

Chapter 4

Validation of the Multiple Language Versions of the Reactions of Adolescents to Traumatic Stress Questionnaire

Abstract

The objective of this study was to provide the preliminary psychometric properties of the Reaction of Adolescents to Traumatic Stress questionnaire (RATS) for refugee adolescents. Four independent heterogeneous adolescent population samples ($N = 3535$) of unaccompanied refugee minors, immigrants and native Dutch and Belgian adolescents were assessed at school. The confirmatory factor analyses, per language version, support the three-factor structure of intrusion, avoidance/numbing and hyperarousal. The total and subscales of the RATS show good internal consistency, and good (content, construct and criterion) validity. The RATS, in this study, was found to be a reliable and valid instrument for assessing posttraumatic stress reactions of cultural diverse adolescents.

Introduction

After experiencing a terrifying event, it is “normal” for adolescents to exhibit stress reactions or problem behavior. Experiencing one or more stressful life event(s) such as a catastrophic disaster (Sack et al., 1993; Pynoos et al., 1993, Green, 1991), physical trauma

(Terr, 1983; Briggs & Joyce, 1997; Roth, Newman, Pelcovitz, van der Kolk, & Mandel, 1997) or a combination of daily stressors, relational and financial problems (Kendler, Karkowski, & Prescott, 1999) can lead to severe psychological distress. After experiencing traumatic events, adolescents can become aware that they are vulnerable to death and injury. Caregivers are not always able or choose not to shield them during threatening situations. In addition, refugee and immigrant adolescents must also adapt their former lifestyle and culture to their new surroundings. The adaptation or acculturation process can be emotionally draining and disturbing which might exacerbate traumatic stress reactions (Berry & Sam, 1997).

The symptom clusters of the DSM-IV's Posttraumatic Stress Disorder (PTSD) (American Psychiatric Association, 1994) intrusion, numbing/avoidance and hyperarousal, can affect the biochemical, physiological, psychological and social systems of an adolescent which may compromise resiliency and increase system breakdowns later in life (Charney, 2004). The effects of trauma can be damaging to the dynamic developmental process of adolescence where high performance at school/work and healthy social relations are key. If left unabated, the effects of traumatic stress can severely disrupt or delay the fulfillment of important developmental tasks.

Two epidemiological studies among general populations which assessed the lifetime prevalence rate for PTSD in adolescents documented 3 to 6.3% in the American population (Cuffe et al., 1998; Giacona et al., 1995). This means that the prevalence rate of PTSD among adolescents from western countries at any given time should be very low. However, much higher prevalence rates between 35 to 75% have been established with refugee adolescent populations (see Lustig et al., 2004, for a review). All studies among adolescents in general that have investigated gender as a risk factor have found that females are more likely than males to develop PTSD (Deykin, 1999; Green et al., 1991). There are contradictory findings in the literature concerning age and PTSD reactions. A high comorbidity has been documented between PTSD and other disorders such as depression (Sack et al., 1993; Sack & Clarke, 1996), anxiety (Warshaw et al., 1993), substance abuse (Deykin & Buka, 1997) and ADHD (Famularo, Fenton, Augustyn, & Zuckerman, 1996; Wozniak et al., 1999).

The "pathway" to professional mental healthcare for refugee adolescents, found to be a high risk population (Lustig et al., 2004), has more barriers than for native adolescents in host countries (e.g., Howard & Hodes, 2000). There is sufficient evidence in the literature suggesting that young people in general that are in need of psychological support or treatment do not receive it (US Department of Health and Human Services, 1999; Cuffe, Waller, Cuccaro, Pumareiga, & Garrison, 1995) or only when the symptoms have progressed and are perceived by significant adults in their lives (e.g., Wu et al., 1999). Significant adults (parents, teachers, guardians) in the lives of adolescents are not always adept in detecting early emotional distress signs of adolescents (Yeh & Weisz, 2001) which has been found to have an adverse effect on the developmental process (Ferdinand, Van der Ende, & Verhulst, 2004). The psychological suffering of URM can go completely unnoticed due to the absence of parents or permanent caregivers, language difficulties and living in minimally adult supervised residential settings.

Mental healthcare professionals in host countries are often hindered in acquiring accurate information concerning the mental health status of refugee adolescents due, in part, to language difficulties, no medical/psychological background information and to a lack of translated reliable and valid diagnostic instruments. A brief translated psychological instrument that measures, reliably and validly, the reactions of traumatic stress among culturally heterogeneous refugee adolescents could be of great assistance to mental healthcare professional in the process of screening, diagnosing and monitoring the mental health status of this specific high- risk population.

