

Chapter 6

Validation of the Teacher's Report Form for Teachers of Unaccompanied Refugee Minors

Abstract

The validity and reliability of the Teacher's Report Form (TRF) for Dutch teachers of unaccompanied refugee minors was evaluated in this study. The teachers of the unaccompanied minors that participated in the study all received a TRF to report on the mental health of the unaccompanied minor. The teachers filled in and returned 486 TRF's. Hierarchical confirmative factor analysis and individual confirmatory factor analyses support the a priori structure of the TRF. However, the Thought problems subscale could not be verified in this study suggesting that some of the problem behavior reported by teachers of unaccompanied minors differs from that of parent reports or that the item constellation of the TRF is different for teachers of unaccompanied minors. The total, internalizing and externalizing scales show good internal consistency. The construct and criterion validity of the TRF were found to be good. The results suggest that the TRF is a reliable and valid instrument to assess emotional and behavior problems of unaccompanied refugee minors.

Introduction

Throughout the world, refugee children and adolescents attend schools in their host countries. The number of foreign-born children and adolescents attending schools in European metropolitan areas is equal to or even greater than the number of native born

students (Eurydice, 2004). The school provides structure and an informal environment in which low threshold mental services can be provided to refugee children and adolescents (and their families). During the last few years, there has been an increase in the literature of school-Report Form (TRF) (Achenbach, 1991c)(e.g., Allwood, Bell-Dolan, & Husain, 2002). The TRF is a checklist that is completed by teachers reporting on the maladaptive emotional and behavioral problems of the children they have in their class. Although, the TRF has been used in a few studies with refugee adolescents, the psychometric properties of the TRF have not yet been thoroughly examined for this specific population.

There have been studies carried out which have used teachers as informants in assessing distinct specific populations such as foster children (i.e., Shore, Sim, Le Prohn, & Keller, 2001), severely maltreated children (i.e., Culp, Howell, McDonald-Culp, & Blankemyer, 2001) and delinquent adolescents (i.e., Tranh & Hill, 2000). Furthermore, the TRF has been used in several different countries (predominantly western) with children and adolescents from diverse cultures and ethnicities (Atzaba, Naama, Pike, & Baret; 2004; French, Jansen, Riansari, & Kusdiratri, 2003; Frigerio et al., 2004; Hartman et al., 1999; Lambert, Lyubansky, & Achenbach, 1998; Stevens et al., 2003; Satake, Yoshida, Yamashita, Kinukawa & Takagishi, 2003; Yang, Soong, Chiang, & Chen, 2000) to measure maladaptive behaviors and emotional problems. Again, the psychometric properties, especially, the validity of the TRF for many of these populations are frequently not reported on.

Only a few studies have examined the factorial validity of the TRF. De Groot, Koot, & Verhulst (1996) found mediocre evidence in their confirmatory factor analyses for the applicability of the eight-factor model of the TRF. Three studies that have evaluated the construct validity of the TRF (De Groot et al., 1996; Spijker, Kramer, Constatine, & Bryant, 1992; Hartman et al., 1999) all found greater violation of distribution assumptions in the TRF model than for the associated CBCL, which led to a poorer fit of the TRF model. The Hartman et al., (1999) study further found minimal support for the two-factor second order or eight-factor first order models. Only significant support was found for the one-factor second order model. The second order two-factor model terminology of “internalizing” and “externalizing” behavior is used to describe the results in most studies which utilize the TRF. Macmann and Barnett (1993) in their critical examination of the interpretations of the instruments associated with the Achenbach cross-informant structure also theoretically favored the second order two-factor model. Although, findings concerning the factorial structure and validity of the TRF are weak, the internalizing and externalizing scales have been found to be reliable and valid measures of child psychopathology over time (e.g., Verhulst & van der Ende, 1992). The second order two-factor model was chosen as the basic model for this study because of the vast empirical evidence that has been collected over the past years using the cross-informant structure associated with the instruments of Achenbach.

Due to a dramatic increase in the number (15,000) of unaccompanied minors in the Netherlands in 2000 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. A secondary goal of the project was to validate and standardize screening instruments that measure emotional distress and behavioral problems for this specific population group.

In the Netherlands, there is a national network of schools that offer a two-year educational program to URM and other parental accompanied immigrant and refugee adolescents. In this program, acquisition of the Dutch language and customs of the Netherlands are prominent tasks in the first year. As soon as a basic vocabulary in the Dutch language has been achieved, adolescents can follow normal lessons and gradually integrate into the regular Dutch educational system. The classes are usually small, about 15 adolescents per class, and the adolescents receive a great deal of individual attention.

