

**Assessment of Urban Street Children and Children living in Government  
Institutions in Georgia:  
Development and Testing of a Locally-Adapted Psychosocial Assessment  
Instrument.**

Conducted by  
Save the Children/Georgia  
and  
The Applied Mental Health Research Group  
Boston University  
School of Public Health/Center for International Health and Development  
Johns Hopkins Bloomberg School of Public Health

Authors  
Paul Bolton  
Laura Murray  
Sara Kippen  
Judith Bass

**November, 2007**

Funding for this project is provided by the Displaced Children and Orphans Fund,  
United States Agency for International Development.



**JOHNS HOPKINS  
BLOOMBERG  
SCHOOL of PUBLIC HEALTH**

**TABLE OF CONTENTS**

**EXECUTIVE SUMMARY .....3**

    PROJECT OBJECTIVES..... 3

    METHODS..... 3

    RESULTS..... 4

    DISCUSSION..... 5

    CONCLUSIONS..... 6

    RECOMMENDATIONS..... 7

**ACKNOWLEDGEMENTS .....8**

**INTRODUCTION .....9**

**BACKGROUND .....9**

**OBJECTIVES .....10**

**METHODS.....10**

    STUDY SITES ..... 10

    DEVELOPING THE DRAFT INSTRUMENT FOR FIELD TESTING ..... 11

*Instrument Selection* ..... 11

*Instrument Adaptation* ..... 13

    PILOT STUDY ..... 14

    RELIABILITY AND VALIDITY STUDY ..... 15

    ANALYSIS ..... 16

**RESULTS .....18**

    PILOT STUDY ..... 18

    RELIABILITY AND VALIDITY STUDY ..... 18

*Sample Characteristics*..... 18

*Instrument Reliability*..... 19

*Test-retest/Inter-rater Reliability*..... 19

*Internal Consistency Reliability*..... 21

*Criterion Validity*..... 22

*Epidemiological Analysis of Study Findings*..... 24

**DISCUSSION .....28**

**CONCLUSIONS .....30**

**RECOMMENDATIONS .....31**

**APPENDIX A – VERSION OF THE YSR INSTRUMENT USED IN VALIDITY STUDY (YSR-G) .....32**

**APPENDIX B: STUDY EXPLANATION READ TO INTERVIEWEES PRIOR TO INTERVIEW.....41**

**APPENDIX C: EXPLANATION OF RELIABILITY AND VALIDITY CONCEPTS .....42**

## **EXECUTIVE SUMMARY**

### **Project Objectives**

This report describes the development and testing of an instrument to assess psychosocial problems among urban street children and children living in government institutions in Georgia. Development consisted of generating a draft instrument which reflects the psychosocial problems that emerged in previous qualitative studies among the same target populations. Testing consisted of assessing the instrument's local acceptability, clarity, validity and reliability among the two target populations. The study is part of a series of field-based activities to inform the design, monitoring, and evaluation (DME) of the Rebuilding Lives Project (RLP) being implemented by Save the Children Georgia (SCG).

The objectives of the work described here are:

1. To develop a draft quantitative assessment instrument based on the psychosocial problems that emerged in previous qualitative studies of the two target populations.
2. To test the acceptability, clarity, validity and reliability of this instrument among these same children and finalize the instrument based on these results.
3. To build the capacity of SCF in this type of applied research.

### **Methods**

The first step was to develop a draft instrument that reflected the wide range of problems found through previous qualitative studies at these sites, the RLP program objectives, and the need for an instrument to help assess children served by government institutions. We identified an existing broad-based child measure which matched these criteria - The Youth Self Report (YSR) which has been used and studied in > 65 societies. The YSR was first adapted to more closely match the problems identified by our target populations (street children served by the RLP-supported Centers and children served by government institutions) in previous qualitative studies by adding items based on frequently mentioned issues in the qualitative studies that were not already captured in the YSR. No items were removed, in order to enable future comparisons with data from other populations in other countries. The adapted YSR was translated into Georgian with an emphasis on using the same vocabulary as that found in the qualitative data.

In addition to the adapted instrument, JHU/BU and SCG staff and the interviewers drafted a description of the study to be read to children prior to administering the instrument and prior to asking whether they agreed to be interviewed

The next step was a pilot study of the adapted YSR (now referred to as the YSR-G) in order to detect any problems with the interview procedure, the project description form and the instrument (including data entry) from the point of view of both the interviewers and the interviewees, to determine whether the instrument was acceptable and understandable to our target populations, and to give the interviewers practice in interviewing. Children were

interviewed at nine sites which included RLP-supported Centers and government institutions, as well as a few children found through a Mobile Unit. Based on feedback from interviewers and interviews the instrument and project description form were adjusted as necessary. The resulting instrument and description form were then tested further in the reliability and validity study.

The purpose of the reliability and validity testing was to determine whether the YSR-G could accurately determine if a child had significant emotional and/or behavioral problems. Testing consisted of assessing the YSR-G's internal consistency, predictive validity, inter-rater reliability and test-retest reliability). Interviewees came from the same sites as the pilot study. Twenty percent of the children were reinterviewed 1-3 days after their first interview by a different interviewer in order to assess the YSR-G's test retest and interrater reliability. Internal consistency reliability was also assessed using the Cronbach's alpha measure. Criterion validity was assessed with the help of clinical psychologists employed by the Centers and government institutions and who work with the children on an ongoing basis. These psychologists assessed the children served by these organizations to produce a list of those thought to have significant emotional or behavioral problems (Cases) and those thought not to have them (Non-cases). Criterion validity was assessed by comparing the scores of Cases and Non-cases on the various YSR-G scales

## **Results**

During the pilot study many children complained about being asked the YSR questions on suicide, drug use, and sex. Based on the feedback from the interviewers, and on several informal meetings with these children by the research team, it became clear that the children felt personally affronted and complained to their friends who then refused to be interviewed. The problem was addressed not by changing the instrument but by adding further explanation on this point to the Study Explanation document read before each interview. This document explained that the questions were asked of all children being interviewed, and this was acceptable to the children concerned. Once this change was made the problem did not recur during subsequent interviews. Otherwise, only minor changes were required based on the pilot study results.

For the reliability and validity study, a total of 386 children were interviewed. Test-retest and interrater reliability results for the problem-based scales range from acceptable (for the Georgia problem scale) to good (for the internalizing, externalizing, and total problem scales). In contrast, the prosocial scales all performed poorly. Similarly, internal consistency reliability was very good for the Internalizing, Externalizing and Total Problems Scale, but poor for the prosocial and Georgia scales.

On the basis of the psychologist's ratings, 132 of the 386 children were classified as likely 'cases' and likely 'non-cases' in order to assess criterion validity. Differences in scores between "cases" and "non-cases" were statistically significant on all scales, except for the Georgia prosocial scale. However, the magnitude of the mean differences between cases and non-cases were small for all scales except for the total problem scale. As with the reliability testing, the problem scales (total problems, internalizing and externalizing problems) performed best.

Overall, the results suggest that the total problems scale is the best scale for distinguishing between “cases” and “non-cases” in terms of magnitude of differences and statistical significance.

Some additional epidemiological analyses were done on the study population. Total problem scores for children who are doing well (few problems) fall in the 0-20 range and represent only 2.6% of our sample. The scores of most children fell into the 20-100 range, with a mean score of 60.5, reflecting a substantial degree of difficulty across multiple problem categories. All subscale scores are very similar regardless of whether children are at the Centers or the government institutions, which suggests that neither group is better or worse off than the other, based on this measure. Prosocial scale scores were very high, and approached the maximum scores for each scale – YSR, Georgia and total prosocial scales.

## **Discussion**

Pilot Testing among our sample proved essential to improving our interviewing procedures and adjusting the draft YSR-G to make it acceptable to our target population. Without pilot testing it was clear that the original instrument and interviewing procedures were potentially distressing to the interviewees, which would have affected ability to use the instrument effectively in the future.

Reliability and validity testing showed a distinct pattern in the performance of the YSR-G. On most measures of reliability or validity, we found that the problem-based scales (the YSR internalizing and externalizing scales, and particularly the total problem scale) have solid psychometric properties in this population. However, the prosocial scales performed poorly. These findings were consistent for the entire sample, and for the gender-specific analyses.

The internalizing and externalizing scales have good internal consistency (Cronbach’s alpha >.8), as did the total problem scale (Cronbach’s alpha>.9). The YSR Prosocial scale and both Georgia-specific scales (problem-based and prosocial) showed poorer internal consistency, which is likely due in part to the mixed nature of the concepts assessed by these scales and the small number of items in the Georgia-specific scales. Combined test-retest/inter-rater reliability was good for the internalizing and externalizing problem scales and the total problem scales, marginal for the Georgia problem scales, and poor for all the prosocial scales.

Tests of criterion validity found that the problem scores consistently matched the criterion chosen for this study: local psychologists’ evaluations. The children identified by psychologists as having significant emotional and/or behavioral problems (“cases”) showed significantly higher levels of symptoms on the same internalizing and externalizing YSR scales and on the total problem scale than those children identified as having few emotional and/or behavioral problems (“non-cases”). As in the other analyses, the prosocial scales did not perform well, with little or not significant differences in scale scores based on “Caseness.”

In addition to the main purpose of the study - developing an acceptable, reliable, and valid instrument - we also conducted preliminary epidemiological analyses of the study data for the total sample and separately for those children from the Centers and government institutions. We

found that the total sample and the Center and institution subsample scores were elevated on all the problem scales, suggesting difficulty across the range of problem categories assessed by these scales: anxiety, depression, withdrawal, aggression, social and cognitive problems and problems with family. Most of the children in our sample had been in contact with the Centers and government institutions for some time (a year or more). They would therefore be expected to have already benefitted from the interventions these organizations provide, and therefore to have experienced improvement. This suggests that children living on the street or at risk, and who are not receiving services, are likely to have substantially higher problem scores than those of the children in our study.

In comparing the scores of children at the Centers and in the government institutions, we found little difference in the severity or degree of problems or their prosocial behaviors. There are several possible explanations. One may be that both groups may be similar upon enrolment and the services of both types of organizations may be equally effective. Another explanation may be that one group of children may be worse off at entry but do better once they arrive and receive services, suggesting that one type of organization is receiving more severely affected children but is able to produce more substantial improvement. Without baseline data at entry, it is not possible to know which explanation is correct. At this time the data can only suggest that the pattern of problems and their severity among the two populations are similar.

Scores on all the prosocial scales were very high. Given the poor psychometric properties of the prosocial scales described above, it may be that children are not answering these questions accurately. One possibility is that social desirability effects (a desire to look good for the interviewer) may be inflating their scores. The high levels of these scores and their likely inaccuracy (based on the psychometrics) suggest that programs will have difficulty in substantially affecting these scores through their interventions, and that any changes in the scores will not accurately reflect program impact.

## **Conclusions**

Based on our pilot study, the YSR-G and associated interview materials are acceptable for use among street children served by the RLP Centers and children served by the types of government institutions included in this study.

Overall, the problem scales of the YSR-G shows solid psychometric properties and therefore we believe them to be suitable for use among this type of child population in Georgia.

These results suggest many options for use of the YSR-G within Georgia. The YSR-G could be used as a screener to help identify children who need attention, based on the Total Problems Scale. The Internalizing and Externalizing scales can 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. The availability of trained mental health professionals in Georgia makes this use of the more specific scales (See Table 3) appropriate for understanding individual problems, as well as in formulating treatment planning. The problem scales in the YSR-G can also be used to evaluate the RLP program. They can also be used to assess the need for, and impact of, other

new or existing programs within the Centers or Institutions that address problems measured by the YSR-G.

The prosocial scales showed poor psychometric properties and were generally high, even approaching the maximum scores for each scale. Poor psychometric properties suggest that these scales are not accurate. High mean scores among this population suggest there is a ceiling effect which will cause problems when attempting to use these scores to assess program impact; it will be difficult to improve on scores that are already high. Therefore, our results do not suggest that the prosocial scores are useful either for screening children into programs or assessing their progress. However, since this is the first time this instrument has been tested, and the study was done among a population of children with prolonged exposure to Centers or government institutions (who may have experienced marked improvement in their prosocial skills due to exposure to these programs) we are not yet ready to advocate removing the prosocial items from the instrument. Instead, they should be retained for the time being and their performance reassessed as part of analyses of future data.

Informal reports (data still pending) suggest that the children in our study sample represent a substantial proportion of the children served by the study sites in which we worked. Average problem scores were high. The distribution of scores shows few children with low scores and significant numbers with high scores. This suggests that most children in our sample have a wide range of significant psychosocial problems. Since most of our sample has been receiving services for a year or more, we suspect that children who are not receiving services will have even higher scores.