The first step in providing such an instrument to professionals is making an inventory of the accessible and obtainable instruments in the field. This was done in an earlier study (Bean, 2000). The most well known and frequently used self-report measure that has been developed for adolescents is the UCLA PTSD Reaction Index for DSM-IV (Pynoos, Rodriguez, Steinberg, Stuber, & Frederick, 1998). However, as far as known by the authors, this checklist has not been validated for culturally diverse adolescent populations following the five dimensions of equivalence for cross-cultural validation of an instrument proposed by Flahtery et al. (1988). The five dimensions are (a) content equivalence which determines whether each item is equally relevant for the culture(s), (b) semantic equivalence is an item-by-item

analysis attempting to convey the original meaning of each item in the adapted version(s), (c) technical equivalence refers to whether the data collection method (e.g., self-report survey, in-person interview) yield comparable results in each culture, (d) criterion equivalence is when the interpretation of the measurement remains the same when norms are compared in each culture, and (e) conceptual equivalence refers to whether the same theoretical construct is being measured in each culture. There are also some conceptual problems that would not make the instrument easily understood by non-western adolescents who (a) are not familiar with western questionnaires and (b) may have very short attention spans, possibly due to hyperarousal associated problems and limited access to formal educational where the attention span has been trained and lengthened. The PTSD Reaction Index (Pynoos et al., 1998) does not order its items following the DSM-IV criteria which would provide more clarity of the concepts that the items are trying to measure for non-western adolescents. Furthermore, the layout of the instrument is very “busy” (lots of numbers and letters, difficult terms- “rating sheet”, two pages), is not bilingual and has no visual aids in the rating scale to help non-western adolescents understand how they need to fill in this kind of form. Although, there are other English PTSD measures available for children and adolescents, these instruments have similar limitations (items which fall outside the criteria of the DSM-IV or have a specific emphasis such as measuring the reactions of children/adolescents to sexual trauma). No validated (and translated) PTSD questionnaire which was conceptually and visually appropriate for culturally diverse adolescents was available. Due to the above outlined reasons, the Reactions of Adolescents to Traumatic Stress (RATS) questionnaire was developed (Bean, 2000; Bean, Eurelings-Bontekoe, Derluyn, & Spinhoven, 2004a). The objective of this article is to provide preliminary information concerning the psychometric properties of the RATS for (refugee) adolescents from many different cultural backgrounds.

Method

Context of the validation study

In the years preceding 2001, there was a dramatic increase in the number of URM living in the Netherlands, peaking at 15,000 in 2001. Many practical problems in referring unaccompanied minors to mental healthcare services were reported by the Nidos Foundation (legal guardian of all of the URM living in the Netherlands). Because there was (and still is) a lack of research studies on the mental health and service utilization of URM, 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 their guardians, teachers and professional mental healthcare providers in 2001. The goal of the project was to determine the severity level of psychological distress of unaccompanied minors, their need for mental healthcare, and the availability of mental healthcare services for this group. 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 fulfilling this second goal, it was also possible to utilize the three screening instruments that were developed/modified for the URM sample with additional samples; (a) a Dutch adolescent population that had parental caregivers, (b) accompanied and unaccompanied migrants and refugee adolescents and (c) Flemish adolescents. The mental health of the last two samples was examined in an independent research project by the Department of Orthopedagogics, Ghent University, Belgium. The aim of the Ghent project was to examine whether being unaccompanied is a risk factor for refugee children and adolescents to develop emotional and behavioral problems.

Samples

Dutch URM sample (n = 920)

A national, longitudinal study was carried out with URM living in the Netherlands. Approximately 4000 URM were randomly selected from the 12,000 (total population) URM in the Central Registrar of Nidos. Information about the study and permission waivers (available in translated versions) were both sent to the guardians to discuss with the URM. Both the minor and his/her guardian gave written permission for the URM to participate. Roughly 2300 URM 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 to a different residential setting, and 7% turned out to be untraceable. A total of 920 URM were present for participation. There was a large number of the URM from the original random sample that did not participate. However, there were no significant differences found in gender, age, and country of origin between the URM that did participate and the URM that did not. The final sample was representative in all of the main characteristics (age, gender, country of origin and type of residential setting) of the total URM population aged 12 to 18 year old in 2002 in the Netherlands. The URM came from 48 countries, predominantly Angola (43%), Sierra Leone (10%), and China (8%). Two-thirds of the sample had lived in the Netherlands for a period of 18 months or less. 45% of the URM sample has received 5 years or less of formal education in their country of origin. A follow-up (63% of the original sample participated) was conducted one year after the first assessment. In addition to the questionnaires mentioned above, an interview regarding mental healthcare was individually administered. Three research assistants administered the questionnaires during one hour.

Dutch normative sample (n = 1059)

Pupils from ten secondary and three tertiary trade schools throughout the Netherlands (schools had also taken part in the URM study) participated and functioned as a control group for the URM sample (comparison was based on age). Two weeks prior to administration of the instruments, informed consent letters were sent to the parents and adolescents asking for the voluntary and anonymous participation (27 students abstained from participation). The assessment of the Dutch sample took approximately 15 minutes.

Belgian immigrant /refugee adolescents sample (n = 939)

A large scale study was carried out with non-Dutch speaking immigrant adolescents in Flanders (Belgium) during November 2002 to May 2003. The adolescents came from 111 countries, predominantly Morocco (14%), Ghana (11%), and Turkey (9%). All schools received standard informed consent letters (translated versions were available) asking parents and students for voluntary and anonymous participation. In 2002, there were 42 secondary schools in Flanders which provided education for recently immigrated adolescents. Thirty-four schools were randomly chosen to participate in the study of which none declined. 65% of the number of recently (less than 1 year) immigrated adolescents (immigrants and refugees) in Flanders between 13-18 years of age, participated in the study. Only 1 student abstained from participation that was present on the day of assessment. There was a continuous stream of new students during the year, which render it very difficult to test the entire population. No attempt was made to test adolescents that were not present on assessment day. The assessment took place (1 hour) under supervision of two research assistants.

