Chapter 5

Validation of the Child Behavioral Checklist for Guardians of Unaccompanied Refugee Minors

Abstract

The objective of this study is to validate the Child Behavioral Checklist (CBCL) for the legal guardians of unaccompanied refugee minors residing in the Netherlands. The legal guardians (caseworkers) of the unaccompanied minors ($N = 920$) that participated in the study all received a CBCL to report on the mental health of the unaccompanied minor. The guardians filled in and returned 478 CBCL's. The results of the hierarchical confirmative factor analyses support a one-factor and a two-factor structure of externalizing and internalizing scales equally well. Moreover, the fit of the original individual eight first order factor models of the a priori CBCL subscales were found to be moderate. The total, internalizing, and externalizing scales show good internal consistency. The construct and concurrent validity of the CBCL were also examined and found to be moderate to good. The findings of this study suggest that the CBCL is a reliable and valid measure for use by guardians to assess the maladaptive emotional and behavior problems of unaccompanied refugee minors.

Introduction

The Child Behavior Checklist (CBCL) finds its origin in the United States (Achenbach & Edelbrock 1983; Achenbach, 1991b). It is a checklist that is usually completed by parents to report on the behavioral and emotional problem of their children. The CBCL and its

corresponding measures (e.g., YSR, TRF) have been used in many countries and with children and adolescents from diverse cultures (Kvernmo & Heyerdahl, 1998; McKelvey, Davies, Sang, Pickering & Hoang, 1999; Stevens et al., 2003; Wiesz et al.,1993; Zukauskiene, Pilkaukaite, Malinauskiene, & Krataviciene, 2004) to measure maladaptive behaviors and emotional problems. Furthermore, the CBCL has been utilized with refugee adolescents (Mollica, Poole, Son, & Murray, 1997; Rousseau & Drapeau, 1998; Sourander, 1998) from an array of different cultures.

Although the CBCL was developed for parents to report on the behavioral and emotional problems of their own children, it has been used in research settings with other significant adults such as residential mental health workers (Albrecht, Veerman, Damen, & Kroes, 2001; Wherry et al., 1992), hospital staff (Kazdin & Bass, 1988), foster parents (Shore, Sim, Prohn, & Keller, 2002; Strijker, Zandberg, & van der Meulen, 2002; Tarren-Sweeney, Hazell, & Carr, 2004), clinicians (Dutra, Campbell, & Westen, 2004) and staff members of a refugee reception center in Finland (Sourander, 1998). However, there have been a few studies which have found that when significant adults (caregivers) other than parents report on the emotional and behavioral problems of children the assessment of internalizing problems is less reliable than parental reports (Tarren-Sweeney et al., 2004), can lead to biases originating from placement issues (Garland et al., 1996), and to under-reporting due to limited "familiarity" with the child which is directly associated with the length of time the adult has known the child (Starr, Dubowitz, Harrington, & Feigelman, 1999). Furthermore, it is well known that the concordance between the reports of different informants is frequently low (i.e., Achenbach, McCounaughy, & Howell, 1987) and constitutes a risk factor for the development of (internalizing) psychopathology in adolescents (Ferdinand, van der Ende, & Verhulst, 2004; Rueter, Scaramella, Wallace, & Conger, 1999). Discrepancies in reports from multiple informants can result from specific context related problems (i.e., at home but not at school) and/or differences in how the informants perceive the behaviors/emotions which are manifested (see Shore et al., 2002 for a discussion).

During the last ten years, there has been considerable debate surrounding the factorial structure and validity of the CBCL. Dedrick, Greenbaum, Freidman, Wetherington, & Knoff (1997) and De Groot, Koot, & Verhulst (1994) found modest evidence in their confirmatory factor analyses for the applicability of the original eight-factor model of the CBCL. A recent study by Dumenci, Erol, Achenbach, & Simsek (2004) verified the first order eight-factor model with a very large Turkish sample ($n \approx 5000$). However, there have been a number of contradictory findings regarding the validity of the eight-factor model (cross informant syndromes) of the CBCL, especially with respect to the two syndromes; Social Problems and Attention Problems (Heubeck, 2000). Hartman and colleagues (1999) found minimal support for the two-factor or eight-factor models, only significant support was found for a second order one-factor model. Heubeck (2000), just like Dedrick et al. (1997), found the second order one-factor and first order eight-factor model to both have a modest fit in their confirmatory factor analyses.

In most studies, like the ones mentioned above, the hierarchical two-factor (second order) model of the CBCL is rarely examined. This is quite odd since most of studies that report on the findings of the CBCL frequently use the broadband, "internalizing" and "externalizing" terminology to describe findings. A study from Greenbaum and Dedrick (1998) used a hierarchical confirmatory analysis based on 16 parcel indicators (instead of the 85 items) to evaluate the two-factor second order structure of the CBCL. In the Greenbaum and Dedrick study and a replication thereof that was carried out among clinicians (Dutra et al., 2004), support was found for the two-factor second order model. Additional support was found for the second order two-factor model by Albrecht and colleagues (2001) who conducted a study in the Netherlands among 846 adolescents which lived in residential institutions. Mental healthcare staff workers completed the CBCL for the adolescents. Albrecht et al. conducted a confirmative factor analysis (CFA) on the original items of the CBCL. They concluded that the data best fit the hierarchical two- factor second order model of maladaptive internalizing and externalizing behaviors and modestly fitted the eight-factor first order model. Macmann and Barnett (1993) in their critical examination of the interpretations of the CBCL also theoretically favored the two- factor model and had earlier found in their own evaluation of the second order structure that the CBCL for practical purposes can be best seen as a global index of the emotional and behavioral problems

(Macmann, Barnett, & Lopez, 1993). Although, there are discrepancies with respect to the factorial validity of the CBCL, the two-factor second order structure was better supported with informants other than custodial parents than the eight-factor first order model. Empirically, the internalizing and externalizing scales have also been found to be reliable and valid measures of child psychopathology over time (e.g., Verhulst & van der Ende, 1992).

