 1 (the high Chi-square statistics are due to the large sample sizes). The four-factor model of the CES-D can thus be replicated for the CES-DC. Table 3 shows the calculated standardised regression weights for all three groups in comparison. The loading profiles are rather similar: i.e., almost every loading is high (>0.50). Nonetheless, item 8 (“something good is going to happen”) con-

Table 1 Fit-indices for the four-factor model of the CES-D according to Radloff [24]

	χ^2	χ^2/df	<i>P</i>	RMSEA	SRMR	CFI	GFI
Acceptability ^a	–	≤3	≥0.05	≤0.08	≤0.10	≥0.95	≥0.90
Parent-report							
7–10 years (<i>n</i> = 957)	370.51	2.3	0.000	0.05	0.07	0.99	0.98
11–17 years (<i>n</i> = 1,829)	611.62	3.7	0.000	0.05	0.06	0.99	0.98
Self-report							
11–17 years (<i>n</i> = 1,852)	438.36	2.7	0.000	0.04	0.06	0.99	0.99

For abbreviations see text

^aRules of thumb according to Schermelleh-Engel and Moosbrugger [27]; *df* = 164

stitutes an exception with *a* < 0.30 in the self-report. The loading >1 for item 19 in the parent assessment of the 7–10 years old is acceptable, because this is an estimated parameter and the deviation from 1 lies within the range of the standard error [17]. The squares of the loadings as a measure for the reliability of the indicators are Med (λ^2) = 0.47, 0.53 and 0.48; the worst values by far are shown by item 8 (λ^2 = 0.13, 0.10 and 0.04). The unexplained residual covariances are approximately normally distributed, and are predominantly <0.10. The intercorrelation of the latent variables, corrected for attenuation, are at Med = 0.58 (from *r* = 0.30 for “Positive”–“Interpersonal” to 0.91 for “Somatic”–“Depressed”).

■ Reliability

The item difficulties (item score divided by maximum possible score) are distributed around Med = 0.10 with a minimum of 0.03 for item 13 (more quiet), and a maximum of 0.53 for item 8 (something good going to happen). For all groups, the scores are predominantly densely distributed in the lower range of the response scale. The “Positive” scale constitutes an exception, since really difficult items are missing. Further statistics are presented in Table 4. The average of the item means are quite similar for all groups (M_M = 0.44, 0.43 and 0.49, respectively). Compared to other items, the variability of responses in item 8 is remarkably low for all three groups. The corrected item-scale correlations are distributed around Med_{ris} = 0.45 for all scales and groups. Again, the coefficients for item 8 are remarkably low. The median internal consistency (Cronbach’s α) for all subscales and age group/informant combinations is Med $_{\alpha}$ = 0.67 with similar coefficients in all three groups. The “Depressed” scale consistently showed the best values. None of the scales could be improved in their internal consistency by elimination of further items.