Belgian normative sample (n = 617)

A control group of Belgian adolescents participated between January, 2003 and May, 2003 for the Belgium immigrant/refugee study. From the six Flemish provinces, 17 secondary schools were randomly selected to participate in the study. All schools received standard informed consent letters asking parents and students for voluntary and anonymous participation. To assemble a well-balanced normative sample of the Flanders adolescent population, the same percentage of Belgian adolescents and Immigrant/Refugee adolescents per province took part in the study. In this way, there would not be an overrepresentation of Belgian adolescents living in urban or rural areas. Furthermore, the proportions for the different age and gender groups of the Belgian adolescents were carefully matched with those of the Immigrant/Refugee sample so that the two groups were similar on these variables. Finally, per province the secondary schools that were chosen had students that were following all three educational track levels (trade, occupational and preparatory for university). No Flemish student refrained from participating.

Table 1.
Summary of Sample Characteristics

	<i>n</i>	Gender		Age in years				Group			Type of caregiver		
		% <i>Boys</i>	% <i>Girls</i>	<i>M</i>	<i>SD</i>	<i>Range</i>	% <i>Natives</i>	% <i>I/R</i>	% <i>URM</i>	% <i>Parental</i>	% <i>Other</i>		
Total sample	3535	59.3	40.7	15.72	1.74	8-26	40.7	30.7	28.6	70	30		
Dutch URM	920	72.8	27.2	15.68	1.49	8-20	0	0	100	0	100		
Belgian immigrant/refugee adolescents	939	52.4	45.3	15.47	1.86	10-26	1	84.2	14.8	87.9	12.1		
Dutch adolescents	1059	56.8	43.2	15.72	1.54	13-21	90.1	9.9	0	97.3	2.7		
Belgian adolescents	617	54.6	45.4	16.46	1.92	13-21	97.9	2.1	0	97.6	2.4		

Note. N= Number. M = Mean. SD = Standard deviation. I/R = Immigrants/ Refugees. URM = Unaccompanied Refugee Minor.

Procedures

Four independent studies were conducted, two in the Netherlands and two in Belgium, between 2002 and 2003. The adolescents in each sample were asked to complete at least three short self-report questionnaires, the RATS (Bean et al., 2004a), the Stressful Life Events (Bean et al., 2004b), and the Hopkins Symptom Checklist -37A (Bean et al., 2004c).

Testing of the Belgian and Dutch normative samples took place in small groups (10-25 young people) during school time. The URM were assessed at schools, if possible. Because many URM did not attend schools or were frequently absent, URM were also assessed (in groups of 10) at the regional offices of Nidos, reception centers for refugees and residential settings. Approximately 20% of the URM were not tested at schools. Demographic information on the URM in the Netherlands was supplied by the Nidos Foundation (legal guardian of all of the URM living in the Netherlands). The rest of the participants took part anonymously and answered written questions that provided demographic and social characteristics about themselves, such as gender, age, nationality, time in Belgium and current living situation.

Questionnaires (measurements)

The RATS was developed to render the instrument multi-cultural and adolescent friendly and to attempt to ensure the content and technical equivalence of the RATS. The literal terms of the Likert scale (*not* = 1, *little* = 2, *much* = 3, *very much* = 4) were enhanced by using colored circles of increasing size. Secondly, items were simplified (based on a vocabulary list for foreign students in the Netherlands; Projectbureau OVB Rotterdam, 1992) to adjust the questionnaire to the language abilities of this population (all trauma-related questions come first), and thirdly, the questionnaires were translated and presented in a bilingual form. It was necessary to have the questionnaires in bilingual form because many of the refugee adolescents had limited written knowledge of their own language and learned the Dutch language quickly allowing them to use both languages to be able to better comprehend the item. The RATS, SLE, and HSCL-37A questionnaires were translated into the most prevalent languages of URM in the Netherlands: Albanian, Amharic, Arabic, Badini, Chinese, Dari, Dutch, English, Farsi, French, German, Mongolian, Portuguese, Russian, Servo-Croatian, Soerani, Somali, Spanish, and Turkish. All written forward translations were done by professionally employed translators. Every translation was controlled for grammatical and idiomatic errors on two different occasions by two different translators. The translated questionnaires were reviewed orally with professional interpreters who were regularly involved in treatment sessions of traumatized adult refugees to control the quality of the translations, to ensure that the original meaning was conveyed in the items, and to attempt to achieve semantic equivalence of the RATS. No written back-translations were done in this study. Instead an oral item-by-item analysis took place with trained interpreters from mental health services. All of the instruments were tested in a pilot study.

The *Stressful Life Events* (SLE) (Bean et al., 2004b) was used to assess the number and type of stressful event(s) that was experienced. Adolescents were asked if they had experienced one or more of twelve stressful events commonly experienced by refugee minors (dichotomous yes/no answer). There was one open question and a place for comments at the end of the questionnaire. The overall average total score of 6.5 of the SLE has been validated in 5 independent studies (Bean et al., 2004b).

