DESIGN, IMPLEMENTATION, MONITORING, AND EVALUATION OF CROSS-CULTURAL HIV-RELATED MENTAL HEALTH AND PSYCHOSOCIAL ASSISTANCE PROGRAMS:

A USER'S MANUAL FOR RESEARCHERS AND PROGRAM IMPLEMENTERS

(ADULT VERSION)

MODULE 2:

DEVELOPING QUANTITATIVE TOOLS

Applied Mental Health Research Group
Center for Refugee and Disaster Response
Johns Hopkins University Bloomberg School of Public Health

The USAID | Project SEARCH, Task Order No.2, is funded by the U.S. Agency for International Development under Contract No. GHH-I-00-07-00032-00, beginning September 30, 2008, and supported by the President's Emergency Plan for AIDS Relief. The Research to Prevention (R2P) Project is led by the Johns Hopkins Center for Global Health and managed by the Johns Hopkins Bloomberg School of Public Health Center for Communication Programs (CCP).

ACKNOWLEDGEMENTS

The development of the original version of this manual was supported by a series of grants from the Victims of Torture Program (VOT) at USAID to The Applied Mental Health Research Group (AMHR) at Johns Hopkins University (JHU). VOT also provided support for the use and refinement of the approach described here in field-based studies with implementing organizations assisting populations affected by torture and violence.

This version is an expansion and adaptation of that original manual with specific reference to research on psychosocial and mental health problems related to HIV. This adaptation process was implemented by AMHR through Project SEARCH | USAID: Research to Prevention (R2P), funded by the United States Agency for International Development (USAID) through task order # GHH-I¬00-07-00032.

An adaptation of the original VOT Module 1 for children has also been written. This was supported by World Vision USA under their Children in Crisis initiative. Under that same initiative, World Vision USA has also supported collaborations between JHU and World Vision national offices in the use of these methods to inform programming for street children and those sexually abused in multiple countries. World Vision USA also supported previous work with adults that also contributed to the development of the methods originally described in the adult version of the module.

Without the support of both USAID/VOT and World Vision USA, the development of the methods described here and the production of this manual would not have been possible.

CONTENTS

Acknowledgements 2

Acronyms 5

Introduction to the Manual 7

Layout of the Manual 8

Intended Users 8

The DIME Model 10

A. Introduction To Module 2 15

A.1 Purpose and Overview of Module 2 15

B. Methodology 16

B.1 Steps in the Instrument Development Process 16

B.2 Drafting the Instrument 16

B.2.1 Demographic Section 17

B.2.2 Function section 17

B.2.3 Mental Health and Psychosocial Problem Section 20

B.2.4 HIV-Specific topics and issues 22

B.2.5 Instructions and Response Categories 23

B.2.6 Translate the Instrument 24

B.3 Interviewers and Supervisors for the Pilot and Validity Study 26

B.3.1 Interviewer Qualifications 26

B.3.2 Supervisor Qualifications 27

B.3.3 Study Director Qualifications 27

B.3.4 Training Interviewers and Superviors 27

B.4 Pilot Study 30

B.4.1 Purpose of the Pilot Study 30

B.4.2 Interviewees 30

B.4.3 Process 30

B.5 Reliability and Validity Testing 32

B.5.1 Identifying Interviewees 32

B.5.2 Interviewing Process 36

B.5.3 Analysis 37

B.5.3.3 Test-retest and Inter-rater Reliability 40

B.6 Finalizing the Instrument for Local Use 41

Appendix A: Example Field Schedule 44

Appendix B. Resource List 45

Time 45

Personnel 45

Transport 46

Training and office materials 46

Interviewer materials 47

Appendix C: Verbal Consent Form 48

Verbal Consent Form for Research for Research Study 48

Appendix D: Example of Draft Assessment INSTRUMENT FOR Testing (Kurdistan, Iraq) 50

Section A: Functionality Assessment 62

Section B: Symptom Assessment Instrument 65

Appendix E: Examples of Non Verbal Response Cards 69

Nonverbal Response Card for Locally-Defined Function Questions* 69

Appendix F: Pilot Study Post-Interview Form 71

Appendix G: Example of interviewee Lists (Multiple Problems) 72

ACRONYMS

AIDS	Acquired immunodeficiency syndrome
AMHR	Applied Mental Health Research
ART	Antiretroviral therapy
BA	Behavioral activation
CBI	Components based intervention
CBT	Cognitive Behavior Therapy
CD4	T-helper cell targeted by HIV
CDC	Centers for Disease Control
CPT	Cognitive Processing Therapy
CSA	Child sexual abuse
DHS	Demographic health survey
DIME	Design, implementation, monitoring and evaluation
DRC	Democratic Republic of Congo
EBT	Evidence Based Treatment
FG	Focus Group
FL	Free List
GBV	Gender Based Violence
HIN	Health information network
HIV	Human immunodeficiency virus
IDU	Injecting drug user
IPT	Interpersonal Therapy for Depression
IRB	Institutional Review Board
JHU	Johns Hopkins University
KAP	Knowledge, attitudes and practices
KI	Key Informant
LGBT	Lesbian, gay, bisexual, transgender
LMIC	Low and middle income countries
MEMS	Medication Event Monitoring System
MI	Motivational interviewing
MOH	Ministry of Health
MSM	Men who have sex with men
NGO	Non-governmental organizations
OVC	Orphans and vulnerable children
PE	Prolonged Exposure
PLWHA	People living with HIV/AIDS
POFO	Positive Outcome for Orphans Study
PPS	Probability proportional to size

PRA	Participatory rural appraisal
PTSD	Posttraumatic stress disorder
R2P	Research to Prevention
RCT	Randomized Controlled Trial
REC	Research Ethics Committee
SEARCH	Supporting Evaluation and Research to Combat HIV/AIDS
SES	Social economic status
SMS	Short Message Service
SOW	Scope of Work
SRP	Stress related response
STI	Sexually transmitted infections
SW	Sex worker
USAID	United States Agency for International Development
TFCBT	Trauma Focused Cognitive Behavior Therapy
VCT	Voluntary counseling and testing
VOT	Victims of Torture Program
WHO	World Health Organization

INTRODUCTION TO THE MANUAL

The Manual for Design, Implementation, Monitoring, and Evaluation of Cross-Cultural HIV-Related Mental Health and Psychosocial Assistance Programs: A User's Manual for Researchers and Program Implementers has been written to assist researchers and organizations developing and implementing programs in HIV affected populations to 1) identify and measure the impact and prevalence of mental health and psychosocial problems in the populations they seek to serve; 2) to develop or adapt appropriate interventions to address these problems; and 3) to measure the impact of these interventions. The Manual consists of 6 modules. Collectively, the modules describe a process of program **d**esign, **i**mplementation, **m**onitoring, and **e**valuation (DIME) that has been developed and used by the authors since 2000. The modules may be used in sequence, to follow the life of a project, or as stand-alone units to address a specific project need.

- **Module 1** describes procedures for a qualitative assessment to identify priority problems from the local perspective.
- **Module 2** provides guidance in the development and validity testing of tools to measure these priority problems.
- **Module 3** describes population-based assessments to gauge prevalence and severity of the priority problems using the instrument developed in Module 2.
- **Module 4** describes a process for overall design of a program to address the priority problems, including design of program monitoring and evaluation.
- **Module 5** outlines the selection, adaptation, and implementation of interventions.
- **Module 6** describes procedures for assessing intervention impacts.

Definition Box

Intervention(s): Service(s)/activity(ies) directly benefitting the client

Program: The intervention(s) and all ancillary activities necessary to support the intervention(s): logistics, finance monitoring and evaluation, etc.

LAYOUT OF THE MANUAL

Modules are presented in narrative form, with extensive use of subheadings. With the exception of text boxes, each section and each paragraph is meant to be read sequentially. Additional material that is useful as examples of concepts or expansion on subjects discussed in the text has been included in text boxes. Examples of study materials that may be adapted for use in an actual study are placed separately as appendices.

This symbol indicates that what follows is a critical requirement or constraint.

INTENDED USERS

This manual is primarily intended for researchers and groups responsible for mental health and psychosocial interventions for HIV affected populations, such as government providers and non-governmental organizations (NGOs).

The methods described in each module are intended to be within the typical budget, resources, and time constraints of organizations that normally focus on implementation rather than data collection. The approach is designed to be used in a limited area among a population with a homogenous language, culture, and similar circumstances. In areas containing populations with a variety of languages, cultures, and environments, the approach described in this manual should be used separately with each group. For this reason, the authors have focused on developing a process that is rapid and relatively inexpensive.

This is meant as a **'user'** manual rather than a training manual. It is intended for use in the field by those who have previously received field-based training in its methods (or have similar training experience) and are now leading teams in their own sites. Such persons should either have some prior experience in qualitative and quantitative data collection methods (depending on the module being used) or lead teams with persons who have such experience.

THIS MANUAL IS NOT APPROPRIATE FOR 'OFF THE SHELF' USE WITHOUT PRIOR ON-THE-GROUND TRAINING OR SIMILAR EXPERIENCE. THOUGH WHAT IS PRESENTED HERE REPRESENTS WHAT THE AUTHORS HAVE FOUND TO WORK WELL TO DATE, FIELD SETTINGS VARY. USERS OF THE METHODS PRESENTED HERE NEED FIELD EXPERIENCE TO INTERPRET AND ADAPT THESE METHODS TO DIFFERENT SITUATIONS.

The authors have found that even with prior experience in data collection, individuals and organizations attempting to use the methods described here for the first time will have many important questions during the process that cannot be addressed in the manual itself. Answering these questions as they arise—and developing the skills required for using the approaches in different settings—is best done in a field-based training situation, with direct instruction in the course of supervised use of this approach among a local population. Even after training, organizations using this approach may want guidance and ad hoc assistance.

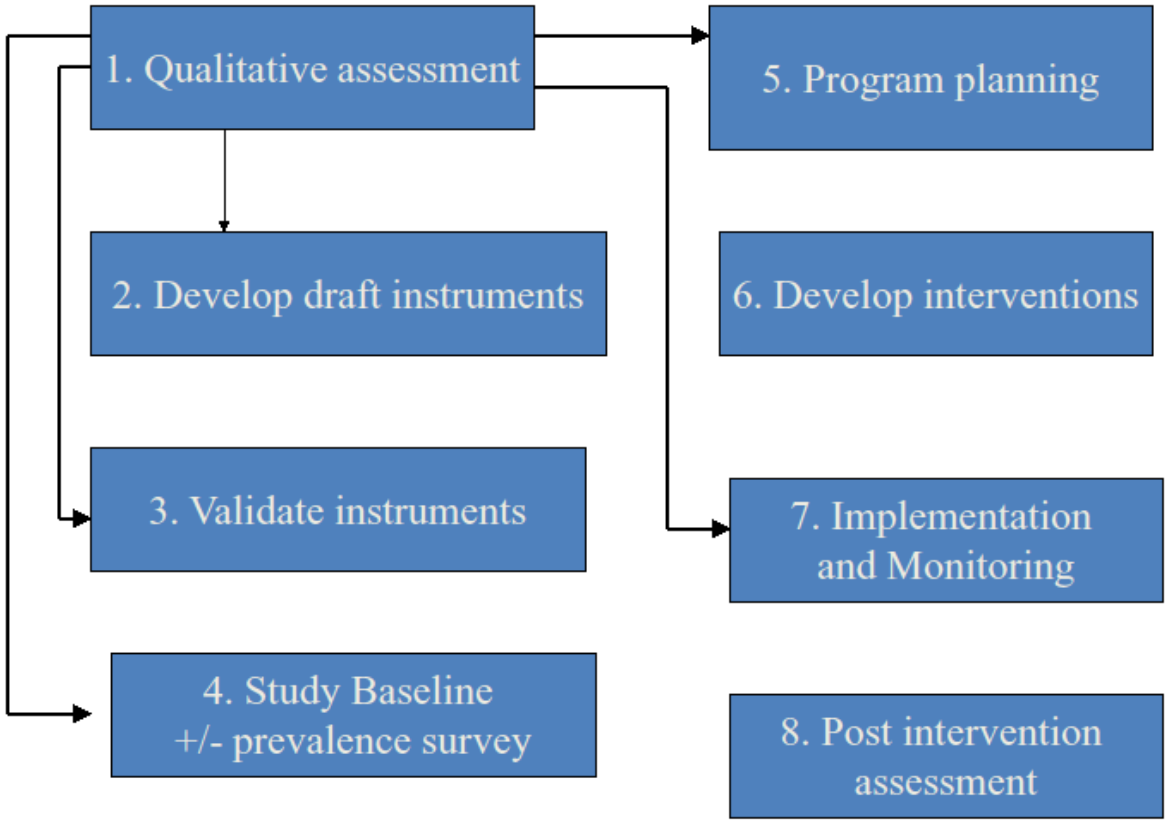
The authors would be pleased to discuss training and technical assistance with any interested organization or individual.