Table 2 Item characteristics of the CES-DC

Item	Parent-report								Self-report				
	7–10 years (<i>n</i> = 957)				11–17 years (<i>n</i> = 1,829)				11–17 years (<i>n</i> = 1,852)				
	<i>M</i>	<i>SD</i>	<i>r_{is}</i>	<i>r_{it}</i>	<i>M</i>	<i>SD</i>	<i>r_{is}</i>	<i>r_{it}</i>	<i>M</i>	<i>SD</i>	<i>r_{is}</i>	<i>r_{it}</i>	
I. Somatic													
1	Bothered by things	0.3	0.58	0.29	0.43	0.3	0.60	0.31	0.38	0.5	0.71	0.41	0.47
2	Not feel like eating	0.3	0.64	0.37	0.28	0.2	0.52	0.35	0.30	0.3	0.67	0.36	0.33
5	Could not pay attention	0.6	0.79	0.26	0.28	0.5	0.73	0.38	0.42	0.6	0.78	0.42	0.40
7	Too tired to do things	0.3	0.60	0.45	0.31	0.4	0.63	0.46	0.38	0.6	0.78	0.43	0.39
11	Sleep was restless	0.3	0.61	0.39	0.35	0.3	0.59	0.37	0.33	0.4	0.79	0.39	0.39
13	More quiet	0.1	0.38	0.41	0.48	0.2	0.50	0.43	0.48	0.4	0.76	0.39	0.46
20	Hard to get started	0.4	0.64	0.47	0.41	0.5	0.71	0.48	0.49	0.4	0.61	0.36	0.40
II. Depressed													
3	Was not happy	0.2	0.45	0.46	0.46	0.2	0.51	0.46	0.54	0.3	0.66	0.51	0.51
6	Felt down and unhappy	0.3	0.61	0.67	0.63	0.3	0.64	0.71	0.67	0.4	0.69	0.68	0.64
9	Things did not work out	0.2	0.48	0.49	0.49	0.2	0.51	0.53	0.51	0.3	0.67	0.50	0.53
10	Felt scared	0.3	0.56	0.43	0.46	0.2	0.49	0.52	0.50	0.2	0.54	0.54	0.50
14	Felt lonely	0.2	0.54	0.49	0.52	0.2	0.45	0.47	0.50	0.2	0.50	0.49	0.55
17	Felt like crying	0.5	0.66	0.53	0.42	0.2	0.52	0.59	0.46	0.2	0.54	0.54	0.45
18	Felt sad	0.6	0.62	0.61	0.54	0.4	0.61	0.67	0.56	0.5	0.69	0.64	0.59
III. Positive													
4	Good as other kids (–)	0.8	0.89	0.25	0.31	0.9	0.93	0.31	0.33	1.2	0.96	0.31	0.28
8	Something good going to happen (–)	1.6	1.11	0.33	0.17	1.7	1.06	0.34	0.18	1.7	1.02	0.30	0.07
12	Was happy (–)	0.6	0.64	0.44	0.43	0.8	0.69	0.50	0.52	0.6	0.75	0.51	0.46
16	Had a good time (–)	0.6	0.63	0.45	0.43	0.8	0.69	0.47	0.45	0.5	0.71	0.47	0.46
IV. Interpersonal													
15	Kids not friendly	0.3	0.62	0.53	0.39	0.2	0.47	0.45	0.32	0.2	0.57	0.44	0.29
19	Felt people disliked me	0.2	0.49	0.53	0.54	0.1	0.39	0.45	0.50	0.2	0.49	0.44	0.44

Abbreviated item wording; (–) reversed items; response scale 0–3 (not at all–a lot); *r_{is}* part-whole corrected item-subscale correlation; *r_{it}* part-whole corrected item-total score correlation

The split-half reliabilities are comparable to the internal consistencies.

■ Scale statistics and sociodemographic correlates

Table 4 also contains the descriptive statistics of the CES-DC scales. Their distributions are mainly left-skewed and leptokurtic; only the “Positive” scale is almost symmetrical and slightly platykurtic. In particular the “Interpersonal” scale, as well as the “Depressed” and “Somatic” scales, showed extreme floor effects. The average scale intercorrelation is $Med_r = 0.31$ for the parent-report of the 7–10 years old, 0.34 for the parent-report of the 11–17 years old, and 0.33 for the self-report. The closest correlation exists between “Somatic” and “Depressed” in the self-report with $r = 0.64$.

Age and gender dependencies were analysed for all scales; in the parent-report by a 3×2 ANOVA, and in the self-report by a 2×2 ANOVA (7–10/11–13/14–17 years old; male/female; effect size η^2 , 0.01 = small, 0.06 = medium, 0.14 = large according to Cohen [8]; for all reported effects $P \leq 0.05$). Small main effects can be seen for age in the parent-report with

decreasing scores for “Depressed” and “Interpersonal,” as well as increasing scores for “Positive.” There are small main effects for age in the self-report with increasing scores on all scales except for “Interpersonal.” Gender shows small effects only in the self-report and only on the “Depressed” scale and in the total score, in which girls have higher scores than boys. Statistical interactions are present exclusively as small effects in the self-report as well, namely in the scales “Somatic,” “Depressed” and in the total score (age-increasing scores for girls, age-constant scores for boys). Further ANOVA’s (method as described above, not represented here) showed a small effect for the socioeconomic status (slightly higher scores for children with lower status) on the “Positive” scale and in the total score in the parent-report. In the self-report, a small effect was identified regarding migration (slightly higher scores for migrant children). Furthermore, all scales, except “Depressed,” showed a small statistical interaction between migration and socioeconomic status in the self-report (migrants showed constant or higher depressiveness with higher status, whereas non-migrants show decreasing depressiveness with higher status).