Posttraumatic stress reactions were assessed with the RATS (Bean et al., 2004c). The 22 items are derived from the seventeen core symptoms of the B, C, and D clusters for the diagnosis of PTSD as defined by the DSM-IV (APA, 1994). The criteria B3, C1, C5, D1, and D2 have been divided into two items to better measure both symptoms of PTSD that appear in one criteria (for example; criteria D1 is "difficulty falling" or "staying asleep"). Great care was taken in the formulation of the items by comparing every word with a Dutch vocabulary list for foreign students and considering how to prevent the item from being misinterpreted in other languages. The checklist is scored using the three clusters of the DSM-IV criteria; intrusion, avoidance/numbing and hyper-arousal. Items 1-6 (range; min.6 - max. 24) correspond to the intrusion symptom cluster, items 7-15 correspond to the avoidance/numbing symptom cluster (range; min. 9 - max. 36), and items 16-22 (range; min.7- max. 28) correspond to the hyperarousal symptom cluster. Separate sub-scores for PTS reactions can be calculated for each symptom cluster. The total score can be calculated adding the points of all

of the 22 items (range; min. 22 – max. 88). Percentile scores and severity classifications are available in the user's manual (Bean et al., 2004c).

The combined use of the SLE and the RATS makes it possible to classify a probable PTSD diagnosis based on the A1, B, D and C criteria of the DSM-IV. One needs to have experienced at least one stressful life event (A1; SLE), one intrusion item, three avoidance/numbing items and two hyper-arousal items (RATS; B, D, and C) to meet the criteria requirements. An item qualifies for scoring (receives a 1) if it has been scored as *much* or *very much*. If the item is scored as *not* or *little*, the item receives a 0. A total sum score of 7 (at least 1 stressful life event, 1 intrusion item, 3 avoidance/numbing items, 2 hyperarousal items) is the minimal score needed for a classification of PTSD.

The Hopkins Symptom Checklist-37 for Adolescents (HSCL-37A (Bean et al., 2004a) (an adaptation of the HSCL-25; Winokur, Winokur, Rickles, & Cox, 1984) measures anxiety symptoms, depression symptoms and externalizing behavior (trauma-related “acting-out”). The psychometric properties appear to be good (Bean et al., 2004a). For the URM sample for the total scale, internalizing, and externalizing behaviour subscales was respectively .91, .92, and .69. 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.

Psychopathology

The criteria “referral” and “utilization of MHC” have been documented as being important in the evaluation of psychopathology in children and adolescents (Anderson, FRANZCP, Williams, McGee, & Silvav, 1987; Cuffe et al., 1995; Verhulst & Van der Ende, 1997). For this reason, (a) self-reported need for mental healthcare (MHC), (b) need for professional MHC for the URM; evaluated by the legal guardian, (c) need for professional MHC for the URM; evaluated by the teacher, (d) self-reported utilization of MHC by URM, and (e) referral to MHC services by a legal guardian was utilized as external criteria of psychopathology. Several studies have also shown that the number of experienced stressful events (dose-effect relationship) to be related to psychopathology, also being a good predictor of psychopathology in adolescents (Allwood, Bell-Dolan, & Husain, 2002; Berthold, 1999; Deykin, 1999; Tiet et al., 1998). The URM were individually interviewed in Dutch about their needs and mental health use. They were also able to read the questions in one of the languages that have been mentioned above. Guardians and teachers received short questionnaires in which they filled-in and returned by mail.

Theoretically, there should be strong significant and positive relationships between the RATS total score and the internalizing score of the of the HSCL-37A because as reported earlier in the introduction of the high co-morbidity that has been found between PTSD on the one hand and general anxiety /depression on the other. The correlation between the externalizing score (measuring trauma-associated acting out behaviour) of the HSCL-37A and the RATS scores should be present but weaker than with the internalizing scale. The total SLE score should be positively related to the total score of the RATS and subscales, since it was reported earlier in this article that trauma is a predictor of psychopathology.

Data Analysis

Descriptive statistics were used to give summary descriptions of the socio-demographic characteristics of the sample (Table 1). Confirmatory factor analyses, per language version, were calculated using the Multiple Group Method (MGM) procedure of the Simultaneous Components Analysis (SCA) (Kiers, 1990) to verify the factorial validity of the RATS (all cases with missing data were removed). The MGM (Guttman, 1952) has been propagated by authors such as Nunnally (1978) who describes the method as “simple, direct, and understandable”. MGM is closely related to the rotation of component weights to perfect congruence and the cross-validation of components weights (Ten Berge, 1986). In this method, the factors which are obtained with this sample are compared with the theoretical three cluster structure of PTSD. SCA is based on the *same* set of weights for the variables in all populations enabling conclusions on the common components found across the samples. It is not a formal statistical test, such as the Maximum Likelihood estimation method. However, this is not a serious objection because the null hypothesis of a factor model based on a small number of factors is invariably false as has been known since Browne (1969, p. 385). Failure

to reject it merely means that the sample size has been too small (see McCrae, Zonderman, Costa, Bond, & Paunonen, 1996 for a discussion).

