SLAVERY TODAY JOURNAL
A Multidisciplinary Journal of Human Trafficking Solutions
Volume 2, Issue 1 January 2015

Toward Assessment of Child Survivors of *Restavèk* in Haiti: Development and Evaluation of a Locally Adapted Psychosocial Assessment Instrument

by

Cara L. Kennedy, PhD

Research Development Specialist in the College of Arts and Humanities at the University of Maryland

Toward Assessment of Child Survivors of *Restavèk* in Haiti: Development and Evaluation of a Locally Adapted Psychosocial Assessment Instrument

by Cara L. Kennedy

Research Development Specialist in the College of Arts
and Humanities at the University of Maryland

Acknowledgements

I would like to thank Equitas Group for financing this project; Beyond Borders for contributing staff time, organizational resources, and project oversight; the collaborating organizations, their site directors and staff; and the children and caregivers who provided their time and participation to make this work possible.

Abstract

Restavèk is a form of child domestic slavery in Haiti that affects an estimated 300,000 children. This article describes the development and evaluation of an instrument to assess mental health and psychosocial problems among survivors of restavèk living in Port au Prince, Haiti. The Youth Self-Report was adapted to reflect the mental health problems that emerged in a previous qualitative study among the same target population. Internal consistency reliability scores were acceptable to good for all scales. Test-retest reliability scores were adequate for all scales, and good for the internalizing and total problems scales. Criterion validity could not be assessed.

INTRODUCTION

The *restavèk* system in Haiti is a system of child trafficking and forced labor that meets criteria for slavery (Bales 1999): *restavèk* children are completely controlled through violence and exploited by the heads of the households for whom they work. Often far from home, isolated and excluded within their own environments, *restavèk* children have no viable options. They are under physical control, and thus cannot walk away. *Restavèk* children are abused physically, verbally, emotionally, and sexually. They are forced to do chores that are inappropriate for their age, often not sent to school, and treated as inferior to the children of the family. Drivers of the *restavèk* practice ("push factors") include

social and economic disparities, social exclusion, lack of educational access and the illusion of opportunity for education and social mobility, loss of parents, and the perception that the "employer" is extended family and a protected environment. "Pull factors" include low-resource environments where daily demands for household work exceed household members' capacity and cultural norms allow that a child is an appropriate choice to carry this labor. In the largest field survey of human rights violations in Haiti, the Pan American Development Foundation (PADF) defined *restavèk* as "an unpaid child servant living and working away from home. . . [who is] treated in a manner distinctly different from children born to the household" (Pierre, Smucker, and Tardieu 2009). Though there remains some controversy regarding the definition of *restavèk*, the PADF definition guides our understanding of *restavèk*, in which the "distinctly different treatment" involves some form(s) of abuse, exploitation, neglect, and/or humiliation.

In 2007-2008, more than 225,000 children, of whom an estimated two-thirds were girls, were living in urban *restavèk* slavery (Pierre, Smucker, and Tardieu 2009). More than one-third of households in Port-au-Prince reported housing *restavèk* children with percentages as high as 44 percent in the urban slum of Cité Soleil. As *restavèk* is not an exclusively urban phenomenon, the total prevalence of *restavèk* in Haiti at the time of that survey was likely closer to 300,000 children nationwide. In 2012, the US Department of Labor conducted a qualitative study including current and former *restavèk* children and the sending and receiving families to understand the root causes and factors that maintain the *restavèk* system, the sending and receiving families' participation and relationship to one another, and the programs and policies that exist to address the root causes (Cooper, Diego-Rosell and Gogue 2012). Most relevant to the current study were the findings that the living and working conditions of the children were "unrelentingly bleak" with children deprived of food, sleep, education, and time for study and play. They were also subjected to excessive labor and physical and sexual abuse.

Mental Health and *Restavèk*

Mirroring the global response to child slavery and trafficking, there have been few examples of intervention that have attended to the mental health needs of children coming out of *restavèk*; those that exist are often limited to basic interventions that create the conditions for healing (e.g., providing shelter, safety, opportunity for play and distraction), but do not address the chronic traumas of abuse, humiliation, exploitation, neglect, and abandonment. Despite the lack of response, anecdotal evidence from service organizations working with *restavèk* in Haiti suggests that the scale and consequences of *restavèk* are significant, with

social costs to subsequent generations stemming from abuse and absence of affection among an entire class of citizens.