The manual does not contain detailed descriptions of commonly done research activities, such as quantitative interviewing, partly due to the expectation that organizations have persons experienced in these activities and partly because there are many other manuals available that describe these activities. Instead, the manual focuses on research activities or methods that are different from commonly used approaches. For example, Module 1 contains much more information on interviewing than the other modules because the qualitative methods used in Module 1 are less commonly used than quantitative methods.

THE DIME MODEL

The diagram below outlines the steps of the **d**esign, **i**mplementation, **m**onitoring, and **e**valuation (**DIME**) process described in this manual. Qualitative data collection (Module 1) is the first step in the process and the diagram indicates which of the subsequent steps (2-8) are informed by qualitative data. A brief description of each step follows.

Figure 1: Steps of the DIME Process



1. **Qualitative Assessment to identify and describe priority HIV-related mental health and psychosocial problems: (Module 1)**

   Variations in culture and environment affect how people understand the mental health and psychosocial problems related to HIV. By *understand,* we mean how these problems are described, how they are prioritized, their perceived causes, and how people currently cope with them. This information is vital in selecting problems that are important to local people, accurately communicating with them about these problems, and identifying interventions that are likely to be acceptable and feasible for local people and therefore effective and sustainable.

2. **Develop draft instruments to assess priority HIV-related mental health and psychosocial problems: (Module 2)**

   Having decided which problems the program will address, we then draft quantitative assessment instruments to address these problems. These instruments have various uses, depending on the program: conducting community or clinic-based surveys; screening persons for inclusion in a specific intervention program (for programs where not all people will be served); identifying those with severe problems who may need specialized services including referral; and monitoring and evaluating the effectiveness of services by tracking changes in severity and/or prevalence of the problems identified.

   The process of drafting appropriate instruments includes reviewing the published literature for measures that have already been developed for the selected problems and comparing available measures with the qualitative data to select the measure or measures that best match how local people describe the problem. These measures are then adapted to better fit local concepts.

   Drafting includes translation. Terminology suggested by translators often differs from that used by local populations, particularly by poor and uneducated people. Therefore, qualitative data is preferred as the best source for translating key concepts. Employing the words and phrases that local people actually use (as identified in the qualitative data) will improve the clarity of the instruments, thereby improving their acceptability and accuracy. The translators are instructed to utilize the qualitative data to directly translate all signs, symptoms, problems and topics in the instruments that were mentioned by interviewees in the qualitative study using the same words found in the qualitative data. Only where concepts are not mentioned in the qualitative data do the translators themselves choose the appropriate terms.

3. Validate draft instrument(s): (Module 2)

Once translated, the draft instrument(s) must be piloted and tested for ease of use, clarity, acceptance (both by interviewers and interviewees), and accuracy in the field. Accuracy refers to reliability and validity, which in turn refer to whether the instrument gives the same result with repeated use or use by different interviewers (reliability), and whether it measures what it is supposed to measure (validity). Testing involves interviews with members of the target population using the assessment instrument and analyzing the results.

Validity and reliability testing are particularly important with psychosocial and mental health measures, where assessment is based on the interview alone (i.e., there are no laboratory or other tests). A tool that is not accurate can lead to inappropriate inclusion/exclusion of intervention participants and also provide incorrect conclusions about need and program impact.

4. Study baseline +/-prevalence surveys: (Module 3)

Both baseline assessments and prevalence surveys are based on the instruments developed in steps 2 and 3. Baseline assessments refer to interviews done using the instrument in order to establish the eligibility of individuals for participation in an intervention program. Prevalence surveys perform the same function at the population level to measure the percentage and numbers of eligible (i.e., affected) persons in the population as well as giving some indication about the variation in severity of problems at the population level.

5. Overall program planning: (Module 4)

This includes planning the program goals and objectives and the strategy and the type of intervention(s) for achieving these. It also includes the development of process and impact indicators and the overall program work plan**.**

6. Develop interventions to address the identified HIV-related mental health and psychosocial problems: (Module 5)

The qualitative data on the perceived causes of problems and how those affected cope with the problems are critical to intervention design. Interventions need to address the perceived causes of priority problems (or explain why they do not) in order to make sense and therefore inspire both confidence and cooperation. The more closely interventions can match the ways in which people currently think about and address the

selected problems, the more likely the interventions are to be acceptable to them. Where there are differences, they need to be explained and agreed upon by the local population. For example, using counseling to address a problem that is thought to be caused by poverty will take some explaining.

7. **Implementation and Monitoring: (Modules 4 and 5)**

   This refers to the implementation and monitoring of the intervention and the overall program. It includes procedures for iterative changes in the planned activities as needed, according to the monitoring data.

8. **Post intervention assessment: (Module 6)**.

   Upon completion of the intervention, participants are interviewed using qualitative methods to identify potentially important unexpected impacts of the program. They are also re-interviewed using the baseline quantitative instrument, to measure changes in the outcome indicators such as problem severity and function. Where possible, the amount of change is compared with the amount of change experienced by a control group, to determine the true program impact.

MODULE 2:

DEVELOPING QUANTITATIVE TOOLS

A. INTRODUCTION TO MODULE 2

A.1 PURPOSE AND OVERVIEW OF MODULE 2

Module 2 describes a process for developing a quantitative instrument(s) to assess the presence and severity of the priority mental and psychosocial problems identified in the preceding qualitative study (See Module 1). The resulting instrument is intended for use in surveys (Module 3) and as a screening tool for inclusion in programs (Module 4) and impact studies (Module 6).

Upon completing the methods described in this manual, program staff will be able to:

1. Draft a quantitative instrument
2. Translate the instrument
3. Pilot test the instrument
4. Evaluate the instrument's validity and reliability
5. Finalize the instrument for local use

The approach described in this module has been field-tested in a variety of countries since the original version was first developed in 1998. The approach is intended to be simple, rapid, efficient, and suitable for use in low resource environments. It is specifically intended for use in these environments by both implementing organizations and researchers.

B. METHODOLOGY

B.1 STEPS IN THE INSTRUMENT DEVELOPMENT PROCESS

Instrument Development consists of the following steps:

1. Drafting and translating the instrument
2. Pilot study
3. Study of validity and reliability
4. Instrument revision and finalization

The remainder of this module describes each of these steps in detail.

B.2 DRAFTING THE INSTRUMENT

The basic draft instrument has three standard sections:

1. Demographic Data

   The demographic data section of the instrument provides information to track the interviewees, to describe the population participating in the program research, and for sub-analyses of various individual characteristics.

2. Function assessment

   Reduced function is the major means by which mental and psychosocial problems adversely affect both the people who are themselves suffering from these problems as well as those around them. Therefore, assessing function and the ability to perform locally relevant activities is a major means of evaluating the negative impact of mental health and psychosocial problems, while improvement in this area can be considered one of the most important expected outcomes of mental health and psychosocial programs.

3. Assessment of mental and psychosocial problems

   This section is made up of the symptoms related to the priority mental health and psychosocial problems identified in the preceding qualitative study (See Module 1). Questions asked in this section of the instrument explore the presence and severity of the signs and symptoms that make up these problems.

In addition to the three standard sections described above, which are common to all instruments, additional HIV-specific sections can be added, depending on the target population

or the target issue being studied. An HIV-specific section may include questions on topics such as:

- Knowledge, attitude and practices (KAP)
- HIV infection status
- Health care utilization and health seeking behaviors
- Stigma
- Substance abuse

An example of a **draft** instrument is provided in Appendix D. This example was taken from a study in Kurdistan, Iraq and includes locally-defined function items for men and women, and questions on depression, anxiety, trauma, and traumatic grief as well as general demographic questions.

B.2.1 DEMOGRAPHIC SECTION

The choice of what demographic information to collect depends on the purpose of the study instrument. This information may include age, gender, location, ethnicity, education, and other characteristics of the interviewee. The demographic data has multiple uses: tracking program participants (i.e., double checking that the demographic information is consistent for the same ID number throughout the program); describing the characteristics of participants (i.e., double-checking that the demographic information is consistent for sub-analyses of outcomes based on specific characteristics of interest, such as whether there are gender differences in HIV-treatment adherence rates). Other tracking data includes the identity of the interviewer, the ID of the participant, and the date of the interview.

B.2.2 FUNCTION SECTION

Like mental health instruments, many of the existing function assessment instruments were created in developed countries. Others, like the WHO Disability Assessment Schedule (WHO, 2010) were developed specifically for cross-cultural use. The WHO DAS II (WHO, 2010) and SF-36 (Ware & Sherbourne, 1992) have been found to have good psychometric properties across many populations (for examples of studies see: Buist-Bouwman et al., 2008; Hoopman et al., 2009; Lam et al., 2008; Luciano et al., 2010)). However, being fixed instruments, they do not necessarily include or focus on the priority activities that are important to local people because these can vary among and within populations. Priority activities can vary by culture, situation, age, and gender. This requires the development of locally-defined function assessment instruments for different situations and/or cultures. This locally-defined function instrument can then be used alongside standard function instruments to assess both broad functional

categories comparable across cultures (standard instruments) and activities of local importance (the locally-defined measures). Examples of standard instruments (the WHO DAS II and the SF-36) are well described elsewhere (WHO, 2010; Ware & Sherbourne, 1992) and so are not described here.

The locally-defined function instruments that the authors have developed have generally focused on the activities that local people do to **care for or contribute to the wellbeing** of themselves, their families, and their communities. For each community the instrument is created using the six (6) composite lists of important activities obtained from the free listing exercise—lists of activities important for self, family and community– and the data from the focus groups (FGs) on functioning. The interviewers and study coordinator together review the free lists (FL) and FG data. For each of the categories of self, family, and community activities they select the most frequently mentioned activities that also meet the following four criteria:

1. Inability to do the activity will clearly affect others

   This criterion is optional as programs may decide that activities that clearly affect individuals are priorities regardless of their impact on others. On the other hand, if there are many activities reported and not all need be included in the instrument, priority should be given to those which affect the most people.