Internal consistency of the total scale and subscales of the RATS was calculated with Cronbach's α . Test-retest reliability was calculated for a twelve month interval for the URM sample only ($n = 519$). Pearson's product-moment correlations (two-tailed) were used to study the association between total and subscale scores of the RATS and the scores on the remaining questionnaires (SLE, HSCL-37A). Differences between groups were determined by using t-tests for independent groups; ANOVA's and effect sizes (d). Effect sizes were calculated using Cohen's d (Cohen, 1988). The Chi-square test with the odds ratio statistic was calculated per group, per event, and per cluster of total number of events to establish which group or event(s) lead to a greater risk for being classified with a PTSD. A maximum of ten percent of missing items was allowed to still be able to extrapolate the total or subscale scores.

Results

Factorial Validity

The factor structure of the RATS was tested with the MGM-SCA which tests for variables measured in two or more populations. The items of the RATS were divided into three a priori factors based on the three criteria clusters of the PTSD diagnosis in the DSM-IV. For the total sample, a principal component analysis was used with Varimax rotation (oblique) to simple structure which allowed for correlation between the factors (Kiers, 1990). In this analysis, the three factors explained 49% of the total variance. The three a priori factors based on the structure of the DSM-IV PTSD criteria explained 47.3% of the total variance as a result of the SCA-Multiple Group Method (MGM). This difference of 1.7% indicates an acceptable discrepancy.

Separate MGM analyses were conducted on the Portuguese, French, Chinese, English, Turkish, Dutch and Russian language versions (approximately 100 completed questionnaires per version). The amount of variance that was lost in enforcing the a priori factor structure in comparison to the results of the PCA in the separate language versions was very limited, ranging from 4% in the Arabic version to 1.2% in the Dutch version. Due to the limited number ($n < 100$) of completed questionnaires in Badini, Servo-Croatian, Albanese, Turkish, Soerani, Dari, Farsi, Amharic, Somali and Mongolian, no individual MGM's could be conducted for these languages. The three-factor model is confirmed in all the separate MGM analyses per language (Table 2).

Internal consistency

The internal consistency (Cronbach's alpha), which measures the homogeneity of the RATS (Myers & Winters, 2002), supports distinct scales in all language versions and per subscale (see Table 2). The internal consistency of the total scale of the RATS in the total sample was .91 and ranged from .81 to .93 for the individual language version. These are exceptionally high alphas, despite the high degree of heterogeneity in the samples.

Temporal Stability

The test-retest scores are used to provide an indication of scale stability and consistency over time (Myers & Winters, 2002). The test-retest reliability of the RATS was determined in a subgroup of 519 unaccompanied refugee minors, who completed the questionnaire twice. The time interval was twelve months. The stability coefficients were .61, $p < .001$, for the total RATS score, .63, $p < .001$, for intrusion, .44, $p < .001$, numbing/avoidance, and .55, $p < .001$ for the hyperarousal scale. The coefficients show the RATS scales to be reasonably stable in measuring traumatic stress reactions and do not differ greatly from other studies with the same time interval (see Cheng & Nicholas, 1998 for a discussion on stability of self-report measures).

Table 2.
Summary of Confirmatory Factor Analyses and Reliability Analyses per Language Version

Language	Total scale			Intrusion			Numbing/Avoidance			Hyperarousal								
	<i>n</i>	EV	LV	α	r_{ii}	r_{it}	<i>n</i>	α	r_{ii}	r_{it}	<i>n</i>	α	r_{ii}	r_{it}				
1. Total sample	3096	49.0%	1.7%	.91	.34	.09-.67	3304	.87	.51	.34-.67	3249	.81	.32	.13-.60	3325	.76	.32	.14-.63
2. Dutch	1712	44.7%	0.4%	.89	.28	-.07-.59	1752	.82	.42	.25-.57	1741	.77	.28	-.09-.60	1758	.76	.32	.20-.58
3. Portuguese	348	44.4%	3.3%	.87	.23	-.05-.65	378	.85	.48	.22-.65	369	.66	.17	-.03-.50	384	.74	.29	.01-.64
4. English	236	45.9%	4.5%	.88	.25	-.11-.59	271	.85	.48	.30-.58	275	.72	.22	-.06-.59	272	.70	.25	.06-.61
5. French	171	44.8%	2.6%	.88	.24	-.10-.72	198	.83	.43	.21-.71	186	.69	.19	.02-.55	200	.74	.28	.06-.58
6. Arabic	47			.93	.35	-.13-.74	52	.90	.60	.33-.75	50	.80	.29	-.13-.64	56	.77	.31	-.02-.65
7. Turkish	126	51.0%	3.5%	.90	.30	-.12-.64	145	.86	.50	.38-.60	136	.71	.22	-.04-.48	144	.81	.39	.24-.57
8. Russian	126	50.1%	2.3%	.90	.29	-.15-.64	141	.84	.47	.34-.60	134	.77	.27	-.05-.61	143	.82	.39	.24-.61
9. Chinese	98	54.0%	2.5%	.92	.32	-.10-.74	101	.84	.45	.13-.74	102	.84	.36	-.04-.69	102	.78	.33	.09-.72
10. Spanish	46			.81	.17	-.18-.69	48	.52	.14	-.13-.37	48	.72	.23	-.15-.69	51	.65	.21	-.17-.62
11. Farsi	43			.87	.22	-.21-.67	46	.83	.44	.12-.63	45	.72	.21	-.18-.59	46	.66	.22	-.03-.47
12. Albanese	26			.90	.28	-.34-.90	36	.84	.45	.13-.76	32	.74	.24	-.33-.77	31	.70	.24	-.19-.68
13. Servo-Croatian	18			.88	.27	-.48-.85	23	.84	.19	.19-.77	24	.58	.14	-.31-.72	22	.80	.35	.00-.67
14. Dari	24			.88	.24	-.43-.82	28	.83	.46	.19-.71	27	.73	.22	-.33-.77	28	.74	.30	-.01-.53
15. Amharic	22			.93	.37	-.43-.95	27	.88	.56	.19-.80	24	.82	.33	-.12-.82	29	.69	.25	-.26-.61
16. Somali	19			.88	.26	-.47-.91	19	.86	.52	.03-.77	21	.74	.23	-.48-.77	22	.56	.18	-.28-.88
17. German	17			.92	.37	-.33-.87	19	.90	.60	.15-.81	18	.86	.47	-.05-.86	18	.82	.15	-.26-.79
18. Mongolian	11			.91	.30	-.40-.90	11	.67	.26	-.20-.75	11	.84	.34	-.21-.90	11	.72	.27	-.40-.71