Because of the uniqueness of this study, Dutch teachers reporting on the mental health of unaccompanied adolescents from many different countries, it was prudent to examine the psychometric properties of the TRF for this specific population. The value which can be attached to results of a study is, of course, predetermined 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

TRF, for use with a specific research population, the teachers of unaccompanied refugee minors.

Method

Sample

Demographic information on the unaccompanied minors in the Netherlands was supplied by the Nidos Foundation. Approximately 4000 unaccompanied minors were randomly selected from 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; 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. A total of 920 unaccompanied minors were present for participation (20% of the URM were not tested at school but in their residential setting or at a Nidos regional office). 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. Ethical approval for this study was given by the Medical Ethics Committee of the Leiden University Medical Center, Leiden University.

Measures

TRF

The Dutch version of the TRF; 1991 Profile-(Achenbach, 1991c) Dutch translation (Verhulst, van der Ende, & Koot, 1997) was used to standardize the assessment of the behavior and emotional problems of unaccompanied minors through the observations of teachers. The TRF has a three point rating scale; 0 = *not true*, 1 = *somewhat true*, and 2 = *very true*. The TRF has 118 problem behavior items, yet 101 items fall under the internalizing or externalizing scales or Social, Thought and Attention subscales of the TRF.

The 118 problem items of the TRF were explanatory factor analyzed to empirically identify the constructs of psychopathology that occur in adolescents (Achenbach, 1991c). The TRF yields three possible types of scores; a total score, a score for second-order internalizing and externalizing scales and scores for the eight first-order scales-(withdrawn, somatic, anxious/depressed, social problems, thought problems, attention problems, delinquent and aggressive). The validity and reliability for the normative and clinical populations of the Dutch version of the TRF is thoroughly described in Verhulst et al. (1997).

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 psychosocial 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 teachers were also asked to fill in 10 questions about themselves and their experience with working with adolescents.

Mental Health Questionnaire for guardians

The need for mental healthcare among the unaccompanied minor as perceived by the guardian, the referral process to mental health services and the satisfaction of 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.

Stressful Life Events

The *Stressful Life Events* (SLE) (Bean, Derluyn, Eurelings-Bontekoe, & Spinhoven, 2004b) checklist (available in 19 different languages) 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?” or “Has someone ever hit, kicked, shot at or some other way tried to physically hurt you?”). The overall average total score of 6.5 of the SLE for URM has been replicated in 5 independent studies and was significantly higher than the total scores for parental accompanied immigrant/refugee adolescents, Dutch and Belgium adolescents (Bean et al., 2004).

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 individually (in Dutch) conducted with the URM after they had filled in the other three questionnaires to ensure that the questionnaire would be filled in properly. The questionnaire was available in 19 different languages so the adolescents could read along in their own language. 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 Procedures

There is one foundation that has the legal guardianship of all of the unaccompanied minors that reside in the Netherlands, The Nidos Foundation. This foundation has offices throughout the entire country and has almost 20 years of experience in working with unaccompanied minors. Two information packages (one for guardian and one for teacher) were sent to the supervisors of each regional office for each guardian that was responsible for one of the 920 unaccompanied minors that took part in the study. The guardians received their own questionnaires and information package and those for the teachers of the URM from their supervisors. The guardian was responsible to send the information package to the teacher. Enclosed in the information package for the teacher, was a letter describing the project, questionnaires and a stamped and addressed enveloped in order to enable the teacher to return the completed questionnaires directly. The teachers received a letter with the questionnaires informing them about the study and providing instructions as to how the questionnaires should be filled in. For the first assessment period, 496 questionnaires were returned. From the questionnaires that were returned for the assessment, 486 TRF (problem items) questionnaires were filled in. The 10 other returned questionnaires only pertained to the questions concerning mental healthcare of the URM.

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. Bandalos (2002) stated that the parcel method is often used with highly skewed, categorized data to obtain distributions that are more normal and continuous in which CFA's can be generated from. The latent factors represent the subscales in the parcel approach. 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. As far as is known by the authors of this article, no study used the parcel approach in evaluating the hierarchical structure of the TRF.