In conclusion, previous qualitative studies have resulted in the selection and adaptation of an existing instrument which has, in the study reported here, proved to provide an accurate assessment of the psychosocial problems of this population.

## **Recommendations**

If feasible, Save the Children Georgia should use the YSR-G to evaluate the impact of the RLP program on children who are new or relatively new to the program. Whether or not this is feasible at this late stage in the program (which will end in Sept, 2008) should be determined as soon as possible. BU/JHU faculty can provide technical assistance in carrying out these assessments.

The YSR-G should be used by other organizations working with these populations (and other child populations at risk), including the Georgian government, other NGOs and private groups. The goals of these assessments should include:

- a) Assessing the nature and severity of needs (by using the instrument as a survey tool).
- b) Using this information to target resources and design appropriate interventions.
- c) Assessing the impact of these interventions

In conducting a-c, there should be a focus on building local capacity in program design, monitoring, and evaluation.

When the YSR-G is used in the future, analysis of the resulting data should include further characterization of the accuracy of the instrument, particularly with regard to the performance of the prosocial scales.

The methods used in this part of the RLP project – qualitative methods resulting in instrument selection and adaptation, followed by instrument piloting and validity testing (as reported here) - should be repeated in other contexts. As with the RLP project, these methods can be used to improve need and impact assessments for other populations, both children and adults, and to assess both psychosocial and other problems. The methods are particularly useful in situations where need has not been well characterized and where the impact of interventions has not been demonstrated.

### **ACKNOWLEDGEMENTS**

This project was funded by the Displaced Children and Orphans Fund (DCOF) at USAID. The authors would like to thank Save the Children Georgia for their collaboration, and all the interviewers who worked on this study. Finally, a special acknowledgement to all the site directors, psychologists, staff and children who welcomed us and helped immensely in the conduct of this study.



## **INTRODUCTION**

This report describes the development and testing of an instrument to assess psychosocial problems among urban street children and children living in government institutions in Georgia. Development consisted of generating a draft instrument which reflects the psychosocial problems that emerged in previous qualitative studies among the same target populations. Testing consisted of assessing the instrument's local acceptability, clarity, validity and reliability among the two target populations.

The report describes the background to the activities described here, and places them in the wider context of an ongoing technical support collaboration between Save the Children Georgia, (SCG) Boston University, (BU) and Johns Hopkins University (JHU). The report also describes the methods used to develop and test the instrument and the results of these tests. Conclusions based on the results are also included, as well as specific recommendations for future activities. Finally, the report ends with a series of appendices containing the final instrument and associated interview materials. There is also an appendix with explanation of some of the technical terms used in the report referring to important elements of reliability and validity.

## **BACKGROUND**

These activities described here (instrument development and testing) form part of a planned series of field-based activities to inform the design, monitoring, and evaluation (DME) of the Rebuilding Lives Project (RLP) being implemented by Save the Children Georgia (SCG). The purpose of the RLP is to assist urban street children in the capital city of Tbilisi and nearby cities. These DME activities to support the RLP are being conducted by SCG in collaboration with faculty from Boston and Johns Hopkins Universities. The collaboration is supported by the Displaced Children and Orphans Fund (DCOF) at USAID.

The DME process consists of the following stages:

1. Qualitative study of the problems affecting the target population (ie, locally important activities and abilities).
2. Development of a locally appropriate quantitative instrument (questionnaire) to assess the major psychosocial problems emerging from the qualitative study.
3. Evaluation of the acceptability, clarity, validity and reliability of the instrument among the target population, with subsequent revision of the instrument based on the results.
4. Use of the final version of the instrument to conduct baseline assessments among children recruited to the program (in this case, street children recruited into the RLP program).
5. Repeat use of the instrument after participation in the program, to assess program impact.

This report describes the process and results of stages 2 and 3, described above, and includes recommendations for future activities. Reports on two previous qualitative studies which constitute stage 1 (and upon which the current work has been based) are available at <http://dec.usaid.gov/> ('Causes of Children Living on the Street in Urban Georgia: A Qualitative Assessment' & 'Problems of Children in Urban Georgia: A Qualitative Assessment of Centers

and Orphanages'). Details on the background to SCG's work in Georgia, the overall DME project and methodology; its rationale, theoretical basis, and explanations of the technical concepts, are described in detail in the qualitative study reports. Therefore much of this information has been omitted from this report. The contents here are limited to a description of the methods used to develop the quantitative instrument and its testing in the field.

When this project began the focus was solely on the target population of the RLP program: street children living in Tbilisi and nearby cities. In recent years the Georgian government has developed plans to remove children from orphanages and other government institutions, which was thought likely to result in at least some children living on the street. After the original qualitative study SCG, USAID, and BU/JHU faculty discussed expanding the focus of the instrument development and testing phases of the project. It was agreed to develop a single instrument that could assess the problems not only of street children but of currently institutionalized children as well. The first qualitative study had focused on the causes of children being on the street, but the results suggested that children faced many other significant problems which should also be assessed. Because of this finding, and the addition of a second target population to the instrument development and testing phases, a second qualitative study was conducted that focused on the problems (rather than the causes) of children both on the street and in institutions. The resulting draft instrument, whose development and testing is described here, was mostly based on the findings of the second qualitative study.

## **OBJECTIVES**

The objectives of the work described here are:

1. To develop a draft quantitative assessment instrument based on the psychosocial problems that emerged in the qualitative studies of the two target populations.
4. To test the acceptability, clarity, validity and reliability of this instrument among these same children and finalize the instrument based on these results.
5. To build the capacity of SCF in this type of applied research.

## **METHODS**

### **Study Sites**

This study took place at 9 institutions in and around the cities of Tbilisi, Gori, and Rustavi. The study sample also included 8 children who were located on the streets of Tbilisi by the Save the Children Mobile Team but were otherwise not known to be served by any Center or institution. The sites are listed and briefly described below:

**Table 1: Study Sites**

Site Name	Site Description
1. Tsisartkela/Rainbow (Tbilisi)	Day center for at-risk children (children from poor families, including IDP families), children on the street and children of the street
2. Beghurebi/Sparrows (Tbilisi)	Night shelter and day center for children of the street, children on the street and at-risk children
3. Sapovnela (Rustavi)	Day center for street and at-risk children
4. Biliki (Gori)	Day center for at-risk and street children
5. Skhivi (Gori)	Day center for street and at-risk children
6. Dighomi Orphanage (Tbilisi)	Institution for children who are orphans, from poor families, or various vulnerable families
7. Gldani Social Adaptation Center for Children (Tbilisi)	Shelter for children in emergencies and crises: children victims of domestic violence, abuse, neglect and exploitation, including street children and juvenile offenders
8. Momavlis Sakhli/ House of the Future (Tbilisi)	Institution for children from poor, homeless and socially vulnerable families, including street children
9. Kojori Boarding School (Kojori)	Boarding school with an emphasis on sports that serves orphans and children from poor families
10. Mobile Team sites (Tbilisi)	Mobile outreach team serving street children on the streets of Tbilisi.

Sites 1-5 are RLP-supported Centers that serve children who spend time on the streets or who are considered at-risk. Sites 6-8 are government run institutions that serve street children as well as other children whose families are poor and cannot support them. Kojori is a boarding school, specializing in sports, and not necessarily serving children who spend time on the street. However, it does include some children with similar problems to the street children and (like the Centers and government institutions) has its own clinical psychologist. Twenty-three of the children interviewed at Gori actually belong to the Gori Club. This is an organization of children who are not on the streets or at risk of being so.

Data from the Gori Club children were included in the reliability analyses, since it is not critical for these tests that all children come from the target populations. However, this is not true for validity testing and these children were therefore excluded from the validity analyses (See Results).

## **Developing the Draft Instrument for Field Testing**

### *Instrument Selection*

In developing a quantitative instrument the major issue 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. Therefore, the main consideration is whether such an instrument already exists. In reviewing existing instruments, we based our decision on whether each one was locally appropriate based on: 1) Selecting an instrument that reflected the important psychosocial problems that emerged from the qualitative studies of both target populations; and 2) Selecting an instrument that includes (but was not be limited to) assessment of those psychosocial issues that the RLP program is trying to address.

The qualitative studies showed that children in SCG Centers and government institutions experience a wide range of psychosocial problems. These problems can be categorized as emotional problems (also referred to as internalizing symptoms) such as sadness, crying, nervousness, and feeling lonely; behavioral problems (or externalizing symptoms) such as aggression, smelling glue, fighting, or being rude; study problems; and relationship issues such as staying alone or not getting along with others. The qualitative results suggested that no 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 broadly-based measure that spanned the range of these problems, rather than one that focuses on a particular disorder or group of disorders, in order to assess the wide range of problems and symptoms reported by children.

The Youth Self Report (YSR) was ultimately selected as an existing instrument developed in the the United States but suitable for use among our target populations. The YSR assesses a broad range of psychosocial problems that closely matches those emerging from the qualitative studies. It also includes most of the psychosocial problems that the RLP expects to address. The YSR is part of a set of instruments developed by the Achenbach System of Empirically Based Assessment (ASEBA)<sup>TM</sup>. ASEBA instruments were designed to obtain standardized data on a broad range of problems from multiple sources (e.g., parents, teachers, and youth). ASEBA instruments have been translated into more than 75 languages, and studied in over 65 societies (Berube & Achenbach, 2007). An entire supplemental manual now exists on the multi-cultural use of ASEBA instruments, including variations in norms across different countries (Achenbach & Rescorla, 2007). The YSR is completed by youths themselves to describe their own functioning and problems. The first 2 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 2 pages contain 112 symptom and behavior specific items with responses rated on a 0-2 scale (0=not true, 1=somewhat or sometimes true, and 2 very true or often true).

The table below provides examples of RLP psychosocial objectives, and how the YSR can be used to evaluate these objectives.

**Table 2: Comparison of RLP psychosocial objectives with YRS content domains.**

RLP psychosocial objectives	⇒	YSR domain
Increased interpersonal relations/communication skills;	⇒	Most items would come from social problems, although some also perhaps from withdrawn, or anxious/depressed items
Increased self-esteem;	⇒	Items from different domains that capture the specific operationalization of this concept such as "I feel no one loves

		me”, I feel that others are out to get me”, I feel worthless or inferior”, “I am not liked by other kids”... etc.
Increased quality relationships with parents;	⇒	Items from the existing YSR relating to relationships with family.
Reduced/lowered destructive behavior – glue sniffing/ smoking;	⇒	Delinquent behaviors mainly, such as “I smoke, chew or sniff tobacco”, “I use drugs for nonmedical purposes”
Reduced/lowered aggression	⇒	Aggressive behavior scale on YSR
Increased constructive behavior – study	⇒	Thought problems; attention problems

### *Instrument Adaptation*

Once the YSR was chosen, BU/JHU faculty and SCG discussed what adaptations might be useful in order to improve the content of the instrument. We decided not to use the demographic and competency questions (those on the first two pages of the standard instrument) reflecting a common practice in cross-cultural studies with these measures (personal communication with Dr. Achenbach). The items in this section include questions on the different sports and clubs youth participate in, jobs/chores youth have, performance in school subjects, and relationships with siblings. Many of these items were identified as inappropriate for our target populations.

None of the remaining 112 symptom and behavior items were removed in order to maintain the empirical base of the instrument, and retain future comparability with results from children in other countries. However, several additional items were added to the YSR based on frequently mentioned issues in the qualitative studies that were not already well captured in the YSR. For example, many children mentioned problems related to family such as “not feeling loved or warmth” or “being forced to beg”, and problems directly related to being street children such as “feeling oppressed because of situation”. Additions included four positive family items and six negative family items, interspersed among the original items. Each added item is identified by a letter next to the number; e.g, 17a (see Appendix A). This initial draft YSR-G (Youth Self Report- Georgia) was then translated by a local SCG staff member who is also a clinical psychologist.

BU/JHU staff then returned to Georgia to assist SCG with further development of the YSR-G instrument. BU/JHU and SCG staff met with 20 local interviewers and two translators to review the initial draft. These interviewers included 13 workers who had been interviewers on at least one of the two previous qualitative studies, and therefore were familiar with the qualitative data. Most of the interviewers have a social science background and have had experience working with at-risk children. Some of these interviewers work at various sites involved in the study, and some are employees of the Ministry of Education. Hence, the group of interviewers were well qualified to comment on the appropriateness of the initial draft instrument for use among our target populations.

