

Development and Aging

Reliability, validity and normative data for the Danish Beck Youth Inventories

MIKAEL THASTUM,¹ KRISTINE RAVN,² SØREN SOMMER³ and ANEGEN TRILLINGSGAARD^{1,2}

¹*Department of Psychology, University of Aarhus, Denmark*

²*Aarhus University Hospital, Regional Centre for Child and Adolescent Psychiatry, Denmark*

³*Psychiatric Hospital for Children and Adolescents, Region of Southern Denmark, Denmark*

Thastum, M., Ravn, K., Sommer, S. & Trillingsgaard, A. (2009). Reliability, validity and normative data for the Danish Beck Youth Inventories. *Scandinavian Journal of Psychology*, 50, 47–54.

This study examines reliability and validity and establish Danish norms for the Danish version of the Beck Youth Inventories (BYI) (Beck, Beck & Jolly, 2001), which consists of five self-report scales; Self-Concept (BSCI), Anxiety (BAI), Depression (BDI), Anger (BANI) and Disruptive Behavior (BDBI). A total of 1,116 school children and 128 clinical children, aged 7–14, completed BYI. Internal consistency coefficients were high. Most test-retest correlations were >0.70. A test-retest difference was found for BAI. Exploratory and confirmatory factor analysis indicated that the five factor structure of the instrument was justified. The BSCI, BAI and BDI discriminated moderately between the norming sample and the clinical group, and the latter group included more children who exceeded the 90th percentile of the norming sample. Diagnostic groups scored higher on relevant scales than norms. Only BSCI and BDI differentiated between diagnostic groups. The BYI showed acceptable internal consistency and test-retest stability, except for BAI. The BYI did not adequately differentiate between internalizing disorders.

Key words: Assessment, children, scale, reliability, validity.

Mikael Thastum, Department of Psychology, University of Aarhus, Jens Chr. Skous Vej 4, 8000 Aarhus C, Denmark. E-mail: mikael@psy.au.dk

INTRODUCTION

Psychosocial rating scales for children have gained popularity due to an increasing focus on scientific measurement, development of new models of juvenile psychopathology, need for outcome measures in clinical trials, the recognition of internalizing disorders in children, and a declining interest in projective measures. Rating scales have multiple applications, such as screening groups in normative settings, monitoring symptoms in high-risk groups, evaluating intervention effects, and ensuring systematic coverage of behaviors. Rating scales, however, are not diagnostic instruments, and should not substitute for diagnostic evaluation (Myers & Winters, 2002a).

In Denmark there is a shortage of standardized and validated self-report tools for assessing children with social and emotional problems. Danish norms are available for the Child Behavior Check list (CBCL) and the Youth Self Report (YSR) (Bilenberg, 1999), and the newly developed FTF (Five To Fifteen) (Kadesjo, Janols, Korkman *et al.*, 2004; Trillingsgaard, Damm, Sommer *et al.*, 2004). The CBCL is based on parental report, and the YSR, which is the self-report version of the CBCL, is limited to children from 11 years and up. The FTF is a parent questionnaire for elicitation of symptoms and problems typical of ADHD and its co-morbidities. Several validated self-rating scales are available for measuring children's self-concept (e.g. the Piers-Harris Children's Self-Concept Scale (Piers, 1984) and the Self-Perception Profile for Children (Harter, 1985)), anxiety (e.g. Multidimensional Anxiety Scale for Children (March, 1997),

Screen for Child Anxiety Related Disorders (Birmaher, 1999)), depression (e.g. Reynolds Child Depression Scale (Reynolds, 1989), Children's Depression Inventory (Kovacs, 1992)), anger (e.g. State-trait Anger Expression Inventory (Spielberger, 1988)), and disruptive behavior (e.g. the Self-Report Delinquency Scale (Elliot, Huizinga & Ageton, 1985)). No Danish norms or studies of the psychometric properties in a Danish context are available for self report questionnaires specifically assessing any of the aforementioned constructs.

The newly developed Beck Youth Inventories of Emotional and Social Impairment (BYI) (Beck, Beck & Jolly, 2001) includes subscales of self-concept, anxiety, depression, anger, and disruptive behavior, and thus both addresses co-morbidity of disorders of negative affect and, if subscales are used individually, is a specific measure of functioning. Due to the general tendency toward co-morbidity in child and adolescent psychiatric problems (Angold, Costello & Erkanli, 1999; Lewinsohn, Rohde & Seeley, 1995), a multipurpose scale is to be preferred over a scale measuring only one problem. None of the existing self-rating scales of negative affect in children are without problems, thus there is no "gold standard" questionnaire available (for reviews of the psychometric properties of self-rating scales for children, see Collett, Ohan & Myers, 2003; Myers & Winters, 2002a, 2002b; Winters, Myers & Proud, 2002). The BYI provides self-report from children beginning age 7 years, and it is brief and easy to administer. American norms, based on a stratified standardization sample, as well as scores obtained from a clinical outpatient sample are available,

but to our knowledge the measure has not yet been evaluated in languages other than English.

The purpose of this study was to establish norms and examine the reliability and validity of the Danish version of the BYI by assessing: (1) Reliability in a sample of school children (internal consistency and test-retest reliability). (2) Validity as judged by (a) internal structure (exploratory and confirmatory factor analysis and inter-correlations among the five inventories), and (b) ability to discriminate between a representative sample of school children and a clinical sample, and between diagnostic groups of the clinical sample.