Example: *In Rwanda 'praying' was frequently mentioned in FLs as an activity that people did to care for themselves. However, it was not included as a function question because it was not clear that inability to pray would adversely affect others. On the other hand, 'washing oneself'* ***was*** *included as it was clear that if a person were unable to do this activity it would affect others by requiring others to do it for the affected person.*

2. The activity is not actually the same as another activity on the list (or part of another activity)

   This is best explained by the example below. Avoiding this issue is done by getting a description in the qualitative study of what the person actually does (in this case, earning money) rather than the end result (the children go to school).

Example: *In a previous study some men listed 'sending children to school' as an important activity in caring for the family. But their descriptions made it clear that their major*

activity in this regard was to earn enough the money to pay the fees. 'Sending children to school' was therefore removed since it was covered under the 'earning money' activity.

3. The activity is clearly done by a large majority of the target population
   The qualitative study identifies particular tasks that people often do. As with any specific task, there will be some people who do not do that task. For example, caring for one's children is a very important task for women in most countries, but does not apply to women who do not yet have children. If the target population consisted of many unmarried women this might not be a useful item for a function instrument. The example below describes how this principle can often result in the generation of separate function instruments by gender.

Example: *In a study of HIV-affected adults it was necessary to include function sections specific to men and women since they had different responsibilities and engaged in different activities. In a study that focused specifically on perinatal mental health problems, only women were interviewed and thus, only a women's function section was included.*

4. The activity is done frequently, such as daily or at least monthly.

The most frequently mentioned activity in each category (self, family and community) that meet the above criteria are then inserted into a template to form the local function assessment section (see example in Appendix D). The number of activities chosen for each category depends on the desired length of the instrument and the number of activities that are **frequently mentioned**. For example, if only three self-care activities were mentioned by more than a few FL interviewees or agreed to by FG participants, then only those three would be included in the instrument.

This manual has been developed based on a strong interest in how mental health and psychosocial problems affect individual, family, and community development. Therefore, the major focus in assessing function has been on the activities that people do to **care for or promote the wellbeing** of these groups. Aspects of HIV-specific functioning, such as coping with stigma, health seeking behaviors, etc., would be included in additional sections that focus on the specific outcomes of interest.

How to define 'healthy' functioning

The purpose of the local function section is to assess important activities that someone living under the same conditions but **not** affected by mental or psychosocial problems would regularly be able to do. In the context of HIV, the definition of healthy functioning may not be as simple. For example if the target population includes people suffering from severe mental illness or later stages of HIV infection, the level of functioning they could hope to achieve may not be the same as someone from the general population. For this reason, it is important to ensure that the FL and FG data (see Module 1) that is used to generate the function items is collected from populations similar to, or knowledgeable about, those being targeted for the research and programming. As the function assessment will be used to not only identify who is in need of services but also as an outcome measure, it is necessary to include activities that are achievable by the target population.

B.2.3 MENTAL HEALTH AND PSYCHOSOCIAL PROBLEM SECTION

The mental and psychosocial problem section assesses the priority problems identified in the preceding qualitative study. If the descriptions from that study are consistent with specific Western mental health and psychosocial concepts, this is initial evidence that these Western concepts are locally valid/applicable, and that existing instruments based on these Western concepts can be adapted for local use. Existing instruments are then reviewed and the one whose content most closely matches that of the qualitative data is selected for inclusion in the instrument. When there are important symptoms from the qualitative study that are not in the standard instrument, questions on these symptoms should be added using the existing question format. Typically, the selected instrument will include symptoms that did not emerge in the qualitative data. These symptoms are retained (see Note below). The final result is an expanded version of the existing instrument. An instrument is selected and adapted in this way for **each** of the priority problems identified in the qualitative study (i.e., If depression and anxiety were both identified as priority problems, then both a depression instrument and an anxiety instrument would be adapted.

1. A local syndrome identified in a qualitative study among a population includes the following symptoms: lack of energy, depressed mood, loss of interest in things, poor appetite, and reduced function. This syndrome is suggestive of the Western concept of *depression,* and that an existing depression instrument can be adapted for local use.

2. Informants from a qualitative study of torture survivors in northern Iraq described combinations of symptoms consistent with the following mental disorders described in other countries: Traumatic Grief, Major Depression, Generalized Anxiety Disorder, and PTSD, as well as psychosocial problems related to poor relationships within the family and marginalization from the wider Kurdish society. Based on these findings we included three existing questionnaires in our draft instrument section on mental and psychosocial problems (See Appendix D) that together reflect most of the symptoms they described:

a) The Hopkins Symptom Checklist for anxiety and depression (HSCL-25)
b) The symptom section of the Harvard Trauma Questionnaire (HTQ)
c) The Traumatic Grief Questionnaire.

NOTE:

There are a wide range of assessment tools used worldwide to assess various mental health and psychosocial problems. Each tool has its own strengths and weaknesses, and most have been validated among at least some populations. Some instruments are freely available while others require permission of the creators and/or a paid license. Literature and internet searches on the mental health problem to be addressed will provide the names of the most common instruments as well as the populations for which the instruments have been validated and used (in the case of the published literature).

A review of published articles related to the identified mental health and psychosocial problems among HIV-affected populations will turn up individual and groups of researchers that can be contacted for suggestions on measures and assessments.

The initial qualitative study is a powerful tool for determining what local people think, and particularly for identifying new local concepts, but it is not appropriate for determining which content to exclude. The fact that an issue/concept does not appear in the interviews or is

mentioned only rarely is not proof that the issue/concept is not important. The weak exclusionary power of qualitative studies is largely due to their use of small sample sizes and reliance on convenience or non-random sampling (increasing the likelihood of non-representative samples).

B.2.4 HIV-SPECIFIC TOPICS AND ISSUES

In addition to the three standard sections described above, which are common to all mental health instruments, additional sections relevant to HIV-affected populations may be added, depending to the target population or the issue being studied. For example:

a) **Knowledge, attitude and practice (KAP) questions**
Most HIV-related KAP surveys include questions on interviewees' *knowledge* of HIV transmission, risks, and outcomes; *attitudes* towards people with HIV and perceptions of one's own risk; and *behaviors,* which can includes both sexual risk behaviors (such as unprotected sex) and protective behaviors (such as using a condom). The behavior section can also include health-seeking behaviors, like getting tested and initiating treatment for those who are HIV-positive.

In HIV research and programming, KAP surveys can be used to pinpoint specific risk groups by identifying those with limited or incorrect knowledge or those who are engaging in high-risk behaviors. KAP Surveys can also provide baseline information for monitoring and evaluation of prevention and intervention programs, with outcomes being improvements in all three domains (see Modules 4 and 5).

b) **Health care utilization and health seeking behaviors**
A mixture of standard and locally-developed measures may be used to gather information on health care utilization. This domain is of particular importance if a study is attempting to understand who is using HIV-related services and where they are going to get HIV-related care. Standard measures that ask about general health care usage (i.e., types of problems people seek care for, frequency) may be relevant. However, for an understanding of specific care usage it is necessary to first understand what kind of health care services are available locally. A simple checklist may be used to find out which kind of services the study participant has used over a defined period of time (last month, year, ever). More specific information about frequency of use can also be included. When considering HIV-related health care services, it may be useful to review the qualitative data results to see if non-traditional (i.e., non-Western clinics) services are mentioned as being important locally. For mental health and psychosocial services,

identifying community-based services, both formal and informal, will result in a more complete picture of the services people may be accessing.

c) Stigma

HIV/AIDS-related stigma and discrimination are barriers to HIV prevention, voluntary counseling and testing uptake, and access to care in many international settings. HIV/AIDS-related stigma acts at both the societal and at the individual level. Programs can be developed to address stigma in order to promote adequate, accessible, and acceptable HIV/AIDS services at multiple levels. Defining and assessing stigma presents a challenge. It can be assessed from the perspective of HIV-infected individuals and/or the larger population to identify stigmatized and discriminatory attitudes towards those infected with HIV. The goals and aims of a program or research must be considered when reviewing existing stigma-related questionnaires and choosing the type and range of stigma-related questions to be included in the final instrument.

d) Substance abuse

Many studies have documented that substance abuse problems significantly increase the risk of adverse health outcomes in patients with HIV. Substance abuse has also been identified as a risk factor for worse clinical outcomes in HIV-infected individuals. To identify appropriate measures to assess substance abuse, a process similar to that used to identify mental health measures can be utilized. Based on the qualitative study, relevant substance use problems can be identified (i.e. alcohol use, injection-drug use, etc.). Once identified, standard measures can be reviewed from the literature to see if there are measures that capture the problems described locally, focusing on any measures that have been used in similar populations. Depending on the goals of the research or program, measures that go in depth on the types and frequency of usage may be important while for other projects, getting a more general overview of usage may be sufficient.

B.2.5 INSTRUCTIONS AND RESPONSE CATEGORIES

Most existing instruments assess the presence and severity of each symptom in the recent past using a Likert response scale based either on how much they bother the person (severity) or on the frequency of symptoms. Typically, categories of severity/frequency are *none/never, a little/sometimes,* moderate/frequently, *a lot/often, and extremely/all the time*. Depending on how much the local language discriminates between levels of response, the type and number of response categories in the original version of the instrument may be retained or adapted for local use. For example, where there is little difference between (for example)

moderately/frequently and a lot/often, these categories would best be combined. Whether the original response categories are appropriate or should be changed (and how) should be decided in consultation with translators, local persons, and the interviewers as part of the translation process (See below).

A Likert response scale is also included in the locally-defined function section template. Here the response choices are *none, little, a moderate amount, a lot, or often cannot do* as measures of level of difficulty with an activity. As for the symptom questions, some of these response categories may be combined according to their similarity in the local language.[1]

Sometimes adaptations require special attention to the hurdles specific to a population, such as high rates of illiteracy. To assist illiterate interviewees the interview materials often include a non-verbal response card. This provides a visual metaphor of the various responses, so that interviewees can keep track of the possible responses and choose by pointing at the pictures. The authors have used the image of a person holding an increasingly heavy stone to represent increasing severity or frequency for either the symptom or function sections. Appendix E contains an example adapted for both problem and function sections.

B.2.6 TRANSLATE THE INSTRUMENT

The study team reviews the selected instruments together with the qualitative data. When a symptom, activity, or other concept is described in the qualitative data, the words/idiom are cut and pasted into the instrument, with minor adjustments for grammar as needed.[2] The process continues in this way for all symptoms and concepts in the instrument that have an equivalent in the qualitative data, thereby ensuring that, as much as possible, the vocabulary of the target population is represented in the adapted instrument.

[1] There is also an additional function non-Likert response category of 'not applicable' or 'n/a.' This is used when the interviewee states that the activity is not relevant to her/him. Examples might include a women asked about caring for her children when she does not have children, or a retired elder asked about earning money.

[2] This is why qualitative analysis focuses on identifying the best language for each concept (See Module 1).

WHY USE QUALITATIVE DATA FOR TRANSLATION?

Humanitarian programs and public health research usually focus on poorer and less-educated populations that often do not speak a second language. Members of these populations tend to have different vocabularies than translators, who are more likely to have a higher level of education. Because of this difference, methods of translation that rely solely on translators to both generate and check translations (such as individual, group, and translation-back-translation) are not sufficient. Therefore, we supplement base translation with input from local and other monolingual workers in the form of the qualitative data.