Under the direction of the BU/JHU and SCG staff, the interviewers reviewed each translated item in the draft instrument. The interviewers each had a copy of the qualitative data and compared the translation of each item with the wording used by the qualitative study respondents. Since the draft instrument was selected and adapted to match the qualitative data, a description was found in the qualitative data for most items. Where the draft translation and the

terminology from the qualitative study were different, the translation was changed to reflect the vocabulary from the qualitative study. For items in the YSR-G that were not reflected in the qualitative study, interviewers used their own knowledge and experience to decide whether the language was appropriate and would be understood by the children.

This process of review took place over 2 days. This included a discussion among the BU/JHU and SCG staff and the interviewers regarding whether there were important psychosocial issues described in the qualitative data but yet represented in the draft instrument. As a result, five additional questions were added to the YSR-G at the end of the main section of the instrument and numbered as B112a-e.

In previous studies conducted by BU/JHU faculty in other parts of the world, part of the validity testing procedure had been to compare responses to individual items with the overall opinion of the child and caregivers as to whether or not they thought they had a problem. Children who stated that they did have psychosocial problems should (if the instrument is valid) report more symptoms and greater severity of symptoms compared with children who stated they did not have psychosocial problems. In these previous studies we identified broad cover terms or phrases in the qualitative study that described a state of having a psychosocial problem. These terms were then used in the identification as to whether the child had a psychosocial problem or not.

In the qualitative studies in Georgia we did not identify any suitable general cover terms. Nor could the interviewers agree on one. Therefore, we decided to include questions asking children whether they had either emotional or behavioral problems, using the Georgian translation of these concepts provided by the interviewers. We included these questions in the draft questionnaire to test in the pilot study whether these questions were understood by the child and how they were understood, as a way of testing whether these questions could be used to test validity.

Following the translation and review activities, the draft YSR-G was ready for piloting in the field. The complete questionnaire consisted of three sections: A brief series of demographic questions (Section A), the main body of the instrument that consisted of the adapted YSR (Section B), and specific questions about whether the child had emotional or behavioral problems as described immediately above (Section C). (See Appendix A for a copy of the finalized YSR-G.)

In addition, the study team (JHU/BU and SCG staff and the interviewers) drafted a description of the study to be read to children prior to administering the instrument and prior to asking whether they agreed to be interviewed. (See Appendix B for a copy of the project description).

## **Pilot Study**

A pilot study of the YSR-G was then conducted. The objectives of the pilot study were to detect any problems with the interview procedure, the project description form and the instrument (including data entry) from the point of view of both the interviewers and the interviewees, to

determine whether the instrument was acceptable and understandable to our target populations, and to give the interviewers practice in interviewing.

On the first day of the pilot study each interviewer interviewed at least one child from across the study sites, using the draft instrument. The interviewing process was implemented as it would be during a normal interview. Once this was completed, the interviewer then asked additional questions about what the interviewee liked and did not like about the interview process and whether they had difficulty understanding any questions. At the end of the first day, the interviewers returned to the training site to review the experience with the research team and each other, including a review of the reactions of the interviewees and their responses to the additional questions. Any problems that emerged from the pilot interviews were discussed. In cases where the same problem recurred in multiple interviews, likely solutions were discussed and implemented. In cases where a problem occurred only once, there was discussion as to whether the problem was likely to recur again. In cases where this was thought likely, solutions were also discussed and implemented by the group.

The following day interviewers returned to the study sites with the revised YSR-G and interview procedures. Each interviewer again interviewed at least one child, and returned to the training site at the end of the day to discuss the experience and make any additional changes, as per the process described above for day 1. The process was repeated on a third day, at which time it was decided that no new changes were required.

### **Reliability and Validity Study**

Following the pilot study interviewers returned to the study sites to commence interviewing children for the reliability and validity study, using the procedures and instrument finalized at the end of the pilot study. As with the pilot study, these interviews were conducted at all the study sites as well as street children attended by the mobile clinic.

The purpose of the reliability and validity study was to determine if the YSR-G could accurately assess the presence and severity of significant emotional and/or behavioral problems. Reliability and validity testing included assessment of the following instrument characteristics<sup>1</sup>:

1. Combined test retest and interrater reliability
2. Internal consistency reliability
3. Criterion validity

Evaluation of test-retest reliability was done by re-interviewing approximately 20 percent of the children 1-3 days after their first interview. For each child who was re-interviewed the interviewers on the first and second occasions were different. Therefore, comparisons of the results were a measure of a combination of test retest and interrater reliability.

The main focus of the validity testing was to explore criterion validity. The criterion on which the instrument was compared was whether or not the child being interviewed had a significant psychosocial problem as assessed by Georgian psychologists currently working with the children. Typical training for a Psychologist profession in Georgia includes 4 years of University

---

<sup>1</sup> Brief explanations of each of these parameters are provided in Appendix C.

to obtain a BA and 2 years to obtain a MA with special study in Psychology. The psychologists who worked on this study were those permanently attached to the Centers and institutions, and who therefore had opportunity to assess the children on an ongoing basis.

The psychologists working at the study sites were asked to provide lists of children they considered to have "significant problems" and lists of children that have "few or no problems". The BU faculty, a trained clinical psychologist with extensive child experience (LM), met with the psychologists and reviewed with them their assessments of each child. During these meetings, the local psychologists were asked to describe the problems each child had and rate the child on a scale of 0-10 (0=no problems at all, 10=very significant behavioral and emotional problems). In their assessments the local psychologists were instructed to focus on emotional and behavioral problems rather than medical problems or mental retardation. For example, children listed as having "significant problems" because they had cerebral palsy or were severely mentally retarded but did not present with emotional or behavioral symptoms were removed from the lists because they had problems not assessed by the YSR-G. In this way, two complete lists of children were generated: One of children with "significant emotional and/or behavioral problems," defined as those who the psychologists rated between 7 and 10. These were referred to as "CASES". The second list consisted of children with "no problems," defined as those who the psychologists rated between 0-1. these were referred to as "NON-CASES". The meetings with the local psychologists occurred while the validity study interviews were being conducted and interviewers were kept unaware of the results of the psychologists' assessments.

At the end of the instrument were questions asking respondents if they felt they had a behavioral problem (and what they understood by that term). The purpose of this question was to provide an alternative measure of criterion validity. The intent was to compare opinions of the child themselves as to whether or not they had a problem with the results of main instrument.

## **Analysis**

To ensure the accuracy of data recording, all data from the validity study was double entered, with all subsequent data analysis conducted using Stata statistical software. Analysis of validity and reliability included measurement of internal consistency reliability using Cronbach's alpha and of combined test-retest/interrater correlation using the pearson correlation coefficient. Criterion validity was assessed by comparing the scale scores of children identified as having significant problems by local psychologists ("cases") with the scores of those children said to have no problems ("non-cases").

Based on extensive research (see ASEBA manual by Achenbach & Rescorla, 2001 for a complete review) syndrome scales, or lists of problems that tend to co-occur, have been constructed based on the YSR instrument. We have followed these established YSR groupings. The YSR internalizing scale includes symptoms of anxiety, depression, withdrawal, and somatic complaints. The YSR externalizing scale includes rule breaking and aggressive behaviors. Additional YSR items in the total symptoms scale include social, attention and thought problems. The YSR also includes several prosocial behaviors (or positive behaviors such as "I can be pretty friendly") which were grouped together into a general prosocial scale.



We also added two scales consisting of items added specifically for this Georgian population based on the qualitative data and not already represented by an existing YSR item. (See Table 3 for scale information.)

**Table 3: Scales and Subscales of the YSR-G.**

Syndrome Name	Items Included	Summary Scales	Original YSR or Added
Anxious/Depressed	14, 29, 30, 31, 32, 33, 35, 45, 50, 52, 71, 91, 112	Internalizing	YSR
Withdrawn/Depressed	5, 42, 65, 69, 75, 102, 103, 111	Internalizing	YSR
Somatic complaints	47, 51, 54, 56a-g	Internalizing	YSR
Rule Breaking Behavior	2, 26, 28, 39, 43, 63, 67, 72, 81, 82, 90, 96, 99, 101, 105	Externalizing	YSR
Aggressive Behavior	3, 16, 19, 20, 21, 22, 23, 37, 57, 68, 86, 87, 89, 94, 95, 97, 104	Externalizing	YSR
Social Problems	11, 12, 25, 27, 34, 36, 38, 48, 62, 64, 79	Additional	YSR
Attention Problems	1, 4, 8, 10, 13, 17, 41, 61, 78	Additional	YSR
Thought Problems	9, 18, 40, 46, 58, 66, 70, 76, 83, 84, 85, 100	Additional	YSR
Prosocial	6, 15, 49, 59, 60, 73, 80, 88, 92, 98, 106, 107, 108, 109	Prosocial	YSR
Georgia Positive Family Items	17a, 57a, 72a, 104a	Georgia Positive	Added
Georgia Negative Family Items	32a, 47a, 112b, 112c, 112d, 112e	Georgia Negative	Added

The reliability and validity analyses were done on the following summary scales (described in Table 3):

A. Original YSR Scales:

1. Internalizing items
2. Externalizing items
3. Prosocial items

B. Georgia Scales (composed of additional items from the qualitative studies not already included in the YSR Scales):

4. Georgia problems
5. Georgia Prosocial items

C. Combinations of YSR and Georgia Scales:

6. Total Problems: Internalizing + Externalizing + Additional YSR problem questions + Georgia Problem Items.
7. Total Prosocial: YSR Prosocial + Georgia Prosocial Items.

## **RESULTS**

### **Pilot Study**

During the first day of piloting a significant problem emerged with the interview process. Many children complained about being asked the standard YSR questions on suicide, drug use, and sex. In several sites those children interviewed on the first day refused to be interviewed. These children told other children of their concerns and many of these children also refused to be interviewed. Based on the feedback from the interviewers, and on several informal meetings with these children by the research team, it became clear that the children felt personally affronted. When it was explained that these questions were asked to all children (because we did not know for which children they would be an issue) this was accepted by the children. The problem was addressed not by changing the instrument but by adding further explanation on this point to the Study Explanation document read before each interview (See Appendix B for final project description).

For the validity study we also included a section which asked children about whether or not they felt they had emotional or behavioral problems (Section C). The intent of this section was to conduct some additional validity analyses by comparing the children's opinions of whether or not they felt they had these types of problems with the opinion of the child psychologist as to whether or not the child was a 'case'. However, during the piloting it was clear that the children did not understand the concept of 'emotional problem' and so this question was removed. The question on behavioral problems seemed to be better understood (and so was retained in the questionnaire). However, our confidence that children understood this question was low, and so the results for this question were not included in the final analysis.

No other significant problems emerged with the interview process or the instrument during the pilot study.

### **Reliability and Validity Study**

#### *Sample Characteristics*

A total of 386 children were interviewed for the study. As described in the Methods Section, psychologists working at the institutions in which the children were found provided brief current assessments of the presence and severity of emotional and behavioral problems among these children. On the basis of the scoring method used by the psychologists (See Methods), 57 children were classified as likely 'cases' and 75 as likely 'non-cases.' Comparison of the characteristics of this subset with that of the entire sample (Table 4) suggests that these 132 children are demographically similar to the total sample, except for the proportions in boarding school or a day club.

**Table 4: Study Sample Characteristics**

	<b>Total Sample</b> (n=386)	<b>Validity Sample</b> (n=132)
<b>Gender</b>		
Boys, N (%)	175 (45%)	58 (44%)
Girls, N (%)	211 (55%)	74 (56%)
<b>Ages</b>		
7-9, N (%)	46 (12%)	14 (11%)
10-11, N (%)	65 (17%)	25 (19%)
12-13, N (%)	91 (23%)	37 (28%)
14-15, N (%)	103 (27%)	33 (25%)
16-18, N (%)	81 (21%)	23 (17%)
<b>Sites</b>		
Orphanages <sup>a</sup> , N (%)	67 (17%)	20 (15%)
Save Centers <sup>b</sup> , N (%)	198 (51%)	97 (55%)
Boarding School/Day Club <sup>c</sup> , N%	97 (25%)	8 (6%)
Social Adaptation Center <sup>d</sup> , N%	16 (4%)	7 (5%)
Mobile Team, N%	8 (2%)	0
<b>Currently School Enrolled</b>		
Yes, N (%)	308 (80%)	102 (77%)
No, N (%)	78 (20%)	30 (23%)
<b>Years in Institution</b>		
Average (SD)	2.4 (2.2)	2.5 (2.2)
Range	0-14	0-12

a. Dighomi & Momavlis Sakhli/ House of the Future

b. Tsisarkela/Rainbow, Beghurebi/Sparrows, Sapovnela/Rustavi, Biliki, & Skhivi

c. Kojori and Gori Club

d. Gldani social adaptation center for children

### ***Instrument Reliability***

#### *Test-retest/Inter-rater Reliability*

Table 5 shows combined test-retest/inter-rater reliability analysis results, based on the approximately 20% of the total sample who were re-interviewed by different interviewers 1-3 days after the first interview. Results are shown for the total study sample (N=386) and separately for boys and girls. Combined test-retest/inter-rater reliability is assessed using the Pearson correlation coefficient, which provides a measure of how similar each subscale score is on the first and second interviews. This in turn provides an indicator of the extent to which children tend to give the same answer to the questions constituting the scale when asked on different occasions and by different interviewers.