The CBCL was used in the present study for the first time with legal guardians, which function as caseworkers for unaccompanied refugee minors (URM). Despite the fact that there are some reliability and validity issues surrounding the use of the CBCL with informants other than parents, a substantial amount of documented information is available regarding the affirmative use of the CBCL with other informants than parents. The objective of the present study was to validate the CBCL for legal guardians of unaccompanied minors. There is one foundation that has the legal guardianship of all of the unaccompanied minors that reside in the Netherlands, The Nidos Foundation. The minors come from more than 100 different countries. This foundation has offices throughout the entire country and has almost 20 years of experience in working with unaccompanied minors. The guardian functions as a caseworker for the welfare of the unaccompanied minor and as a rule has a degree in social work. He/she is responsible for the emotional, developmental and educational needs, housing, allowance and asylum procedure of the unaccompanied minor. Furthermore, the guardian has all of the legal responsibilities and parental authority of the minors just as a custodial parent would have (for example; legal permission is needed from the guardian for a minor younger than 16 years of age for medical treatment, opening a bank account, and applying for a passport).

On the average, 20 unaccompanied minors will be assigned to one guardian. A guardian has at least one appointment (1 hour) with an unaccompanied minor per month (more if needed) and is available at the office for the unaccompanied minor. The guardian discusses with the minor how things are going and sorts out any problems that need to be discussed. If the minor is transferred to a different residential setting or Nidos office, they will receive a new guardian and their file will be transferred with them so that no information is lost. Although the guardian has limited personal contact with the unaccompanied minor, he/she is kept up-to-date regarding the functioning of the unaccompanied minors from the housing staff, group worker or teachers on a daily to weekly basis and is readily available in crisis situations.

Because of the uniqueness of this study, Dutch guardians, reporting on the mental health of adolescents from a wide variety of countries and ethnical backgrounds, and the previous conflicting findings regarding the validity and reliability of the CBCL, it was prudent to examine the psychometric properties of the CBCL for this specific population. In addition, much too often in research with adolescents from other cultures, no attempt is made to validate the instruments that were utilized resulting in uncertainty surrounding the results of the study (Drotar, Stein, & Perrin, 1995). The value which can be attached to results of a study is, of course, determined by the degree of reliability and validity of the instrument that has been utilized. In this study, the endeavor was undertaken to evaluate the psychometric properties of a well known existing psychological instrument, the CBCL, for use with a specific research population, the guardians of unaccompanied refugee minors.

Method

Context of study

Due to a dramatic increase in the number (15,000) of unaccompanied minors in the Netherlands in 2001 and problems in referring unaccompanied minors to mental healthcare services, a national and longitudinal research project "Unaccompanied Refugee Minors and Dutch Mental Healthcare Services" was started among unaccompanied refugee minors living in The Netherlands and among their guardians, teachers and professional mental healthcare providers in 2001. Ethical approval for this study was given by the Medical Ethics Committee of the Leiden University Medical Center, Leiden University.

The goal of the project was to determine the severity level of psychological distress of unaccompanied minors, their need for mental healthcare, the availability of mental healthcare services for this group and finally, the associations between all of these factors. A secondary

goal of the project was to validate screening instruments that measure emotional distress and behavioral problems for this specific population group.

Sample

Demographic information on the unaccompanied minors in The Netherlands was supplied by the Nidos Foundation. Approximately 4000 unaccompanied minors were randomly selected in 2002 from the total population of 12,000 in the Central Registrar of Nidos. Information about the study and permission waivers (available in translated versions) were sent to the guardians to discuss with the unaccompanied minors. Both the minor and his/her guardian gave written permission for the unaccompanied minors to participate. Roughly 2300 unaccompanied minors' permission waivers were returned; 1300 (57%) wished to participate, 15% refused, 12% did not participate for a wide range of practical reasons, 9% were transferred, and 7% turned out to be untraceable. However, there were no statistical significant differences found between the URM that did participate and the URM that did not in gender, age, and country of origin. A total of 920 unaccompanied minors were present for participation. The final sample was representative in all of the main characteristics of the total unaccompanied minors population aged 12 to 18 year old in 2002 in the Netherlands. The unaccompanied minors came from 48 countries. Two-thirds of the sample had lived in the Netherlands for a period of 18 months or less.

Measures
CBCL

The Dutch version of the CBCL-4/18; 1991 Profile-(Achenbach, 1991b; Dutch translation: Verhulst , Van der Ende, & Koot, 1996b) was used to standardize the assessment of the behavior and emotional problems of unaccompanied minors through the observations of guardians. The CBCL has a three point rating scale; 0 = *not true*, 1 = *somewhat true*, and 2 = *very true*. The 118 problem items of the CBCL were explanatory factor analyzed to empirically identify the constructs of psychopathology that occur in adolescents (Achenbach, 1991b). These items are presented in the English version in alphabetical order to reduce the bias that might occur as a result of informants' preconceived notions regarding the presence or absence or a particular disorder. The Dutch items follow the same order as the English; however they are no longer alphabetical. The CBCL can be scored in three ways; (1) eight first order scales-withdrawn, somatic, anxious/depressed, social problems, thought problems, attention problems, delinquent and aggressive; (2) two second order scales, internalizing (consisting of the withdrawn, somatic, anxious/depressed scales) and externalizing (consisting of the delinquent and aggressive) scales, and (3) a Total score. The validity and reliability of the Dutch CBCL for normative and clinical populations is thoroughly described by Verhulst et al. (1996b). Although the CBCL has 118 problem behavior items, only 85 items are utilized in the subscales of the CBCL. The scales were originally classified on the basis of exploratory factor analysis (Achenbach, 1991b).

Mental Health Questionnaire for guardians

The perceived need for mental healthcare for the unaccompanied minor by the guardian, the referral process to mental healthcare (MHC) services and the satisfaction with the utilized mental healthcare was measured using a checklist of 23 items. Examples of some questions are; Do you find that this minor needs professional psycho-social mental healthcare? , Did you refer this minor to a mental healthcare facility that provides psychosocial assistance? , Did this minor want to go to the MHC facility? Did you go with the minor to the facility? Have you seen a change in the symptoms/behaviors of the minor after treatment? The guardians were also asked to fill in 10 questions about themselves and their experience with working with adolescents and specifically URM.