Fueling this lack of response, until recently no systematically gathered data existed on the psychological, social, economic, or educational consequences of *restavèk* on the children who are forced to live in these conditions, or on their families in cases where they return home. Prior to the current study we conducted two linked qualitative studies with children who had been in *restavèk* and their caregivers (Kennedy 2014) to systematically understand the consequences of *restavèk*. Studies were designed to understand the range of challenges facing children who have been in *restavèk* and to probe more deeply into those challenges that relate to children's mental health. Children who had been in *restavèk* exhibited a diverse range of signs and symptoms of mental health problems that are consistent with other findings of children in high stress and abusive family environments (Moylan et al. 2010) and in other abusive forms of trafficking (Flowers 2001; Raymond and Hughes 2001), and included both internalizing (crying, sadness, rumination, remembering the bad moments) and externalizing (stealing, fighting, being unruly or disobedient) symptoms, several of which may be markers of trauma in children and youth.

In addition to mental health symptoms, qualitative study findings indicated a number of social and environmental experiences children face after leaving *restavèk* that further complicate their risk of mental health problems (Kennedy 2014). While both children and adults consistently reported that families are happy to have their children return, and children are more comfortable and happy back at home, the extremely challenging economic situation of families, in addition to the stigma the child carries home from the experience of *restavèk*, were cited as causes for the child to experience a range of serious mental health symptoms as well as social challenges. The interlocking stigma of poverty and *restavèk* combine forcefully to place the children at risk for mental health problems as well as to be placed in a position where they are not accepted and are humiliated by peers and other adults.

Mental Health Assessment with Former *Restavèk* Children

Given the evidence for mental health problems among children who have been in *restavèk*, it is important that local service providers and organizations begin to address these needs. However, a substantial challenge to mental health intervention in child slavery globally and in Haiti in particular is the lack of local, valid measures for the assessment of mental health symptomatology (Betancourt et al. 2010, 2008), and for the measurement of intervention outcomes. Translated measures developed in other, Western contexts to assess mental health syndromes

or disorders are often utilized; however, they are typically not validated for use in the context or with the population of interest (Hollifield et al. 2002; Mollica et al. 2004), and the use of such measures can lead to erroneous conclusions about the mental health of a community or population of interest (Hobfoll et al. 2011). For numerous reasons, psychometric properties of reliability and validity cannot be assumed to be maintained in translated measures used in other contexts, and must be evaluated locally (Hall et al 2014): instruments may not adequately capture what they are intended to measure, they may include irrelevant items, and/or exclude the local cultural idioms and expressions of psychological distress and resilience that will serve as the most refined indicators of the construct of interest. The reliability and validity of measures developed for one population and used in another continues to be a central issue in cross-cultural mental health research (Betancourt et al. 2009; Hall et al. 2014). Reliability refers to the degree to which empirical measures result in reproducible results across different assessments or with different interviewers (Shrout 1998). Validity refers to the degree to which an empirical tool measures the construct it was designed to measure (Goldstein and Simpson 1995).

One of the most challenging and most important psychometric properties to evaluate in cross-cultural mental health research is criterion validity – the degree to which the instrument measures what it was intended to measure. Criterion validity is measured by comparing its performance with that of a "gold standard," or another method of assessing mental health problems that is known to be accurate (Bolton 2001). For studies of mental health across cultures, the gold standards for criterion validity are typically assessment by a psychiatrist or psychologist or comparison with another instrument known to have high validity among the local population (Bolton 2001). In Haiti, however, an existing valid instrument or a psychiatrist or psychologist familiar with the local culture (or any psychiatrist or psychologist) are not available. Without a method of known accuracy for comparison, an alternative standard that involves comparing data from the instruments with self-assessments by the survivors and adults close to them has been developed and used in other non-Western mental health validation studies (Bass et al. 2008; Betancourt et al. 2009; Bolton 2001). While this is not a standard of known accuracy, and both the survivor and other adult may be incorrect in many cases, the assumption is that a survivor is more likely to have or not have each of these problems if the survivor and the other adult agree. Those said to have such problems should therefore have higher scores than those said not to have such problems, if both the instrument and the opinions of survivors and other adults are accurate assessment of the presence and severity of these syndromes.