MATERIAL AND METHODS

Instrument

The BYI consists of five self-report measures, which assess perceptions of self-concept, anxiety, depression, anger and disruptive behavior in children age 7–14 years. The Beck Self-Concept Inventory (BSCI) includes items about self-perceptions, such as competence and positive self-worth. The Anxiety Inventory (BAI) reflects fear, worrying, and physiological symptoms associated with anxiety. The Depression Inventory (BDI) includes items that reflect children's negative thoughts about themselves, their lives, and their future; feelings of sadness; and physiological indications of depression. The Anger Inventory (BANI) includes perceptions of mistreatment, negative thoughts about others, feelings of anger and physiological arousal. The Disruptive Behavior Inventory (BDBI) measures behaviors and attitudes associated with conduct disorders and oppositional defiant disorders.

Each inventory contains 20 statements written at a second grade level. An inventory takes up to 10 minutes to complete and can be administered individually or in groups. The children are asked to indicate the extent to which they think that a sentence characterizes them best on a four-point scale (0 = never, 1 = sometimes, 2 = often, 3 = always). Each inventory is scored by adding the 20 ratings. Inventories may be used separately or in combinations.

The BYI were translated from the original American-English version to Danish by the first and fourth authors. The two translations were compared, and a native English speaker back translated the version upon which they agreed. The back translation was compared with the original and revisions were made according to inconsistencies found (Bravo, Woodbury-Farina, Canino & Rubio-Stipec, 1993).

Sampling

Beck *et al.* (2001) consider two age groups: 7–10 and 11–14, but in a Danish context of later compulsory school start at age 7 (compared to US start age of 6), 7-year-old responses may be problematic. As the goal of our analysis was to validate an adaptation of BYI to Denmark, a sample of 7-year-old school children was included in our study and analyzed separately. However the comparison between the clinical and population based samples was carried out for two groups of 8–10 and 11–14 year olds.

The normative sample

Recruitment and participation rates. The sample was selected from school children in the county of Aarhus. Parents of 1,716 children attending classes at two elementary public schools from the city, one suburban public school, one public school in a rural area and two private schools were mailed information about the project.

Testing took place in the children's classrooms and two research assistants administered the questionnaires by following a testing protocol.

Of the children whose parents were mailed information material, 65% participated in the study. The sample consisted of 1,119 children (50.5% females) from grade 1 to 8, aged 7 through 14 years. Table 1 shows the distribution of the children across ages and gender. Children from private schools (25.6%) as well as public schools (city (46.3%), suburban (19.2%) and a rural (8.2%)) were represented.

Missing items. A maximum of two missing items out of the twenty were accepted for each inventory, as recommended in the manual (Beck *et al.*, 2001). Missing items were replaced by intrapersonal means of each inventory before analysis.

Socio-economic classification (SES). Prior to analysis, SES was recoded on a five-point scale in accordance with guidelines formulated by the Danish National Institute for Social Research (Hansen, 1977). Families were classified by the status of the highest-ranking parent in the household. Students (under education) were classified one group lower than the group the person would belong to after graduation.

Analyses of representation. There was an even distribution by age and gender. Data on socio-economic status was obtained from parents of 1,088 children (97.5% of the total sample). Compared to a nationally representative sample (Andersen, 2003) social groups 1 and 2 were over-represented and groups 3–5 were underrepresented in our sample (group 1 is highest level).

The clinical sample

Sample selection. The clinical sample consisted of 128 outpatient children (39.8% females) age 8–14 (mean = 11.7, *SD* = 1.9 years.) from three regional psychiatric centers for children and adolescents. They were tested with the BYI between spring 2003 and spring 2005.

Procedures for BYI-assessment. A clinical psychologist administered the BYI inventories with the child alone and without parental involvement. The BYI was not scored before the child was diagnosed and the diagnostic evaluation was not informed by the BYI results. The consultant psychiatrist leading the diagnostic team was not involved in the BYI study.

Diagnostic procedure

All children with their parents went through a thorough clinical assessment and diagnostic evaluation (ICD-10) conducted by a multidisciplinary team with child and adolescent psychiatrists and psychologists and headed by a consultant psychiatrist. The assessment included semi structured diagnostic interviews, psychological testing, somatic examination, and observation of child in school or home.

The following diagnoses were represented: Anxiety disorders (*N* = 27), mood disorders (*N* = 21), attention and hyperactivity disorders (ADHD) (*N* = 29), specific developmental disorders (*N* = 12), disorders of social functioning (*N* = 8), conduct disorders (*N* = 9), autism spectrum disorders (*N* = 12), other disorders (*N* = 11). These were the primary diagnoses. Most of the children had co-morbid diagnoses.