Mental Health Questionnaire for teachers

The need for mental healthcare of the unaccompanied minor perceived by the teacher and the referral process to mental health services were measured using a checklist of 6 items. Examples of some questions are; Do you find that this minor needs professional psycho-social mental healthcare?, Did you have contact with the guardian of this minor about the psychosocial problems of the minor?, and Did you contact the school doctor about the

psychosocial problems of this minor? The teachers were also asked to fill in 10 questions about themselves and their experience with working with adolescents in general and specifically with URM.

Teacher's Report form (TRF)

The Dutch version of the TRF 4/18; 1991 Profile-(Achenbach, 1991c)-Dutch translation (Verhulst, van der Ende, & Koot, 1997) was used to standardize the assessment of the behavioral and emotional problems of unaccompanied minors through the observations of teachers. The validity and reliability of the Dutch TRF for normative and clinical populations is thoroughly described by Verhulst et al. (1997). The psychometric properties for the TRF in this study did not differ from those of Verhulst and colleagues (1997).

Stressful Life Events

The *Stressful Life Events* (SLE) (Bean, Eurelings-Bontekoe, Deluyn, & Spinhoven, 2004b) checklist was used to assess the number and type of stressful event(s) that was experienced. The SLE consists of 12 dichotomous (yes/no) questions and an open question on the occurrence of stressful life events of relevance for adolescent refugee minors (e.g. Have you ever experienced a war or an armed military conflict going on around you in your country of birth?, Has someone ever hit, kicked, shot at or some other way tried to physically hurt you?, Have you ever been separated from your family against your will? Have you been involved in a serious accident? (for example involving a car,)Has someone ever tried to touch your private sexual parts against your will or forced you to have sex?). The overall mean total score of 6.5 on the SLE for URM has been replicated in 5 independent studies. The overall mean of URM is significantly higher than the total mean SLE scores for parental accompanied immigrant/refugee adolescents, Dutch and Belgium adolescents (Bean et al., 2004b).

HSCL-37A

The Hopkins Symptom Checklist-37 for Adolescents (HSCL-37A) (Bean et al., 2004a) measures internalizing distress and externalizing behavior (trauma-related "acting-out"). The psychometric properties have been investigated among a culturally diverse adolescent population and appeared to be satisfactory to good (Bean et al., 2004b). Internal reliability for the URM sample for the total scale internalizing distress, and externalizing behavior subscales was respectively .91, .92, and .69. Twelve-month test–retest reliability for the total scale was .63 ($p < .001$). Inter-measure correlations with the total scores of the RATS and SLE were respectively .77 ($p < .001$) and 0.38 ($p < .001$). Using a confirmatory factor analysis, the two-factor (internalizing and externalizing) structure was verified in the URM sample with a loss of only .4% of the explained variance.

The *Reactions of Adolescents to Traumatic Stress* (RATS) is a self-report questionnaire developed to assess posttraumatic stress reactions defined in the DSM-IV (APA, 1994) with culturally diverse adolescents. The RATS consists of 22 items that correspond directly to the B (intrusion), C (numbing/avoidance), and D (hyper-arousal) criteria of the DSM-IV for PTSD. Items were adapted to measure symptoms of intrusion, numbing/avoidance and hyper-arousal in adolescents, especially adolescent refugees. The psychometric properties have been investigated among culturally diverse adolescent populations and per language version of the RATS and appear to be satisfactory to good (Bean et al., 2004c). Internal reliability for the URM sample for the total scale, and intrusion, numbing/avoidance and hyper-arousal subscales was respectively .88, .85, .69, and .73. Twelve-month test-retest reliability for the total scale was .61 ($p <.001$). Using a confirmatory factor analysis, the three-factor structure was verified in the URM sample with a loss of only 3% of the explained variance (Bean et al., 2004c).

Mental Health Questionnaire for adolescents

The self-perceived need for, knowledge of and satisfaction with MHC services was measured using an interview of 23 items. The interview was individuality conducted with the URM after they had filled in the other three questionnaires to ensure that the questionnaire would be filled in properly. The research assistants always stressed that the questions were

about receiving help for problems regarding "thinking and feeling" and not about practical problems. This interview was translated in the above mentioned languages so that the URM could read along in their own language if that was necessary or for clarification. Examples of the questions are: Do you think that you have problems (emotional) that you need help for?, Would you like to contact someone that could help you (with your emotional problems)?, Have you already been to a "(mental) health professional" (for your emotional problems)?. The answer categories were specific to the nature of each question.

Assessment procedure

The CBCL and a short questionnaire regarding the mental healthcare for the unaccompanied minors was sent to the supervisors of each regional office for each guardian that was responsible for at least one of the 920 unaccompanied minors that took part in the study. The guardians received a letter with the questionnaires informing them about the study and giving instructions concerning how the questionnaires should be filled in. The guardians were instructed in the letter and by their supervisors that they could fill in the questionnaire or ask a staff member of the living unit/foster parent of the unaccompanied minors to do so. However, the guardian remained responsible for retuning the completed questionnaires to their supervisors which in turn sent all the completed questionnaires back from the regional office. For the first assessment period, 557 questionnaires were returned. From the 557 questionnaires that were returned for the first assessment 478 CBCL (118 items) questionnaires were filled in of which 421 completed all of the 85 core items. The rest of the returned questionnaires pertained only to the questions concerning the mental healthcare of the URM. Twenty-two percent of the completed CBCL's were filled in by someone else than the guardian (e.g., staff at residential setting).

Procedures for the confirmatory factor analyses (CFA)

The minimum CFA sample requirements are 10 cases per item (i.e., five cases for the factor loadings and five cases for the residual; Kline, 1998). A simpler procedure that can be used involves a scale-based CFA (e.g., Byrne, 1988; McCreary, Newcomb, & Sadava, 1998), in which parcels of items are used as indicators to form the latent factors that represent the subscales. Bandalos (2002) stated that the parcel method is often applied to highly skewed, categorized data to obtain distributions that are more normal and continuous in which CFA' s can be generated from. The parcels can vary in the number of items they contain, and typically three parcels are created for each latent factor (Nasser & Wisenbaker, 2003). However, there have been two studies (Greenbaum & Dedrick, 1998; Dutra et al., 2004) in which two parcels have been used per sub-scale (total of 16 parcels) to evaluate the hierarchical structure of the CBCL because two subscales (Social problems ;8 items) and Thought problems (7 items) have not enough items to be divided into three parcels. The present study used the same methods (random selection of items for parcel, summing of items) as Greenbaum and Dedrick to create the parcels which will be used as indicators for the hierarchical confirmatory analysis.