When assessing test-retest reliability alone (i.e. when the same interviewer interviews the same child at both times), Pearson correlation coefficient scores of .7 are considered to be acceptable. Since we assessed both test-retest and inter-rater reliability at the same time, there is likely to be

more variation in responses and acceptable Pearson scores would be therefore be lower.

The results in Tables 5 suggest that the scores for the problem-based scales range from acceptable (.69 for the Georgia problem scale) to good (>.80) for the internalizing, externalizing and total problem scales. In contrast, the prosocial scales all performed noticeably less well. Overall, interviews with boys were more reliable than those with girls although the reasons for this are not clear. We speculated that the differences could reflect possible greater sensitivity of girls' responses to the gender of the interviewer. However, analysis of girls' interviews in which we included only first and second interviews done by female interviewers did not demonstrate any improvement in reliability (analysis not shown).

Based on these results, the problem-based scales were found to have acceptable test-retest and interrater reliability and the prosocial scales to have fair or poor test-retest and interrater reliability.

**Table 5: Combined Test-Retest/Inter-rater Comparison**

Total Sample	First Interview		Repeat Interview		Correlation*
	N	Mean (sd)	N	Mean (sd)	
<b>YSR Subscales</b>					
Internalizing score, mean (sd)	66	21.0 (11.6)	67	19.2 (7.5)	.80
Externalizing score, mean (sd)	57	19.0 (10.7)	57	17.8 (11.0)	.85
Prosocial score, mean (sd)	71	21.9 (4.0)	69	21.9 (4.1)	.58
<b>Georgia Subscales</b>					
Georgia problems, mean (sd)	67	2.4 (2.3)	69	1.9 (2.0)	.69
Georgia prosocial, mean (sd)	65	6.1 (1.6)	63	6.5 (1.5)	.64
<b>Total Scales<sup>a</sup></b>					
Total problems, mean (sd)	56	74.5 (33.6)	57	66.6 (34.5)	.83
Total prosocial, mean (sd)	65	28.0 (4.9)	62	28.3 (4.9)	.62

Girls Sample	First Interview		Repeat Interview		Correlation*
	N	Mean (sd)	N	Mean (sd)	
<b>YSR Subscales</b>					
Internalizing score, mean (sd)	35	22.9 (11.5)	35	21.0 (11.4)	.73
Externalizing score, mean (sd)	30	18.7 (10.2)	30	16.9 (10.3)	.75
Prosocial score, mean (sd)	36	21.6 (4.2)	34	22.2 (3.8)	.51
<b>Georgia Subscales</b>					
Georgia problems, mean (sd)	35	2.5 (2.2)	35	1.9 (2.0)	.60
Georgia prosocial, mean (sd)	34	5.9 (1.6)	31	6.3 (1.6)	.74
<b>Total Scales<sup>a</sup></b>					
Total problems, mean (sd)	29	76.4 (32.7)	30	68.5 (33.5)	.71
Total prosocial, mean (sd)	34	27.4 (4.9)	30	28.2 (4.9)	.57

<b>Boys</b>	<b>First Interview</b>		<b>Repeat Interview</b>		<b>Correlation*</b>
	<b>N</b>	<b>Mean (sd)</b>	<b>N</b>	<b>Mean (sd)</b>	
<b>YSR Subscales</b>					
Internalizing score, mean (sd)	31	18.9 (11.5)	32	17.3 (9.6)	.89
Externalizing score, mean (sd)	27	19.4 (11.3)	27	18.8 (11.8)	.93
Prosocial score, mean (sd)	35	22.1 (3.8)	35	21.7 (4.4)	.67
<b>Georgia Subscales</b>					
Georgia problems, mean (sd)	32	2.3 (2.5)	34	1.8 (2.1)	.77
Georgia prosocial, mean (sd)	31	6.4 (1.6)	32	6.8 (1.4)	.52
<b>Total Scales<sup>a</sup></b>					
Total problems, mean (sd)	27	72.4 (35.2)	27	64.5 (36.2)	.94
Total prosocial, mean (sd)	31	28.6 (4.9)	32	28.4 (5.0)	.69

\* Pearson correlation coefficient

### *Internal Consistency Reliability*

Internal consistency reliability measures the extent to which questions that assess the same underlying concept agree or disagree. If these questions disagree this suggests that either the questions themselves are unreliable, or they are not really measuring the same concept. As described in Appendix C, internal consistency reliability is measured using Cronbach's alpha. Scores should be at least .7 and ideally >.8.

Table 6 shows the Cronbach's alpha scores on each of the scales for boys, girls, and the total sample (N=386). Alpha scores for the internalizing, externalizing, and total problems subscales are very good. Since the internalizing and externalizing subscales are part of the total problem subscale, high scores on all three suggest that all three scales may be measuring the same concept or (more likely) they are measuring separate concepts that are very highly correlated. In other words, children who have any problems tend to have many diverse but related problems, as measured by the different scales. This is consistent with findings from children in other populations: children rarely have a single problem but instead tend to have a wide variety of issues.

It is more difficult to interpret the Cronbach's alpha scores for the other scales. While the negative items in the YSR are divided into separate scales on the basis of underlying themes (such as internalizing and externalizing symptoms) all prosocial items (which represent various underlying concepts) are included in a single scale and therefore may be showing low Cronbach's alpha scores for that reason. Similarly, the Georgia problem and Georgia prosocial scales each contain items whose only common link with each other is that they were not already included in one of the YSR scales. Finally, Cronbach's alpha scores are affected by the number of items in the subscale. Few items tend to be associated with artificially lower scores. The Georgia prosocial scale has only 4 items and the Georgia problem scale has only 6, which would tend to depress the alpha scores for both.<sup>2</sup>

<sup>2</sup> The much greater number of items in the total problem scale compared with the internalizing and externalizing scales probably accounts for the higher Cronbach's alpha score of the total problem scale. In reality, the internal consistency reliability of all three scales is probably similar.

**Table 6: Cronbach's alpha scores**

	Total Sample (N=386) <sup>b</sup>	Boys (N=175) <sup>b</sup>	Girls (N=211) <sup>b</sup>
<b>YSR Subscales</b>			
Internalizing score, mean (sd)	.86	.86	.85
Externalizing score, mean (sd)	.86	.86	.83
Prosocial score, mean (sd)	.42	.50	.37
<b>Georgia Subscales</b>			
Georgia problems, mean (sd)	.61	.62	.61
Georgia prosocial, mean (sd)	.46	.57	.35
<b>Total Scales<sup>a</sup></b>			
Total problems, mean (sd)	.94	.94	.94
Total prosocial, mean (sd)	.50	.58	.44

a. Total scales include all items from the instrument. The total problems score includes 113 items; the total prosocial score this includes 18 items.

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

### *Criterion Validity*

Table 7 examines the criterion validity of the scales by comparing scale scores of the subsample of 132 children defined as likely “cases” and “non-cases” based on the assessments by the local psychologists (See Methods).

For all 132 children, differences in scores between “cases” and “non-cases” were statistically significant on all scales, except for the Georgia prosocial scale. Differences tended to be in the direction expected: higher problem scores and lower prosocial scores among “cases”. However, the absolute difference in most scale scores between “cases” and “non-cases” are not large: i.e. < 10 points difference for all the larger scales, except the total problems scale where the difference was near 25 points regardless of gender. Differences among the prosocial scores were particularly small, even for the total prosocial scale. The magnitude and direction of change were similar across the scales for boys and girls. Overall, the results suggest that the total problems scale is the best scale for distinguishing between “cases” and “non-cases” in terms of magnitude of differences and their statistical significance.

**Table 7: Comparison of Cases versus Non-Cases**

	Total Sample N=132				Girls only (N=74)			Boys only (N=58)		
	Score Range (Min, Max)	Cases (n=57) <sup>b</sup>	Non-Cases (n=75) <sup>b</sup>	Difference (p-value <sup>c</sup> )	Cases (n=26) <sup>b</sup>	Non-Cases (n=48) <sup>b</sup>	Difference (p-value <sup>c</sup> )	Cases (n=31) <sup>b</sup>	Non-Cases (n=27) <sup>b</sup>	Difference (p-value <sup>c</sup> )
Female		26 (46%)	48 (64%)	P=.04						
Male		31 (54%)	27 (36%)							
<b>YSR Subscales</b>										
Internalizing score, mean (sd)	0, 62	22.2 (12.3)	16.3 (10.4)	<b>5.9 (.005)</b>	24.4 (12.5)	17.6 (11.4)	<b>6.8 (.03)</b>	20.3 (12.1)	13.8 (8.1)	<b>6.5 (.03)</b>
Externalizing score, mean (sd)	0, 64	20.3 (11.2)	13.2 (8.9)	<b>7.1 (.0003)</b>	19.8 (2.6)	12.0 (8.1)	<b>7.8 (.003)</b>	20.9 (10.3)	15.6 (10.1)	<b>5.3 (.10)</b>
Prosocial score, mean (sd)	0, 28	21.9 (3.3)	23.0 (3.2)	<b>-1.1 (.05)</b>	22.2 (3.2)	22.7 (3.2)	<b>-0.5 (.56)</b>	21.5 (3.4)	23.6 (3.3)	<b>-2.1 (.02)</b>
<b>Georgia Subscales</b>										
Georgia problems, mean (sd)	0, 12	2.7 (2.4)	1.3 (1.7)	<b>1.4 (.0002)</b>	2.5 (2.5)	1.4 (1.9)	<b>1.1 (.05)</b>	2.9 (2.4)	1.1 (1.5)	<b>1.8 (.002)</b>
Georgia prosocial, mean (sd)	0, 8	6.3 (1.6)	6.5 (1.4)	<b>-0.2 (.42)</b>	6.2 (1.6)	6.4 (1.4)	<b>-0.2 (.55)</b>	6.4 (1.7)	6.7 (1.6)	<b>-0.3 (.48)</b>
<b>Total Scales<sup>a</sup></b>										
Total problems, mean (sd)	0, 226	78.3 (36.1)	53.7 (27.8)	<b>24.6 (.0001)</b>	79.5 (35.7)	54.5 (30.1)	<b>25.0 (.005)</b>	77.2 (36.4)	52.1 (23.0)	<b>25.1 (.009)</b>
Total prosocial, mean (sd)	0, 36	28.1 (4.1)	29.7 (3.8)	<b>-1.6 (.02)</b>	28.4 (3.5)	28.4 (3.5)	<b>0.0 (.30)</b>	27.8 (4.6)	30.5 (3.9)	<b>-2.7 (.02)</b>

a. Total scales include all problem/prosocial items in the instrument. The total problem score includes 113 items; the total prosocial score includes 18 items.

b. Only for respondents with complete data.

c. P-value for the statistical significance of the difference in scale scores by caseness.

## Epidemiological Analysis of Study Findings

Our study sample of 386 children constituted all the children who were available for interview at each site at the time of the study (See Table 8.). The last column in Table 8 gives the percentage of children we interviewed at each site, based on figures provided to Save by the institutions. These data suggest that that our interviewees constitute a substantial (though likely biased) sample of the total child population that attends these centers and institutions.

Therefore, in addition to the validity and reliability analyses presented above, we conducted additional exploratory analysis of this population, to gain some impressions of their general psychosocial well-being at the time of the study.

**Table 8: Number and percentage of interviewees at each site**

Site Name	Number children interviewed	Number of children served by the Center/institution	% of total number of youth at each site who were interviewed
Tsisartkela/Rainbow (Tbilisi)	57	58	98.2
Beghurebi/Sparrows (Tbilisi)	32	37	86.4
Sapovnela (Rustavi)	31	37	83.7
Biliki (Gori)	64	76	84.2
Skhivi (Gori)	14	25	56.0
Dighomi (Tbilisi)	32	41	78.0
Gldani (Tbilisi)	16	35	45.7
Momavlis Sakhli/ House of the Future (Tbilisi)	35	55	63.6
Kojori (Kojori)	74	105	70.4
Mobile	8	25	32.0
Gori Club	23	32	71.8

Table 9 describes the scale scores for the entire sample, and for those attending government institutions compared with children attending the RLP supported Centers. Table 10 shows the same results separated by gender.