Note. EV = Explained Variance with PCA; LV = Loss of Explained Variance with MGM; α = Alpha coefficient; r_{ii} = Mean inter-item correlation; r_{it} = Range item-total correlations

Table 3.
Intermeasures Correlations between RATS scales with the HSCL-37A and SLE

RATS subscales	internalizing HSCL-37A	externalizing HSCL-37A	Total SLE
URM (n = 771)			
1. Total RATS	.79***	.32***	.46***
2. Intrusion	.70***	.18***	.43***
3. Numbing/Avoidance	.61***	.26***	.37***
4. Hyperarousal	.73***	.40***	.38***
Dutch natives (n = 1058)			
1. Total RATS	.76***	.23***	.50***
2. Intrusion	.64***	.12***	.45***
3. Numbing/Avoidance	.67***	.10***	.39***
4. Hyperarousal	.64***	.34***	.45***
Belgian immigrants/refugees (n = 870)			
1. Total RATS	.66***	.33***	.52***
2. Intrusion	.58***	.23***	.53***
3. Numbing/Avoidance	.55***	.25***	.44***
4. Hyperarousal	.64***	.39***	.38***
Belgian natives (n = 596)			
1. Total RATS	.67***	.31***	.45***
2. Intrusion	.56***	.23***	.44***
3. Numbing/Avoidance	.60***	.27***	.36***
4. Hyperarousal	.58***	.36***	.40***

*** $p < .001$, two-tailed.

Content Validity

Content validity is a measure of the relevance of the items with regard to that behavior which they aim to measure (Morgan, Gliner, & Harmon, 2001). The RATS claims to measure the B, C and D criteria of the PTSD diagnosis defined in the DSM-IV (APA, 1994). All items of the RATS correspond with the DSM-IV criteria for a PTSD diagnosis. In fulfilling the content equivalence dimension of cross-cultural validation as proposed by Flaherty et al. (1988), the wording of the items was made equivalent to the reading level of a 12 year old in Dutch. In addition, the items were semantically made as concrete and “universal” as possible so that they were relevant for adolescents coming from a variety of different cultures and therefore could also facilitate the translation process. The content validity of the RATS can be considered good.

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. In construct validation, three processes are used to establish construct validity: (1) convergent validity – high correlations between one scale and others that in theory measure the same construct; (2) discriminant validity – low associations between the scale under study and other measures that should theoretically not be related; and (3) factorial validity – supports the theory-based grouping of items when a particular construct is complex (Morgen et al., 2001). In this study, the factorial validity of the RATS (reported earlier) has been examined. Table 3 shows the measure intercorrelations (with two-tailed test) between the RATS total and subscale scores, the HSCL-37A total score and the SLE total score for the URM sample and native Dutch sample. The measures' intercorrelations are presented in Table 3 between the HSCL-37A total score and the SLE total score for the immigrant/refugee and native Belgian sample.

As hypothesized, the RATS total scores show significant strong and positive correlations with the HSCL-37A internalizing scores (see Table 3). The relationship between the total score and subscales on the RATS and the total number of experienced events on the

SLE is positive and significant providing evidence of convergent validity. The relationship between the HSCL-37A externalizing scale and the RATS scales are small (Cohen, 1988), but still significant providing some support of discriminant validity. These findings are applicable to all samples.

The total mean scores of girls are expected to be significantly higher than that of boys. Table 4 shows that girls reported significantly higher total mean scores than boys (medium effect size). Age, in this study, played no role with respect to total mean scores. URM reported significantly higher total mean scores on the RATS than the other groups (very large effect sizes). Adolescents who reported having experienced four or more stressful life events scored significantly higher on the RATS than adolescents reporting less than four events.