The current study aimed to develop a local instrument for the assessment of mental health problems among child survivors of *restavèk*, and to evaluate its reliability and validity for this population.

METHODS

The instrument development and evaluation methods described here are part of a promising methodology developed by the Applied Mental Health Research (AMHR) group of Johns Hopkins University (JHU) in their work in several different non-Western contexts worldwide (Bass et al. 2008; AMHR 2011). To ensure that instrument development adequately captures the local understanding and expressions of psychological distress, the methodology relies on an initial rapid, comprehensive qualitative research phase in which children and parents/ caregivers participate in semi-structured interviews to understand the problems they identify, the types of distress children are experiencing, and the ways their distress is expressed. This preliminary step was the focus of the qualitative studies described previously (Kennedy 2014) and serves as the basis for the current study. In the current study, the AMHR methodology builds upon those findings to develop quantitative assessment instruments, drawing upon published measures for assessing the phenomenon of interest and adapting them based on local concepts and descriptions that emerged in the qualitative research. The instruments are then validated to ensure not only that they are understandable and acceptable, but also that they are reliable (give the same results with repeated use) and valid (measure what they are supposed to measure).

Developing the Draft Instrument for Field Testing

Instrument Selection

In developing a quantitative instrument for use with a new population, one important consideration is whether to adapt an existing instrument already used in other populations, or produce an entirely new instrument for local use. Using an existing instrument is preferable if there is one that adequately reflects the local situation, since use of an existing instrument allows for comparison with other populations. In reviewing existing instruments, we based our choice on whether the instrument reflected those important mental health and psychosocial problems that emerged from the previous qualitative study and were amenable to interventions that are within the resources of the institutions and service providers involved in this collaborative research.

The qualitative study (Kennedy 2014) revealed that child survivors of *restavèk* in the Port au Prince metropolitan area experience a wide range of mental health and psychosocial challenges, which we refer to as "problems" in this report. These problems closely map onto Western concepts of internalizing symptoms, such as sadness, crying, rumination, remembering the bad moments, and being uncomfortable or nervous; externalizing symptoms, such as insulting others, fighting, aggression, or being unruly; and relationship problems such as staying alone, being subjected to insults or humiliation by others, and problems related to adaptation after being in *restavèk*. Because the qualitative study was designed to elicit the range of problems facing children without restricting responses to those related to mental health, a number of other social, economic, and community-level problems for children who have been in *restavèk* were reported. Those problems are not addressed directly by the current study; however, the emotional and behavioral symptoms that result from them are. The qualitative results suggested that no one particular problem was more prominent than the others, and that most children had multiple problems. Therefore, it was decided that any appropriate instrument would have to be a broadly-based measure that spanned the range of these problems, rather than one that focused on a particular symptom or group of symptoms, in order to assess the wide range of problems and symptoms reported by children.

The Youth Self-Report (YSR, Achenbach 1991) was selected. The YSR assesses a broad range of psychosocial problems that closely matches those emerging from the qualitative studies. The YSR is part of a set of instruments developed by the Achenbach System of Empirically Based Assessment (ASEBA)TM, and is designed to obtain standardized data on a broad range of problems from multiple sources. Parents, teachers, and youth ASEBA instruments have been translated into more than 90 languages, and studied in over 80 societies (Berube and Achenbach 2012). An entire supplemental manual now exists on the multi-cultural use of ASEBA instruments, including variations in norms across different countries (Ivanova et al. 2007). The YSR is self-administered by youths. The first two pages include demographic and competency questions on the child's interests, chores, social interactions, performance in academics, and open-ended questions about illness, disabilities, problems and concerns. The following two pages contain 105 symptom and behavior specific items, plus 14 positive qualities items, with responses rated on a 0-2 scale (0=not true, 1=somewhat or sometimes true, and 2 very true or often true).

Instrument Adaptation

Through communication with the developers of the YSR, we obtained a license agreement to translate the YSR and underwent an iterative translation and back-translation process to generate the approved translation of the YSR. After examining the items from the YSR in relation to the problems described in the qualitative study, all of the original YSR items were retained for the validation study in order to maintain the empirical base of the instrument and retain future comparability with results from children in other countries. In addition, several items (34 symptom items and 14 function items) were added to the YSR based on frequently mentioned issues in the qualitative studies that were not already well captured in the YSR.