In addition to the simple comparison of RLP vs government sites (Table 9) we also conducted a 5-way comparison using a different sub-grouping of sites (see Table 11). This was done because of concerns that some RLP sites are more similar to some of the government sites than they are to other RLP sites. The groupings in Table 11 reflect these concerns, with the groupings constructed according to these perceived similarities between sites. Because of the multiple groups we conducted an ANOVA analysis to determine if any of the groups were significantly different from the others. We found that all groups were statistically identical on all scales except the Georgia Problems scale. Only on that scale was at least one of the groups significantly different from the others ( $p=.007$ ).



Table 12 and Figure 1 show the distribution of total problem scores. Total problem scores for children who are doing well (few problems) fall in the 0-20 range. Only 2.6% of our sample was in this category (Table 12). The scores of most children fell into the 20-100 range, with a mean score of 60.5 (Table 9), reflecting a substantial degree of difficulty with multiple problems across multiple scales. 8.4% of the sample had scores above 100, reflecting a high degree of difficulty across many problem categories (Table 12). All subscale scores are very similar regardless of whether children are at the Centers or the government institutions (Tables 9 and 10), which suggests that neither group is better or worse off than the other, based on this measure.

Prosocial scale scores were also very high, and approached the maximum scores for each scale.

**Table 9: Scale Scores for Total Sample, Government Institutions, and RLP Center Sites**

	Total Sample (n=386) <sup>b</sup>	Government Institution (n=67) <sup>b,c</sup>	Center (n=198) <sup>b,d</sup>	p-value <sup>e</sup>
<b>YSR Subscales</b>				
Internalizing score, mean (sd)	17.9 (9.8)	18.0 (7.5)	17.9 (11.0)	.95
Externalizing score, mean (sd)	15.4 (8.8)	14.6 (8.7)	15.9 (9.8)	.40
Prosocial score, mean (sd)	22.4 (3.5)	21.7 (3.8)	22.6 (3.5)	.07
<b>Georgia Subscales</b>				
Georgia problems, mean (sd)	1.6 (1.9)	2.1 (2.1)	1.7 (2.0)	.22
Georgia prosocial, mean (sd)	6.3 (1.5)	6.0 (2.0)	6.4 (1.5)	.16
<b>Total Scales<sup>a</sup></b>				
Total problems, mean (sd)	60.5 (28.1)	61.0 (22.4)	61.4 (32.0)	.93
Total prosocial, mean (sd)	28.8 (4.2)	27.6 (4.8)	29.0 (4.2)	.03

a. Total scales include all items. Total problems score includes 113 items; Total prosocial score includes 18 items.

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

c. Sample does not include Kojori (See Study Sites).

d. Not including the Gori Club (See Study Sites).

e. P-value for the statistical significance of the difference in scale scores comparing government institution and center samples.

**Table 10: Scale Scores for Total Sample, Government Institutions, and RLP Center Sites (by gender)**

YSR Subscales	Total Sample <sup>b,c</sup>			Government Institution Sample <sup>b,d</sup>			Center Sample		
	Boys (n=175)	Girls (n=211)	p-value <sup>e</sup>	Boys (n=28)	Girls (n=39)	p-value <sup>e</sup>	Boys (n=88) <sup>b</sup>	Girls (n=110)	p-value <sup>e</sup>
Internalizing score, mean (sd)	17.2 (9.4)	18.5 (10.1)	.22	17.6 (7.8)	18.3 (7.4)	.72	17.0 (10.2)	18.5 (11.6)	.35
Externalizing score, mean (sd)	17.8 (9.2)	13.6 (8.1)	<.0001	18.1 (2.1)	12.0 (1.3)	.01	18.5 (9.9)	14.1 (9.4)	.004
Prosocial score, mean (sd)	22.2 (3.7)	22.5 (3.4)	.41	20.5 (4.0)	22.6 (3.5)	.03	22.6 (3.7)	22.6 (3.3)	.99
<b>Georgia Subscales</b>									
Georgia problems, mean (sd)	1.9 (2.0)	1.5 (1.8)	.03	2.5 (2.3)	1.8 (1.9)	.22	1.9 (2.0)	1.7 (2.0)	.53
Georgia prosocial, mean (sd)	6.4 (1.6)	6.3 (1.5)	.84	6.0 (1.9)	6.0 (2.1)	.94	6.4 (1.7)	6.3 (1.4)	.58
<b>Total Scales<sup>a</sup></b>									
Total problems, mean (sd)	63.0 (27.9)	58.7 (28.2)	.18	64.6 (22.1)	58.4 (22.7)	.34	64.2 (31.4)	59.5 (32.4)	.35
Total prosocial, mean (sd)	28.7 (4.4)	28.9 (4.1)	.54	26.4 (5.0)	28.5 (4.6)	.09	29.1 (4.6)	29.0 (3.9)	.97

a. Total scales include all items. Total problems score includes 113 items; Total prosocial score includes 18 items.

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

c. Sample does not include Kojori (See Study Sites).

d. Not including the Gori Club (See Study Sites).

e. P-value for the statistical significance of the difference in scale scores by gender.

Table 11: Scale Scores for RLP and Government Sites, Grouped According to Similarities Between Sites.

<b>YSR Subscales</b>	<b>Total Sample (n=386)</b>	<b>Tsisartkela/ Tsisartkela/ Rainbow (Tbilisi) and Biliki (Gori) (n=121)</b>	<b>Sapovnela (Rustavi) and Skhivi (Gori) (n=45)</b>	<b>Beghurebi/ Beghurebi/Spar rows (Tbilisi) and Momavlis Sakhli/ House of the Future (Tbilisi) (n=67)</b>	<b>Dighomi Orphanage (Tbilisi) and Kojori Boarding School (Kojori) (n=106)</b>	<b>Gldani Social Adaptation Center for Children (Tbilisi) and Mobile Team sites (Tbilisi) (n=24)</b>
Internalizing score, mean (sd)	17.9 (9.8)	16.6 (10.1)	18.0 (9.7)	19.6 (11.7)	19.3 (8.5)	21.3 (8.6)
Externalizing score, mean (sd)	15.4 (8.8)	14.7 (8.6)	15.5 (8.9)	18.6 (11.5)	14.8 (7.1)	17.2 (11.5)
Prosocial score, mean (sd)	22.4 (3.5)	22.8 (3.5)	22.3 (3.8)	22.4 (3.3)	21.7 (3.9)	22.2 (3.0)
<b>Georgia Subscales</b>						
Georgia problems, mean (sd)	1.6 (1.9)	1.4 (1.7)	2.3 (2.0)	2.1 (2.3)	1.5 (1.8)	2.6 (2.0)
Georgia prosocial, mean (sd)	6.3 (1.5)	6.5 (1.4)	6.2 (1.7)	6.1 (1.9)	6.4 (1.4)	5.9 (1.3)
<b>Total Scales<sup>a</sup></b>						
Total problems, mean (sd)	60.5 (28.1)	57.1 (27.9)	62.3 (29.3)	69.5 (35.8)	61.0 (22.8)	72.2 (29.1)
Total prosocial, mean (sd)	28.8 (4.2)	29.3 (4.2)	28.6 (4.8)	28.6 (3.8)	28.2 (4.6)	28.2 (4.0)

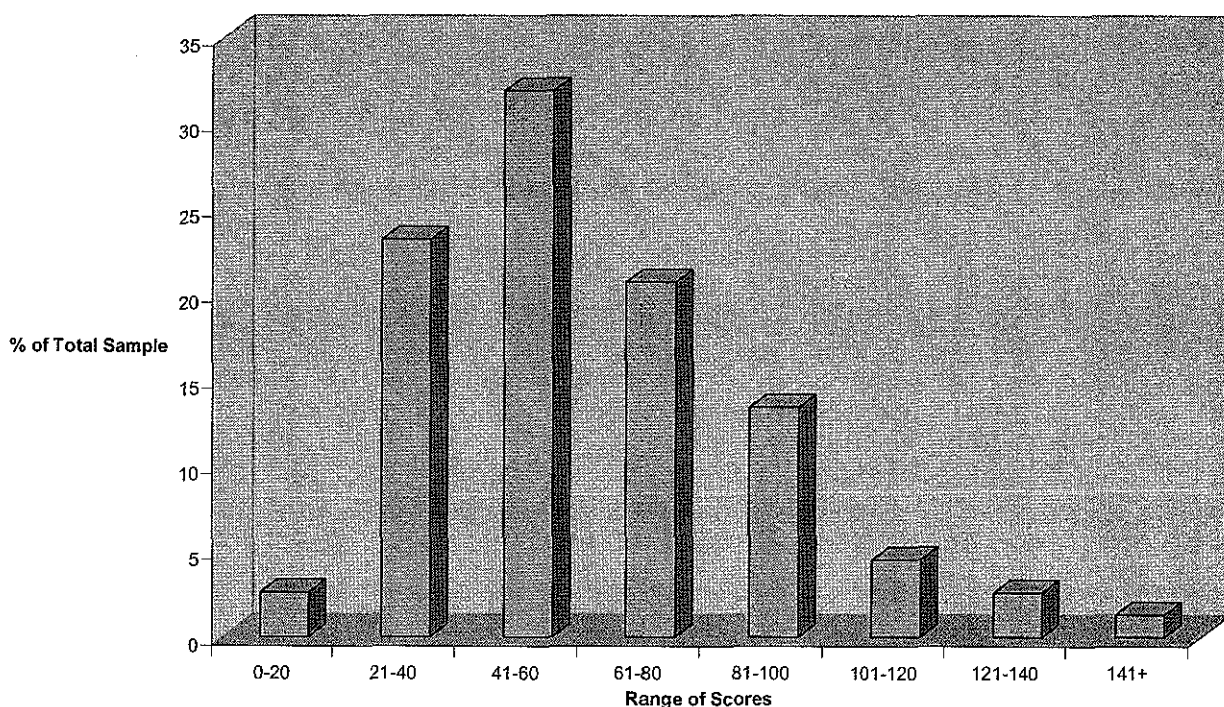
**Table 12: Distribution of Total Problems Scores<sup>a</sup>**

	N=314	%
0-20 points	8	2.6
21-40 points	73	23.2
41-60 points	100	31.9
61-80 points	65	20.7
81-100 points	42	13.4
101-120 points	14	4.5
121-140 points	8	2.6
141+ points	4	1.3

a. Total scales include all items. Total problems score includes 113 items; Total prosocial score includes 18 items. Not all respondents have complete data. Data presented for only those with complete data.

**Figure 1**

**Distribution of Total Problems Score**



## DISCUSSION

Pilot Testing among our sample proved essential to improving our interviewing procedures and adjusting the draft YSR-G to make it acceptable to our target population. Without pilot testing it was clear that the original instrument and interviewing procedures were potentially distressing to

the interviewees, which would have affected ability to use the instrument effectively in the future.

Reliability and validity testing showed a distinct pattern in the performance of the YSR-G. On most measures of reliability or validity, we found that the problem-based scales (the YSR internalizing and externalizing scales, and particularly the total problem scale) have solid psychometric properties in this population. However, the prosocial scales performed poorly. These findings were consistent for the entire sample, and for the gender-specific analyses.

The internalizing and externalizing scales have good internal consistency (Cronbach's alpha >.8), as did the total problem scale (Cronbach's alpha>.9). The YSR Prosocial scale and both Georgia-specific scales (problem-based and prosocial) showed poorer internal consistency, which is likely due in part to the mixed nature of the concepts assessed by these scales and the small number of items in the Georgia-specific scales. Combined test-retest/inter-rater reliability was good for the internalizing and externalizing problem scales and the total problem scales, marginal for the Georgia problem scales, and poor for all the prosocial scales.

Tests of criterion validity found that the problem scores consistently matched the criterion chosen for this study: local psychologists' evaluations: The children identified by psychologists as having significant emotional and/or behavioral problems ("cases") showed significantly higher levels of symptoms on the same internalizing and externalizing YSR scales and on the total problem scale than those children identified as having few emotional and/or behavioral problems ("non-cases"). As in the other analyses, the prosocial scales did not perform well, with little or not significant differences in scale scores based on "Caseness."