Bagozzi and Heatherton (1994) have listed the many advantages of using parcels. Nonetheless, when using parcels as indicators in the model, information is lost at item-level. To supplement the findings of the hierarchical analysis, individual first order and second order factor models will be examined to evaluate how the single items behave in their a priori defined factors.

Models

In this study, the fit for each of the eight first order factor models and the second order factor models will be examined individually to evaluate the behavior of items in each model. Furthermore, individual second order factor models will be examined to see if they can be better described as single models or as a grouping of smaller factors. The second order factor structure of the two a priori internalizing and externalizing factors will be evaluated using parcels as indicators in a hierarchical CFA (see Figure 1). Also, the possibility of a one-factor second order structure will be investigated.

Statistical Analysis

The indicators, parcels and subscales of the CBCL, are skewed and in great violation of normal distribution assumptions (see Table 2). The indicators for all of the models were treated as ordinal. There were additional analyses not discussed in the article in which the indicators for the CBCL were treated as dichotomous, however, no improvement was found in the fit of the models when using dichotomous indicators instead of ordinal indicators. Missing items were deleted listwise since deleting items pairwise resulted in multiple Heywood cases which prevented convergence of the models. It is known that the Maximum Likelihood (ML) method of estimation, which uses product-moment correlations for estimating model parameters, is based on data in which there is no multivariate kurtosis (Browne, 1984). This method is not adequate to use with the URM sample because of the non-normality of the indicators. It was necessary to calculate polychoric correlation matrices and the asymptotic covariance matrices using PRELIS 2 (Jöreskog & Sörbom, 1990; Jöreskog & Sörbom, 1996b) on which all of the confirmatory factor analysis could be calculated so that the model parameters were not underestimated, that the chi-squared statistic would not be inflated and that the standard error estimates would not be downwardly biased (Flora & Curran, 2004). From the matrices, the individual factor models (using items) and hierarchical factorial structure (using parcels) could be calculated with LISREL 8.71 (Jöreskog & Sörbom, 1996a) using the unweighted least squares (ULS), weighted least squares (WLS), or diagonally weighted least squares (DWLS) methods for estimating model parameters.

The ULS method has been used in many CBCL studies (e.g., Albrecht et al, 2001; Dedrick et al., 1997; De Groot et al., 1994; Hartman et al., 1999). Dumenci et al. (2004) recently used the WLS method, an asymptotically distribution-free (ADF) estimator in their study evaluating the eight-factor first order model of the CBCL. They used this method instead of the ULS because they found that the ULS is not the most efficient estimator of the model. However, the WLS method can only be used with very large samples. Jöreskog and Sörbom recommend a sample size $=1.5p$ $(p + 1)$, where p = number of variables. To use the WLS method with the 85 items of the CBCL, a sample of more the 7000 would be necessary! However, using the suggested sample size formula for the 16 parcels, a stable weight matrix could be obtained using the WLS method (minimum of 408) since the effective sample size of the present study is 421 fully completed CBCL's. The DWLS (Muthén, du Toit, & Spisic, 1997) method (estimation capabilities falling in between the ULS and WLS methods [Jöreskog & Sörbom, 1996a]) was applied to the generation of the eight first order a priori individual factor models. Although the DWLS method is not the best estimator of the parameters, it seems to be the best method to use when evaluating the factor models in relation to the URM sample because it has behaved stable among smaller sample sizes (Flora & Curran, 2004) while the WLS method did not.

Browne and Cudeck (1993) and Hu and Bentler (1999) recommend using multiple fit indices to determine how well the hypothesized factor structure will fit the observed data. To examine the models, indices of model fit, model comparison and model parsimony were calculated. The fit indices include (a) Satorra-Bentler chi-square (SB) (values should not be significant, but in larger samples this is often not feasible), (b) [chi square]/df ratio (values should be < 2.0) (c) Incremental indices of fit were examined : the parsimony normed fit index (PNFI), the comparative fit index-takes the non-centrality parameter into consideration (CFI) and Adjusted Goodness-of-Fit Index (AGFI) values should be [greater than or equal to] .80). Root Mean Square Error of Approximation (RMSEA) values less than .08 indicate at least sound fit while values between .08 and .1 reflect mediocre fit (Byrne, 1998). Incremental indices reflect the improvement in fit gained by a given factor model relative to the most restrictive (null or independence) model. All three incremental indices are scaled from 0 (no fit) to 1 (perfect fit). Hu and Bentler (1999) advised that values close to .95 are indicative of good fit. PNFI values range from 0 to 1, values close to 1 (perfect fit) are not expected. Moreover, indices around .5 are not unexpected in sound-fitting models (Byrne, 1998). Parsimony adjusted measures take the number of parameters estimated in the model into account. Models are penalized for each parameter. Multiple fit statistics can be used to compare models with differing number of parameters to determine the impact of adding additional parameters to the model based on theoretical driven assumptions.

Descriptive statistics were used to give summary descriptions of the demographic characteristics of the URM sample (Table 1). Internal consistency of the total scale, subscales, and parcels of the CBCL was calculated with Cronbach's α (Table 2). Pearson's product-moment correlations and Spearman's rho (two-tailed) were used to study the associations (pairwise exclusion of missing data) between raw scores. The kappa statistic was used to assess the agreement between the psychopathology indicators used in this study. Differences between groups (Table 5) were determined by using t-tests for independent groups, ANOVA's and effect sizes [calculated using Cohen's *d*](Cohen, 1988). A maximum of ten percent of the missing items were allowed to still be able to extrapolate the total and subscale scores of all measures.