Data analysis and statistics

All data were analyzed using SPSS version 11.00, except for the confirmatory factor analysis (CFA) where AMOS version 7.0.0 was used. Test-retest reliability of the BYI scales was assessed by Pearson's correlation coefficients, and by paired-sample *T*-tests. Since no rotated exploratory factor analysis (EFA) of the BYI to our best knowledge has

Table 1. Mean scores and standard deviations by age-group and gender for each of the Beck Youth Inventories for the population based sample

	Age 7 years				Age 8–10 years				Age 11–14 years			
	Girls		Boys		Girls		Boys		Girls		Boys	
	<i>N</i>	Mean (<i>SD</i>)	<i>N</i>	Mean (<i>SD</i>)	<i>N</i>	Mean (<i>SD</i>)	<i>N</i>	Mean (<i>SD</i>)	<i>N</i>	Mean (<i>SD</i>)	<i>N</i>	Mean (<i>SD</i>)
BSCI	44	41.0 (10.0)	41	43.5 (12.6) ^b	242	41.3 (10.3) ^a	216	41.9 (8.1)	278	38.8 (7.6)	295	40.4 (6.8)
BAI	44	19.2 (13.2) ^c	41	16.9 (14.3) ^c	241	14.7 (9.2)	215	12.3 (8.9)	278	13.8 (6.4)	295	10.6 (6.6)
BDI	44	16.1 (11.0) ^c	40	13.1 (10.8) ^c	241	11.8 (8.5)	216	9.9 (7.1)	278	12.0 (7.3)	295	8.6 (6.2)
BANI	43	17.3 (12.5) ^c	38	15.5 (15.0) ^c	237	12.9 (8.2)	210	12.0 (8.6)	278	12.9 (6.7)	295	11.2 (6.6)
BDBI	25	7.2 (8.2)	16	13.8 (19.2) ^e	221	4.8 (5.8) ^d	198	7.7 (7.3)	278	6.4 (4.5)	295	8.0 (5.3)

Notes: BSCI = Self-Concept; BAI = Anxiety; BDI = Depression; BANI = Anger; BDBI = Disruptive Behavior.

^a 8–10 year old girls scored higher than 11–14 year old girls ($p < 0.05$).

^b 7 year old boys scored higher than 11–14 year old boys ($p < 0.05$).

^c For both genders 7 year olds scored higher than 8–10 years and 11–14 years old ($p < 0.005$).

^d 8–10 year old girls scored lower than 11–14 year old girls ($p < 0.004$).

^e 7-year-old boys scored higher than the 8–10 year old boys ($p < 0.003$) and 11–14 year old boys ($p < 0.004$) (Tukey HSD tests).

been published and to use the information from all items, an EFA was performed. This procedure first computed intercorrelations among the 100 scale items and then extracted the factors using principal component analysis. Since the factors were correlated, the extracted factors were then rotated to a simple structure using oblique (promax) rotation (Costello & Osborne, 2005). To test whether the five-factor structure of the BYI proposed by Beck *et al.* (2001) could be confirmed, a CFA was subsequently conducted. The following goodness-of-fit indices were used to assess the degree of fit between the model and the data: the chi-square (χ^2) divided by its degrees of freedom (df). Thresholds below 2 or 3 are generally considered adequate. This statistic, however, is easily influenced by the sample size and tends to be large in large samples (Marsh, Balla & McDonald, 1988). The Bentler comparative fit index (CFI) and the Tucker-Lewis coefficient (TLI) assess the adequacy of a specified model in relation to a baseline model. Values >0.90 are considered as acceptable and values >0.95 indicates a good fit (Bentler, 1990; Marsh *et al.*, 1988). The root mean square error of approximation (RMSEA) tests the fit of the model to the covariance matrix and takes model size and sample size into account. With regard to RMSEA values, Browne and Cudeck (1993) state that values <0.08 are acceptable and values <0.05 are excellent.

Group differences were analyzed using two-way and one-way analyses of variance and *T*-tests, and for categorical variables chi-square tests were used. *p*-values below 5% were considered statistically significant.

RESULTS

The normative sample

Age and gender differences. A series of two-way between group analyses of variance were conducted to explore the impact of age and gender on levels of the five BYI. Significant main effects for gender and age were found for all five scales. No significant interaction effects were found.

Girls scored significantly lower on Self-Concept ($p < 0.02$), higher on Anxiety ($p < 0.0005$), Depression ($p < 0.0005$), and Anger ($p < 0.005$), and lower on Disruptive Behavior ($p < 0.0005$). Tukey HSD tests indicated that in general 7 years olds scored significantly higher than all other ages.

Since most 7 year olds cannot read, due to later compulsory school start in Denmark than in the US, the sample was divided

into three age groups with 7, 8–10, and 11–14 year olds. Again for all five inventories there were significant main effects for gender and age but no significant interaction effects. Mean scores and standard deviations for the six groups and differences between groups (Tukey HSD tests) are displayed in Table 1.

Reliability

Internal consistency of each of the scales by gender was calculated using Cronbach's alpha coefficients. Cronbach's alpha ranged from 0.87 to 0.92 (Self-Concept; girls: 0.89, boys: 0.87. Anxiety; girls: 0.90, boys: 0.90. Depression; girls: 0.92, boys: 0.90. Anger; girls: 0.92, boys: 0.92. Disruptive Behavior; girls: 0.89, boys: 0.90).

Test-retest reliability coefficients were computed based on a sub-sample of 104 school children. The mean retest interval was 30 days. The BAI scale had lower average retest scores for 8–10 year olds (time 1: $M = 16.94$, $SD = 9.96$; time 2: $M = 13.27$, $SD = 10.28$, $t = 3.50$, $p < 0.001$), for 11–14 year olds (time 1: $M = 13.11$, $SD = 8.21$; time 2: $M = 11.64$, $SD = 8.23$, $t = 2.85$, $p < 0.01$), and for the total sample (time 1: $M = 15.37$, $SD = 9.70$; time 2: $M = 13.27$, $SD = 10.64$, $t = 3.83$, $p < 0.005$). No other significant differences were found. For test-retest administration over a month, a correlation greater than 0.70 is considered reasonably stable (Myers & Winters, 2002a). Pearson's correlation coefficients were smaller than 0.70 for 7 year olds for self-concept, depression and anger. For all other ages and inventories the correlations were greater than 0.70. Cronbach's alpha ranged from 0.80 (Disruptive Behavior; girls) to 0.96 (Anxiety; girls).