In addition to the main purpose of the study - developing an acceptable, reliable, and valid instrument - we also conducted preliminary epidemiological analyses of the study data for the total sample and separately for those children from the Centers and government institutions. We found that the total sample and the Center and institution subsample scores were elevated on all the problem scales, suggesting difficulty across the range of problem categories assessed by these scales: anxiety, depression, withdrawal, aggression, social and cognitive problems and problems with family. Most of the children in our sample had been in contact with the Centers and government institutions for some time (a year or more). They would therefore be expected to have already benefitted from the interventions these organizations provide, and therefore to have experienced improvement. This suggests that children living on the street or at risk, and who are not receiving services, are likely to have substantially higher problem scores than those of the children in our study.

In comparing the scores of children at the Centers and in the government institutions, we found little difference in the severity or degree of problems or their prosocial behaviors. Little difference was found when comparing all RLP sites with the government sites (Tables 9 and 10) or when grouping sites according to other similarities and comparing these groups (Table 11). There are several possible explanations. One may be that both groups may be similar upon enrolment and the services of both types of organizations may be equally effective. Another explanation may be that one group of children may be worse off at entry but do better once they arrive and receive services, suggesting that one type of organization is receiving more severely

affected children but is able to produce more substantial improvement. Without baseline data at entry, it is not possible to know which explanation is correct. At this time the data can only suggest that the pattern of problems and their severity among the two populations are similar.

Scores on all the prosocial scales were very high. Given the poor psychometric properties of the prosocial scales described above, it may be that children are not answering these questions accurately. One possibility is that social desirability effects (a desire to look good for the interviewer) may be inflating their scores. The high levels of these scores and their likely inaccuracy (based on the psychometrics) suggest that programs will have difficulty in substantially affecting these scores through their interventions, and that any changes in the scores will not accurately reflect program impact.

## CONCLUSIONS

Based on our pilot study, the YSR-G and associated interview materials are acceptable for use among street children served by the RLP Centers and children served by the types of government institutions included in this study.

Overall, the problem scales of the YSR-G shows solid psychometric properties and therefore we believe them to be suitable for use among this type of child population in Georgia.

These results suggest many options for use of the YSR-G within Georgia. The YSR-G could be used as a screener to help identify children who need attention, based on the Total Problems Scale. The Internalizing and Externalizing scales can 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. The availability of trained mental health professionals in Georgia makes this use of the more specific scales (See Table 3) appropriate for understanding individual problems, as well as in formulating treatment planning. The problem scales in the YSR-G can also be used to evaluate the RLP program. They can also be used to assess the need for, and impact of, other new or existing programs within the Centers or Institutions that address problems measured by the YSR-G.

The prosocial scales showed poor psychometric properties and were generally high, even approaching the maximum scores for each scale. Poor psychometric properties suggest that these scales are not accurate. High mean scores among this population suggest there is a ceiling effect which will cause problems when attempting to use these scores to assess program impact; it will be difficult to improve on scores that are already high. Therefore, our results do not suggest that the prosocial scores are useful either for screening children into programs or assessing their progress. However, since this is the first time this instrument has been tested, and the study was done among a population of children with prolonged exposure to Centers or government institutions (who may have experienced marked improvement in their prosocial skills due to exposure to these programs) we are not yet ready to advocate removing the prosocial items from the instrument. Instead, they should be retained for the time being and their performance reassessed as part of analyses of future data.

Informal reports (data still pending) suggest that the children in our study sample represent a substantial proportion of the children served by the study sites in which we worked. Average problem scores were high. The distribution of scores shows few children with low scores and significant numbers with high scores. This suggests that most children in our sample have a wide range of significant psychosocial problems. Since most of our sample has been receiving services for a year or more, we suspect that children who are not receiving services will have even higher scores.

In conclusion, previous qualitative studies have resulted in the selection and adaptation of an existing instrument which has, in the study reported here, proved to provide an accurate assessment of the psychosocial problems of this population.

## **RECOMMENDATIONS**

If feasible, Save the Children Georgia should use the YSR-G to evaluate the impact of the RLP program on children who are new or relatively new to the program. Whether or not this is feasible at this late stage in the program (which will end in Sept, 2008) should be determined as soon as possible. BU/JHU faculty can provide technical assistance in carrying out these assessments.

The YSR-G should be used by other organizations working with these populations (and other child populations at risk), including the Georgian government, other NGOs and private groups. The goals of these assessments should include:

- a) Assessing the nature and severity of needs (by using the instrument as a survey tool).
- b) Using this information to target resources and design appropriate interventions.
- c) Assessing the impact of these interventions

In conducting a-c, there should be a focus on building local capacity in program design, monitoring, and evaluation.

When the YSR-G is used in the future, analysis of the resulting data should include further characterization of the accuracy of the instrument, particularly with regard to the performance of the prosocial scales.

The methods used in this part of the RLP project – qualitative methods resulting in instrument selection and adaptation, followed by instrument piloting and validity testing (as reported here) - should be repeated in other contexts. As with the RLP project, these methods can be used to improve need and impact assessments for other populations, both children and adults, and to assess both psychosocial and other problems. The methods are particularly useful in situations where need has not been well characterized and where the impact of interventions has not been demonstrated.

**Appendix A – Version of the YSR instrument used in Validity Study (YSR-G)**

ახალგაზრდების გამომკითხვა, 2007  
Survey of Youth in Tbilisi Georgia 2007

**წინასწარი ინსტრუქციები**  
Preliminary Instructions

გამარჯობა, ჩემი სახელია \_\_\_\_\_. მე ვმუშაობ ფედერაციასთან „გადავარჩინოთ ბავშვები“ და ბოსტონის უნივერსიტეტთან. ჩვენ ვატარებთ კვლევას, რათა უკეთესი მომსახურება გავუწიოთ ბავშვებს. შეგიძლია, 45 წუთი დაგვითმონ და ჩვენს კითხვებს უპასუხო?

Hello, my name is \_\_\_\_\_. I work for Save the Children and Boston University and we are conducting a study to help us to better serve the children here. Would you have 45 minutes to answer some questions?

თუ რესპონდენტი გიპასუხებთ, რომ არა, მაშინ ჰკითხეთ მას, შეგიძლიათ თუ არა, რომ სხვა დროს მიხვიდეთ ინტერვიუს ჩასატარებლად. თუ რესპონდენტი კითხვებს დაგისვამთ, უპასუხეთ, მაგრამ ნუ მიაწვდით ისეთ ინფორმაციას, რაც კითხვარზე მათ მიერ გაცემულ პასუხებზე მოახდენს გავლენას. თუ რესპონდენტი დადებით პასუხს გაგცემთ, უთხარით შემდეგი:

If the respondent answers no ask them if you can come back another time and interview them. If the respondent asks questions, answer them but do not provide information that could affect their answers to the questionnaire. If the respondent answers yes, then say the following:

სანამ დავიწყებდეთ, მინდა წაგიკითხო ფორმა, რომელიც უფრო კარგად აგიხსნის, თუ რის გაკეთებას ვაპირებთ ერთად.

Before we start I want to read a form to you that explains more about what we are going to do together.

ამ წაკითხვით კვლევის განმარტება:

Read explanation of study form here.

დაუინებოთ მოითხოვეთ, რომ ინტერვიუ ერთი-ერთზე ჩატარდეს. თუ ამას ვინმე შეეკამათება, აუხსენით მას, რომ ეს ჩვენი პროცედურის მნიშვნელოვანი ნაწილია და ჩვენი დაკვირვებით, ზოგი ადამიანი განსხვავებულად გეპასუხობს სხვა ადამიანების თანდასწრებით.

Insist that the interview be conducted in private. If this is questioned by anyone, explain that this is an important part of our procedure, and that we have found that some people give different answers when there are other people present.

რესპონდენტის სახელი, გვარი: \_\_\_\_\_

Respondent Name

ადგილი: \_\_\_\_\_

Location

**ავსებს სუპერვაიზორი:**

რესპონდენტის ტიპი: \_\_\_\_\_ (1) პილოტირება \_\_\_\_\_ (2) ვალიდიზაცია  
Type of Respondent Pilot Study Validity Study

ვალიდიზაციის სტატუსი: \_\_\_\_\_ (0) აღნიშნულია, როგორც პრობლემის არმქონე  
Validity Study Id'd as not having problem

\_\_\_\_\_ (1) აღნიშნულია, როგორც პრობლემის მქონე  
Id'd as having problem

ნაწილი A: დემოგრაფიული მონაცემები



Part A: Demographics

A01. ასაკი \_\_\_\_\_ (წლის)  
Age (years)

A02. სქესი \_\_\_\_\_ (0) მამრობითი \_\_\_\_\_ (1) მდედრობითი  
Sex male female

A03. ამჟამად დადიხარ სკოლაში? \_\_\_\_\_ (0) არა \_\_\_\_\_ (1) დიახ  
Currently enrolled in school no yes

A04. რომელ კლასში ხარ? რომელ კლასში იყავი ბოლოს? \_\_\_\_\_ (კლასი)  
Which grade are you in? Which grade have you been last? (grade)

A05. რამდენი ხანია, რაც ინსტიტუციაში/დაწესებულებაში იმყოფება:  
Time at institution/location:

\_\_\_\_\_ დღე \_\_\_\_\_ თვე \_\_\_\_\_ წელი  
Days Months Years

ნაწილი B- ადაპტირებული ახალგაზრდების თვითშეფასების კითხვარი (YSR-G)  
Part B- Adapted Youth Self Report (YSR-G)

მე წავიკითხავ იმ გრძნობებისა და ქცევების ჩამონათვალს, რომლებიც ზოგჯერ ბავშვებს აქვთ ხოლმე. თითოეული მათგანის წაკითხვისას მე გკითხავ, რამდენად ხშირად გამოგიცდია თითოეული ბოლო ერთი კვირის განმავლობაში, დღევანდელი დღის ჩათვლით. პასუხებია: არასოდეს; ზოგჯერ; ხშირად.

*I am going to read you a list of feelings or behaviors that children sometimes have. For each one I am going to ask you how often you have experienced each one IN THE LAST WEEK, including today. The responses are never, sometimes, often.*

დაასახელებთ სიმპტომები და ჰკითხეთ რესპონდენტს, თუ რამდენად ხშირად აქვს თითოეული მათგანი. თითოეული სიმპტომის შემდეგ გაიმეორეთ პასუხის კატეგორიები და რესპონდენტს საშუალება მიეცით, ერთ-ერთი მათგანი ამოირჩიოს. პასუხი აღნიშნეთ სიმპტომის გასწვრივ შესაბამისი პასუხის შემოხაზვით.

*Say each symptom, and after each one ask how often the respondent has experienced it.. Repeat the categories after each symptom and let the respondent choose one. Record the response by circling the appropriate box next to the symptom.*

ნიშნები, სიმპტომები და ქცევები <i>Signs, Symptoms and Behaviors</i>	არასოდეს <i>Never</i>	ზოგჯერ <i>Sometimes</i>	ხშირად <i>Often</i>	არ მეხება <i>N/A</i>	უარი განაცხადა <i>Refused</i>
B001. ჩემი ასაკისათვის ზედმეტად ბავშვურად ვიქცევი. <i>I act too young for my age.</i>	0	1	2	8	9
B002. ალკოჰოლურ სასმელებს ვსვამ ჩემი მშობლების ნებართვის გარეშე. <i>I drink alcohol without my parents' approval.</i>	0	1	2	8	9
B002a: აღწერეთ <i>Describe</i>					
B003. ბევრს ვკამათობ <i>I argue a lot.</i>	0	1	2	8	9
B004. დაწყებულ საქმეს ვერ ვამთავრებ. <i>I fail to finish things I start.</i>	0	1	2	8	9