Table 3 Comparison of the standardised regression coefficients calculated by the ADF procedure

Item	Parent-report		Self-report
	7–10 years (n = 957)	11–17 years (n = 1,829)	11–17 years (n = 1,852)
I. Somatic			
1 Bothered by things	0.63	0.56	0.66
2 Not feel like eating	0.43	0.53	0.49
5 Could not pay attention	0.46	0.58	0.58
7 Too tired to do things	0.50	0.57	0.54
11 Sleep was restless	0.60	0.56	0.55
13 More quiet	0.86	0.77	0.62
20 Hard to get started	0.65	0.68	0.59
II. Depressed			
3 Was not happy	0.71	0.77	0.71
6 Felt down and unhappy	0.85	0.88	0.83
9 Things did not work out	0.70	0.73	0.73
10 Felt scared	0.68	0.72	0.70
14 Felt lonely	0.72	0.74	0.79
17 Felt like crying	0.58	0.74	0.68
18 Felt sad	0.74	0.84	0.76
III. Positive			
4 Good as other kids (–)	0.51	0.53	0.46
8 Something good going to happen (–)	0.36	0.31	0.21
12 Was happy (–)	0.76	0.85	0.83
16 Had a good time (–)	0.77	0.75	0.83
IV. Interpersonal			
15 Kids not friendly	0.68	0.67	0.70
19 Felt people disliked me	1.02	0.99	0.95

Abbreviated item wording; (–) reversed items

■ Inter-rater agreement

Table 5 presents the intra-class correlations of the parent- and self-reports for the 11–17 years old stratified by age and gender (two-way, random, single

rater). The concordance of the absolute scores ranges from ICC = 0.06 (“Interpersonal” for the 11 to 13-year-old boys) to 0.40 (“Interpersonal” for the 11 to 13-year-old girls); its Median is Med = 0.28. When the differences in the rater means are controlled (not represented), the average rater consistency is still Med = 0.28 (Min = 0.06, Max = 0.40). Hence, parents’ and children’s judgements are based on similar anchors. The ratings are similarly inconcordant for all scales, with the “Positive” scale showing the lowest concordance. A closer look at the single coefficients reveals that the inter-rater agreement is largely independent from the age and gender of the child. The only exception is the scale “Interpersonal”—it shows the lowest agreement for boys, whereas the concordance between girls and the parent-report is rather high.

The comparison of means between the parent- and self-reports stratified by age and gender for the 11–18 years old by a 2 × 2 × 2 ANOVA (the same groups as above, not represented) reveals a small but significant effect for all scales (exception: medium effect for “Somatic”) with higher scores for the self-report. Small interactions are observed for informant and age on “Somatic,” “Depressed” and the total score, with age-related constant or decreasing scores in the parent-report, and increasing scores in the self-report. Furthermore, low interaction effects between informant and gender are found in all scales—except for “Interpersonal”—with higher scores for girls in the self-report, whereas boys show constant or even lower scores in the parent-report. Second-order interactions are not observed.

Table 4 Psychometric properties of the CES-DC scales

Scale	n_i	α	r_{tt}	Med _{ris}	M	SD	Skewness	Kurtosis	Floor effect
Parent-report (7–10 years; n = 957)									
I Somatic	7	0.65	0.65	0.39	2.3	2.46	1.79	4.25	26.8
II Depressed	7	0.79	0.76	0.49	2.2	2.64	2.25	7.23	31.3
III Positive	4	0.56	0.53	0.39	3.6	2.20	0.35	–0.28	7.1
IV Interpersonal	2	0.68	0.68	0.53	0.5	0.97	2.70	9.02	71.3
Total score	20	0.82	0.82	0.43	8.6	6.12	1.83	5.37	1.6
Parent-report (11–17 years; n = 1,829)									
I Somatic	7	0.69	0.73	0.38	2.3	2.54	1.63	3.30	26.1
II Depressed	7	0.82	0.78	0.53	1.7	2.62	2.63	9.34	48.5
III Positive	4	0.60	0.55	0.41	4.3	2.32	0.21	–0.26	5.5
IV Interpersonal	2	0.61	0.61	0.45	0.3	0.73	3.15	12.46	80.7
Total score	20	0.84	0.84	0.47	8.6	6.25	1.83	5.29	2.3
Self-report (11–17 years; n = 1,852)									
I Somatic	7	0.69	0.67	0.39	3.2	3.01	1.34	2.22	19.3
II Depressed	7	0.82	0.77	0.54	2.1	2.97	2.33	6.80	41.5
III Positive	4	0.60	0.57	0.39	4.1	2.35	0.38	–0.15	6.0
IV Interpersonal	2	0.60	0.60	0.44	0.4	0.90	2.59	7.97	74.5
Total score	20	0.83	0.84	0.46	9.8	6.93	1.62	3.76	1.2