Validity

Internal structure. Pearson's inter-correlation coefficients among the five inventories for the total sample were calculated. Overall inventory scores were significantly correlated with each other.

Table 2. Pearson's intercorrelation coefficients of the Beck Youth Inventories for the full normative sample

	Beck Youth Inventories			
	Self-concept	Anxiety	Depression	Anger
Self-concept				
Anxiety	-0.25			
Depression	-0.44	0.71		
Anger	-0.33	0.59	0.76	
Disruptive behavior	-0.27	0.29	0.45	0.61

Particularly high correlations were found between depression and anxiety, and between anger and disruptive behavior (Table 2).

Factorial validity

Exploratory factor analysis. Using Cattell's (1966) method for looking for the elbow in the curve of the scree, in combination with the theoretical question of whether the five inventories did fit the factor solution, a promax rotated exploratory factor analysis with five components was performed on the 100 BYI items. On component one all of the BAI items loaded above 0.31. On component two 18 of the BDBI items loaded above 0.31. On component three 17 of the BANI items loaded above 0.31. Two cross loadings above 0.31 were found (Tabachnick & Fidell, 2001). On component four all of the BSCI items loaded above 0.31. On component five 11 of the BDI items loaded above 0.31. Three cross loadings were found. Six items did not load above 0.31 on any component. The five components accounted for 39.5% of the variance.

Confirmatory factor analysis. Factor structures are difficult to confirm when more than 5 to 8 items are free to load on each latent variable (Floyd & Widaman, 1995). Item parcels are often used to reduce model complexity and to reduce the number of parameters estimated without having to eliminate items and losing information that may contribute to the meaning of a latent variable. A parcel is a sum of several items that are assumed to be conceptually similar and psychometrically unidimensional and that are used to assess the same construct (Nasser & Takahashi, 2003). Scree plots of the eigenvalues of each of the five BYI inventories showed that the inventories were unidimensional. Since the BYI consists of five inventories each comprising 20 items, parcels were therefore made by randomly creating four five-item parcels for each inventory and then specifying four parcels to load on each of the five inventories. The results showed that although the chi-square was significant and the chi-square divided by its degree of freedom did not indicate an appropriate fit for the data ($\chi^2(160) = 915.26, p = 0.0005, \chi^2/df = 5.7$), the CFI and the TLI indicated a good to acceptable fit for the data (CFI = 0.96, TLI = 0.94), and the RMSEA value was at an acceptable level (RMSEA = 0.065).

Table 3. Differences between mean scores of the norm group and the clinical sample for the Beck Youth Inventories using *t*-test (equal variances not assumed)

	Norm group		Clinical sample		<i>p</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>	
Girls					
BSCI	39.9	8.1	31.9	9.8	0.0005
BAI	14.2	7.9	18.7	10.4	0.004
BDI	11.9	7.9	18.1	11.7	0.001
BANI	12.9	7.4	17.1	11.5	0.02
BDBI	5.7	5.1	8.7	7.8	0.01
Boys					
BSCI	41.0	7.3	34.7	10.5	0.0005
BAI	11.3	7.6	13.9	7.6	0.008
BDI	9.2	6.6	13.5	10.8	0.002
BANI	11.5	7.3	14.9	11.2	0.02
BDBI	7.9	6.2	8.9	7.3	0.36

Notes: BSCI = Self-Concept; BAI = Anxiety; BDI = Depression; BANI = Anger; BDBI = Disruptive Behavior.

SES differences. Spearman's inter-correlation coefficients between SES and each of the five inventories for the total sample were calculated. SES was significantly ($p < 0.01$) correlated with all inventories (Self-Concept -0.12, Anxiety 0.13, Depression 0.12, Anger 0.13, Disruptive Behavior 0.12).

Validity judged by comparing norms and the clinical group. Independent sample *t*-tests were conducted to compare the five inventories for norms and the clinical group using a Bonferroni adjusted alpha level of 0.01 (0.05 divided by 5). Since there were no 7 year olds in the clinical group, 7 year olds were omitted from the norm group in this specific analysis. As seen in Table 3, for both boys and girls significant differences were found for self-concept, anxiety, and depression. No significant differences were found for anger. For disruptive behavior, a significant difference was found only for girls. The magnitudes of the differences in the means were small (eta squared ranging from 0.05 to 0.01).

The proportion of children in the clinical group exceeding the 90th percentile of the BYI (obtained score higher than 90% of norm children of same gender) was significantly different from norms for all inventories (Chi-square ranging from 33.6, $p < 0.0005$ (Self-Concept, girls) to 3.9, $p > 0.05$ (Disruptive Behavior, boys)).