Table 1.
Summary of Sample Characteristics of unaccompanied minor and guardians

		Unaccompanied minor	Guardians
N		(920) 478*	(557) 478*
Gender			
	Male	71.3%	22.4%
	Female	28.7%	77.6%
Age in years			
	Mean	15.48	36.27
	SD	1.52	8.96
	Range	10-18	20-64
Country of Origin			
Netherlands		0.0%	78.9%
Angola		43.9%	
Iran/Afghanistan/Iraq		4.4%	
Eritrea/Ethiopia		2.7%	
Somalia		2.1%	
Sierra Leone		7.9%	
Guinea		6.7%	
Other African Countries		14.0%	
China/Tibet		8.6%	
Other Countries		9.6%	21.1%
Most frequent reported level of education		1-5 years of education 44.9%	Bachelor's Degree 92.7%
Most frequent reported Occupation		N/A	Social worker (93%)

Note. ( ) Total number of completed/returned questionnaires for the first assessment
* number of CBCL's filled in and returned

Results

Demographic information about unaccompanied minors and guardians
Table 1 represents the demographic background information for the 478 unaccompanied minors (of the 920 of whom there was a CBCL completed) and their guardians. For thirty-five percent of the all of the 478 minors, one guardian filled in a questionnaire for one unaccompanied minor that had taken part in the study, in 22% of the cases one guardian filled in questionnaires for 2 minors., in 18% of the cases guardians filled in questionnaires for 3 minors and in 25% of the cases one guardian filled in questionnaires for more than 4 minors.

The sample of unaccompanied minors consisted mostly of boys (71.3%). The mean age was approximately 16 years and the most frequent countries of origin were Angola, Sierre Leone, Guinea and China. Most of the unaccompanied minors came from Africa (69.4%). 55.1% of the unaccompanied minors sample had received educational training for more than 5 years. The guardian population consisted of mostly females (77.6%) that had a mean age of 36 years. The greater majority of guardians were born in the Netherlands. A large portion of the guardians had received the equivalent of a Bachelor's degree in Social Work. Almost seventy-five percent of the guardians reported being "very well informed" about the kinds of behavioral and emotional problems that URM frequently experience. The guardians also reported on their work experience with youth. Forty-five percent of the guardians said to have worked 5 years or less with youth in general, 63% said to have had 5 years or less working experiences with non-Dutch youth and 51% reported having 3 years or less experience working with URM.

Table 2.
Descriptive statistics for indicator variables (Listwise exclusion of missing cases)

CBCL scales	CBCL items	α	*M*	*SD*	Skewness	Kurtosis
Withdrawn	All items in subscale	.78	3.25	3.03	1.05	.70
Somatic	All items in subscale	.76	1.61	2.35	2.36	7.46
Anxious/Depressed	All items in subscale	.84	4.24	4.24	1.48	2.72
Social	All items in subscale	.64	1.08	1.65	1.90	3.73
Thought	All items in subscale	.65	.83	1.49	3.73	20.96
Attention	All items in subscale	.80	2.71	3.08	1.57	2.75
Delinquent	All items in subscale	.70	1.43	2.12	2.15	5.54
Aggressive	All items in subscale	.88	3.11	4.47	1.98	3.93
Internalizing	All items in subscale	.89	8.40	7.72	1.46	2.69
Externalizing	All items in subscale	.90	4.55	6.12	1.94	3.71
Total	All 85 items	.94				
Withdrawn parcel 1	65, 69, 75, 80, 88, 102	.67	2.06	2.04	.99	.44
Withdrawn parcel 2	42, 103, 111	.62	1.20	1.31	1.21	1.30
Somatic parcel 1	54, 56a, 56b, 56d, 56c, 56f	.69	1.39	1.98	2.03	4.71
Somatic parcel 2	51, 56e, 56g	.62	.23	.61	3.33	14.00
Anx/dep parcel 1	14, 31, 32, 35, 71, 103, 112	.72	2.47	2.36	1.04	.90
Anx/dep parcel 2	12, 33, 34, 45, 50, 52, 89	.76	1.80	2.26	1.87	4.29
Social parcel 1	1, 11, 55, 64,	.36	.63	.99	1.73	2.80
Social parcel 2	25, 38, 48, 62	.69	.46	1.01	2.81	9.13
Thought parcel 1	40, 66, 70, 85	.59	.13	.58	7.29	68.51
Thought parcel 2	9, 80, 84	.52	.71	1.11	2.11	5.46
Attention parcel 1	1, 10, 13, 41, 45, 61	.66	1.51	1.86	1.51	2.19
Attention parcel 2	8, 17, 46, 62, 80	.63	1.24	1.50	1.43	2.19
Delinquent parcel 1	39, 43, 63, 67, 72, 96, 105	.55	.71	1.21	2.20	5.35
Delinquent parcel 2	26, 81, 82, 90, 101, 106	.47	.74	1.15	1.97	4.47
Aggressive parcel 1	7, 16, 19, 20, 37, 87, 93, 95, 97, 104	.77	1.50	2.29	2.00	4.09
Aggressive parcel 2	3, 21, 22, 23, 27, 57, 68, 74, 86, 94	.81	1.63	2.42	1.81	3.13

Individual Confirmatory and Hierarchical Confirmatory Factor Analysis
    The fit of all of the eight lower order factor models are sound to mediocre (not shown, but available). The greater majority of the item factor loadings were estimated above .40, ranging in mean estimates from .51 to .72. Although the SB chi-square goodness-of-fit statistics for all of the individual factors except Social and Thought, lacked fit, the alternative

measures of fit supported the six other factor models. These findings do not diverge from the results of previous studies regarding the lower-order factor model of the CBCL that have been reported earlier in this article.

The SB chi-square goodness-of-fit test revealed for the hierarchical structure of the CBCL (Figure 1) that the model lacked fit for the observed data in the present study. However, when the alternative measures of fit were examined, the results indicated that the fit of the model was good (SB-χ^2(92)= 340; AGFI = .97; PNFI = .74; CFI = .98; RMSEA = .08). The results of the model fit of the present study are very similar to those of the Greenbaum and Dedrick study findings. None of the modification indices suggested further refining of the model. In the model, the Thought factor loaded quite high on internalizing (.75) while the Social factor (.71) loaded quite high on the externalizing model.

The correlation between the second order internalizing and externalizing factors was .84. Because of the traditional cross-loadings of the Social, Thought, and Attention syndromes on both internalizing and externalizing factor models, the correlations between the second order factors could have been inflated. To test this hypothesis, the two factor second order model was recalculated without the Social, Thought, and Attention factors. The new model yielded a slightly lower correlation of .70 between internalizing and externalizing factors.