ნიშნები, სიმპტომები და ქცევები <i>Signs, Symptoms and Behaviors</i>	არასოდეს <i>Never</i>	ზოგჯერ <i>Sometimes</i>	ხშირად <i>Often</i>	არ მკებია <i>N/A</i>	უარი განაცხადა <i>Refused</i>
B005. ცხოვრება არ მისარია. <i>I don't enjoy the life.</i>	0	1	2	8	9
B006. მიყვარს ცხოველები. <i>I like animals.</i>	0	1	2	8	9
B007. ვტრიალებ. <i>I brag.</i>	0	1	2	8	9
B008. მიჭირს კონცენტრირება ან ყურადღების მიქცევა. <i>I have trouble concentrating or paying attention.</i>	0	1	2	8	9
B009. ზოგიერთ ფიქრს თავიდან ვერ ვიგდებ. <i>I can't get my mind off certain thoughts.</i>	0	1	2	8	9
B009a: აღწერეთ <i>Describe</i>					
B010. მიჭირს უმოქმედოდ ჯდომა. <i>I have trouble sitting still.</i>	0	1	2	8	9
B011. ზედმეტად დამოკიდებული ვარ უფროსებზე. <i>I'm too dependent on adults.</i>	0	1	2	8	9
B012. მარტო ვარ. <i>I am lonely.</i>	0	1	2	8	9
B013. დაბნეული ვარ. <i>I am confused.</i>	0	1	2	8	9
B014. ბევრს ვტირი. <i>I cry a lot.</i>	0	1	2	8	9
B015. საკმაოდ გულახდილი ვარ. <i>I am pretty honest.</i>	0	1	2	8	9
B016. სხვების მიმართ აგრესიული ვარ. <i>I am mean to others.</i>	0	1	2	8	9
B017. ბევრს ვოცნებობ. <i>I daydream a lot.</i>	0	1	2	8	9
B017a. ვგრძნობ, რომ ჩემი ოჯახის წევრებს ვუყვარვარ. <i>I feel loved by my family.</i>	0	1	2	8	9
B018. განზრახ ვცდილობ საკუთარი თავის დაზიანებას ან მოკვლას. <i>I deliberately try to hurt or kill myself.</i>	0	1	2	8	9
B019. ვცდილობ, სხვების ყურადღება მივიპყრო. <i>I try to get a lot of attention from others.</i>	0	1	2	8	9
B020. ვაფუჭებ ჩემს ნივთებს. <i>I destroy my own things.</i>	0	1	2	8	9
B021. ვაფუჭებ სხვის ნივთებს. <i>I destroy things belonging to others.</i>	0	1	2	8	9
B022. არ ვემორჩილები მშობლებს. <i>I disobey my parents</i>	0	1	2	8	9
B023. სკოლაში დაუმორჩილებელი ვარ. <i>I disobey at school.</i>	0	1	2	8	9
B024. არ ვიკვებები ისე კარგად, როგორც უნდა ვიკვებებოდე. <i>I don't eat as well as I should.</i>	0	1	2	8	9
B025. ბავშვებთან არ ვმეგობრობ. <i>I don't make friends with kids.</i>	0	1	2	8	9

ნიშნები, სიმპტომები და ქცევები <i>Signs, Symptoms and Behaviors</i>	არასოდეს <i>Never</i>	ზოგჯერ <i>Sometimes</i>	ხშირად <i>Often</i>	არ მეხება <i>N/A</i>	უარი განაცხადა <i>Refused</i>
B026. თუ რამეს ვაშაგებ, თავს ღამნაშაგედ არ ვგრძნობ. <i>I don't feel guilty after doing something I shouldn't.</i>	0	1	2	8	9
B027. მშურს სხვების. <i>I am jealous of others.</i>	0	1	2	8	9
B028. ცუდად ვიქცევი. <i>I misbehave.</i>	0	1	2	8	9
B029. skolis gareT რაღაცების მეშინია. I am afraid of certain things, outside school.	0	1	2	8	9
B029a: აღწერეთ: <i>Describe</i>					
B030. მეშინია სკოლაში სიარულის. <i>I am afraid of going to school.</i>	0	1	2	8	9
B031. მეშინია, რომ შეიძლება რაიმე ცუდი გავიფიქრო ან გავაკეთო. <i>I am afraid I might think or do something bad.</i>	0	1	2	8	9
B032. უნდა ვიყო საუკეთესო. <i>I have to be the best.</i>	0	1	2	8	9
B032a. ოჯახი მაიძულებს, რომ ვიმათხოვრო. <i>My family forces me to beg.</i>	0	1	2	8	9
B033. ვგრძნობ, რომ არავის ვუყვარვარ. <i>I feel that no one loves me.</i>	0	1	2	8	9
B034. მგონია, რომ სხვები ჩემს წინააღმდეგ არიან. <i>I feel that others are out to get me.</i>	0	1	2	8	9
B035. ვგრძნობ, რომ სხვებზე უარესი ვარ და ჩემგან არაფერი გამოვა. <i>I feel worthless and inferior to others.</i>	0	1	2	8	9
B036. SeiZleba შემთხვევით რამე ვიტკინო. <i>I may accidentally get hurt.</i>	0	1	2	8	9
B037. ჩხუბისთავი ვარ. <i>I get in fights.</i>	0	1	2	8	9
B038. მაღიზიანებენ ხოლმე. <i>I get teased.</i>	0	1	2	8	9
B039. მე ისეთ ბავშვებთან ერთად ვარ, რომლებიც რაიმე ხიფათში ეხვევიან. <i>I hang around with kids who get in trouble.</i>	0	1	2	8	9
B040. მე ისეთი ხმაური და ხმები ჩამესმის, რაც სხვა ადამიანებს არ ესმით. <i>I hear sounds or voices that other people think aren't there.</i>	0	1	2	8	9
B040a: აღწერეთ: <i>Describe</i>					
B041. დაუფიქრებლად ვიქცევი. <i>I act without prior thinking.</i>	0	1	2	8	9
B042. არ მსიამოვნებს სხვებთან ერთად ყოფნა. <i>I don't enjoy being with others.</i>	0	1	2	8	9
B043. ვიტყუები. <i>I lie.</i>	0	1	2	8	9

ნიშნები, სიმტომები და ქცევები <i>Signs, Symptoms and Behaviors</i>	არასოდეს <i>Never</i>	ზოგჯერ <i>Sometimes</i>	ხშირად <i>Often</i>	არ მეხება <i>N/A</i>	უარი განაცხადა <i>Refused</i>
B044. ფრჩხილებს ვიკენეტ. <i>I bite my fingernails.</i>	0	1	2	8	9
B045. განერვიულებული ვარ. <i>I am nervous.</i>	0	1	2	8	9
B046. მაყვს თიკები (sxexilis uneblie moZraobebi). <i>I have tics (involuntary movements).</i>	0	1	2	8	9
B046a: აღწერეთ: <i>Describe</i>					
B047. მესიზმრება კოშმარები (ცული სიზმრები). <i>I have nightmares (bad dreams).</i>	0	1	2	8	9
B047a. ჩემი ოჯახის წევრ(ებ)ი მცემენ. <i>I am beaten by a family member(s).</i>	0	1	2	8	9
B048. სხვა ბავშვებს არ მოეწონვარ. <i>I am not liked by other kids.</i>	0	1	2	8	9
B049. ზოგი რამეს სხვა ბავშვებზე უკეთესად ვაკეთებ. <i>I do certain things better than other kids.</i>	0	1	2	8	9
B050. ძალიან მშიშარა ან შფოთიანი ვარ. <i>I am too fearful or anxious.</i>	0	1	2	8	9
B051. თავბრუს მახვევა. <i>I feel dizzy or lightheaded.</i>	0	1	2	8	9
B052. თავს ძალიან დამნაშავედ ვგრძნობ. <i>I feel too guilty.</i>	0	1	2	8	9
B053. ძალიან ბევრს ვჭამ. <i>I eat too much.</i>	0	1	2	8	9
B054. უმიზეზოდ ვიღლები. <i>I feel overtired without good reason.</i>	0	1	2	8	9
B055. ზედმეტი წონა მაქვს. <i>I am overweight.</i>	0	1	2	8	9
B056. მაქვს ფიზიკური ჯანმრთელობის პრობლემები და არ ვიცი რის გამო: <i>I have physical health problems and I don't know why.</i>	0	1	2	8	9
B056a. ტკივილი (მუცლისა და თავის ტკივილის გარდა) <i>Aches or pains (not stomach or headaches)</i>	0	1	2	8	9
B056b. თავის ტკივილი <i>Headaches</i>	0	1	2	8	9
B056c. გულისრევის შეგრძნება <i>Nausea, feel sick</i>	0	1	2	8	9
B056d. მხედველობის პრობლემები (არ ითვლება, თუ სათვალეებს იყენებს) <i>Problems with eyes (not if corrected by glasses)</i>	0	1	2	8	9
B056da: აღწერეთ: <i>Describe</i>					
B056e. გამონაყარი ან კანის სხვა პრობლემები <i>Rashes or other skin problems.</i>	0	1	2	8	9

ნიშნები, სიმპტომები და ქცევები <i>Signs, Symptoms and Behaviors</i>	არასოდეს <i>Never</i>	ზოგჯერ <i>Sometimes</i>	ხშირად <i>Often</i>	არ მეხება <i>N/A</i>	უარი განაცხადა <i>Refused</i>
B056f. მუცლის ტკივილი <i>Stomachaches.</i>	0	1	2	8	9
B056g. პირღებინება <i>Vomiting, throwing up</i>	0	1	2	8	9
B056h. სხვა <i>Other</i>	0	1	2	8	9
B056ha: აღწერეთ <i>Describe</i>					
B057. გცემ სხვა ადამიანებს. <i>I physically attack people.</i>	0	1	2	8	9
B057a. ჩემს მშობლებს (მშობელს) ვუყვარვარ და თბილად მექცევა. <i>My parent(s) love me and show me warmth.</i>	0	1	2	8	9
B058. კანს ან სხეულის სხვა ნაწილებს ვიწიწკნი. <i>I pick my skin or other parts of my body.</i>	0	1	2	8	9
B058a: აღწერეთ <i>Describe</i>					
B059. შემიძლია, საკმაოდ მეგობრული ვიყო. <i>I can be pretty friendly.</i>	0	1	2	8	9
B060. მიყვარს ახალი რაღაცის კეთება და განცდა. <i>I like to try and do new things.</i>	0	1	2	8	9
B061. ცუდად ვსწავლობ. <i>My school work is poor.</i>	0	1	2	8	9
B062. მოუხერხებელი და მოუქნელი ვარ. <i>I am poorly coordinated or clumsy.</i>	0	1	2	8	9
B063. <i>mirCevnia, Cemze ufros bavSvebTan erTad viyo, vidre Cems TanatolebTan.</i> <i>I would rather be with older kids than kids my own age.</i>	0	1	2	8	9
B064. <i>mirCevnia, Cemze umcros bavSvebTan erTad viyo, vidre Cems TanatolebTan.</i> <i>I would rather be with younger kids than kids my own age.</i>	0	1	2	8	9
B065. <i>lparakze uars vacxadeb.</i> <i>I refuse to talk.</i>	0	1	2	8	9
B066. <i>erTsa da imave moqmedebas bevrjer vimeoreb.</i> <i>I repeat one and the same things over and over.</i>	0	1	2	8	9
B066a: აღწერეთ <i>Describe</i>					
B067. სახლიდან გავრბივარ. <i>I run away from home.</i>	0	1	2	8	9
B068. ბევრს ვყვირი. <i>I shout a lot.</i>	0	1	2	8	9
B069. ჩაკეტილი ადამიანი ვარ და საიდუმლოს სხვას არ ვანდობ. <i>I am secretive or keep things to myself.</i>	0	1	2	8	9

ნიშნები, სიმპტომები და ქცევები <i>Signs, Symptoms and Behaviors</i>	არასოდეს <i>Never</i>	ზოგჯერ <i>Sometimes</i>	ხშირად <i>Often</i>	არ მესება <i>N/A</i>	უარი განაცხადა <i>Refused</i>
B070. ისეთ რაღაცებს ვხედავ, რაც სხვებს ჰგონიათ, რომ იქ არ არის. <i>I see things that other people think aren't there.</i>	0	1	2	8	9
B070a: აღწერეთ <i>Describe</i>					
B071. მორიდებული ვარ და ადვილად ვვარდები უხერხულ მდგომარეობაში. <i>I am self-conscious or easily embarrassed.</i>	0	1	2	8	9
B072. ცეცხლს ვუქვამ ხოლმე რაღაცებს. <i>I set fires.</i>	0	1	2	8	9
B072a. ჩემს მშობლებს (მშობელს) აქვთ ჩემთან ურთიერთობა. <i>My parent(s) communicate with me.</i>	0	1	2	8	9
B073. მარჯვე ხელები მაქვს. <i>I can work well with my hands.</i>	0	1	2	8	9
B074. თავს ვიწონებ ან ვმაიმუნობ. <i>I show off or clown around.</i>	0	1	2	8	9
B075. ძალიან მორცხვი და მორიდებული ვარ. <i>I am too shy or timid.</i>	0	1	2	8	9
B076. მე უფრო ნაკლები მძინავს, ვიდრე სხვა ბავშვების უმრავლესობას. <i>I sleep less than most kids.</i>	0	1	2	8	9
B077. დღის და/ან ღამის განმავლობაში უფრო მეტი მძინავს, ვიდრე სხვა ბავშვებს. <i>I sleep more than most kids day and/or night.</i>	0	1	2	8	9
B077a: აღწერეთ <i>Describe</i>					
B078. უყურადღებო ვარ და ყურადღება ადვილად მეფანტება. <i>I am inattentive or easily distracted.</i>	0	1	2	8	9
B079. მეტყველება მიჭირს. <i>I have a speech problem.</i>	0	1	2	8	9
B079a: აღწერეთ <i>Describe</i>					
B080. ვიცავ საკუთარ უფლებებს. <i>I stand up for my rights.</i>	0	1	2	8	9
B081. სახლში ვკურდობ. <i>I steal at home.</i>	0	1	2	8	9
B082. სახლის გარდა სხვა ადგილებში ვკურდობ. <i>I steal from places other than home.</i>	0	1	2	8	9
B083. ბევრ ისეთ ნივთს ვინახავ, რომელიც არ მჭირდება. <i>I store up too many things I don't need.</i>	0	1	2	8	9
B083a: აღწერეთ <i>Describe</i>					