Response scale 0–3 (not at all–a lot); α Cronbach’s α ; r_{tt} split-half reliability (Spearman–Brown); Med_{ris} Median of item-scale correlations

Table 5 Inter-rater agreement of the parent- and self-report for 11–18 years old stratified by age and gender

Scale		Boys		Girls		Total
		11–13 (<i>n</i> = 363)	14–17 (<i>n</i> = 535)	11–13 (<i>n</i> = 344)	14–17 (<i>n</i> = 530)	
I	Somatic	0.25	0.19	0.23	0.24	0.22
II	Depressed	0.38	0.30	0.28	0.36	0.34
III	Positive	0.11	0.17	0.11	0.28	0.19
IV	Interpersonal	0.06	0.11	0.40	0.27	0.22
	Total score	0.34	0.31	0.32	0.37	0.34

Intraclass correlation (two-way, random, single rater, absolute agreement); $P \leq 0.05$ for all coefficients

Standardisation

One possibility that may account for the heterogeneity of data is group wise standardisation separated for age, sex, and source of information. The chosen age categories 7–10, 11–13 and 14–17 years represent a compromise between age dependency, absolute cell size and uniform group sizes. Because of the significant deviation from normal distribution in all scales, percentiles should be computed. T-scores can then be calculated by assigning z-scores to percentiles according to McCall [21]. In our preliminary analysis such scores were also calculated for the unreliable scales “Somatic”, “Positive” and “Interpersonal” in order to enable age and gender adjusted analyses on a group statistical level. Mentally, psychosomatically and/or somatically ill children and adolescents are included in the sample according to their population prevalence.

Discussion

The four-factor model after Radloff [24] can be confirmed for all three groups and shows no weakness with regard to the total fit indices. The criteria of one-dimensionality, uncorrelated error variances and congenerity are thus fulfilled. On the indicator level, item 8 shows the worst performance. The loading >1 of item 19 in the parent-report of the 7–10 years old goes along with a negative residual variance—this phenomenon, known as Heywood-Case, often occurs when only few and high correlating indicators are considered for latent variables [10]. A significant improvement of the model fit according to the modification indices could be achieved by specification of the error covariance between group wise different pairs of items. However, this would indicate deviation from the assumption of unidimensionality. In order to evaluate the robustness of the given results, different initial matrices and discrepancy functions were used in addition (e.g. dummy-coding, Pearson correlation, maximum likelihood function, unweighted least square). All analyses confirmed the solution

represented here—which does not mean that more economic alternative models do not exist.

All reliability-based criteria were on average comparable between the three groups in the analysis. No items had to be eliminated because of theoretical or contextual reasons (violation of cultural norms, redundancy, etc.). The item difficulties were predominantly too high or unequally distributed. The item characteristic curves (not presented) increased strictly monotonously. The item-scale correlations can be considered good in view of the high item difficulty. Item eight shows somewhat deviating properties and should be taken into special consideration in future evaluations. The internal consistencies are similarly moderate for the subscales in all three age-informant combinations. All scales reach an acceptable level for group wise analyses with $\alpha \geq 0.60$, except “Positive.” Only the “Depressed” scale and the total score reach the reliability of ≥ 0.70 necessary for individual diagnostic purposes.