Validity judged by comparing the clinical groups. One-way between groups analyses of variance were conducted to explore the impact of diagnostic category on each of the inventories (in this analysis the autism spectrum disorder group and the other disorders group were omitted). Significant effects were found for BSCI: $F(5, 103) = 2.5, p = 0.04$, and BDI: $F(5, 96) = 3.6, p = 0.006$. No other significant effects were found. Although the *N* in three of the groups was small, we repeated the analyses adjusting for age and gender. Near significant effects were now

Table 4. Mean scores and standard deviations for BYI across 6 diagnostic groups. Post hoc comparisons (Tukey HSD test) of the scores for the diagnostic groups

	Anxiety disorders (N = 27)	Mood disorders (N = 20)	ADHD (N = 29)	Specific developmental disorders (N = 12)	Disorders of social functioning (N = 8)	Conduct disorders (N = 9)
BSCI	36.9 (9.1) ^a	27.4 (11.0) ^a	33.8 (9.0)	35.0 (10.1)	35.0 (8.2)	30.3 (10.8)
BAI	15.4 (6.1)	20.8 (11.2)	14.1 (8.5)	17.8 (9.4)	15.0 (10.9)	12.9 (6.7)
BDI	13.8 (11.1) ^b	24.9 (13.2) ^b	13.2 (8.5) ^b	14.2 (6.9)	13.6 (10.6)	17.0 (10.7)
BANI	14.7 (10.3)	19.7 (13.1)	15.9 (10.2)	16.5 (10.6)	12.4 (6.6)	20.0 (15.8)
BDBI	6.7 (6.4)	10.5 (5.6)	9.1 (7.7)	9.4 (6.7)	6.4 (5.1)	13.6 (15.2)

Notes: BSCI = Self-Concept; BAI = Anxiety; BDI = Depression; BANI = Anger; BDBI = Disruptive Behavior.

^a The mood disorder group scored significantly lower than the anxiety disorder group ($p < 0.02$).

^b The mood disorder group scored significantly higher than the anxiety disorder and the ADHD disorder groups ($p < 0.005$).

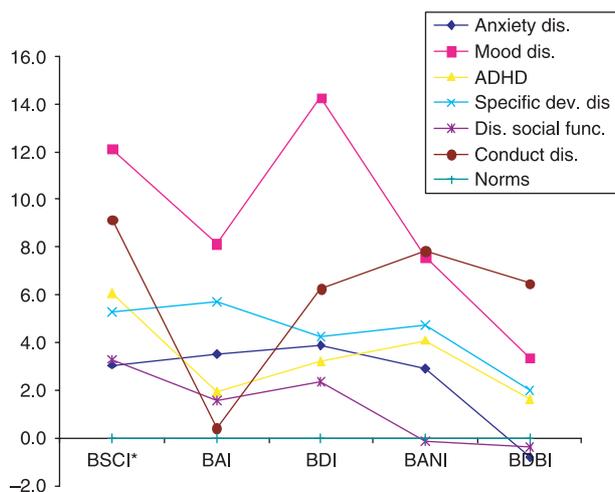


Fig. 1. Age and gender adjusted scores for the six diagnostic groups. The 0.0 line marks the means in the norm sample.

BSCI = Self-Concept; BAI = Anxiety; BDI = Depression; BANI = Anger; BDBI = Disruptive Behavior.

* The BSCI scale has been reversed in this figure.

found for BSCI: $F(5, 103) = 2.2$, $p = 0.055$, and significant effects were again found for BDI: $F(5, 96) = 3.3$, $p = 0.009$. Post-hoc comparisons using the Tukey HSD test indicated that the mean score of the mood disorder group was significantly lower than the mean score of the anxiety disorder group on BSCI, and that the mean score of the mood disorder group was significantly higher than the mean score of the anxiety disorder group and the ADHD disorder group on BDI. No other significant differences were found (Table 4).

To compare the scores on the five inventories of each of the six diagnostic categories with the population based sample, the mean score for each inventory for the individual clinical child was adjusted by subtracting it from the corresponding mean score among children with the same gender and in the same age group in the population based sample. Figure 1 shows the age and gender adjusted scores for the six diagnostic groups. The 0.0 line in the graph marks the mean in the norm sample.

One-sample t -tests were conducted to compare the differences between the diagnostic groups and the normative sample for

each inventory. On BSCI, children with mood disorder ($t = -4.85$), ADHD ($t = -3.64$), and conduct disorder ($t = -2.60$) had significantly lower scores. On BAI, children diagnosed with anxiety ($t = 2.85$) or mood disorder ($t = 3.42$) had significantly higher scores. On BDI, BANI and BDBI children with mood disorder ($t = 5.04$; 2.44 ; 2.61) scored significantly higher. No other significant differences were found

DISCUSSION

The current study demonstrates that the internal consistency and the one-month test-retest reliability and internal consistency of the Danish BYI appears to be acceptable for most inventories and norm groups, except for the anxiety inventory, where scores were lower at the second administration as compared to the first. With respect to the test-retest reliability of social-emotional child self-report instruments in general, the stability of internalizing disorders and self-concept in children is influenced by developmental changes (e.g. intellectual, social, perceptual, and affective) as well as contextual factors, and many internalizing symptoms fluctuates naturally over time (Michael & Merrell, 1998), why a correlation greater than 0.70 over a month is considered reasonable stable (Myers & Winters, 2002a). With respect to our findings for the anxiety inventory, there are some possible explanations. First, decreased reporting of symptoms (attenuation) has been cited to be a potential source of unreliability in psychometric studies (Piacentini, Roper, Jensen *et al.*, 1999), especially with a close test-retest interval, in community samples, and with lower age. Second, anxiety comprises a normal part of development and fluctuates in a community sample (Myers & Winters, 2002b). Third, lower scores at the second administration have been found for other anxiety scales as well, although for longer test-retest intervals (10–12 months) (Storch, Masia-Warner, Dent, Roberti & Fisher, 2004). The attenuated mean anxiety scores may be due to an over-endorsement of anxiety in distressed children upon initial testing, and/or better understanding of the assessment task during the second testing. It would have been preferable to examine test-retest interval both in the short term (2 weeks and 4 weeks) and in the long term (12 months), to assess the variability of reports over time. Information about the temporal stability of measures is essential in determining the

extent to which differences in scores over time are due to actual or error variance (Storch *et al.*, 2004). Further studies are needed to determine whether the low test-retest reliability of the anxiety scale will be replicated over a shorter and longer time span and in children with an anxiety diagnosis.