The present study used the same methods (random selection of items for parcels, summing of items) as Greenbaum and Dedrick to create the parcels for the TRF which will be used as indicators for the hierarchical confirmatory analysis. Using two parcels per factor has been found to result in a less reliable CFA model fit than three or more parcels per factor (Nasser & Wisenbaker, 2003). Dividing the 101 TRF items into three parcels, per subscale, was possible instead of two used in Greenbaum and Dedrick because the TRF two-factor model contains more items (101) than the CBCL (85). It was necessary to add one item to the Thought subscale, item 80 (Stares blankly), to allow three indicators for each parcels, thus the three parcel minimum found to be necessary by Nasser & Wisenbaker (2003) to generate reliable parameter estimates was fulfilled. Item 80 has been shown to have a complex relation to the cross-informant structure of the TRF and CBCL (Hartman et al., 1999) and also strengthened the reliability of the Thought scale from .53 to .58 (see Table 2). Bagozzi and Heatherton (1994) have listed the many advantages of using parcels. Nonetheless, when using parcels as indicators in the CFA model, information is lost at item-level. To supplement the findings of the hierarchical analysis, individual first order and second order factor models were 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 of the TRF will be examined individually to evaluate the behavior of items in each model. Individual second order factor models were also examined to see if they are better described as single models or as a grouping of smaller factors. The second order factor structure of the two a priori factors, internalizing and externalizing, 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 TRF, are skewed (West, Finch, & Curran (1995) recommend concern if skewness > 2 and kurtosis > 7) and in great violation of distribution assumptions (see Table 2). As reported earlier, the violation is greater than found in studies which have evaluated the validity of both the CBCL and TRF. Therefore, the indicators for all of the models need to be considered as categorical. 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 nonnormality of the indicators. It was necessary to calculate tetrachoric correlation matrices (instead of polychoric due to the large skewness and kurtosis) and the asymptotic covariance and variances matrices using PRELIS 2 (Jöreskog & Sörbom, 1990; Jöreskog & Sörbom, 1996b) on which all of the confirmatory factor analysis could be generated 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 in their study evaluating the eight-factor first order model of the CBCL, the WLS, an asymptotically distribution-free (ADF) estimator. 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 101 items of the TRF, a sample of more the 15000 would be necessary. Using the suggested sample size formula for the 24 parcels (created for this study), a stable

weight matrix could also not be obtained using the WLS method (minimum sample size needed of 900) since the effective sample size of the URM study is 461 (fully completed TRFs). Therefore, the individual factor models and the hierarchical CFA will be evaluated using the DWLS (Muthén, du Toit, & Spisic, 1997) method (estimation capabilities falling inbetween the ULS and WLS methods; Jöreskog & Sörbom, 1996a). Although, DWLS 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 (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 sample (Table 1). Internal consistency of the total scale, subscales and parcels of the TRF used in this study were calculated with Cronbach's α . Pearson's product-moment correlations (two-tailed) were used to study the association between total and subscale scores of the TRF. Differences between groups were determined by using ANOVA's and effect sizes, 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.

Results

Demographic information about unaccompanied minors and teachers

Table 1 represents the demographic background information for the 486 unaccompanied minors (of the 920 of whom there was a TRF completed) and their teachers. The teachers could reply anonymously if they chose to do so. Therefore, it was not possible to establish the exact number of teachers that took part in the study. We were able to estimate that approximately 400 teachers took part in the study. The unaccompanied minors in this sample consisted mostly of boys (71.3%). The mean age was approximately 15 years and the most frequent countries of origin were Angola, Sierre Leone, Guinea and China. Most of the unaccompanied minors came from Africa (80.8%). 44.9% of the unaccompanied minors sample had received educational training for more than 5 years. The teacher population consisted of mostly females (68.1%) that had a mean age of 46 years. The greater majority of teachers were born in the Netherlands. Most of the teachers had received the equivalent of a Bachelor's degree in Teaching.

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

	Unaccompanied minor	Teachers
N	(920) 486*	(496) 486*
Gender		
Male	73.5%	31.9%
Female	26.5%	68.1%
Age in years		
Mean	15.44	45.97
SD	1.58	8.72
Range	9-18	23-64
Country of Origin		
Netherlands	0.0%	90.3%
Angola	47.3%	
Iran/Afghanistan/Iraq	3.9%	
Eritrea/Ethiopia	1.6%	
Somalia	1.9%	
Sierra Leone	8.2%	
Guinea	7.8%	
Other African Countries	14.0%	
China/Tibet	8.2%	
Other Countries	7.0%	7.8%
Questionnaire completed by		Mentor 86.1%
Time spent with URM per week		7 hours or longer 72.0%
How long did the teacher know the URM		6 months or longer 59.8%
Most frequent reported level of education	1-5 years of education 44.9%	Bachelor's Degree 98.2%

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

Individual and Hierarchical Confirmatory Factor analyses

The fit of all of the individual eight lower order factor models, except for the Thought subscale, of the TRF are mediocre (data not shown). The Thought subscale fits the observed data poorly (even after attempting to change the model based on modification indices). The greater majority of the items were estimated above .40, ranging in mean estimates from .58 to .90 (except for Thought). These findings do not deviate from previous studies regarding the lower-order factor model of the TRF that have been reported on earlier in this article.