The remainder of the instrument (i.e., instructions and any concepts not mentioned in the qualitative study) is then translated. The final complete version is reviewed by representatives of the target population, usually in a meeting or FG, to provide feedback and correct any mistakes in content and grammar. A similar review is then conducted with the interviewers as part of their training prior to testing the instrument. The version resulting at the end of the translation and review process is the draft that will be tested in the pilot and reliability/validity studies.

Example: Translating an instrument in Tbilisi Georgia

After drafting the assessment instrument, the research team for a study of street children in Tbilisi, Georgia met with 20 local interviewers and two translators to review the initial translation of the draft. These interviewers included 13 workers who had been interviewers on at least one of the two previous qualitative studies, and were familiar with the qualitative data. Most of the interviewers had a social science background and had experience working with at-risk children. Some of these interviewers worked at various sites involved in the study, and others were employees of the Ministry of Education. Therefore, the interviewers were well qualified to comment on the appropriateness of the initial draft instrument for use among the target populations.

The interviewers and translators reviewed each translated item in the draft instrument. Everyone had a copy of the qualitative data and compared the translation of each item with the wording used by the qualitative study interviewees. 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 draft that were not reflected in the qualitative study, interviewers and translators used their own knowledge and experience to determine whether the language was appropriate and would be understood by the children.

This review process took two days. It included a discussion among the research team, local staff, interviewers, and translators regarding whether there were important concepts described in the qualitative data not yet represented in the draft instrument. As a result, five additional questions were added to the draft.

B.3 INTERVIEWERS AND SUPERVISORS FOR THE PILOT AND VALIDITY STUDY

B.3.1 INTERVIEWER QUALIFICATIONS

Each interviewer should have the following qualifications:

- ✓ Fluent and literate in the language(s) of the local population where the study will be conducted
- ✓ Able to commit to the study timeline (full time for up to 12 working days)
- ✓ In good health and able to walk long distances if needed
- ✓ Acceptable to the target population (in terms of reputation, where they are from, gender, age, ethnicity)

While the interviewers need to be acceptable to the sample being interviewed, they should not be personally known by the interviewees. This is because the interviewees may feel uncomfortable sharing information with someone they know. They may also neglect to mention some topics because they know that this person knows about those topics already.

B.3.2 SUPERVISOR QUALIFICATIONS

- ✓ Fluent and literate in language(s) of the local population where the study will be conducted
- ✓ Fluent and literate in language of the project/study director (in order to act as a liaison between study director and interviewers when they do not share a common language)
- ✓ Able to commit to the study timeline (full time for up to 12 working days), plus any additional preparation time
- ✓ In good health and able to walk long distances if needed
- ✓ Acceptable to local people (in terms of reputation, where they are from, gender, age, ethnicity)

Supervisors provide the link between the study director and the interviewers. Like the interviewers, they need not have interviewing experience although prior experience working on a study of any type is helpful. As a supervisor, they will need to conduct some interviews and/or sit in when an interviewer becomes unavailable or additional supervision is needed, and also may need to re-interview random interviewees to assess reliability of findings. Thus, they share the same qualifications as the interviewers and undergo the same training, with the additional requirement that they are able to communicate verbally

B.3.3 STUDY DIRECTOR QUALIFICATIONS

- ✓ Preferably team leader or manager for the service organization, or someone with similar experience
- ✓ Available to direct pre-study planning and activities and post-study use and sharing of data
- ✓ Available for duration of study itself (full time for 12 working days) plus any additional preparation time.
- ✓ Can also be the trainer or otherwise speaks same language as the trainer and the supervisors (and interviewers if possible)

B.3.4 TRAINING INTERVIEWERS AND SUPERVIORS

The interviewer and supervisor roles may be filled by people identified from the implementing organization, by outside hires, or a combination of both. Staff may be used if there is an interest in building capacity (particularly if future studies are anticipated) and/or in order to save costs. Often there is not sufficient staff to cover all positions, so interviewers and supervisors are typically a mix of staff and outside hires. It is important to ensure that the selected interviewers will be acceptable to the population being interviewed. This is particularly so for HIV-affected persons who have been stigmatized by (and therefore are mistrustful of) other sections of the population. For example, to enhance cooperation, we have used former drug users as interviewers in a study of HIV-related behaviors related to current drug use. Consultation with community leaders and stakeholders is useful in thinking through who would be appropriate interviewers.

Potential interviewers and supervisors must be informed of the importance of working every day during the study. Once hired, they will be expected to prioritize the study over other work (an issue that often comes up when interviewers come from the implementing organization and are pulled in many different directions). This includes either the review of the questionnaire or the practice periods. Without the training she/he cannot be expected to use the interviewing methods correctly. It is understood that emergencies and/or unexpected but important events can occur that can oblige those involved to miss a day or more. Under such circumstances, an interviewer can leave briefly and return to the study as soon as possible. Potential interviewers should arrange childcare, as it is inappropriate to have infants and children present in the interview.

If a person misses any of the classroom training, however— either the review of the questionnaire or the practice periods — s/he cannot continue to work on the study regardless of the cause. Without the training s/he cannot be expected to use the interviewing methods correctly.

Between 10 and 20 interviewers and half that number of supervisors are needed to pilot and test the validity of the instrument for a given site. Each supervisor is responsible for two interviewers. If the study is to be done in multiple sites that have different characteristics (i.e. urban vs. rural; different languages; etc.) in ways that might significantly affect the results, it will be necessary to multiply the number of interviewers and supervisors by the number of different types of sites, effectively resulting in a separate study in each site.

If available, the qualitative study interviewers can be used since they have prior training in general interviewing techniques. Supervisors and interviewers are trained together. Training consists of 2-3 days of didactic training including standardized interviewing methods and procedures as well as specific orientation and practice with the draft instrument among

themselves. During the training, interviewers and supervisors also discuss any special considerations that need to be considered when interviewing vulnerable populations, such as HIV-positive adults, active drug users, etc.

B.4 PILOT STUDY

B.4.1 PURPOSE OF THE PILOT STUDY

The purpose of the pilot study is to:

- Test the clarity and ease of use of the instrument and interview procedures, for both the interviewees and interviewers
- Detect and address other problems with the instrument and interview procedures
- Provide preliminary data on the types and ranges of responses to the instrument
- Provide practice for the interviewers
- Test the data handling, entry, and analysis procedures

B.4.2 INTERVIEWEES

As one of the goals of the pilot study is to assess the comprehensibility and acceptability of the questions in the instrument, it is necessary to conduct the pilot study with a sample from the target population. For example, if the research and programming will be done among HIV-infected adults, the pilot study should be done with a convenience sample of HIV-infected adults. In cases when it is simply not feasible to pilot the instrument with the actual target population (e.g., if the study will be conducted in a difficult to access location), then identifying a very similar population will allow for testing at least the comprehensibility and clarity of the questions, giving the interviewers practice and testing of the data systems. However, preliminary data on the ranges of responses may not be accurate.

There should be sufficient numbers of respondents for each interviewer and each supervisor to interview at least two respondents. The two respondents should differ on key demographics that vary within the target population. The decision on what demographics to consider should be made by the study team and may include gender, HIV-infection status, socioeconomic status, etc.

B.4.3 PROCESS

The pilot interviews have three parts:

1. The interviewer begins with an introduction, including reading of a consent form. (See Appendix B for example)
2. The interviewer asks the instrument questions. The interview is performed using the complete instrument, just as it will be done in the main study.
3. The interviewer asks about the interview experience.

As soon as the interview with the complete instrument is completed, the interviewer asks the interviewee about the interview experience using the following questions:

i. *What did you like about the interview? Please explain.*
ii. *What did you not like about the interview? Please explain.*
iii. *Were there any questions that you found difficult to understand? Which ones were they? What was it that was difficult to understand?*
iv. *Were there any questions that you did not like being asked? Which ones were they?*
v. *Overall, how did you feel about being interviewed? Please explain.*

Responses to each of these questions are recorded by the interviewer on a separate form from instrument (see Appendix F). With the addition of these questions, it is possible that a pilot interview may take an hour or more. It is necessary to complete the full instrument and the pilot-specific questions during the same day. Therefore, the interview should continue through completion even if it runs longer than an hour.

After the pilot interviews are completed, all of the interviewers, supervisors, and the project director(s) meet and review the experience. The group reviews each question with interviewers and supervisors, raising any problems they encountered. The comments of interviewees are also reviewed, particularly in reference to problems of understanding or discomfort with any part of the interview. Where there are obvious problems, changes to the questions are made by consensus. Where significant changes are made, the entire process is repeated until only minor changes or no changes need to be made.

NOTE:

'Significant changes' consist of substitution of new words for key concepts, the removal or addition of questions, or changes in the content of the explanations that precede any of the sections. If further interviews are required, each interviewer needs to do only one interview with one new interviewee, unless the changes are extensive (three or more significant changes), in which case, the whole pilot study should be repeated. The pilot study is complete **only** when both interviewers and respondents are happy with the instrument and the process.

B.5 RELIABILITY AND VALIDITY TESTING

Reliability and validity testing is typically carried out immediately after completion of the pilot study, using the same interviewers and supervisors. Reliability and validity testing measure whether the instrument consistently and accurately measures what it was designed to measure. This is particularly important for concepts that are difficult to measure, such as mental health and psychosocial problems. Items that can be clearly verified (for example gender, HIV-status when using biological assessments) do not need to be validated through the process described below.

B.5.1 IDENTIFYING INTERVIEWEES

The first step in identifying interviewees is to forms lists of potential interviewees among the target population. Two lists are created for each of the major mental health and psychosocial problems being assessed: a list of those who are likely to have the problem and another list of those who are not likely to have it.

Preparing the lists of interviewees who do and do not have each of the select problem(s) under study is the task of the study supervisors. Supervisors begin this process by revisiting those knowledgeable persons (key informants (KIs)) from the prior qualitative study (Module 1), and/or other community members who know the community well. These informants are asked to provide:

- ✓ A list(s) of names and contact information for local people who they are confident **have** the local syndrome(s) most similar to the selected mental health and psychosocial problem(s)
- ✓ a separate list for people who they are confident **do not have** the problem(s)

The supervisors visit each person on these lists and ask if they themselves believe that they have or do not have the locally-described problem or syndrome depending on which list they are on. Those who disagree with the informant are removed from the list.

WHY IS THIS DONE?

These lists provide a standard against which the instrument will be compared. Therefore, it is important that the persons on each list are **as likely as possible** to actually have/not have the problem of interest. Removing persons about whom there is disagreement among informants increases the likelihood that those who remain really have/don't have the problem.

The goal is to generate lists with at least 50 respondents on each list. If there are two mental health or psychosocial problems being validated, there will be three lists (one for each problem and one for those without either problem), with about 150 total respondents. In reality there will likely be fewer total respondents because of overlaps in the problem lists, with some respondents identified as having both problems (This is very common; see the note below). Interviewers can usually complete 5-6 interviews per day, so data collection typically takes 3-5 days, depending on the number of interviewers.

NOTE:

Where multiple problems are being studied, a separate list of 50 persons will be required for **each** problem, resulting in a larger sample size and more time required. However, each additional problem does not increase the interviewee requirement by 50 because most mental health problems are co-morbid; persons tend to have more than one. The more frequently two mental health problems are co-morbid, the smaller the number of additional interviews required. For example, depression and PTSD-like illness commonly occur together. One list for depression and a separate list for PTSD will include many of the same names.