The result was a draft version of the Youth Self-Report – Haiti (YSR-H). The team of Haitian staff conducting the validation study reviewed each item of the draft instrument to ensure clarity, comprehensibility, and tolerability of the items. Suggested changes to improve clarity or comprehension of the translation of the YSR items were made without changing the meaning of the item or its back-translation, so as to maintain fidelity to the approved translation.

In addition to the adapted instrument, study staff and the interviewers drafted an informed consent procedure to be read to children prior to administering the instrument. A pilot study of the YSR-H was conducted in order to detect any problems with the interview procedure, the informed consent procedure and the instrument from the point of view of both the interviewers and the interviewees, and to ensure that the instrument was acceptable and understandable to our target population. Based on feedback from interviewers and interviewees in the pilot study, minor changes to the wording of some items and a more extensive protocol for responding to participants who endorsed items related to self-harm and/or suicidal ideation were required.

Instrument Validation

The purpose of the reliability and validity study was to determine if the YSR-H could accurately assess the presence and severity of the relevant mental health and psychosocial problems reported in the qualitative study. All data analyses were conducted using STATA statistical software. Internal consistency reliability was assessed using Cronbach's alpha (Cronbach 1951). Test-retest reliability was assessed using the Pearson correlation coefficient and was evaluated by re-interview of approximately 39% (n=27) of the participants 2-12 days (mean = 5.7 days) after their first interview. Criterion validity was assessed by comparing scale scores for the survivors said by themselves and an adult to have affective/

behavioral problems with scores of survivors said by themselves and others to not have them.

Study Team and Preparation

The study team consisted of the study author, a Haitian co-investigator, and five interviewers who were local staff of local Haitian non-governmental organizations working with children. In addition, Dr. Paul Bolton, Associate Scientist at JHU, trained the study author in the quantitative methodology and provided distance supervision periodically throughout data collection and analysis. Interviewers took part in a two-day training on instrument administration and consent procedures, involving didactic training and role-play, and a two-day pilot study to increase interviewer's ease and efficiency in administration.

Study Sites

This study took place in four institutions in the Port au Prince metropolitan area. All four institutions provide transitional, residential care to children at risk, providing housing and education for children for a period of months to years. Two of them work exclusively with children who have been in *restavèk*, one of which is for girls only. The remaining two work with children in difficulty for a variety of reasons, including children who have been in *restavèk*.

RESULTS

Sample Characteristics

Interviewers assessed 75 children/youth and the institution staff with closest relationships to them across the four study sites. Of the 75 children/youth, one child left the center prior to the time of the validity study, one refused the assessment, three refused less than halfway through the instrument, and the data for an additional three were deemed invalid based on the response patterns. The characteristics of the remaining 67 participants are provided in Table 1.

Scale Characteristics

Although the complete YSR-H was administered, the reliability and validity analyses are reported only for the symptom subscales and totals. The developers of YSR at ASEBATM recognize the culture-specific nature of the YSR competence items, and have authorized cross-cultural use of the YSR excluding the competence

items. Thus, the reliability and validity analyses are based on seven scales. These include the original YSR Scales (Internalizing Problems, Externalizing Problems, and Total Problems), the Haiti Symptoms Scale, and the YSR-H Scales (Internalizing Problems, Externalizing Problems, and Total Problems). The descriptive statistics for the seven scales are provided in Table 2.

Internal Consistency Reliability

Internal consistency reliability refers to how well questions measuring the same underlying concept on the same occasion agree with each other. Agreement is measured quantitatively by correlations, and Cronbach's alpha is a statistical measure that provides this. It is a single figure that summarizes the average correlation between all pairs of questions in a questionnaire. Cronbach's alphas should be above 0.7 and ideally between 0.8-0.9. If questions believed to be measuring the same concept disagree this suggests that either the questions themselves are unreliable, or they are not really measuring the same concept. Table 3 shows the Cronbach's alpha scores on each of the scales for males, females, and the total sample. Alpha scores for all scales are in the acceptable (>.7) to good (>. 8) range.