The SB chi-square goodness-of-fit test indicated the second order model did not fit the observed data very well. However, when the alternative measures of fit were examined, the results indicated that the fit of the model was acceptable (SB- $\chi^2(240) = 958$; AGFI = .97; PNFI = .84; 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 for the CBCL. Adjusting the model based on modification indices did not lead to better refining of the model. In the model, the Thought (.86) and Social factors (.73) loaded quite high on the secondary hierarchical internalizing factor.

The correlation between the second order internalizing and externalizing factors was .52. 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 examine if this hypothesis was correct, the two factor second order model was recalculated without the Social, Thought, and Attention factors. The recalculated model yielded almost the same correlation of .50 between the internalizing and externalizing factors.

Because of this rather high correlation and the previous findings that have been documented in CBCL and TRF 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 study ($SB-\chi^2(244) = 1205$; AGFI = .95; PNFI = .85; CFI = .97; RMSEA = .09). There was a small difference found in model fit between the one-factor second order model and the two-factor second order factor for the TRF. However, the two-factor second order model of the TRF fit the data slightly better than the one-factor second order model .

Table 2.

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

TRF scales	TRF items	α	M	SD	Skewness	Kurtosis
Withdrawn	All items in subscale	.82	3.63	3.46	.99	.52
Somatic	All items in subscale	.74	1.58	2.29	2.23	6.87
Anxious/Depressed	All items in subscale	.84	4.59	4.62	1.45	2.27
Social	All items in subscale	.78	1.73	2.58	2.34	7.91
Thought	All items in subscale (+ item 80)	.55 (.58)	.95	1.53	2.43	7.33
Attention	All items in subscale	.92	6.16	6.82	1.44	1.65
Delinquent	All items in subscale	.72	1.86	2.36	1.74	3.44
Aggressive	All items in subscale	.95	4.18	7.15	2.67	5.07
Internalizing	All items in subscale	.89	8.97	8.09	1.25	1.75
Externalizing	All items in subscale	.94	6.03	8.86	2.14	4.55
Total	All 101 items	.95				
Parcel						
Withdrawn parcel 1	42, 65, 69	.65	1.31	1.40	.90	-.02
Withdrawn parcel 2	80, 88, 102	.58	.91	1.23	1.36	1.35
Withdrawn parcel 3	75, 103, 111	.63	1.41	1.47	1.07	.64
Somatic parcel 1	56a, 56b, 56d	.46	.76	1.15	1.70	2.88
Somatic parcel 2	51, 54, 56f	.52	.69	1.04	1.79	3.28
Somatic parcel 3	56c, 56e, 56g	.50	.16	.58	5.10	33.29
Anx/dep parcel 1	31, 33, 35, 71, 81	.65	1.33	1.62	1.56	2.71
Anx/dep parcel 2	12, 14, 34, 45, 47, 50	.66	1.31	1.83	1.83	3.50
Anx/dep parcel 3	52, 89, 103, 112, 106, 108	.62	2.03	1.94	1.01	.85
Social parcel 1	1, 11, 12, 33, 64	.53	.76	1.20	1.93	3.95
Social parcel 2	14, 25, 36, 38	.50	.41	.84	2.57	7.91
Social parcel 3	34, 35, 48, 62	.60	.62	1.10	2.32	6.96
Thought parcel 1	2, 109, 85	.51	.49	.98	2.78	9.65
Thought parcel 2	40, 18, 70	.41	.06	.35	7.18	59.26
Thought parcel 3	80, 84, 66	.42	.52	.91	2.13	4.78
Attention parcel 1	1, 2, 4, 10, 13, 15, 41	.82	1.50	2.42	2.10	4.32
Attention parcel 2	22, 45, 49, 60, 61, 62	.77	2.21	2.40	1.12	.51
Attention parcel 3	8, 17, 80, 72, 78, 92, 100	.81	2.59	2.80	1.28	1.26
Delinquent parcel 1	26, 39, 90	.63	.40	.92	3.07	11.28
Delinquent parcel 2	43, 82, 63	.57	.41	.87	2.95	10.70
Delinquent parcel 3	98, 101, 105	.62	1.16	1.41	.93	-.36
Aggressive parcel 1	3, 6, 7, 16, 19, 20, 21, 23	.86	1.58	2.68	2.12	4.45
Aggressive parcel 2	24, 27, 37, 53, 57, 67, 68, 74	.87	1.31	2.53	2.40	5.83
Aggressive parcel 3	76, 77, 86, 87, 93, 94, 95, 97, 101	.89	1.84	3.18	2.22	4.70