In comparison to the means and standard deviations calculated here, previously published distributional statistics of the CES-DC total score are significantly lower [11, 13, 23]. This might be due to differences between the samples (see “Introduction”), an effect of the telephone mode, and/or intercultural differences. The remarkable floor effects indicate a low discriminatory power of true individual differences for respondents with few depressive symptoms, and suggest that the 1-week period prevalence should be extended to 2 or 4 weeks. The average intercorrelation of scales reveals a moderate dependency of the different dimensions of depressive disorders, which goes along with the results for the CES-D [30]. The scales show different age and gender related trends, particularly between the parent- and self-report form with the beginning of puberty. The increasing scores of the self-report correspond to the known age-dependent increment of depression and to the increasing convergence of child psychopathology and adult-oriented criteria of measurement [7]. The gender-specific effects correspond to the typically expected tendencies resulting from the biological and social changes of puberty (such as the beginning of menstruation and the assumption of male role behaviour, [7]).

Because the CES-DC measures subjectively experienced depressive symptoms, the self-report can be seen as the more valid information, while the parent-report can be understood as a predictor or an approximation. From a statistical point of view, the goodness of this concordance has to be classified as poor; however, it is on a usual level for clinical questionnaires regarding internalised psychopathologies [22]. The detected differences can be explained by the latent and subjective nature of the phenomenon, as well as by possible tendencies of dramatisation or trivialisation in either adolescents or parents. This applies particularly for the large gender difference on the “Interpersonal” scale.

The main limitations of the study are the collection of questionnaire data by telephone interview, the absence of repeated measurements to determine the test-retest reliability, and the absence of clinical diagnoses for depressive disorders according to DSM-IV and/or ICD-10.

Conclusion

In summary, several conclusions can be drawn from the existing data. The German version of the CES-DC

can be considered as factorially valid in the four-factor version of the parent- and self-report form according to Radloff [24] for population-based samples of 7–17 years old. Whether this can be applied to clinical samples must be shown in future studies. The main problems of the questionnaire are the high difficulties of the items, strong floor effects of the scales (in particular for “Interpersonal”) and the low cross-sectional reliability, which is just acceptable for screening purposes (exception: “Positive” for the 7–10 years old). Test-retest reliability and sensitivity to change must be evaluated in repeated measurement designs. Comparative evaluations of alternative factor models, analyses of divergent and diagnostic validity, as well as Rasch scaling are important issues for future research. The parent-report can serve as a proxy for the self-rating only to a limited degree and, whenever possible, should be complemented by the self-report. Particular advantages of the CES-DC, in comparison to similar tests, are the existence of the parent-report form and the adult version, as well as its multifactorial structure.

■ **Conflict of interest** All authors declare no conflict of interest.

References

1. Achenbach T (1979) The child behavior profile: an empirically based system for assessing children's behavioral problems and competencies. *Int J Ment Health* 7:24–42
2. Aguilar G, Berganza CE (1990) Validación del CES-DC-M. Para el Diagnóstico de la Depresión en Adolescentes Guatemaltecos. *Av Píscol Clin Latinotot* 8:75–84
3. Aguilar G, Berganza CE (1993) Confiabilidad test-retest de la Escala de Depresión para Niños y Adolescentes del Centro de Estudios Epidemiológicos (CES-DC-M), en el diagnóstico de la depresión en adolescentes guatemaltecos. *Av Píscol Clin Latinotot* 11:59–63
4. Beaton D, Bombardier C, Guillemin F, Bosi Ferraz MB (2002) Recommendations for the cross-cultural adaptation of health status measures. Institute for Work and Health, Illinois
5. Beck AT, Ward C, Mendelson M (1961) Beck depression inventory (BDI). *Arch Gen Psychiatry* 4:561–571
6. Berganza CE, Aguilar G (1992) Depression in Guatemalan adolescents. *Adolescence* 27:771–782
7. Birmaher B, Ryan ND, Williamson DE, Brent DA, Kaufman J, Dahl RE, Perel J, Nelson B (1996) Childhood and adolescent depression: a review of the past 10 years. Part I. *J Am Acad Child Adolesc Psychiatry* 35:1427–1439
8. Cohen J (1988) *Statistical power analysis for the behavioral sciences*. Erlbaum, Hillsdale
9. Costello EJ, Angold A (1988) Scales to assess child and adolescent depression: checklists, screens and nets. *J Am Acad Child Adolesc Psychiatry* 27:726–737
10. Dillon RD, Kumar A, Mulani N (1987) Offending estimates in covariance structure analysis: comments on the causes of and solutions to Heywood cases. *Psychol Bull* 101:126–135
11. Doerfler LA, Felner RD, Rowlison RT, Raley PA, Evans E (1988) Depression in children and adolescents: a comparative analysis of the utility and construct validity of two assessment measures. *J Consult Clin Psychol* 56:769–772
12. Faulstich ME, Carey MP, Ruggiero L, Enyart P, Gresham F (1986) Assessment of depression in childhood and adolescence: an evaluation of the Center for Epidemiological Studies depression scale for children (CES-DC). *Am J Psychiatry* 143:1024–1027
13. Fendrich M, Weissman MM, Warner V (1990) Screening for depressive disorder in children and adolescents: validating the Center for Epidemiologic Studies depression scale for children. *Am J Epidemiol* 131:538–551
14. Finkbeiner C (1979) Estimation for the multiple factor models when data are missing. *Psychometrika* 44:409–420
15. Harrington R, Clark A (1998) Prevention and early intervention for depression in adolescence and early adult life. *Eur Arch Psychiatry Clin Neurosci* 248:32–45
16. Jöreskog KG, Sörbom D (1996) *Preliis2: user's reference guide*. Scientific Software International, Lincolnwood
17. Jöreskog KG (1999) How large can a standardized coefficient be? <http://www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf>. Accessed 22 June 1999