Test-Retest Reliability

Table 4 shows test-retest reliability results, based on the 27 (39%) interviews that were repeated 2-12 days (mean=5.7 days) after the preliminary interview. In ten (37%) of the 27 cases, the instrument was administered by a different interviewer at retest. Test-retest reliability is assessed using the Pearson correlation coefficient, which provides a measure of how similar each scale score is on the first and second interviews. This provides an indicator of the extent to which respondents tend to give the same answer to the questions constituting the scale when asked on different occasions. For each comparison, a scatterplot of the scale scores on the first interview was compared with those on the second interview in order to determine whether there was a linear relationship and therefore whether the Pearson correlation coefficient was an appropriate measure. For all comparisons the scatterplot suggested a linear relationship, confirming that the Pearson correlation co-efficient was an appropriate measure of test-retest reliability.

When assessing test-retest reliability, Pearson correlation coefficient scores of .7 are considered to be desirable, .6 adequate. On that basis, all scales demonstrated adequate test-retest reliability. Internalizing scales and total problems

scales, for both YSR and YSR-H demonstrated higher test-retest reliabilities (>.8) than externalizing scales (>.6).

Criterion Validity

One major focus of validity testing was to explore criterion validity. We attempted to sort participants into those who have mental health problems ("cases") and those who don't ("non-cases") based on conversations with children and the staff who know them well. However, in the majority of cases, there was no agreement between the child report and the adult report in the sorting assessments. Conventionally, all cases in which there is no agreement between child and caregiver would be excluded from the study, so that only those where both parties agreed would be used. Thus, due to the nature of our methods and the low agreement between children and caregivers in this study regarding emotional and behavioral problems, we were unable to proceed with a test of criterion validity.

DISCUSSION

On this study's measures of reliability and validity, we found that the YSR and YSR-H symptom-based scales have solid psychometric properties in this population. Internal consistency reliability results were good for YSR externalizing, YSR total problems, Haiti symptoms, and the three YSR-H scales (internalizing, externalizing, and total problems), and adequate for YSR internalizing problems, indicating that these items on these scales perform well in terms of measuring the same underlying concept.

Test-retest reliability results ranged from good (for the YSR internalizing, YSR total problems, YSR-H internalizing, and YSR-H total problems scales), to adequate (for the YSR externalizing, and YSR-H externalizing scales). It is important to point out that while the YSR is a self-administered scale, for our purposes interviewers read the items aloud and completed the forms based on the participants' verbal responses. We were unable to ensure that the same interviewer was present for each re-interview, and thus the test-retest results may be confounded by inter-interviewer differences as well. Analyses of distributions indicated very low frequencies on a number of the rule-breaking behavior items, and it is possible that the externalizing scales were more affected by social desirability as a function of their content, and/or that the strong cultural norms prohibiting certain behaviors influenced responses when in the presence of different interviewers.

Study Limitations

Though we attempted to assess criterion validity, the low agreement between the children and the staff who know them well in the sorting process precluded our proceeding with this test. There are a number of reasons why this may have occurred. First, research on the use of multiple informants, including meta-analyses of studies using the Achenbach scales for self, parent/caregiver, and teacher report have indicated low to modest levels of agreements between pairs of informants, especially between self ratings and ratings of others, the case in our study (Achenbach, McConaughy and Howell 1987; Renk and Phares 2003). These results vary based role of the informants and situation in which they observe the child (self, mother, father, teacher, etc.), the age of the child and type of behavioral/ emotional problem. This may result from differences in informants' observations as well as differences in children's behavior in different contexts. The low to modest correlations across reporters have provided a strong basis for the use of data that combines multiple informants, and/or assesses children on multiple axes that are designed to reflect and capture the variations that surface across multiple informants. A sorting of cases and non-cases in our current study could neither capture these variations nor capitalize on the different perspectives of multiple informants.

Second, a good test of criterion validity using the approach in the current study is dependent upon respondents being willing to provide a definite answer to the question being asked, and that the question being asked is fairly highly correlated with the underlying concept being assessed by the instrument. In our case, because there was no commonly known and used term for either internalizing or externalizing problems, nor for mental health or psychosocial problems more generally, we had to use a phrase that described problems related to behavior and affect, without being more specific about the nature of the problems. This is a limitation to our ability to assess criterion validity, as an unclear criterion for comparison, or a criterion open to a wide range of interpretation, would be unlikely to yield to a strong test of criterion validity.