ნიშნები, სიმპტომები და ქცევები <i>Signs, Symptoms and Behaviors</i>	არასოდეს <i>Never</i>	ზოგჯერ <i>Sometimes</i>	ხშირად <i>Often</i>	არ მესება <i>N/A</i>	უარი განაცხადა <i>Refused</i>
B084. ისე ვიქცევი, რომ ეს სხვებს უცნაურად მიაჩნიათ. <i>I do things other people think are strange.</i>	0	1	2	8	9
B084a: აღწერეთ <i>Describe</i>					
B085. ისეთი აზრები მაქვს, რაც სხვა ადამიანებს უცნაურად მოეჩვენებოდათ. <i>I have thoughts that other people would think are strange.</i>	0	1	2	8	9
B085a: აღწერეთ <i>Describe</i>					
B086. ჯიუტი ვარ. <i>I am stubborn.</i>	0	1	2	8	9
B087. ჩემი ხასიათი და გრძობები მოულოდნელად იცვლება. <i>My moods or feelings change suddenly.</i>	0	1	2	8	9
B088. მიხარია ადამიანების გარემოცვაში ყოფნა. <i>I enjoy being with people.</i>	0	1	2	8	9
B089. ეჭვიანი ვარ. <i>I am suspicious.</i>	0	1	2	8	9
B090. ვიგინები ან ცუდ სიტყვებს ვხმარობ. <i>I swear or use dirty language.</i>	0	1	2	8	9
B091. ვფიქრობ თავის მოკვლაზე. <i>I think about killing myself.</i>	0	1	2	8	9
B092. მიყვარს სხვების გაცინება. <i>I like to make others laugh.</i>	0	1	2	8	9
B093. ძალიან ბევრს ვლაპარაკობ. <i>I talk too much.</i>	0	1	2	8	9
B094. სხვებს ხშირად ვაღიზიანებ. <i>I tease others a lot.</i>	0	1	2	8	9
B095. ფიცხი ვარ. <i>I have a hot temper.</i>	0	1	2	8	9
B096. სექსის შესახებ ძალიან ბევრს ვფიქრობ. <i>I think about sex too much.</i>	0	1	2	8	9
B097. ვიმუქრები, რომ სხვა ადამიანებს ზიანს მივაყენებ. <i>I threaten to hurt people.</i>	0	1	2	8	9
B098. მიყვარს, როცა სხვებს ვეხმარები. <i>I like to help others.</i>	0	1	2	8	9
B099. სიგარეტს ვწევი. <i>I smoke cigarettes.</i>	0	1	2	8	9
B100. ცუდად მძინავს. <i>I have trouble sleeping.</i>	0	1	2	8	9
B100a: აღწერეთ <i>Describe</i>					
B101. ვაცდენ გაკვეთილებს, სკოლას. <i>I cut classes or skip school.</i>	0	1	2	8	9
B102. ადვილად ვიღლები. <i>I easily get tired.</i>	0	1	2	8	9

ნიშნები, სიმპტომები და ქცევები <i>Signs, Symptoms and Behaviors</i>	არასოდეს <i>Never</i>	ზოგჯერ <i>Sometimes</i>	ხშირად <i>Often</i>	არ მესება <i>N/A</i>	უარი განაცხადა <i>Refused</i>
B103. ვარ მოწყენილი, სევდიანი ან დეპრესიული. <i>I am unhappy, sad or depressed.</i>	0	1	2	8	9
B104. მე უფრო ხმაურიანი ვარ, ვიდრე სხვა ბავშვები. <i>I am louder than other kids.</i>	0	1	2	8	9
B104a. მე და ჩემი მშობლები (მშობელი) ერთად ვთამაშობთ ან ვერთობით. <i>My parent(s) and I play or have fun together</i>	0	1	2	8	9
B105. წამლებს და სხვა ნივთიერებებს, მაგ: წებოს ან სხვა, ვხმარებ არასამედიცინო დანიშნულებით (აქ არ შედის თამბაქო და ალკოჰოლი). <i>I use drugs, for example: glue and other things, for non-medical purposes (don't include alcohol or tobacco).</i>	0	1	2	8	9
B105a: აღწერეთ <i>Describe</i>					
B106. მიყვარს, როცა სხვების მიმართ სამართლიანი ვარ. <i>I like to be fair to others.</i>	0	1	2	8	9
B107. მსიამოვნებს კარგი ხუმრობები. <i>I enjoy a good joke.</i>	0	1	2	8	9
B108. მიყვარს, როცა ცხოვრებას იოლად ვუყურებ. <i>I like to take life easy.</i>	0	1	2	8	9
B109. ვცდილობ, დავეხმარო სხვებს, როცა შემიძლია. <i>I try to help other people when I can.</i>	0	1	2	8	9
B110. ვნატრობ, ბიჭი/გოგო ვიყო. <i>I wish I were of the opposite sex.</i>	0	1	2	8	9
B111. ვცდილობ თავი ავარიდო ადამიანებთან ახლო ურთიერთობებს. <i>I keep from getting involved with others.</i>	0	1	2	8	9
B112. ბევრს ვნერვიულობ. <i>I worry a lot.</i>	0	1	2	8	9
B112a. მინდა თავისუფალი ცხოვრება. <i>I want to live a free life.</i>	0	1	2	8	9
B112b. ყურადღება მაკლია. <i>They lack attention.</i>	0	1	2	8	9
B112c. მჩაგრავენ ჩემი მდგომარეობის გამო. <i>I am oppressed because of my situation</i>	0	1	2	8	9
B112d. ცხოვრებაზე ხელი მაქვს ჩაქნეული. <i>I have slipped hand at life.</i>	0	1	2	8	9
B112e. რაღაცებს მაბრალებენ. <i>They blame me for things.</i>	0	1	2	8	9



## Appendix B: Study Explanation Read to Interviewees Prior to Interview

### კვლევის განმარტება *Explanation of Study*

თქვენ კვლევაში მონაწილეობის მიღება გთხოვეს. აღნიშნული კვლევის მიზანია იმ ინფორმაციის მოპოვება, რომლის მეშვეობითაც ორგანიზაციები, რომლებიც ბავშვებს ეხმარებიან, უკეთ შეასრულებენ თავიანთ სამუშაოს. ეს ორგანიზაციები ცდილობენ დაეხმარონ ბავშვებს, რომლებიც ქუჩაში, ან ბავშვთა სახლებში ცხოვრობენ, ან სხვა სახის პრობლემები აქვთ. თქვენ სწორედ ამიტომ შეგარჩიეს.

*You are being asked to be included in a study. The purpose of the study is to provide information that will enable organizations that assist children to improve their work. These organizations try to help children who live in the street, or in orphanages, or who otherwise have problems. This is why you were chosen.*

შეგიძლიათ უარი თქვათ ინტერვიუზე და ეს გავლენას არ მოახდენს იმ დახმარებაზე, რასაც ფედერაციიდან „გადავარჩინოთ ბავშვები“, ან სხვა ორგანიზაციიდან იღებთ. თუ თანახმა ხართ ინტერვიუზე, შეგიძლიათ ნებისმიერ დროს შეწყვიტოთ ეს ინტერვიუ, ან არ გასცეთ პასუხი რომელიმე შეკითხვას.

*You can refuse to be interviewed and this will not affect any assistance you receive from Save the Children or any other organization. If you agree to be interviewed you can stop the interview at any time or refuse to answer any questions.*

თქვენგან მიღებულ ინფორმაციას საიდუმლოდ შევინახავთ და არავის ვეტყვით.  
*We will keep the information received from you secret, and will not tell it to anybody.*

ზოგიერთი კითხვა, რომელსაც დაგისვამთ, ეხება საკითხებს, რომლებიც შეიძლება სასაცილოდ ან უცნაურად მოგეჩვენოთ, ან თავი უხერხულად იგრძნოთ, მაგრამ ეს საკითხები პრობლემატურია ზოგიერთი ბავშვისათვის. თუმცა ჩვენ არ ვიცით კონკრეტულად რომელი ბავშვისათვის არის ისინი პრობლემატური, ამიტომ ამ კითხვას ყველა ბავშვს ვეკითხებით.

ეცადე უპასუხო ყველა კითხვას, მაგრამ გახსოვდეს, რომ შეგიძლია უარი თქვა კითხვაზე პასუხის გაცემაზე.

*Some questions which we are going to ask you may sound to you funny or strange or you may feel embarrassed. But these topics may be problems for other children, but we don't know to which children exactly. So we have to ask these questions to all children.*

*Try to answer all questions, but remember, that you can refuse to answer any question.*

## **Appendix C: Explanation of Reliability and Validity Concepts**

### Reliability

Reliability refers to the extent to which different measures of the same concept agree with each other. It can refer to measurements taken at the same time, or different times. To be useful an instrument must have good local reliability, which must therefore be tested whenever a questionnaire is changed (including translation) or used among a new population.

#### *Test-Retest and Interrater Reliability*

Testing reliability over time is also useful. This is called test-retest reliability. The questionnaire is given to the same subject on two different occasions. It is usually done at least a day later, to reduce the effect of memory on the responses, but not too long because what is being measured may actually change (mood, for example). Therefore, the repeat interview is usually done 1-7 days after the first interview. For this study the second interview was done by a different interviewer. Comparison of the results of the first and second interviews is therefore a measure of both test-retest and interrater reliability. To make this comparison, a summary scale is first created using all the questions on the same topic (in this case each of the YSR subscales) and calculated for both the first and second interview. Test-retest/interrater reliability is tested by measuring correlations between these scores. Opinions vary as to what is an acceptable score, although correlations above 0.7 are considered desirable for test-retest reliability alone. A problem arises in interpreting low scores. These may be due to a poor instrument, or to using different interviewers, or because the concept being measured has changed. Partly for these reasons, test-retest and interrater reliability are not generally considered as important as internal consistency reliability (Streiner et al 1995).

#### *Internal Consistency Reliability*

This refers to how well questions measuring the same underlying concept on the same occasion agree with each other. For example, two questions that measure different aspects of depression should agree with each other in that the same individual should score high or low on both. Agreement is measured quantitatively by correlations. For questionnaires with many questions measuring the same concept, a large number of correlations would be required to check the agreement of every question with every other question, and some summary of these correlations would be needed. Cronbach's's alpha is a statistical measure which provides this. It is a single figure which summarizes the average correlation between all pairs of questions in a questionnaire. Cronbach's's alphas should be above 0.7 and ideally between 0.8-0.9.<sup>3</sup> The reliability of each question can be assessed by calculating the alpha with and without it. Significant increases in alpha without the question would suggest that the question is not measuring the same thing as the other questions, and should be removed. Studying the effect of each question in this way is called Item Analysis.

---

<sup>3</sup> Above 0.9 suggests that the questionnaires has too many questions and some could be eliminated (Streiner et al, 1995).

## Validity

Validity refers to the extent to which the measurement provided by an instrument agrees with the correct measurement. Instruments may be reliable but not valid, if they consistently give the same (but wrong) measurement and so both reliability and validity must be measured to assess instrument accuracy. There are two aspects of validity to be considered when testing a questionnaire:

### *Content validity*

This refers to whether the instrument is considered by experts to be appropriate for measuring what it is supposed to measure. Part of content validity is whether experts believe that the questionnaire covers all the important aspects of the concept being studied. In the course of this study and the previous qualitative studies we consulted two groups of 'experts.' The first group were psychologists and clinicians (including the BU/JHU faculty) who helped us choose the YSR as an appropriate child measure for this population. The second group was the local population, through the qualitative studies; the YSR was chosen to match as closely as possible the psychosocial issues that emerged in those studies.

### *Criterion validity*

This refers to the agreement between the questionnaire and an external measure (criterion) of the same construct known to be accurate. In other words, comparing the questionnaire with a 'gold standard.' In this study the 'gold standard' was the assessment of the psychologists at each of the study sites. We also used an alternative local standard -- assessment by the child themselves as to whether they have a behavioral problem.