Because of this high correlation and the previous findings that have been documented in CBCL studies indicating that the one-factor second order model fits the data better than a two-factor model, a final one-factor second order model was calculated (SB-χ^2(96)=380; AGFI=.97; PNFI=.77; CFI=.98; RMSEA=.08). There was a small difference found in model fit between the one-factor second order model and the two-factor second order model. The two-factor second order model fit the data slightly better than the one-factor second order model.

Reliability

The reliability of the total CBCL scale is .94. The remaining Cronbach's alpha values for the rest subscales of the present study are located in Table 2 and are consistent with those calculated for Dutch parents (Verhulst et al., 1996), American parents (Achenbach, 1991), residential staff workers (Albrecht et al., 2001), and clinicians (Dutra et al., 2004).

Inter-measure correlations

The correlations for the first assessment between all the scales of the CBCL are presented in Table 3. Using Cohen's effect magnitude for correlations, correlations above .10 are considered small, above .30 are considered medium and correlations above .50 are considered large (Cohen, 1988). These correlations are comparable to those found for Dutch parents (Verhulst et al., 1996). It can be observed from Table 4 that the externalizing scale and internalizing scale show a correlation of .40. Accordingly, this means that these two scales are not totally independent of each other.

Table 3.
Intercorrelations of the scales of the CBCL

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Withdrawn										
2. Somatic	.47									
3. Anxious/Depressed	.67	.55								
4. Social	.41	.28	.49							
5. Thought	.54	.49	.63	.33						
6. Attention	.59	.44	.69	.63	.62					
7. Delinquent	.39	.15	.34	.49	.24	.58				
8. Aggressive	.30	.17	.42	.59	.30	.62	.68			
9. Internalizing	.85	.75	.92	.49	.67	.70	.36	.38		
10. Externalizing	.35	.18	.43	.61	.30	.66	.84	.97	.40	
11. Total score	.72	.60	.84	.68	.69	.86	.63	.72	.88	.75

Note. All correlations reached the .01 significance level. Two-tailed.

Inter-informant agreement

In Table 4, the correlation matrix for inter- and intra-informant correlations can be found. The intra-informant specific correlation between internalizing and externalizing (raw) scores were all significant and strong. The internalizing correlations per informant pair were found to be significant but small (minors & guardian, $r = .23$; minors & teacher, $r = .13$; teacher & guardian, $r = .26$). The externalizing correlations per informant pair were smaller between minors and guardian ($r = .18$), and minors and teachers ($r = .14$), than between guardians and teachers ($r = .47$). These findings do not deviate extensively from the inter-informant correlations found in previous studies (Achenbach et al., 1987; see Tarren-Sweeney et al., 2004 for an overview). There were strong significant and positive correlations between URM's self-reported total number of SLE's and internalizing, externalizing and traumatic stress reactions (small to medium effect sizes). However, there were no significant correlations found between the guardians' or teachers' report of psychological distress or behavioral problems and URM report of total number of SLE's. Older age was significantly negatively related to the guardians (and teachers) reports of externalizing problems, however significant, positively related to the unaccompanied minors reports (negligible effect size).

Validity
Confounding factors for the validity of the CBCL among guardians

In this specific study, several factors such as someone other than the guardian completing the CBCL (22%), one guardian completing multiple CBCL's for individual URM, or experience in working with URM could have in some way confounded the validity results. Therefore, the effect of the confounding factors will be first addressed and then the actual validity findings.

First, the reliability of the reports of the guardians was compared with those of non-guardians to investigate if there were differences. There was no indication that the guardians reports (internalizing - minors & guardian, $r (338) = .23$, $p <.01$; teacher & guardian, $r(210) = .30$, $p <.01$; externalizing- minors & guardian, $r (345) = .18$; $p <.01$; teacher & guardian, $r(200) = .45$, $p <.01$); were less reliable than the reports from mentors/residential staff workers who have daily contact with the minors (internalizing- minors & guardian, $r (77)=.25$, $p <.05$; teacher & guardian, $r(44) = .24$, $p <.05$; externalizing - minors & guardian, $r (77)=.24$; ns; teacher & guardian, $r(44) = .53$, $p <.01$).

Furthermore the effect of one guardian completing multiple CBCL's was examined. Previously was reported that 75% of the CBCL's were filled in by one guardian for 3 or less minors. There was a small but significant and negative relationship found between the number of CBCL's filled in by the same guardian and the reported CBCL internalizing scores (rho $(372) = -.11$, $p <.05$). Implying that if a guardian had filled in multiple CBCL's they also had reported lower internalizing scores. There was no correlation found between CBCL externalizing scores and the number of CBCL's filled in by one guardian.

If the guardian said that he/she was "very well informed" of the types of emotional problems that unaccompanied minors experience, there was a larger and significant inter-rater agreement found between the minor and guardian CBCL internalizing report (r $(30) = .47$, $p <.01$) than when the guardian was "well informed" (r $(301) = .20$, $p <.01$) or "average informed" (r $(52) = .19$, ns). However, the results regarding the agreement between minors and guardians concerning externalizing problems were different. If the guardians had said that they were "very well informed" (r $(28) = .26$) the relationship was not significant. Whereas among the guardians who said they were "average informed" (r $(51) = .30$, $p <.05$) or "well informed" (r $(310) = .14$, $p <.01$) stronger and significant relationships were found.

Similarly, the variable "years of work experience with unaccompanied minors" had effect on the agreement between guardian and minor concerning externalizing problems. However, in this case, agreement was better when guardians had reported having many years (7 or more) of experience (r $(60) = .26$, $p <.05$) than when they reported medium (4-6 years) (r $(147) = .18$, $p <.01$) or little (0-3 years) (r $(207 = .12$; ns) experience working with minors. However, if the guardian had reported many years of experience, the agreement between the guardian and minor concerning internalizing problems was moderate (r $(57) = .27$, $p <.05$) and did not differ from the category of guardians with little (0-3 years) (r $(203) =.24$; $p <.01$) or medium (4-6 years) (r $(148) = .20$, $p <.01$) experience.