Third, situational and cultural factors also may have influenced both child and adult responses to the criterion validity sorting process. In particular, interviewers noticed reluctance on the part of some staff to categorize children in what they perceived to be negative terms, even when children themselves reported they had the problems being asked about. In some cases, caregivers wished to place children in an intermediate category, rather than selecting one of the two response options, which interviewers interpreted as a means of politely categorizing the child as one with problems. When forced to choose between two options, those caregivers frequently close to categorize the child as a non-case.

Recommendations

The results of this study suggest that the YSR-H has adequate reliability for use with former *restavèk* children in Haiti, and that criterion validity of the instrument remains to be established. A reassessment of criterion validity that addresses the limitations of the current approach is needed. These may include a more thorough process to define in common terms the criterion of interest, a more in-depth orientation and explanation of the purpose of the sorting to reach a common understanding of the criterion terms used, and/or extra care to ensure that the sorting process is not perceived as stigmatizing. Finally, if local professionals are available and espouse an understanding of the criterion of interest based on local symptoms and expressions, rather than Western concepts learned in academic settings, the judgments of such professionals could be used a "gold standard" in future studies.

Future work on the YSR-H should also move to establish standardized scores and norms for the instrument to provide a basis for interpreting raw scores. ASEBA has a wide base of research support that establishes clinical cutoffs for subscales and syndromes in US samples; however, these may not be used as comparison or for interpretation of the scores of children in Haiti.

In the interim, the YSR-H may be used as a screener to help identify children who need attention, based on the YSR and/or YSR-H Total problems scale. The internalizing and externalizing scales may be used to explore the nature and severity of the problems affecting each child, and therefore help to tailor interventions to the child's needs and to assess impact of interventions provided within the centers where children are residing. Because there were few differences in the reliability estimates between the YSR and the YSR-H, the shorter, standardized version may be used. However, because the local items may contribute to criterion validity, we recommend that all symptom-based items be retained pending an assessment of criterion validity.

CONCLUSION

The availability of culturally-adapted, reliable and valid instruments for the assessment of mental health among survivors of restavèk in Haiti, and for survivors of child slavery and trafficking globally, is an important area for continued study and development. Without such measures, service providers and families are ill-equipped to understand the range of needs of children in their care. In many contexts where modern slavery pervades, attention to mental health in general is a luxury afforded only to the most elite. As this study indicates, children who emerge

from harsh conditions of slavery and/or trafficking are likely to exhibit behavioral and psychological symptoms that may create subsequent difficulties in their reintegration into their families and other contexts.

Equipping service providers and families with access to instruments that accurately and responsively assess children's mental health creates an opportunity for the development and enhancement of interventions designed to improve or stabilize children's mental health post-slavery. The use of such instruments to understand children's baseline levels of functioning, and to assess the impact of existing interventions or contexts on functioning will create the possibility both for redesign of ineffective interventions, and for the promotion and dissemination of interventions found to be effective in improving mental health and well-being. Ultimately, such tools hold promise for shaping the quality and impact of services provided.

Table 1: Study Sample Characteristics

Total N=67	**N (%)**
Gender	
Male	20 (30%)
Female	47 (70%)
Ages*	
9-11	16 (24%)
12-14	34 (51%)
15-17	15 (22%)
Sites	
#1. Mixed sex, residential, *restavèk* only	29 (43%)
#2. Girls only, residential, *restavèk* only	6 (9%)
#3. Mixed sex, residential	23 (34%)
#4. Mixed sex, residential	9(13%)

*Missing age data for two participants

Table 2: Scale Descriptive Statistics

	# of Items	**Mean**	**SD**	**Min**	**Max**
YSR Scales					
YSR Internalizing Problems	31	23.76	8.21	8	44
YSR Externalizing Problems	32	13.83	7.59	0	31
YSR Total Problems	105	68.90	24.42	22	133
Haiti Scales					
Haiti Symptoms Scale*	34	21.14	8.77	2	41
YSR-H Scales					
YSR-H Internalizing Problems*	48	35.45	12.81	10	63
YSR-H Externalizing Problems*	40	17.23	9.65	0	39
YSR-H Total Problems*	139	89.85	31.30	28	165

*Not all respondents have complete data. Data presented for only those with complete data.