Example: Reliability and Validity Testing in Kurdistan, Iraq

A pilot and validity study of an instrument to assess torture survivors took place in four cities in the Suleimaniyah governate of Iraqi Kurdistan: Suleimaniyah, Rania, Halabja, and Kalar. These were the same sites as a previous qualitative study and represent the variety of experiences of torture survivors living in the governate. These experiences include torture and detention, the torture and murder of many friends and relatives during the Anfal (genocide) under the Hussein regime, and gas attacks on Halabja in 1988.

In the month leading up to the study, staff of the local NGO research partner contacted former prisoners' associations working in and around the four study sites. The staff of these organizations were asked to provide three lists of members who they or others considered to have depression, anxiety, or fear, and a list of those considered not to have mental health problems. This was done in order to generate lists of potential interviewees who were likely to have the problems being assessed by the instrument and those not likely to have those same problems.

In order to determine if they did in fact have/not have these problems, the interviewers contacted the persons on these lists, usually by phone. They explained the study and asked each person whether they would agree to be interviewed. For those who agreed, they asked the person whether they felt they were depressed, anxious, and whether they experienced excessive fear. During the same call, or another call soon after, another adult in the house was separately asked the same questions about the survivor. Persons on the list were defined as having depression, anxiety or excessive fear **if they and the other adult agreed that they had it**. Similarly, persons on the list were defined as not having depression, anxiety or excessive **fear if they and the other adult agreed that they did not have it**. If there was disagreement, the subject was eliminated from the list. The initial lists generated by the prisoner associations were designed to have more names than would be required for the study, to allow for persons who subsequently were not available or refused.

Appendix G is an example of the resulting lists (identifying information has been removed). The column marked 'MH problem' refers to the comments by the prisoners' association staff as to whether or not they felt those identified had a significant mental or psychosocial problem. This constitutes the original list of potential interviewees. The columns marked 'self' records the response of the interviewee about whether s/he had each of the three problems to be assessed, while columns marked 'other' refer to the same questions answered by the other adult. Where there was agreement between the person and the informant on a particular problem, the person was included on the final lists for interview to test instrument validity for that problem (either on the list for having the problem or on the list for not having the problem). Since there was often agreement on the presence/absence of more than one problem, a single interviewee could fulfill the sample requirements for multiple lists.

NOTE:

Potential constraint: Issues in the validity testing process; A case study.

Recently, concerns have arisen with the process described above of generating lists of people who are thought to have/do not have the selected problem(s). Members of the Institutional Review Board (IRB) of JHU have raised concern with using this 'cold calling' approach with children. *Cold calling* is contacting someone on the basis of information about them provided by others that they did not authorize researchers to have. There is concern that there is a potential risk to participants of being identified in this manner and that participants therefore need to provide consent first.

In response to this feedback, we have developed an alternative process for generating these lists that eliminates 'cold calling.' In this alternative procedure, the process described above is reversed. Interviewers from the study team go out in the community and talk to members of the target population. During this initial contact, the interviewer uses a script to briefly explain the study and what participation involves. The interviewer then asks the individual if they themselves believe they **have** or **do not have** the locally-described problem(s). The potential participant will then be asked if there is a person (KI) in their community who she/he trusts and knows her/him well, and whether it would be acceptable for the study team to contact this informant and ask about the individual's status regarding the locally-described problem(s). If the individual agrees to provide the contact information for the informant, the study team will obtain informed consent from the individual to contact the informant regarding the mental health status of the individual.

Once consent to contact is granted, the study team locates and contacts the informant and explains that the individual suggested them as someone who knows the individual well. The interviewer will provide the informant with a description of the problem(s) and ask if they believe the individual in question **has** or **does not have** the relevant problem(s). The interviewer will report these results back to the study team, where they will be compared to the self-classification.

Those individuals whose self-classifications disagree with the informant's assessment are dropped from the list. Practically, this would mean that there would be one visit to the individual to determine mental health status and obtain consent to contact another trusted individual in the community; and one visit to this informant to obtain her/his opinion on the mental health status of the individual. Finally, once both the informant's classification and the individual's opinion are compared, the list of individuals with agreement are contacted, consented and interviewed. Just as before, all interviewers performing the actual study interview will be blinded to the participants' problem(s) status.

B.5.2 INTERVIEWING PROCESS

Once the interviewee lists are finalized, supervisors assign each interviewee a study ID number. The supervisor then distributes the names, contact information, and ID numbers of the interviewees among the interviewers without revealing the reported status of the participants (i.e., which lists they are on). The interviewers then implement the full interview with each respondent on their list using the pilot-tested instrument.

For the purposes of reliability testing, a random sample of at least 30 interviews is repeated 1-3 days after the first interview. These are conducted by a different interviewer than the first.

At the end of each day supervisors collect all completed instruments and review them for completeness, correctness, and clarity before passing them on for data entry and analysis. Supervisors also note at the top of the instrument the list (with/without the select problems) the respondent came from so this information can be entered into the validity database. This can be done by inserting a table for this information at the top of the first page (see Figure 2 below). This information ('diagnosis by self and other/KI') is entered into the database along with the rest of the instrument and used in the analysis. The supervisor would write 'Yes' or 'No' for each box, representing whether the respondent themselves indicated they had the syndrome and whether the KI or other person indicated the respondent had the syndrome.

Figure 2: Example Validity Data Table

	Depression	**Anxiety**	**Fear**
Self			
KI/Other			

B.5.3 ANALYSIS

The basic analysis strategy involves generating scale scores for each study participant for each of the targeted mental health and psychosocial problems. Symptom scale scores are usually created by summing up the responses for all the items in that scale. The same is done for the locally-defined function questions to generate a function score. In the case of function, some interviewers will record 'N/A' for some activities for respondents who do not do that activity. This is coded as missing for the data analysis. Prior to any analyses, any missing function responses are replaced with the respondent's average score on the function questions they did answer.

Alternatively, the average score on all items that are answered can be used as the scale score instead of the sum of scores. This has the advantage that it does not require replacement of missing data with scale averages followed by summation.

Example: Kurdistan, Northern Iraq.

The reliability and validity analyses were based on subscales within the instrument referring to depression, anxiety, fear (an English translation of the concept for trauma-like symptoms) and function. Each scale score was generated by summing the responses to the individual questions that comprised that scale. For the symptoms comprising the depression, anxiety, and trauma scales interviewees were asked how much they experienced each symptom: never (score =0); sometimes (score=1); often (score=2); or always (score=3). Scores for each scale were obtained by summing the scores for each symptom in the scale. For the locally-defined function scale, interviewees were asked how much difficulty they had with each of a series of activities derived from the qualitative data: none (0); very little (1); a moderate amount (2); a lot (3); or they cannot do the activity (4). Again, a single score was derived by summing the score for each item, after replacing missing values with the average score for items they did answer. These subscales, and their severity scores, were the basis for tests of reliability and validity that explored the following characteristics of each scale:

1. Criterion validity
2. Combined test-retest and inter-rater reliability
3. Internal consistency reliability

B.5.3.1 CRITERION VALIDITY

Criterion validity refers to the extent to which the instrument agrees with an independent method of assessing mental health problems that is known to be accurate (the criterion). In Western psychiatric research, the criterion by which a scale is assessed is usually a psychiatric

interview or other clinical assessment. Since neither is available in many low-resource settings, we instead use KI reports together with the respondent's self-reports to act as the standard (criterion) of comparison. Then we compare the scale scores of those respondents identified by themselves and KIs as having the selected mental health or psychosocial problems with the scale scores of those respondents identified by themselves and KIs as having none of the selected problems. While this is not a standard of known accuracy and both the interviewee and other adult may be incorrect, we assume that the interviewee is more likely to have or not have each of these problems if they and the other informant agree. Criterion validity analysis therefore consists of comparing the mean scores for each problem of those said to have the problem with those said not to have it. For example, criterion validity of the depression scale would be supported if those said to have depression have substantially higher average depression scores than those said not to have depression (we arbitrarily use 20% difference as the cut-off). This will occur if the concept (in this case, depression) is understood and accurately recognized by local people, and if the instrument is valid for assessing depression.

Example: Kurdistan, Northern Iraq.

The table below is an example of the results of criterion validity testing. Results for the Depression, Anxiety, and Trauma sections of the instrument suggest acceptable criterion validity in that the differences between persons said to have those problems are substantially higher than those said not to have them. However, for the traumatic grief scale the results are similar for cases and non-cases. This suggests that either (1) the criterion validity of that scale is poor, (2) local people could not differentiate between persons with and without traumatic grief, or (3) both. Faced with these results, the research team can either revise and retest the scale or proceed with using the scale with the understanding that the validity of the results is questionable. The latter approach may be taken in cases where the other tests of validity and reliability are supportive, suggesting that the problem may lay with the testing procedures, such as the accuracy of the 'diagnoses' by the informants and other adults.

Criterion Validity Testing Results (N=128)

	Score Range (Min, Max)	Cases[a]	Non-Cases[a]	Difference p-value)[b]
Depression symptoms score, mean (sd)	0, 60	28.32 (10.04)	21.13 (10.97)	7.20 (.001)
Anxiety section score, mean (sd)	0, 30	13.04 (5.81)	7.96 (4.74)	5.08 (.000)
Trauma symptoms score, mean (sd)	0, 96	40.78 (14.92)	28.83 (12.84)	11.96 (.000)
Traumatic grief symptoms score, mean (sd)	0, 36	9.60 (5.70)	7.30 (4.99)	2.30 (.026)

[a.] 'Cases' refers to survivors who were said by both themselves and other adults close to them to have the problem. 'Non-Cases' refers to survivors who were said by both themselves and other adults close to them to NOT have the problem.
[b] P-value for the statistical significance of the difference in scale scores by case.

B.5.3.2 INTERNAL CONSISTENCY RELIABILITY

Internal consistency reliability refers to the extent to which responses to questions that assess the same underlying concept are similar: If a respondent scores high on one symptom they are likely to score high on other symptoms as well. If they are not similar this suggests that either the questions are unreliable or that they are not actually measuring the same concept. Internal consistency reliability is measured using Cronbach's alpha. Scores should be at least .7 and ideally >.8 for each scale. Cronbach's alpha scores are also used to conduct Item Analysis. Scores are calculated with and without each item in the scale. An increase in the Cronbach's alpha score when an item is removed suggests that the item is performing poorly and could be removed.

Example: Kurdistan, Northern Iraq.

The table below shows the Cronbach's alpha scores for five different scales. Alpha scores for all scales are in the acceptable range and are good for the depression, trauma, and anxiety scales. The traumatic grief scales show only acceptable internal reliability.

Item analysis did not support the removal of any items from any of the scales. Questions from the qualitative studies that were added to the various scales performed well, with correlations to the total scale as high as, or higher, than most of the standard items. This suggests that these local symptoms are indeed useful indicators of these syndromes among this population.

Cronbach's alpha scores (N=128)

Total Sample	Total Sample (N=128)[a]	Males (N=91)[a]	Females (N=37)[a]
Depression symptoms score	.881 (120)	.884 (87)	.882 (33)
Anxiety section score	.837 (128)	.834 (91)	.791 (37)
Trauma symptoms score	.904 (122)	.891 (87)	.928 (35)
Traumatic grief symptoms score	.775 (126)	.738 (89)	.840 (37)
Locally-defined Function Scale	----------	.799 (91)	.725 (37)

[a.] Not all interviewees have complete data. Data presented for only those with complete data.