Table 4 .
Relationship of RATS Total Scores with Demographic Characteristics

Characteristic	<i>df</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	Comparison Groups for <i>d</i>	<i>d</i>
Gender*							
1. Girls	2408	1081	36.41	10.66	8.93***		.37
2. Boys		1329	32.74	9.25			
					<i>F</i>		
Age*							
1. ≥14 years	3;2360	609	34.67	10.18	1.83		
2. 15 years		509	33.60	9.70			
3. 16 years		454	34.02	10.35			
4. ≤ 17 years		792	34.80	10.09			
Group							
1. Natives	2;3359	1560	32.63	9.18	748.69 ***	1 vs. 2 =	.51
2. Immigrants/Refugees		868	37.59	10.80		1 vs. 3 =	1.62
3. URM		934	49.09	11.57		2 vs. 3 =	1.03
Total number of reported SLE's							
1. 0 events	3;3363	235	27.82	6.21	258.40***	1 vs. 2 =	.58
2. 1-3 events		1387	32.70	8.74		1 vs. 3 =	1.29
3. 4-7 events		1301	41.73	11.42		1 vs. 4 =	2.45
4. 8-13 events		444	52.45	11.60		2 vs. 3 =	.89
						2 vs. 4 =	2.08
						3 vs. 4 =	.94

Note. *analyzed without URM group. *** $p < .001$.

Criterion Validity

Criterion validity refers to validating the instrument with some form of external criterion in which a relationship is established between the instrument and a measurable external or outside criterion. (Morgan et al., 2001). Criterion validity can be further divided into two types; predictive and concurrent. Only the concurrent validity of the RATS was addressed in this study. The five indicators of psychopathology are discussed in the Methods section of the article. URM themselves ($M = 51.98$, $SD = 11.08$), $F(2, 737) = 45.74$, $p < .001$, $d = .89$, their guardians ($M = 55.02$, $SD = 10.84$), $t(477) = 5.51$, $p < .001$, $d = .65$, and their teachers ($M = 52.40$, $SD = 11.76$), $t(388) = 3.77$; $p < .001$, $d = .43$, reported significant higher RATS total mean scores for URM who needed professional care than for URM that did not need professional care. URM that had utilized mental health services ($M = 51.80$, $SD = 15.87$), $t(128) = 2.09$, $p < .05$, $d = .17$, and the URM which were referred by their legal guardian to mental health services ($M = 56.84$, $SD = 11.71$), $t(482) = 5.42$, $p < .001$, $d = .77$, also had significant higher mean total scores on the RATS than did URM who had never used or had

been referred to mental health services. The RATS questionnaire discriminated well, consistently, and significantly between URM that do and do not have a need for psychosocial help.

Odds ratio

The risk estimate for a PTSD classification for different groups, total number of SLE's and individual events can be found in Table 5. Dose-response relationships were evident between experiencing many SLE's (4 or more) and being classified with a possible PTSD diagnosis. URM have a 10 times greater risk for developing PTSD reactions than natives and 5 times greater risk for immigrant/refugees. If an adolescent had experienced being separated from their family, they were six times more likely of being classified with the diagnosis of PTSD than if they had not been separated from family. Furthermore, reporting having experienced physical (OR = 5.03) or sexual maltreatment (OR = 4.47) greatly increased the likelihood of being classified as having a possible PTSD. If an adolescent had reported experiencing a drastic change in the family situation last year or a war/armed conflict they were approximately 4 times more likely of for developing a PTSD than adolescents who have not experienced these events. Finally, if an adolescent had experienced losing a loved one, being seriously (life- threatening) ill or having witnessed someone being physically maltreated they had approximately a 3 times greater risk of being classified with a possible PTSD.

Table 5 .
Odds Ratio for Possible PTSD Classification

Stressful life event	χ^2	Odds Ratio	95% Confidence interval
Drastic change in family during last year (no/ yes)	235.81*	4.18	3.45 - 5.06
Separated from family (no/ yes)	430.13*	6.18	5.14 - 7.42
Loss of loved ones (no/ yes)	130.53*	3.56	2.84 - 4.48
Serious Illness (no/ yes)	187.53*	3.72	3.06 - 4.53
Serious Accident (no/ yes)	21.14*	1.64	1.33 - 2.03
Disaster (no/ yes)	93.14*	2.54	2.09 - 3.08
War or Armed Conflict (no/ yes)	287.28*	4.39	3.67 - 5.25
Physical maltreatment (self) (no/ yes)	335.22*	5.03	4.19 - 6.04
Physical maltreatment (witnessed) (no/ yes)	207.41*	3.88	3.20 - 4.71
Sexual maltreatment (no/yes)	195.65*	4.47	3.57 - 5.58
Other (self) (no/ yes)	270.74*	4.50	3.72 - 5.43
Other (witnessed) (no/ yes)	142.25*	2.94	2.45 - 3.53
SLE Total (0-3) / (4-13)	424.48*	8.95	7.06 - 11.35
Group (Native/ Immigrant/refugee)	23.00*	1.93	1.47 - 2.53
Group (Immigrant/refugee / URM)	216.74*	5.09	4.05 - 6.39
Group (Native/URM)	469.36*	9.80	7.79 - 12.34

* $p < .001$, two-tailed .

Discussion

The RATS is a psychometrically sound screening measure to assess posttraumatic stress reactions of adolescents. The preliminary data has been collected among four independent populations stretched across the Netherlands and Flemish Belgium. The psychometric properties of the RATS demonstrate invariance of factor structure in a heterogeneous sample, strong reliability and good validity which is remarkable considering the diversity of the populations.