Table 4.
Intermeasures Correlations

	1.	2.	3.	4.	5.	6.	7.	8.
1. internalizing HSCL-37A (n)								
2. externalizing HSCL-37A (n)	.42** (825)							
3. internalizing CBCL (n)	.23** (415)	.20** (426)						
4. externalizing CBCL (n)	-.05 (411)	.18** (422)	.40** (434)					
5. internalizing TRF(n)	.13* (401)	.20** (413)	.26** (254)	.06 (254)				
6. externalizing TRF (n)	-.04 (401)	.14** (415)	.03 (247)	.47** (244)	.40** (396)			
7. SLE total (n)	.41** (812)	.12* (835)	.07 (445)	-.06 (442)	-.05 (425)	-.10* (428)		
8. RATS total (n)	.79** (767)	.32** (780)	.23** (404)	-.03 (399)	.10* (390)	-.03 (390)	.46** (799)	
9. Age (n)	.24** (828)	.07* (852)	.04 (455)	-.32** (453)	.06 (433)	-.18** (435)	.22** (894)	.25*** (799)

*p <.05 . **p <.01 . ***p <.001 . Two-tailed.

Construct validity

Construct validity is a measure of the relationship between the instrument and variables that, on theoretical grounds, are expected to correlate with the measured variable. Factorial validity of the CBCL of the internalizing and externalizing scales was found to be moderate, as reported earlier. In this study, guardians did report higher internalizing mean scores for girls than boys (t (453) = 2.01, p <.05) but there were no significant differences between girls and boys regarding externalizing mean scores (t (451) = .79, *ns*). There are contradictory findings in the literature concerning age and emotional distress. Age, divided into 4 categories in the present study (14 years and younger, 15 years, 16 years and 17 years or older), did not seem to play a role with respect to internalizing mean scores reported on the CBCL (F (3,454) = 1.39, *ns*) but did play a role in CBCL externalizing scores (F (3,449) = 17.35, p <.001), 14 years and younger minors having significant higher scores than older minors.

Several studies have shown the number of experienced stressful events (dose-effect relationship) to be a good predictor of psychopathology (e.g., Tiet et al., 1998). Nevertheless, the number of stressful life events that the unaccompanied minor reported did not play any role in the perceptions of emotional or behavioral problems of the unaccompanied minors by guardians (internalizing , F (3,441) = .52, *ns*; externalizing , F (3,438) = 1.09, *ns*).

Criterion validity

Criterion-based validity shows whether the test scores can be used to predict future behavior or to diagnose symptoms. Ideally, a standardized diagnostic interview is used in combination with questionnaires to determine the presence and severity of psychopathology. It was not feasible in the present study to administer a diagnostic interview. Five indicators of psychopathology were utilized as external criteria; (1) self-reported need for mental healthcare (MHC) by the unaccompanied minors, (2) need for professional MHC for the unaccompanied minors; evaluated by the legal guardian, (3) need for professional MHC for the unaccompanied minors; evaluated by the teacher, (4) self-reported utilization of MHC by unaccompanied minors, and (5) referral to MHC services by a legal guardian. Using the kappa statistic, the inter-rater agreement between the indicators was examined to estimate the extent to which each of these variables provided unique information. There was poor inter-rater agreement between all indicators suggesting that each indicator reveals additional and important information (highest kappa = .36, between guardian reported referral and URM reported service use; lowest kappa's = .03, between teacher reported need and URM reported need). Agreement between guardian reported need and guardian referral was good (kappa = .68). However, agreement between URM reported need and URM report service use was poor (kappa = .02).

The criterion "referral" and "utilization of MHC" are important in the evaluation of psychopathology (Anderson, FRANZCP, Williams, McGee, & Silvav**,** 1987; Cuffe et al., 1995; Verhulst & van der Ende, 1997). For this reason, unaccompanied minors themselves, their guardians and their teachers were asked to evaluate the need for professional MHC. An unaccompanied minor was asked if he/she had seen a MHC professional and the legal guardian was asked if he/she had referred the unaccompanied minor to MHC services. The findings presented in Table 5 show that the CBCL can discriminate well between unaccompanied minors whose guardians and teachers report that he or she needs professional help but not between unaccompanied minors with and without a self-reported need for MHC themselves. The CBCL internalizing mean scores for the unaccompanied minors of which the guardians (M =19.30; SD = 10.80) or teachers (M = 14.45; SD = 11.09) reported that they needed mental health services were higher or in the clinical borderline range (*T-score > 60*) that has been established for Dutch adolescents by Verhulst et al. (1996). The externalizing mean score, however, fell below the clinical borderline range (*T-score > 60*) for both guardians (M =8.39; SD=8.37) and teachers (M = 6.58; SD = 7.77).

Table 5.
External criteria influencing CBCL internalizing and externalizing scores

			internalizing							externalizing			
	N	Mean	SD	F(df)	p	d	N	Mean	SD	F(df)	p	d
Unaccompanied minor: Need for MHC												
Need for MHC	234	10.03	8.98	.43 (2,403)	.65		232	4.36	6.20	4.16(2;397)	<.05	.36
No need for MHC	86	9.85	10.14				84	6.79	8.29			
Uncertain of need	84	11.00	10.19				82	5.58	6.81			
	N	Mean	SD	t (df)	p	d	N	Mean	SD	t (df)	p	d
Guardian: Need for MHC												
Need for MHC	90	19.30	10.80	9.64 (108)	<.001	1.46	87	8.39	8.37	4.23(111)	<.001	.61
No need for MHC	351	7.75	6.99				352	4.34	6.23			
Teacher: Need for MHC												
Need for MHC	82	14.45	11.09	4.59 (119)	<.001	.70	81	6.58	7.77	2.35(132)	<.05	.34
No need for MHC	175	8.22	7.67				173	4.25	6.35			
Unaccompanied minor: MHC Utilization												
Utilization of MHC	49	12.57	11.76	1.66 (56)	.10		51	4.77	6.19	.25(368)	.81	
No utilization of MHC	327	9.68	8.62				319	5.03	6.98			
Referral: MHC by guardian												
Referred to MHC	56	20.74	11.99	7.39 (61)	<.001	1.48	58	8.54	8.03	3.59(69)	<.001	.59
Not referred MHC	394	8.57	7.55				390	4.58	6.53			

Discussion

The results of this study indicate that the CBCL can be utilized as screening instrument to assess the global emotional distress and maladaptive behaviors that are reported by guardians based on their observations of unaccompanied refugee adolescents. However, there are several findings related to validity and reliability that should be considered when using the CBCL to assess the mental health of unaccompanied minors.