B.5.3.3 TEST-RETEST AND INTER-RATER RELIABILITY

Reliability over time (*test-retest reliability*) is normally tested by carrying out a second interview with an interviewee after a short period of time (usually 1-3 days). A subset of all interviews are identified for reliability testing by random selection from the list of all interviews. The repeat interviews are done at least a day later. This time lapse is long enough to allow for a reduced effect of memory on the responses, but not long enough after the first interview for the concept being measured to substantially change in reality. Results for both interviews are then compared. *Inter-rater reliability* refers to interviewing the same interviewee at the same or nearly the same time by two different interviewers. These forms of reliability can be assessed separately or together, by means of assigning a different interviewer conduct the second interview 1-3 days later. In the latter case, the analysis would therefore reflect a

Example: Kurdistan, Northern Iraq.

The table below shows combined test-retest and inter-rater reliability analysis results, based on 33 repeat interviews (26% of total) conducted 1-3 days after the original interviews conducted by different interviewers. Combined test-retest/inter-rater reliability was measured using the Pearson correlation coefficient, which provides an indicator of the extent to which interviewees tend to give the same answer to the questions constituting each scale when asked on different occasions by different interviewers.

Based on a standard cutoff Pearson score of 0.7 as indicating acceptable reliability, each of the scales performed well except for the least reliable scale (anxiety), which was only acceptable. The function scales showed the highest repeat reliability.

Test-Retest Comparison (N=33)

Total Sample	Mean (sd) First Interview	Mean (sd) Repeat Interview	Correlation*
Depression symptoms score, mean (sd)	25.70 (11.18)	23.88(12.61)	.832
Anxiety section score, mean (sd)	10.76 (5.68)	9.67 (6.23)	.728
Trauma symptoms score, mean (sd)	36.88 (14.68)	33.58 (15.01)	.840
Traumatic grief symptoms score, mean (sd)	9.76 (6.77)	7.79 (6.23)	.833
Locally-defined Function Scales	31.91 (15.55)	29.76 (14.76)	.864

* Pearson correlation coefficient. Whether this is an appropriate measure depends on whether there is a linear relationship between the variables being measured (in this case, scores on first and second interviews for each scale). For all comparisons the scatterplot suggested a linear relationship, confirming that the Pearson correlation co-efficient was an appropriate measure of reliability.

combination of test-retest reliability and inter-rater reliability.

Test-retest and inter-rater reliability can be measured using Pearson's correlation coefficient. This requires a normal distribution of scores for both first and repeat interviews. It also assumes a linear relationship between first and repeat scores. This can be assessed by generating a scatterplot of the scale scores on the first interview plotted against the second interview to determine whether there is a linear relationship. Opinions vary as to what is an acceptable correlation, with acceptable cutoffs tending to be around 0.6 or 0.7. Low reliability scores suggest that the instrument and/or interviewer training procedures are unreliable.

Definitions:

Reliability: refers to the extent to which the results of using the same instrument vary under different circumstances.
Test-retest reliability*:* refers to agreement between results collected at different times by the same interviewer.
Inter-rater reliability*:* refers to agreement between results by different interviewers.
Internal consistency*:* refers to agreement between different questions in a questionnaire that are assessing the same thing at the same time.

B.6 FINALIZING THE INSTRUMENT FOR LOCAL USE

Where the above measures suggest significant problems with any of the scales, further investigation is required to determine the source of the problem. As described above, item analysis using Cronbach's alpha can determine whether individual items appear to be measuring different concepts than the rest of a scale and should be removed. Similarly, poor test retest reliability is investigated by comparing the mean scores of each question in the scale between the first and second interview, to determine which questions are the main sources of variation. If variation occurs mainly among a few questions, they should be removed. Similar variation across many or most questions would suggest that the entire scale needs to be revised and retested.

Poor results for criterion validity suggest either that the scale does not measure the concept or that the interviewees and/or the KIs are mistaken in their assessment. The latter is possible, given that the assessment by self and others is not a gold standard known to be accurate across populations. This is likely to be the cause of apparent poor criterion validity when the scale performs well on all other measures of validity and reliability. In this case, the criterion validity testing is regarded as having failed and the criterion validity testing results are ignored. If other measures of validity and reliability are sound, and interviewees did not appear to have difficulty

in understanding or answering the questions, this suggests that the scale is adequate even though criterion validity is unknown. Note that this is not true for criterion validity testing using known gold standards. In such cases poor criterion validity in the face of good results for other validity and reliability testing would indicate the need for further instrument adaptation and retesting.

In addition to the analyses described above, other parameters are reviewed on a question-by-question basis. Questions that result in many refusals or inabilities to answer are generally removed because they do not provide useful data. Locally-defined function questions for which many interviewees respond *'N/A'* are also removed for similar reasons. Questions for which average responses suggest very low severity are also removed, since this suggests that they are not assessing real problems in the population.

In our experience, the use of methods in this module usually results in the removal of some individual items. On the whole, most items (and scales) tend to perform well and the analysis and procedures described here result in mild to moderate changes.

REFERENCES

Aday, L.A. & Cornelius, L.J. (2006). *Designing and Conducting Health Surveys: A Comprehensive Guide: 3rd edition.* Josey-Bass: San Francisco.

American Psychiatric Association (APA). (1994). *Diagnostics and Statistical Manual of Mental Disorders: 4th Edition.* Washington, D.C.

WHO | WHO disability assessment schedule 2.0 WHODAS 2.0. (2010). Retrieved 8/12/2012, from http://www.who.int/classifications/icf/whodasii/en/index.html

APPENDIX A: EXAMPLE FIELD SCHEDULE

A typical schedule of working days is proposed in the table below. It may be adapted as needed. This schedule **assumes** that the distance from training site to community is less than one hour, and both training and data collection in the community can occur in the same day. **When the training site and communities are further apart, it will be necessary to take into consideration travel time** and whether interviewers and supervisors will need to spend the night (or several nights) at the site of the interviews and then return to the training site for analysis and further training.

Day	Activities
Prior to study	Adapt and translate instruments. Develop interview lists.
1-2	Welcome interviewers and provide interviewer training for supervisors and interviewers at the training site. The group reviews and (if necessary) corrects the draft instrument translation.
3-4	Interviewers and supervisors proceed directly to the community to conduct pilot interviews. After completing the interviews they return to the training site where the group discusses the interview experience and makes any necessary changes to instrument and procedures. Later the same day/evening the pilot study data is entered into the computer as a test of data entry methods. New copies of the instrument are printed which incorporate the agreed changes
5	At the training site the group reviews the new version of the instrument and procedures. The director explains the validity study. Supervisors and interviewers proceed to the community to conduct the validity study interviews. The supervisors meet with the director at the end of the day to discuss any problems. Reliability and Validity study data entry begins.
6-10	Interviews and data entry continue.
11-12	Complete data entry and begin analysis.
After field study	Complete analysis of validity and reliability studies. Finalize the instruments for local use.

APPENDIX B. RESOURCE LIST

TIME

Preparation usually requires one month prior to the study on a part-time basis. Activities completed during this preparation period include:

- Preparations for the assessment
- Meeting with the community
- Preparing logistics
- Arranging personnel

Actual implementation of the validity study, including training, interviews, and analysis, takes up to 12 working days. (See typical time schedule in Appendix A)

PERSONNEL

- One study director for four weeks, including approximately two and a half weeks implementing the study.
- 5-10 Supervisors
- 10-20 Interviewers

Study Director Qualifications

- ✓ Preferably team leader or manager for the provider, or someone with similar experience
- ✓ Available to direct pre study planning and activities and post study use and sharing of data
- ✓ Available for duration of study itself (12 working days)
- ✓ Can also be the trainer or otherwise speaks same language as the trainer and the supervisors (and interviewers if possible)

Supervisor Qualifications

- ✓ NGO or Ministry of Health (MOH) local staff or students (e.g., nursing students)
- ✓ At least a high school education (Preferred = college level education and good knowledge of English
- ✓ Available for duration of study (12 working days)
- ✓ Ability to read and write in the local language
- ✓ Able to ride a motorbike, drive car, or otherwise use transportation as needed

Interviewer Qualifications

- ✓ Less important that they be NGO or MOH local staff
- ✓ NGO or MOH local staff or students (e.g., nursing students)
- ✓ At least a high school education (college level education and good knowledge of English preferred
- ✓ Available for duration of study (12 working days)
- ✓ Ability to read and write in the local language
- ✓ Able to ride a motorbike, drive car, or otherwise use transportation as needed.
- ✓ Capable of walking long distances
- ✓ Understand they will be supervised by supervisors

TRANSPORT

The project director and all staff will meet at the training site at the beginning of each day before proceeding to the study site. Daily transport between the training and study site for all workers (interviewers and supervisors) is required throughout all phases of the project.

During the pilot and validity study, each supervisor will require a motorbike or similar individual transportation. These are necessary to enable supervisors to move between interviewers while returning to sites visited on previous days to repeat interviews and check on refusals. The lack of such independent transport for each supervisor will considerably slow down the assessment process, particularly during the study, so they are well worth the trouble to arrange.

TRAINING AND OFFICE MATERIALS

Site Requirements

- ✓ Large enough to seat all study personnel
- ✓ Quiet
- ✓ Available power (electric or generator)
- ✓ Available exclusively for use by the team throughout the project (24 hr access)
- ✓ Able to be locked at night

Equipment

- ✓ Blackboard/whiteboard/easel and chalk/marker pens
- ✓ Overhead projector or LCD projector
- ✓ Printer and 2 reams of paper
- ✓ Reliable photocopier capable of printing thousands of pages a day when required (this includes printing copies of the instrument

- ✓ Three toner cartridges
- ✓ 20 reams of paper
- ✓ Laptop computer with word processing and statistical programs.

INTERVIEWER MATERIALS

For completing all modules, each interviewer and supervisor should have the following:

- ✓ Exercise book
- ✓ Two pencils, eraser and sharpener
- ✓ Waterproof carrying bag
- ✓ Daily portable lunch (or per diem and options for purchasing lunch)
- ✓ Sufficient copies of instruments and consent forms
- ✓ Cell phones or other method of communication with study director and supervisors

APPENDIX C: VERBAL CONSENT FORM

VERBAL CONSENT FORM FOR RESEARCH FOR RESEARCH STUDY

Instructions for the Interviewer

The following sections printed in italics are to be read to the subject prior to the interview. If the subject then agrees to participate, you must sign on the line marked 'Witness to Consent Procedures' at the end of this form. Also mark the date on the appropriate line.

Purpose of the Study

You are being asked to be part of a research study. We want to find out about the problems affecting people in this area. By learning about the problems of local people [organization you are working for] hopes to design better programs to assist local people. This research is being done by [organization you are working for]. We would like to invite you to participate.

Procedures

To obtain this information we are talking with some people in the community who we selected by chance/because we heard that you are knowledgeable about this topic. This is how we selected you. If you agree to help us, I will ask you some questions. These questions are about your health. We may also want to return and talk with you again later.

Risks and Discomfort

Each interview will take about __ __ minutes. It is possible that you may not like some questions or that some questions may upset you. You may refuse to answer these questions, or any questions, if you wish. You may stop the interview at any time.

Benefits

*This information will help (your organization) to provide better programs to improve the health of the people in this area. However, there may be no direct benefit to you personally***.**

Confidentiality

*During the interview I will write down the information you tell me. This is the information we will use for our study. The record of this information will not have any information which can be used to identify you. I will also record your name and address, but this will be stored separately from the record, and will be locked in the project director's office. The project director will have the key. Only the research team will be able to see this information***.** *Nothing that you tell us will be shared with anyone outside the research team. We will not allow anyone else to find out who*

gave a particular answer unless we think that you or someone else might be in danger. In that case, we cannot keep that information private and we will take steps to make sure that all people involved are kept safe. Every effort will be made to protect the confidentiality of this information as far as is legally possible.