Table 3: Internal Consistency Reliability

	Total Sample (N=67)	**Males (N=20)**	**Females (N=47)**
YSR Scales			
YSR Internalizing Problems	.753	.740	.776
YSR Externalizing Problems	.802	.777	.802
YSR Total Problems	.834	.866	.853
Haiti Scales			
Haiti Symptoms Scale*	.817	.801	.830
YSR-H Scales			
YSR-H Internalizing Problems*	.844	.830	.871
YSR-H Externalizing Problems*	.846	.816	.844
YSR-H Total Problems*	.882	.902	.897

*Not all respondents have complete data. Data presented for only those with complete data.

Table 4: Test-Retest Comparison

	Mean (sd) First Interview	**Mean (sd) Repeat Interview**	**Correlation**
YSR Scales			
YSR Internalizing Problems	24.15 (9.29)	19.67 (9.76)	.835
YSR Externalizing Problems	13.33 (6.78)	11.78 (8.93)	.628
YSR Total Problems	70.48(23.86)	53.19 (27.40)	.834
Haiti Scales			
Haiti Symptom Scale	20.52 (8.38)	20.15 (10.15)	.689
YSR-H Scales			
YSR-H Internalizing Problems	35.22(14.10)	29.74(15.60)	.827
YSR-H Externalizing Problems	17.15 (8.84)	16.00 (12.13)	.617
YSR-H Total Problems	91.70(30.51)	78.85 (39.47)	.834

REFERENCES

Achenbach, Thomas M., Stephanie H. McConaughy, and Catherine T. Howell. 1987. "Child/Adolescent Behavioral and Emotional Problems: Implications of Cross-Informant Correlations for Situational Specificity." *Psychological Bulletin* 101 (2): 213-232.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont Department of Psychiatry.

AMHR (Applied Mental Health Research Group). 2011. *Design, Implementation, Monitoring, and Evaluation of Cross-Cultural HIV-Related Mental Health and Psychosocial Assistance Programs: A User's Manual for Researchers and Program Implementers Module 2: Developing Quantitative Tools*. http://www.jhsph.edu/research/centers-and-institutes/research-to-prevention/dime-manual/DIME-MODULE-2.pdf.

Bales, Kevin. 1999. *Disposable People: New slavery in the Global Economy*. Berkeley: University of California Press.

Bass, J. K., R. W. Ryder, M.-C. Lammers, T. N. Mukaba, and P. A. Bolton. 2008. Post-Partum Depression in Kinshasa, Democratic Republic of Congo: Validation of a Concept Using a Mixed-Methods Cross-Cultural Approach. *Tropical Medicine and International Health* 13 (12): 1534-1542.

Berube, R. L. and T. M. Achenbach. 2012. *Bibliography of Published Studies Using ASEBA Instruments: 2007 Edition*. Burlington, VT: University of Vermont, Research Center for Children, Youth and Families.

Betancourt, Theresa S., Jessica Agnew-Blais, Stephen E. Gilman, David R. Williams, and Heidi Ellis. 2010. "Past Horrors, Present Struggles: The Role of Stigma in the Association Between War Experiences and Psychosocial Adjustment Among Former Child Soldiers in Sierra Leone." *Social Science & Medicine* 70:17-26.

Betancourt, Theresa S., Judith Bass, Ivelina Borisova, Richard Neugebauer, Liesbeth Speelman, Grace Onyango, and Paul Bolton. 2009. "Assessing Local Instrument Reliability and Validity: A Field-Based Example from Northern Uganda." *Social Psychiatry and Psychiatric Epidemiology* 44: 685-692.

Betancourt, Theresa, Ivelina Borisova, Julia Rubin-Smith, Tara Gingerich, Timothy Williams, and Jessica Agnew-Blais. 2008. "Psychosocial Adjustment and Social Reintegration of Children Associated with Armed Forces and Armed Groups: The State of the Field and Future Directions." Austin: Psychology Beyond Borders. Accessed March 14, 2011. http://psychologybeyondborders.com/PublicReports.aspx.

Bolton, Paul. 2001. "Cross-Cultural Validity and Reliability Testing of a Standard Psychiatric Assessment Instrument without a Gold Standard." *Journal of Nervous and Mental Disease* 189 (4): 238–242.