Reliability

The reliability of the 101 TRF items is .95. The alpha for the internalizing scale is .89 and for externalizing scale the alpha is .94. The Cronbach's alpha's for each parcel and subscale have been reported on in Table 2. The alpha's for the scales of the present study are as high, or higher than those calculated for the Dutch normative population (Verhulst et al., 1997).

Intermeasure correlations

The correlations for the first assessment between all the scales of the TRF are presented in Table 3. These correlations are comparable to those found in the Dutch normative population (Verhulst et al., 1997). It can be observed from this table that the externalizing scale and internalizing scale show a correlation of .40. Accordingly, this means that these two scales are not independent of each other, implying that the two constructs partly overlap.

Table 3.

Intercorrelations of the subscales of the TRF

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Withdrawn										
2. Somatic	.39									
3. Anxious/Depressed	.64	.47								
4. Social	.53	.39	.75							
5. Thought	.36	.37	.50	.54						
6. Attention	.56	.45	.55	.66	.49					
7. Delinquent	.36	.30	.40	.48	.38	.65				
8. Aggressive	.23	.21	.43	.62	.48	.65	.59			
9. Internalizing	.84	.67	.91	.71	.51	.63	.43	.35		
10. Externalizing	.27	.27	.45	.62	.47	.70	.74	.98	.40	
11. Total score	.62	.52	.69	.78	.62	.83	.65	.71	.76	.77

Note. All correlations are significant to the .01 level. Two-tailed.

Validity

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 structure and validity of the TRF of the internalizing and externalizing scales have been found to be good to mediocre, with the exception of the individual Thought subscale, reported on earlier.

Usually, teachers report a higher number of emotional problems for girls and higher externalizing behavioral for boys. In this study however, teachers did not report higher internalizing mean scores for girls than boys ($F(1,432) = 3.12, ns$) but there were significant differences between externalizing mean scores ($F(1,433) = 6.29, p < .05; d = .09$) for girls and boys, boys having higher mean scores. There are contradictory findings in the literature concerning age and emotional distress. Age did not seem to play a role with respect to internalizing mean scores reported on the TRF ($r(n = 435) = .06; ns$) but did play a role in TRF externalizing scores ($r(n = 438) = -.18; p < .001$), younger minors having significant higher scores than older minors.

Several studies have shown the number of experienced stressful events (dose-response relationship) to be a good predictor of psychopathology (e.g. Papageorgiou et al., 2000; Tiet et al., 1998). Nevertheless, the number of stressful life events that the unaccompanied minor reported were not significantly associated with the perceptions of emotional or behavioral problems of the unaccompanied minors by teachers (internalizing, $r(n = 425) = -.05, ns$; externalizing, $r(n = 428) = -.10, ns$).

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

Table 4 .
External criteria influencing TRF 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	218	9.85	8.94	.21 (2,385)	.81		218	5.20	7.55	2.65(2,388)	.07	
Need for MHC	86	9.34	9.02				87	7.62	11.31			
No need for MHC	84	9.21	7.65				86	5.12	8.43			
Uncertain of need												
Guardian: Need for MHC	58	14.75	12.19	15.32 (1,296)	<.001	.57	56	8.06	11.31	1.43(1,293)	.23	
Need for MHC	240	9.52	8.24				239	6.35	9.24			
No need for MHC												
Teacher: Need for MHC	106	18.50	10.02	194.02 (1,379)	<.001	1.60	101	11.79	12.22	55.09(1,385)	<.001	.86
Need for MHC	275	6.75	6.07				286	4.32	7.06			
No need for MHC												
Unaccompanied minor: MHC Utilization	41	11.20	10.29	1.41 (1,354)	.24		41	6.58	9.27	.32(1,352)	.57	
Utilization of MHC	316	9.47	8.58				314	5.75	8.87			
No utilization of MHC												
Referral: MHC by guardian	35	17.26	13.21	22.13 (1,301)	<.001	.85	34	10.40	12.92	6.14(1,299)	<.05	.45
Referred to MHC	268	9.65	8.30				267	6.10	9.05			
Not referred MHC												

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.