Voluntariness

It is your decision whether or not to be in this study. You can stop being in this study at any time. This will not affect any assistance you get from (your organization) or any other organization.

Whom to Contact

If you have any questions now you can ask me. If you have any questions after I leave, you can ask (insert name of project director). He/she is in charge of the study and can be contacted through the (your organization) Office in _____, telephone _____. In the future if you have any questions about the study, you should ask (local contact for your organization). He/she and the other researchers will tell you if they learn anything new that they think will affect you.

Do you have any questions?

Do you agree to participate in this study? **Yes** (PROCEED) **No** (STOP)

_____ _____

Signature of Interviewee Date

[I have explained this research study to the subject.]

_____ _____

Interviewer or other witness to consent. Date

(to be signed only if subject has verbally consented).

_____

Signature of Investigator

APPENDIX D: EXAMPLE OF DRAFT ASSESSMENT INSTRUMENT FOR TESTING (KURDISTAN, IRAQ)

_____	**Depression**	**Anxiety**	**Fear**
Self			
Other			

Date _____/_____/_____

Site _____

Interviewer _____

Age of Interviewee _____

Gender of Interviewee _____ M _____ F

_____ 1 _____ 2

Name of Interviewee

_____

_____-

Section A: Functionality Assessment

There are two versions of this section - one for men and one for women. If the interviewee is a man, use only the version for men. If the interviewee is a woman, use only the version for women.

I am going to read a list of activities and duties. These are activities and duties that other torture survivors have told us are important to them. For each one I am going to ask you how much more difficulty you are having doing it THAN MOST OTHER MEN/WOMEN OF YOUR AGE. You should tell me whether you are having no more difficulty, a little more, a moderate amount more, or a lot more, or you often cannot do that activity.

Now say each activity, and after each one say:

Are you having no more difficulty than most other men/women of your age, a little more, a moderate amount more, a lot more, or you often cannot do this activity?

NOTE:

The text of the original language has been deleted from the following tables to simplify the table.

Each activity is identified by *AM* (for male activity) or *AF* (for female activity) + the activity's number in the activity list. Each symptom is identified by *B* + the symptom's number in the list. Some symptoms are identified by more than one number because they are included in multiple symptoms scales. For example, 'blaming self is part of both a depression scale (as B12) and a trauma scale (as B44).

Record the response by marking the appropriate box next to the symptom in the table below.

Male Functionality	**Amount of difficulty doing the activity**					
Activities	**None**	**Very little**	**A moderate amount**	**A lot**	**Cannot do**	**N/A**
AM01 Providing for the family	***0***	***1***	***2***	***3***	***4***	***9***
AM02 Looking after family behaviors	***0***	***1***	***2***	***3***	***4***	***9***
AM03 Labor	***0***	***1***	***2***	***3***	***4***	***9***
AM04 Giving advice to family members	***0***	***1***	***2***	***3***	***4***	***9***
AM05 Giving advice to other community members	***0***	***1***	***2***	***3***	***4***	***9***
AM06 Exchanging ideas with others	***0***	***1***	***2***	***3***	***4***	***9***
AM07 Having harmonious relationship with wife and family	***0***	***1***	***2***	***3***	***4***	***9***

Tasks/Activities	**None**	**Very little**	**A moderate amount**	**A lot**	**Cannot do**	**N/A**
AM08 Bringing up children correctly	***0***	***1***	***2***	***3***	***4***	***9***
AM09. Doing things to improve the community	***0***	***1***	***2***	***3***	***4***	***9***
AM10. Sympathizing with others	***0***	***1***	***2***	***3***	***4***	***9***
AM11 Visiting and socializing with others in community	***0***	***1***	***2***	***3***	***4***	***9***
AM12 Asking for or getting help when you need it	***0***	***1***	***2***	***3***	***4***	***9***
AM13 Making decisions	***0***	***1***	***2***	***3***	***4***	***9***
AM14 Taking part in family activities or events	***0***	***1***	***2***	***3***	***4***	***9***
AM15 Taking part in community activities/events	***0***	***1***	***2***	***3***	***4***	***9***
AM16. Learning new skills or knowledge	***0***	***1***	***2***	***3***	***4***	***9***
AM17. Concentrating on your tasks or responsibilities	***0***	***1***	***2***	***3***	***4***	***9***
AM18. Interacting or dealing with people you do not know	***0***	***1***	***2***	***3***	***4***	***9***
AM19. Attending mosque or religious gathering	***0***	***1***	***2***	***3***	***4***	***9***
AM20. Assisting others	***0***	***1***	***2***	***3***	***4***	***9***

Female Functionality	**Amount of difficulty doing the activity**					
Activities	**None**	**Very little**	**A moderate amount**	**A lot**	**Cannot do**	**N/A**
AF01 Housework	*0*	*1*	*2*	*3*	*4*	*9*
AF02 Cooking	*0*	*1*	*2*	*3*	*4*	*9*
AF03 Other types of manual labor	*0*	*1*	*2*	*3*	*4*	*9*
AF04 Caring for family members	*0*	*1*	*2*	*3*	*4*	*9*
AF05 Giving advice to others	*0*	*1*	*2*	*3*	*4*	*9*
AF06 Exchanging ideas with others	*0*	*1*	*2*	*3*	*4*	*9*
AF07 Having harmonious relationship with husband+ family	*0*	*1*	*2*	*3*	*4*	*9*
AF08 Bringing up children correctly	*0*	*1*	*2*	*3*	*4*	*9*
AF09. Doing things to improve the community	*0*	*1*	*2*	*3*	*4*	*9*
AF10. Sympathizing with others	*0*	*1*	*2*	*3*	*4*	*9*
AF11 Visiting and socializing with others in community.	*0*	*1*	*2*	*3*	*4*	*9*
AF12 Asking for help or getting help when you need it	*0*	*1*	*2*	*3*	*4*	*9*
AF13 Making decisions	*0*	*1*	*2*	*3*	*4*	*9*
Activities	**None**	**Very little**	**A moderate amount**	**A lot**	**Cannot do**	**N/A**

AF14 Taking part in family activities or events	*0*	*1*	*2*	*3*	*4*	*9*
AF15. Taking part in community activities or events	*0*	*1*	*2*	*3*	*4*	*9*
AF16. Learning new skills or knowledge	*0*	*1*	*2*	*3*	*4*	*9*
AF17. Concentrating on your tasks or responsibilities	*0*	*1*	*2*	*3*	*4*	*9*
AF18. Interacting or dealing with people you do not know	*0*	*1*	*2*	*3*	*4*	*9*
AF19. Attending mosque or religious gathering	*0*	*1*	*2*	*3*	*4*	*9*
AF20. Assisting others	*0*	*1*	*2*	*3*	*4*	*9*

Section B: Symptom Assessment Instrument

Hopkins Symptom Checklist -25

Listed below are some problems that people sometimes have. For each one state how often you have felt like this in the last 2 weeks including today.

Symptom	Never or No	Sometimes	Often	Always
B01. Suddenly feeling scared for no reason	*0*	*1*	*2*	*3*
B02. Feeling fearful	*0*	*1*	*2*	*3*
B03. Feeling faintness	*0*	*1*	*2*	*3*
B04. Nervousness	*0*	*1*	*2*	*3*
B05. Heart pounding or racing	*0*	*1*	*2*	*3*
B06. Trembling	*0*	*1*	*2*	*3*
B07 and B34. Feeling tense	*0*	*1*	*2*	*3*
B08. Headaches	*0*	*1*	*2*	*3*
B09. Episodes of terror or panic	*0*	*1*	*2*	*3*
B10. Feeling restless, can't sit still	*0*	*1*	*2*	*3*
B11. Feeling low in energy, slowed down	*0*	*1*	*2*	*3*
B12 and B44. Blaming yourself for things	*0*	*1*	*2*	*3*
B13. Crying easily	*0*	*1*	*2*	*3*
B14. Loss of interest in sex or loss of sexual pleasure	*0*	*1*	*2*	*3*

Symptom	**Never or No**	**Sometimes**	**Often**	**Always**
B15. Poor appetite	*0*	*1*	*2*	*3*
B17, B39, and B46. Feeling hopeless about the future	*0*	*1*	*2*	*3*
B18. Feeling depressed	*0*	*1*	*2*	*3*
B19. Feeling lonely	*0*	*1*	*2*	*3*
B20. Thinking about ending your life	*0*	*1*	*2*	*3*
B21. Feeling not free or caught	*0*	*1*	*2*	*3*
B22. Worrying too much about things	*0*	*1*	*2*	*3*
B23 and B38. Loss of interest in things	*0*	*1*	*2*	*3*
B24. Feeling that everything you do is difficult	*0*	*1*	*2*	*3*

Harvard Trauma Questionnaire (Part IV – Symptoms)

The following are problems that people sometimes have after experiencing hurtful or terrifying events in their lives. For each problem please state how often you have had it in the last 2 weeks including today.

Symptom	**Never or No**	**Sometimes**	**Often**	**Always**
B26. Recurrent thoughts or memories of the hurtful or terrifying events	*0*	*1*	*2*	*3*
B27. Feeling as though these events were happening again	*0*	*1*	*2*	*3*
B28. Nightmares	*0*	*1*	*2*	*3*
B30. Able to feel emotions	*3*	*2*	*1*	*0*
B31. Feeling jumpy, easily startled	*0*	*1*	*2*	*3*
B32. Difficulty concentrating	*0*	*1*	*2*	*3*
B33 and B16. Trouble sleeping	*0*	*1*	*2*	*3*
B35. Feeling irritable or having outbursts of anger	*0*	*1*	*2*	*3*
B36. Avoiding activities that remind you of the traumatic or hurtful events	*0*	*1*	*2*	*3*
B37. Inability to remember parts of the traumatic or hurtful events	*0*	*1*	*2*	*3*
B40. Avoiding thoughts or feelings associated with the traumatic or hurtful experience	*0*	*1*	*2*	*3*
B41. Suddenly feeling very different emotionally or physically when reminded of the traumatic or hurtful events	*0*	*1*	*2*	*3*

Symptom	Never or No	Sometimes	Often	Always
B42. Feeling that people do not understand what happened to you	*0*	*1*	*2*	*3*
B43. Difficulty performing your work or daily activities	*0*	*1*	*2*	*3*
B45. Feeling guilty for having survived	*0*	*1*	*2*	*3*
B47. Feeling ashamed of the hurtful or traumatic events that happened to you	*0*	*1*	*2*	*3*
B48. Spending time thinking about why these events happened to you	*0*	*1*	*2*	*3*
B49. Feeling as if you are going crazy	*0*	*1*	*2*	*3*
B50. Feeling that you are the only person who has suffered these events	*0*	*1*	*2*	*3*
B51. Feeling that others are hostile to you	*0*	*1*	*2*	*3*
B52, Feeling that you have no-one to rely on	*0*	*1*	*2*	*3*
B53. Finding out or told by other people that have done something that you cannot remember	*0*	*1*	*2*	*3*
B54. Feeling as if you are split into two people, one of you is watching what the other is doing	*0*	*1*	*2*	*3*
B55. Feeling betrayed	*0*	*1*	*2*	*3*