Cooper, Allen, Pablo Diego-Rosell, and Christelle Gogue. 2012. "Child Labor in Domestic Service (Restavèks) in Port-au-Prince, Haiti." Accessed February 12, 2013. http://www.dol.gov/ilab/reports/pdf/2012RestavekHaiti.pdf.

Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests. *Psychometrika* 16: 297-334.

Flowers, R. Barri. 2001. "The Sex Trade Industry's Worldwide Exploitation of Children." *Annals of the American Academy of Political and Social Science* 575: 147-157.

Goldstein, Jill M. and John C. Simpson. 1995. "Validity: Definitions and Applications to Psychiatric Research." In *Textbook in Psychiatric Epidemiology*, edited by Ming T. Tsuang and Mauricio Tohen, 229-242. New York: Wiley

Hall, Brian J., Eve Puffer, Laura K. Murray, Abdulkadir Ismael, Judith K. Bass, Amanda Sim, and Paul A. Bolton. 2014. "The Importance of Establishing Reliability and Validity of Assessment Instruments for Mental Health Problems: an Example from Somali Children and Adolescents Living in Three Refugee Camps in Ethiopia. *Psychological Injury and Law* 7: 153-164.

Hobfoll, Stevan E., Daphna Canetti, Brian J. Hall, Danny Brom, Patrick A. Palmieri, Robert J. Johnson, Ruth Pat-Horenczyk, and Sandro Galea. 2011. "Are Community Studies of Psychological Trauma's Impact Accurate? A Study Among Jews and Palestinians." *Psychological Assessment* 23(3): 599–605.

Hollifield, Michael, Teddy D. Warner, Nityamo Lian, Barry Krakow, Janis H. Jenkins, James Kesler, Jayne Sevenson, and Joseph Westermeyer. 2002. "Measuring Trauma and Health Status in Refugees: A Critical Review." *Journal of the American Medical Association* 288(5): 611–621.

Ivanova, M. Y., T. M. Achenbach, L. A. Rescorla, L. Dumenci, F. Almqvist, N. Bilenberg, H. Bird, A. G. Broberg, A. Dobrean, M. Döpfner, N. Erol, M. Forns, H. Hannesdottir, Y. Kanbayashi, M. C. Lambert, P. Leung, A. Minaei, M. S. Mulatu, T. Novik, K. J. Oh, A. Roussos, M. Sawyer, Z. Simsek, H. C. Steinhausen, S. Weintraub, C. Winkler Metzke, T. Wolanczyk, N. Zilber, R. Zukauskiene, and F. C. Verhulst. 2007. "The Generalizability of the Youth Self-Report Syndrome Structure in 23 Societies." *Journal of Consulting and Clinical Psychology* 75 (5): 729-728.

Kennedy, Cara L. 2014. "Toward Effective Intervention for Haiti's Former Child Slaves." *Human Rights Quarterly* 36 (4): 756-778.

Mollica, Richard F., B. Lopes Cardozo, H. J. Osofsky, B. Raphael, A. Ager, and P. Salama. 2004. "Mental health in Complex Emergencies." *The Lancet* 364(9450): 2058–2067.

Moylan, Carrie A., Todd I. Herrenkohl, Cindy Sousa, Emiko A. Tajima, Roy C. Herrenkohl, and M. Jean Russo. 2010. "The effects of child abuse and exposure to domestic violence on adolescent internalizing and externalizing behavior problems." *Journal of Family Violence* 25: 53-63.

Pierre, Yves-Francois, Glenn R. Smucker, and Jean-Francois Tardieu. 2009. "Lost Childhoods in Haiti: Quantifying Child Trafficking, Restavèks, and Victims of Violence." Accessed March 10, 2011. http://b.3cdn.net/padf/d0b483e2777248284b_mlbrz3lll.pdf.

Raymond, Janice G. and Donna M. Hughes. 2001. "Sex Trafficking of Women in the United States: International and Domestic Trends." Accessed March 14, 2011. http://www.ncjrs.gov/pdffiles1/nij/grants/187774.pdf.

Renk, K., and V. Phares. 2003. Cross-Informant Ratings of Social Competence in Children and Adolescents. *Clinical Psychology Review* 24(2): 239-254.

Shrout, Patrick E. 1998. "Measurement Reliability and Agreement in Psychiatry." *Statistical Methods in Medical Research* 7: 301-317.