Norming

The normative sample was underrepresented by the lower SES children, and significant higher symptom scores and lower Self Concept scores were found among lower SES children. That rates of psychopathology are higher among children of lower SES than among children of higher SES is a common finding (Wadsworth & Achenbach, 2005), but the bias in our study should be considered when interpreting results. Mean scores may be underestimated for the negative affect scales and overestimated for the self-concept scale.

The gender differences consistently found for all inventories correspond well with a host of other studies. It is a general finding that girls have higher scores on internalizing self-report scales (Ivarsson, 2006; Ivarsson, Svalander & Litlere, 2006), and that externalizing symptoms are more prevalent in boys (Collett *et al.*, 2003). In accordance with results from a Danish standardization of the Child Behavior Checklist (CBCL) where no age (but marked gender) differences were found in a population-based sample (Bilenberg, 1999), scores did not differ markedly between age groups (with the exception that the 7 year olds scored differently from other ages).

Validity

Factorial validity of the BYI was supported by the EFA and the CFA, which confirmed the existence of five dimensions in a Danish context, corresponding to the five BYI inventories suggested by Beck *et al.*

Existent internalizing self-report measures for children demonstrates in general good convergent validity but poor discriminant validity (Stark & Laurent, 2001). We investigated the validity of the BYI in two ways. First, to which extent the BYI scales were related and second, whether the BYI scales differentiated between diagnostic groups within the clinical sample and between norms and the clinical sample.

Concerning the relations between the scales they were all in the expected directions, and with particularly strong relations between the depression, anxiety and anger scales. Self-report measures of anxiety and depression generally are not very sensitive to differences among types of internalizing disorders, with correlations typically within the 0.50 to 0.70 range (Stark & Laurent, 2001), which may be due to several factors. First, these measures often share the same or similar items, which is also the case for the BYI, e.g. item 39 (anxiety scale): "I am afraid that something bad might happen to me", and item 50 (depression scale): "I think that bad things happen to me". Sharing similar items presents difficulties for self-report measures that purport to measure different constructs. Second, the BYI scales

may tap into the same general construct, negative affect or subjective distress (Bose-Deakins & Floyd, 2004). Recent studies indicate that anxiety and depression is related but separate, in that anxiety might be associated with high negative and normal levels of positive affect associated with high physiological arousal, while depression is associated with high negative and low levels of positive affect and normal levels of physiological arousal (Chorpita, 2002). The high correlations in the present study may be caused by lack of specificity in the BYI anxiety and depression total scales, as well as the lack of positively worded items in the scales. Third, anxiety and depression may be distinct conditions that frequently occur together, because similar etiological mechanisms are involved in both phenomena, or because one disorder increases risk of the other (Cole, Peeke, Martin, Truglio & Seroczynski, 1998). Whether the high inter-correlation among the internalizing scales reflects low discriminant validity or high intercorrelations among the concepts that the scales measure remains to be investigated.

Only the Self Concept and Depression scales differentiated between the diagnostic groups within the clinical sample. The mood disorder group scored higher on the scales (lower on Self-Concept) than children diagnosed with other disorders, between which no significant differences were found. The poor ability of particularly the Anxiety scale to discriminate between children diagnosed with anxiety and children with other forms of psychopathology may be a limitation of the BYI in its present form. Studies assessing the discriminative validity of other self-report measures for children designed to measure anxiety (e.g. the Revised Children's Manifest Anxiety Scale, the State Trait Anxiety Inventory for Children, and the Fear Survey Schedule For Children-Revised) indicate that clinically anxious children can be differentiated from non-anxious controls, but that they in general fail to discriminate anxious children from those with other psychiatric diagnoses, e.g. depression and ADHD (Schniering, Hudson & Rapee, 2000). The limited ability to discriminate between anxious children and other psychiatrically disturbed children is thus a general problem of the available anxiety measures, as well as the BYI anxiety scale, and may be due to overlap in symptomatology among the disorders, or that anxiety may be a component in several disorders.

The self-concept, anxiety and depression inventories moderately but significantly discriminated between the norm sample and a combined clinical sample for both genders. For disruptive behavior, differences were found only for boys. The clinical sample consisted of a variety of diagnoses, of which several (for example ADHD and adjustment disorders) are not associated with high degrees of negative affect, which might explain the relatively low discriminative power in the present sample. The two groups differed most in self-concept, which suggest that this construct needs more attention in assessment and treatment of children. For all inventories the combined clinical group had significantly more children exceeding the 90th percentile than the children in the norm group.