18. Kovacs M (1992) The children's depression inventory (CDI) manual. Multi-Health Systems, New York
19. Kurth BM, Bergmann KE, Hölling H, Kahl H, Kamtsiuris P, Thefeld W (2002) Der bundesweite Kinder- und Jugendgesundheitsurvey. Das Gesamtkonzept. *Gesundheitswesen* 64: S3-S11
20. Lang M, Tisher M (1987) Childrens depression scale manual, North American Edition. Consulting Psychologists Press, Palo Alto
21. McCall WA (1939) Measurement. Macmillan, New York
22. Myers K, Winters N (2002) Ten-year review of rating scales. II: scales for internalizing disorders. *J Am Acad Child Adolesc Psychiatry* 41:634-659
23. Olsson G, von Knorring AL (1997) Depression among Swedish adolescents measured by the self-rating scale Center for Epidemiology Studies-depression child (CES-DC). *Eur Child Adolesc Psychiatry* 6:81-87
24. Radloff LS (1977) The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1:385-401
25. Ravens-Sieberer U, Kurth B-M, KiGGS study group, BELLA study group (2008) The mental health module (BELLA study) within the German Health Interview and Examination Survey of Children and Adolescents (KiGGS): study design and methods. *Eur Child Adolesc Psychiatry* 17(Suppl 1):10-21
26. Rossmann P (2005) Depressionstest für Kinder (DTK). Hans Huber Verlag, Bern
27. Schermelleh-Engel K, Moosbrugger H (2003) Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods Psychol Res Online* 8:23-74
28. Schoenbach VJ, Kaplan BH, Wagner EH, Grimson RC, Miller FT (1983) Prevalence of self-reported depressive symptoms in young adolescents. *Am J Public Health* 73:1281-1287
29. Shafer AB (2006) Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton and Zung. *J Clin Psychol* 62:123-146
30. Sharp LK, Lipsky MS (2002) Screening for depression across the lifespan: a review of measures for use in primary care settings. *Am Fam Physician* 66:1001-1008
31. Stiensmeier-Pelster J, Schürmann M, Duda K (2000) Depressions-Inventar für Kinder und Jugendliche (DIKJ), Handanweisung. Hogrefe, Göttingen
32. Weissman MM, Orvaschel H, Padian N (1980) Children's symptom and social functioning self-report scales: comparison of mothers' and children's reports. *J Nerv Men Dis* 168:736-740
33. West SG, Fich JF, Curran PJ (1995) Structural equation models with non-normal variables: problems and remedies. In: Hoyle RH (ed) *Structural equation modeling: concepts, issues, and applications*. Sage, Thousand Oaks, pp 56-75