The criterion “referral” and “utilization of MHC” are important in the evaluation of psychopathology (Cuffe et al., 1995; Verhulst & Van der Ende, 1997). For this reason, unaccompanied minors themselves, their legal 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 were asked if he/she had referred the unaccompanied minor to MHC services. Table 4 shows that the TRF can discriminate well between unaccompanied minors whose teacher’s report that he or she needs professional help but not between unaccompanied minors with and without a self-reported need for MHC themselves. Furthermore, using the *T*-scores which have been established for Dutch adolescents by Verhulst et al. (1997) the TRF Internalizing mean scores for the URM of which the guardians (Boys $M = 13.96$; $SD = 12.04$; T -score = 66) (Girls $M = 16.64$; $SD = 12.70$; T -score = 65) or teachers (Boys $M = 17.65$; $SD = 9.83$; T -score = 68) (Girls $M = 20.87$; $SD = 10.33$; T -score = 68) reported that they needed mental health services was higher than the clinical range (T -score > 63). The TRF Externalizing mean scores, however, fell just below or just above the clinical borderline range (T -score > 60) for both guardians (Boys $M = 9.46$; $SD = 12.12$; T -score = 59) (Girls $M = 5.09$; $SD = 8.97$; T -score = 58) and teachers (Boys $M = 12.98$; $SD = 12.33$; T -score = 61) (Girls $M = 8.68$; $SD = 11.55$; T -score = 61).

Discussion

The results of this study indicate that the TRF can be utilized as a screening instrument to assess the global emotional and problem behaviors that are reported by teachers based on their observations of unaccompanied refugee adolescents. Since the school provides structure and an environment for informal mental health services for refugee adolescents, this measure can help mental health providers to reliably and validly assess when a refugee adolescent is in need of professional care. However, there are several findings that should be considered when using the TRF to assess the mental health of unaccompanied minors.

It is important to consider that the teachers of unaccompanied minors may not be able to observe or perceive all of the emotional distress and behavioral problems that they have because of the internal nature of the psychological problems. Additional information from alternative sources (guardians, residential staff workers, adolescents) should be collected regarding the mental health of the adolescent. This information is crucial in assessing the degree of impairment in daily functioning and the severity of the symptoms.

The factorial hierarchical structure of the two-factor second order model of the TRF's externalizing and internalizing scales were examined in this study and were supported. Although there was only a very small 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 TRF in future studies. For specific populations, such as URM (internalizing problems) or delinquent youth (externalizing behavior) the two-factor terminology gives better insight into which type of psychopathology is predominant. A clear conceptual description of the psychological problems can in turn lead to more tailored interventions and treatments to alleviate the emotional distress of adolescents.

The individual first order and second order factor models also verify the factorial validity of each subscale, except for the Thought subscale. However the fit of the individual factors was only acceptable to mediocre. This result does not deviate from other studies that have investigated the goodness-of-fit of the TRF model which have found the model fit of the TRF to be worse than the model fit of the CBCL. Thought Problems and Social Problems loaded quite high on the internalizing factor. Examining the individual items which make up the Thought subscale (e.g., can't get mind off problem, repeats acts, strange behavior) reveal 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 teachers observe psychological distress that might be related to traumatic reactions which they consider to be strange because they cannot interpret it. The Social Problem subscale of the TRF (unlike the CBCL) contains 6 items from the Anxious/Depressed subscale, explaining most likely, the strong relationship found between internalizing and Social problems in this specific population.

Finally, the total score of the TRF was not a good discriminator for the self-reported need for mental healthcare among the unaccompanied minor themselves. This finding is not new. Bilenberg (1999) found that all of the CBCL related 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 informants is usually low (Achenbach, McCounaughy, & Howell, 1987; Ferdinand, Van der Ende & Verhulst, 2004). That is why alternative sources of information are not interchangeable for the purpose of making treatment decisions, but can simply be complimentary (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 also been confirmed in this study. 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. Cross-informant questionnaires such as the TRF 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 from parents, and 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 where they reside in host countries.

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

Figure 1. Standardized parameter estimates for the TRF hierarchical indicators