Validity was also tested by comparing inventory scores for six diagnostic groups with scores for normal children. All diagnostic

groups except the disorder of social functioning group showed a tendency to have elevated scores on negative affect and deflated scores on self-concept compared with norms. Significant differences were found for the following four diagnostic groups. The mood disorder group scored higher on all scales (lower on self-concept) than norms, and was more elevated in negative affect and deflated in self-concept than the other disorder groups. The conduct disorder group exhibited, to a lesser extent, the same pattern as the mood disorder group, with a significantly deflated self-concept. Scores on disruptive behavior were higher than any other group, but did not reach statistical significance. The anxiety disorder group showed higher scores on anxiety than normal, but numerically lower scores than the mood disorder group. The ADHD group showed deflated scores on self-concept, and a tendency to higher scores on anger. With respect to anxiety, depression, and disruptive behavior the scores of the ADHD group were close to the normal range. The disorder of social functioning group and the specific developmental disorder group did not differ significantly from the norm on any of the inventories. The profile analyses of the diagnostic groups compared to norms thus showed meaningful results, indicating a reasonable convergent validity of the BYI.

Age differences

The test-retest reliability of the self-concept, depression and anger scales for the 7 years olds were lower than the recommended correlation of 0.70 or higher (Myers & Winters, 2002a) and the mean scores of most scales were significantly higher than the mean scores of the rest of the group. However, the internal consistency (alphas) of the scales was very high. Younger children (5–6 years) have been found to have more extreme scores on self-report scales assessing emotional content on a Likert-type scale than older children (9–10 years), perhaps due to more dichotomous thinking at younger ages (Chambers & Johnston, 2002). The higher means of the 7 years olds may be due to a tendency in this age group to respond at the extremes of rating scales, as well as the very high alphas may be explained by a combination of response sets and extreme responding. The low test-retest correlation might be explained by the relatively long interval between the first and second administration (4 weeks). Test-retest correlations have been found to decrease with longer test-retest intervals, for example from 2 weeks to 4 weeks (Michael & Merrell, 1998), and for the very young children a 2-week interval may have been more appropriate than the 4-week interval of the present study. In studies incorporating structured interview formats, consistent age-related increases in the reliability of self-reported symptomatology over 1 to 3 week intervals in children have been found (Michael & Merrell, 1998). Future studies are needed to determine whether the results for the 7 years olds in the present study indicate that emotional problems at this young age are fluctuating more than in older children or can be explained by difficulties in reliably reporting subjective internal states on a self-report scale for children this young.

There are several limitations to the current study. The socioeconomic composition of the norm group was skewed which may have influenced the normative data. The clinical diagnostic groups were small. Thus the results regarding differences between diagnostic groups themselves and between diagnostic groups and norms must be considered as preliminary, and future research with larger samples is warranted. No comorbidity diagnoses were available, and the high degree of comorbidity that can be expected between the diagnostic groups, may be one explanation why a low discriminative validity was found. Finally, the scales were administered in the same order, and all the symptom items were negatively worded, which may have led the children to produce response sets that may have compromised the validity of their scores (Bose-Deakins & Floyd, 2004).

CONCLUSION

The inexpensive, confidential and brief nature of dimensional rating scales makes these instruments attractive tools for research, screening and selection of samples for intervention programs. Danish norms for the five BYI scales are now available. The internal consistency and test-retest reliability of the BYI scales appears to be acceptable, except for 7 years olds and for the anxiety scale. Using the five scales together can provide important insights into possible elevated scores, indicating clinical difficulties as such; however, the BYI in its present form cannot adequately differentiate between internalizing disorders, except for depression. The BYI is a newly developed instrument, and there is still no evidence of the sensitivity to treatment of the BYI scales, which could support their utility in clinical monitoring. Also no parallel forms for parents or an index of social desirability and inconsistency is provided. With the above-mentioned limitations in mind, the BYI might be used as a screening instrument e.g. in school psychology settings, but as for other self-rating scales, the BYI is not a diagnostic instrument, and should not substitute for a diagnostic evaluation. In addition, more than one scale should be used to evaluate a specific construct, thereby assuring a more robust assessment.

REFERENCES

- Andersen, B. H. (2003). *Udviklingen i befolkningens levevilkår over et kvart århundrede*. København: Socialforskningsinstituttet.
- Angold, A., Costello, E. J. & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry*, 40, 57–87.
- Beck, J. S., Beck, A. T. & Jolly, J. (2001). *Beck Youth Inventories™ of Emotional and Social Impairment*. The Psychological Corporation.
- Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, 107, 238–246.
- Bilenberg, N. (1999). The Child Behavior Checklist (CBCL) and related material: standardization and validation in Danish population based and clinically based samples. *Acta Psychiatrica Scandinavica Supplementum*, 398, 2–52.
- Birmaher, B. (1999). *Screen for child anxiety related emotional disorders (SCARED)*. Pittsburgh: Division of Child Psychiatry, Western Psychiatric Institute and Clinic.