In this study, the layout of the instrument made the RATS comprehensible for adolescents that had not mastered the Dutch language. However, in individual cases, lengthy

explanations of the meaning/nuances of the items were necessary, especially with “almost” illiterate adolescents. It is not clear if errors in understanding the questions might not be visible in the data. The Spanish version of the RATS had obvious less internal consistency on the total scale and all subscales. This could be due to the fact that the translation was in European Spanish and adolescents came from South American countries which speak a different dialect of Spanish.

The brevity of the RATS and the SLE takes into account the importance of not overburdening apprehensive adolescents and allows for quick, repeated measurements to assist with determining initial and enduring trauma-related symptomatology. Early symptom detection can lead to the initiation of appropriate therapeutic interventions. The RATS is to be used by trained professionals who are capable of proficiently assessing the well-being of adolescents. In all settings, one must be aware that the instrument may trigger emotional distress. Therefore, adequate crisis and/or follow-up care should be arranged prior to administration to protect the integrity of the adolescents.

Clinical observations and additional assessment are important in establishing a valid diagnosis and making treatment recommendations. Furthermore, it is important that clinicians are aware that URM in host countries are at an extremely high risk of experiencing severe posttraumatic reactions, thus the RATS should be used for screening purposes with this specific population.

Study Limitations

There are several limitations to the findings of this study. Although it appears from the preliminary findings on the multiple language versions that the cross-cultural equivalence of the RATS has been verified, written back-translations of the language versions were not done, deviating from standard protocol which can be seen as a limitation of the study. Back-translation is the method that is usually used to verify semantic equivalence of translated measures (see Mallinckrodt & Wang, 2004 for a discussion). However, a back-translation does not implicitly guarantee that the content equivalence of the translated instrument has been established (Flaherty et al., 1988). A great amount of effort in this study was spent on ensuring the content equivalence of the items of the RATS for different cultures.

The large amount of time and intense effort that the refugee adolescents needed to fill in the instruments limited the number of measures that could be utilized in the study (such as additional measure for PTSD or instruments to measure thoroughly the discriminate and divergent validity of the RATS). The number of instruments that were used were limited to a minimum for a number of reasons; (a) the short attention spans of the refugee adolescents, (b) the amount of time needed to explain and administer (getting the right language version to the right adolescent) the three instruments took around 15 minutes of the testing time, (c) the substantial amount of time and effort used by the refugee adolescents to complete only three questionnaires, and (d) the ethical issue of administering long instruments with severely traumatized individuals which might induce emotional distress. Additional measures would have enhanced the quality of the study and would have been useful in determining the divergent validity of the RATS. The divergent validity of the RATS will need to be further investigated in future studies.

The stability (test-retest) of the RATS was calculated over a longer interval (12 months) than the usual 8-week interval. This resulted in poor temporal stability than is typically expected or desired. However, it could be expected after one year that many changes (i.e., developmental, stressful life events, transfers, residential status, and therapeutic interventions) would have taken place in the constantly changes lives of URM which could have led to much lower stability levels. The significant stability coefficients found for traumatic stress reactions among URM could imply continuity of psychopathology among this sample. Nonetheless, the stability of the RATS for short and long periods will need to be further evaluated in future studies to properly establish the temporal stability of the measure.

Furthermore, the fact that the instruments that were used in the study all have been developed or modified by the authors could have in some way affected the results. RATS and SLE together do not measure the A2 and E criteria of a PTSD defined by the DSM-IV (APA, 1994) meaning that caution should be used when using these two measures to classify for a PTSD. Because no standardized diagnostic interview was utilized in this study, the sensitivity and specificity of the RATS could not be evaluated, nor the PTSD classification confirmed.

Ideally, a standardized diagnostic interview is used in combination with questionnaires to determine the presence and severity of psychopathology. However, “referral” of children and adolescents to psychiatric services has been used as a “gold standard” instead of a diagnostic interview (e.g., Nolan et al., 1996). It was not feasible in the URM study to administer a diagnostic interview for the reasons that have been listed above and that there is no validated psychiatric diagnostic interview available in all of the languages of the (refugee) adolescents who took part in this study. It would have first been necessary to test the validity and reliability of a psychiatric interview in the 19 languages before it could have been utilized, which is a very time consuming and expensive process. Furthermore, the use of diagnostic interviews in cross-cultural studies invokes itself a host of methodological issues such as classifying culture-specific disorders and ensuring “the semantic and psycholinguistic equivalence of psychiatric symptoms across cultures” (Cheng, 2001). Nonetheless, the preliminary validity findings suggest that the RATS is able to discriminate between adolescents that do and do not need to utilize mental health services.

Self-report questionnaires such as the RATS yield less diagnostic information than extensive structured interviews and therefore should be used only to indicate whether a refugee adolescent is experiencing global posttraumatic reactions due to a traumatic event and not to diagnosis a PTSD. Additional information should be collected regarding the mental health of the adolescent from the viewpoint of significant adults (caregivers/ teachers) in the environment of the adolescent. This information is crucial in assessing the degree of impairment in daily functioning and the severity of the symptoms of adolescents.