First, it is important to consider that the guardians or other caregivers of an unaccompanied minor may not always be able to observe the emotional distress and behavioral problems as well as parents simply because they spend much less time interacting with the adolescents. Another reason why emotional distress among URM might go unnoticed is due to the internal nature of their maladaptive problems which are difficult to observe. Additional information from alternative sources (teachers, residential staff workers, adolescents) is essential to make an adequate assessment of the mental health of the adolescent because of the low agreement between informants. Information from each informant is crucial in assessing the degree of impairment in daily functioning and the severity of the symptoms. In this study, it appears that guardians are reliable informants on the psychological well-being of URM.

Second, the factorial validity of the second order two-factor model of CBCL's externalizing and internalizing scales has been verified for this specific, culturally unique, population. The two-factor model was chosen based on previous findings concerning the reliability and validity of the internalizing and externalizing constructs. Although there was no actual difference in fit between the one-factor and two-factor solution, it is advised for theoretical and conceptual reasons that the two-factor model be used when reporting on the findings of the CBCL in future studies. For specific populations, such as URM (internalizing problems) or delinquent youth (externalizing behavior) the two-factor solution gives better insight into which type of psychopathology is predominant. A clear conceptual description of the psychological problems for specific adolescent populations can in turn lead to better tailored interventions and treatments to alleviate their emotional distress.

It appears that for this specific population, the Thought problem subscale is strongly associated with the internalizing problems whereas the Social problem subscale was strongly related to the externalizing problems. Examination of the individual items which make up the Thought subscale (e.g., Can't get mind off problem, repeats acts, strange behavior) reveals that many items could indicate observed behavior resulting from internal traumatic stress reactions of the URM. Due to the great exposure of URM to traumatic experiences (Bean et al., 2004), it would seem logical to expect that guardians might observe psychological distress that could be related to traumatic reactions and which they consider to be strange because they are difficult to interpret. Heubeck (2000) defined social problems that were strongly related with the externalizing scale as being an indication of overt antisocial behavior. All of the original first order individual eight-factor models of the CBCL appear to moderately fit the observed data in this study.

Furthermore, it appeared that when guardians had many years of experience with working with URM and were well-informed about the type of psychological problems that URM can exhibit, the concordance between their reports and the reports of URM regarding externalizing behavior was better than for guardians with less experience and not as well informed. If a guardian was well-informed about the type of internalizing distress URM can experience there was also better agreement between URM and their guardian's reports.

The total score of the CBCL was not a good discriminator for the self-reported need for mental healthcare by the unaccompanied minor themselves. This finding is not new. Bilenberg (1999) found that the CBCL material has never provided good diagnostic validity, however is useful as a guideline for early diagnostic purposes. It is widely known that the agreement between cross-informants is usually low (Achenbach et al., 1987; Ferdinand et al., 2004; Weissman et al., 1987; Yeh & Weisz, 2001). That is why alternative sources of information are not interchangeable for the purpose of making treatment decisions, but can simply be complementary (Macmann & Barnett, 1993). Macmann and Barnett (1993) further indicated that "the composition of core syndromes may vary depending on the items sampled, subjects sampled, and methods of analysis used". This finding has been confirmed among this very specific population (i.e., internalizing problems are more prevalent and frequently

recognized among unaccompanied refugee minors than externalizing problems which seem to be more prevalent among western youth in foster care (Tarren-Sweeney et al., 2004). Great care needs to be taken in the decision making process in determining when professional mental healthcare services need to be consulted for unaccompanied minors. In this study, it was not feasible to administer standardized diagnostic interviews to the URM because of the lingual diversity of the sample which could have established specificity and sensitivity of the CBCL for this specific population. Cross-informant questionnaires such as the CBCL yield less diagnostic information than extensive structured interviews and therefore cannot be used to determine a psychiatric disorder. Considering the multiple risk factors (exposure to multiple traumatic experiences, separation form parents, uncertainty of residential status) that unaccompanied minors are faced with in their lives, it is crucial to their well-being that they receive adequate and appropriate psychosocial care in the residential settings, reception or detention centers were they reside in host countries. One technique to accomplish this objective is by using psychological instruments, such as the CBCL to screen URM for their emotional distress.

In future studies, multiple informants (including the minor themselves) should be included in the assessment of the mental health of unaccompanied minors. It is apparent from this study that each informant contributes unique and important information about the mental health of the URM. Comparing the concordance between different informants that have different relationships (i.e., staff workers and guardians) with the minors themselves could give better insight into the reliability of the reports of significant adults in the lives of minors. Furthermore, measuring the effect of "familiarity" of the minor by the guardian more systematically than in this study could reveal if the "quality" or "quantity" of time spent with the minor is more important in accurately assessing behavioral and emotional problems. Finally, studying the effect of training guardians to accurately perceive emotional and behavioral problems of unaccompanied minors should be investigated to improve the mental healthcare services and the referral process. In this study, it appears that if guardians are knowledgeable about the types of psycho-social problems URM experience, agreement between URM and guardian reports is improved.

The results of this study, in which the mental health of unaccompanied refugee minors was reported on by their legal guardians, demonstrate consistency with previous studies which have evaluated the factorial structure of the CBCL. The two-factor model of internalizing and externalizing is supported in spite of the fact that other methods were used (parcels in the CFA) and that the adolescents were a culturally heterogeneous population coming from 48 different countries. Furthermore the internal consistency was found to be good to moderate and the criterion validity was found to be moderate when significant adults in the lives of the URM were used as informants. The present findings on the psychometric properties of the CBCL suggest that the CBCL can be used by guardians to reliably assess the global maladaptive emotional and behavior problems of unaccompanied refugee minors.

Figure 1. Standardized parameter estimates for the CBCL hierarchical factor model