- Bose-Deakins, J. E. & Floyd, R. G. (2004). A review of the Beck Youth Inventories of Emotional and Social Impairment. *Journal of School Psychology, 42*, 333–340.
- Bravo, M., Woodbury-Farina, M., Canino, G. J. & Rubio-Stipec, M. (1993). The Spanish translation and cultural adaptation of the Diagnostic Interview Schedule for Children (DISC) in Puerto Rico. *Culture, Medicine and Psychiatry, 17*, 329–344.
- Browne M. V. & Cudeck R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). London: Sage.
- Cattell, R. B. (1966). Scree Test for Number of Factors. *Multivariate Behavioral Research, 1*, 245–276.
- Chambers, C. T. & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology, 27*, 27–36.
- Chorpita, B. F. (2002). The tripartite model and dimensions of anxiety and depression: An examination of structure in a large school sample. *Journal of Abnormal Child Psychology, 30*, 177–190.
- Cole, D. A., Peeke, L. G., Martin, J. M., Truglio, R. & Seroczynski, A. D. (1998). A longitudinal look at the relation between depression and anxiety in children and adolescents. *Journal of Consulting and Clinical Psychology, 66*, 451–460.
- Collett, B. R., Ohan, J. L. & Myers, K. M. (2003). Ten-year review of rating scales. VI: scales assessing externalizing behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 1143–1170.
- Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*, 1–9.
- Elliot, D., Huizinga, D. & Ageton, S. (1985). *Explaining delinquency and drug use*. Beverly Hills, CA: Sage.
- Floyd, F. J. & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286–299.
- Hansen, E. J. (1977). *Socio-economic classification in Denmark*. Copenhagen: Danish Social Research Institute.
- Harter, S. (1985). *Manual for the self-perception profile for children; Revision of the perceived competence scale for children*. Denver: University of Denver Press.
- Ivarsson, T. (2006). Normative data for the Multidimensional Anxiety Scale for Children (MASC) in Swedish adolescents. *Nordic Journal of Psychiatry, 60*, 107–113.
- Ivarsson, T., Svalander, P. & Litalere, O. (2006). The Children's Depression Inventory (CDI) as measure of depression in Swedish adolescents. A normative study. *Nordic Journal of Psychiatry, 60*, 220–226.
- Kadesjo, B., Janols, L. O., Korkman, M., Mickelsson, K., Strand, G., Trillingsgaard, A. et al. (2004). The FTF (Five to Fifteen): The development of a parent questionnaire for the assessment of ADHD and comorbid conditions. *European Child & Adolescent Psychiatry, 13*, 3–13.
- Kovacs, M. (1992). *Children's depression inventory manual*. New York: Multi Health Systems.
- Lewinsohn, P. M., Rohde, P. & Seeley, J. R. (1995). Adolescent psychopathology: III. The clinical consequences of comorbidity. *Journal of the American Academy of Child and Adolescent Psychiatry, 34*, 510–519.
- March, J. S. (1997). *Manual for the multidimensional anxiety scale for children (MASC)*. New York: Multi Health System.
- Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-Fit Indexes in Confirmatory Factor-Analysis – the Effect of Sample-Size. *Psychological Bulletin, 103*, 391–410.
- Michael, K. D. & Merrell, K. W. (1998). Reliability of children's self-reported internalizing symptoms over short to medium-length time intervals. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*, 194–201.
- Myers, K. & Winters, N. C. (2002a). Ten-year review of rating scales. I: Overview of scale functioning, psychometric properties, and selection. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 114–122.
- Myers, K. & Winters, N. C. (2002b). Ten-year review of rating scales. II: Scales for internalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 634–659.
- Nasser, F. & Wisenbaker, J. (2003). A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement, 63*, 729–757.
- Piacentini, J., Roper, M., Jensen, P., Lucas, C., Fisher, P., Bird, H. et al. (1999). Informant-based determinants of symptom attenuation in structured child psychiatric interviews. *Journal of Abnormal Child Psychology, 27*, 417–428.
- Piers, E. V. (1984). *The Piers-Harris children's self-concept scale (The way I feel about myself); Revised manual*. Los Angeles: Western Psychological Services.
- Reynolds, W. M. (1989). *Reynolds child depression scale (RCDS)*. Odessa: Psychological Assessment Resources.
- Schniering, C. A., Hudson, J. L. & Rapee, R. M. (2000). Issues in the diagnosis and assessment of anxiety disorders in children and adolescents. *Clinical Psychological Review, 20*, 453–478.
- Spielberger, C. (1988). *State-trait anger expression inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Stark, K. D. & Laurent, J. (2001). Joint factor analysis of the Children's Depression Inventory and the Revised Children's Manifest Anxiety Scale. *Journal of Clinical Child Psychology, 30*, 552–567.
- Storch, E. A., Masia-Warner, C., Dent, H. C., Roberti, J. W. & Fisher, P. H. (2004). Psychometric evaluation of the Social Anxiety Scale for Adolescents and the Social Phobia and Anxiety Inventory for Children: construct validity and normative data. *Journal of Anxiety Disorders, 18*, 665–679.
- Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.
- Trillingsgaard, A., Damm, D., Sommer, S., Jepsen, J. R. M., Ostergaard, O., Frydenberg, M. et al. (2004). Developmental profiles on the basis of the FTF (Five to Fifteen) questionnaire – Clinical validity and utility of the FTF in a child psychiatric sample. *European Child & Adolescent Psychiatry, 13*, 39–49.
- Wadsworth, M. E. & Achenbach, T. M. (2005). Explaining the link between low socioeconomic status and psychopathology: testing two mechanisms of the social causation hypothesis. *Journal of Consulting and Clinical Psychology, 73*, 1146–1153.
- Winters, N. C., Myers, K. & Proud, L. (2002). Ten-year review of rating scales. III: scales assessing suicidality, cognitive style, and self-esteem. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 1150–1181.

Received 12 October 2007, accepted 23 August 2008