Traumatic Grief

Symptom	Never or No	Sometimes	Often	Always
B56. Hearing the voice of a deceased person speaking to you	*0*	*1*	*2*	*3*
B57. Seeing a deceased person standing in front of you	*0*	*1*	*2*	*3*
B59. Feeling that you have lost your sense of control	*0*	*1*	*2*	*3*
B60. Feeling that the death of someone close to you has changed your view of the world	*0*	*1*	*2*	*3*
B61. (Feeling that you are) having pain the same part of your body or having the same symptoms as people who have died.	*0*	*1*	*2*	*3*
B62. Feeling that moving on with your life (like making new friends, pursuing new interests) would be difficult.	*0*	*1*	*2*	*3*
B63. Feeling envious of others who have not lost someone close.	*0*	*1*	*2*	*3*
B64 and B29. Feeling like you have lost the ability to care about other people.	*0*	*1*	*2*	*3*
B65. Feeling drawn to places and things associated with people who have died.	*0*	*1*	*2*	*3*
B66.imitating some of the same behaviors or characteristics of people who have died.	*0*	*1*	*2*	*3*

Problems specific to Kurdish survivors of torture in Suleimaniyah governate and their families

Symptom	Never or No	Sometimes	Often	Always
B67. Feeling desperate	*0*	*1*	*2*	*3*
B68. Feeling mentally unstable	*0*	*1*	*2*	*3*
B69 and B25. Feeling inferior to others.	*0*	*1*	*2*	*3*
B70. Wishing you were dead	*0*	*1*	*2*	*3*
B71 and B58. Feeling as if you were already dead	*0*	*1*	*2*	*3*
B72. Waiting for your dead relatives to come back	*0*	*1*	*2*	*3*
B73. The brain is tired.	*0*	*1*	*2*	*3*
B74.Able to express your feelings	*3*	*2*	*1*	*0*
B75. Fighting with others	*0*	*1*	*2*	*3*
B76. Poor relationship with family members	*0*	*1*	*2*	*3*
B77. Able to enjoy feasts or other celebrations	*3*	*2*	*1*	*0*
B78. Drinking too much alcohol	*0*	*1*	*2*	*3*
B79. Thinking too much	*0*	*1*	*2*	*3*

Final Instructions

State that this is the end of the questionnaire. Review the questionnaire while still with the interviewee, checking that all questions have been answered and the answers are clear. If not, review the question with the interviewee and insert the missing information. Once you are satisfied that all questions have been answered and the responses are clear, ask the interviewee if it would be OK for you, or your supervisor, to return if they have further questions. If this is OK, ask if there are any restrictions on when you might return and record these in the space below. Finally, ask the interviewee not to discuss the interview with anyone until after the study is over. This is so that their comments do not affect the answers of others who are yet to be interviewed.

Restrictions on Re-visiting?

End the interview

_____	**Depression**	**Anxiety**	**Fear**
Self			
Other			

_____

Date _____/_____/_____

Site _____

Interviewer _____

Age of Interviewee _____

Gender of Interviewee _____ M _____ F

_____ 1 _____ 2

Name of Interviewee

_____

SECTION A: FUNCTIONALITY ASSESSMENT

There are two versions of this section - one for men and one for women. If the interviewee is a man, use only the version for men. If the interviewee is a woman, use only the version for women.

I am going to read a list of activities and duties. These are activities and duties that other torture survivors have told us are important to them. For each one I am going to ask you how much more difficulty you are having doing it THAN MOST OTHER MEN/WOMEN OF YOUR AGE. You should tell me whether you are having no more difficulty, a little more, a moderate amount more, or a lot more, or you often cannot do that activity.

Now say each activity, and after each one say:

Are you having no more difficulty than most other men/women of your age, a little more, a moderate amount more, a lot more, or you often cannot do this activity?

Record the response by marking the appropriate box next to the symptom in the table below.

Note:

The text of the original language has been deleted from the following tables to simplify the table.

Each activity is identified by *AM* (for male activity) or *AF* (for female activity) + the activity's number in the activity list. Each symptom is identified by *B* + the symptom's number in the list. Some symptoms are identified by more than one number because they are included in multiple symptoms scales. For example, 'blaming self is part of both a depression scale (as B12) and a trauma scale (as B44).

Male Functionality	**Amount of difficulty doing the activity**					
Activities	**None**	**Very little**	**A moderate amount**	**A lot**	**Cannot do**	**N/A**
AM01 Providing for the family	***0***	***1***	***2***	***3***	***4***	***9***
AM02 Looking after family behaviors	***0***	***1***	***2***	***3***	***4***	***9***
AM03 Labor	***0***	***1***	***2***	***3***	***4***	***9***
AM04 Giving advice to family members	***0***	***1***	***2***	***3***	***4***	***9***
AM05 Giving advice to other community members	***0***	***1***	***2***	***3***	***4***	***9***
AM06 Exchanging ideas with others	***0***	***1***	***2***	***3***	***4***	***9***
AM07 Having harmonious relationship with wife and family	***0***	***1***	***2***	***3***	***4***	***9***
AM08 Bringing up children correctly	***0***	***1***	***2***	***3***	***4***	***9***
AM09 Doing things to improve the community	***0***	***1***	***2***	***3***	***4***	***9***
AM10. Sympathizing with others	***0***	***1***	***2***	***3***	***4***	***9***
AM11 Visiting and socializing with others in community	***0***	***1***	***2***	***3***	***4***	***9***
AM12 Asking for or getting help when you need it	***0***	***1***	***2***	***3***	***4***	***9***
AM13 Making decisions	***0***	***1***	***2***	***3***	***4***	***9***
AM14 Taking part in family activities or events	***0***	***1***	***2***	***3***	***4***	***9***
AM15 Taking part in community activities/events	***0***	***1***	***2***	***3***	***4***	***9***
AM16. Learning new skills or knowledge	***0***	***1***	***2***	***3***	***4***	***9***
AM17.Concentrating on your tasks or responsibilities	***0***	***1***	***2***	***3***	***4***	***9***
AM18 Interacting or dealing with people you do not know	***0***	***1***	***2***	***3***	***4***	***9***
AM19 Attending mosque or religious gathering	***0***	***1***	***2***	***3***	***4***	***9***
AM20 Assisting others	***0***	***1***	***2***	***3***	***4***	***9***

Female Functionality	**Amount of difficulty doing the activity**					
Activities	**None**	**Very little**	**A moderate amount**	**A lot**	**Cannot do**	**N/A**
AF01 Housework	*0*	*1*	*2*	*3*	*4*	*9*
AF02 Cooking	*0*	*1*	*2*	*3*	*4*	*9*
AF03 Other types of manual labor	*0*	*1*	*2*	*3*	*4*	*9*
AF04 Caring for family members	*0*	*1*	*2*	*3*	*4*	*9*
AF05 Giving advice to others	*0*	*1*	*2*	*3*	*4*	*9*
AF06 Exchanging ideas with others	*0*	*1*	*2*	*3*	*4*	*9*
AF07 Having harmonious relationship with husband+ family	*0*	*1*	*2*	*3*	*4*	*9*
AF08 Bringing up children correctly	*0*	*1*	*2*	*3*	*4*	*9*
AF09. Doing things to improve the community	*0*	*1*	*2*	*3*	*4*	*9*
AF10 Sympathizing with others	*0*	*1*	*2*	*3*	*4*	*9*
AF11 Visiting and socializing with others in community.	*0*	*1*	*2*	*3*	*4*	*9*
AF12 Asking for help or getting help when you need it	*0*	*1*	*2*	*3*	*4*	*9*
AF13 Making decisions	*0*	*1*	*2*	*3*	*4*	*9*
AF14 Taking part in family activities or events	*0*	*1*	*2*	*3*	*4*	*9*
AF15 Taking part in community activities or events	*0*	*1*	*2*	*3*	*4*	*9*
AF16 Learning new skills or knowledge	*0*	*1*	*2*	*3*	*4*	*9*
AF17 Concentrating on your tasks or responsibilities	*0*	*1*	*2*	*3*	*4*	*9*
AF18 Interacting or dealing with people you do not know	*0*	*1*	*2*	*3*	*4*	*9*
AF19 Attending mosque or religious gathering	*0*	*1*	*2*	*3*	*4*	*9*
AF20 Assisting others	*0*	*1*	*2*	*3*	*4*	*9*

SECTION B: SYMPTOM ASSESSMENT INSTRUMENT

Hopkins Symptom Checklist -25

Listed below are some problems that people sometimes have. For each one state how often you have felt like this in the last 2 weeks including today.

Symptom	Never or No	Sometimes	Often	Always
B01. Suddenly feeling scared for no reason	***0***	***1***	***2***	***3***
B02. Feeling fearful	***0***	***1***	***2***	***3***
B03. Feeling faintness	***0***	***1***	***2***	***3***
B04. Nervousness	***0***	***1***	***2***	***3***
B05. Heart pounding or racing	***0***	***1***	***2***	***3***
B06. Trembling	***0***	***1***	***2***	***3***
B07 and B34. Feeling tense	***0***	***1***	***2***	***3***
B08. Headaches	***0***	***1***	***2***	***3***
B09. Episodes of terror or panic	***0***	***1***	***2***	***3***
B10. Feeling restless, can't sit still	***0***	***1***	***2***	***3***
B11. Feeling low in energy, slowed down	***0***	***1***	***2***	***3***
B12 and B44. Blaming yourself for things	***0***	***1***	***2***	***3***
B13. Crying easily	***0***	***1***	***2***	***3***
B14. Loss of interest in sex or loss of sexual pleasure	***0***	***1***	***2***	***3***
B15. Poor appetite	***0***	***1***	***2***	***3***
B17, B39, and B46. Feeling hopeless about the future	***0***	***1***	***2***	***3***
B18. Feeling depressed	***0***	***1***	***2***	***3***
B19. Feeling lonely	***0***	***1***	***2***	***3***
B20. Thinking about ending your life	***0***	***1***	***2***	***3***
B21. Feeling not free or caught	***0***	***1***	***2***	***3***
B22. Worrying too much about things	***0***	***1***	***2***	***3***
B23 and B38. Loss of interest in things	***0***	***1***	***2***	***3***
B24. Feeling that everything you do is difficult	***0***	***1***	***2***	***3***

Harvard Trauma Questionnaire (Part IV – Symptoms)

The following are problems that people sometimes have after experiencing hurtful or terrifying events in their lives. For each problem please state how often you have had it in the last 2 weeks including today.

Symptom	Never or No	Sometimes	Often	Always
B26 Recurrent thoughts or memories of the hurtful or terrifying events	***0***	***1***	***2***	***3***
B27 Feeling as though these events were happening again	***0***	***1***	***2***	***3***
B28 Nightmares	***0***	***1***	***2***	***3***
B30 Able to feel emotions	***3***	***2***	***1***	***0***
B31 Feeling jumpy, easily startled	***0***	***1***	***2***	***3***
B32 Difficulty concentrating	***0***	***1***	***2***	***3***
B33 and B16 Trouble sleeping	***0***	***1***	***2***	***3***
B35 Feeling irritable or having outbursts of anger	***0***	***1***	***2***	***3***
B36 Avoiding activities that remind you of the traumatic or hurtful events	***0***	***1***	***2***	***3***
B37 Inability to remember parts of the traumatic or hurtful events	***0***	***1***	***2***	***3***
B40 Avoiding thoughts or feelings associated with the traumatic or hurtful experience	***0***	***1***	***2***	***3***
B41 Suddenly feeling very different emotionally or physically when reminded of the traumatic or hurtful events	***0***	***1***	***2***	***3***
B42 Feeling that people do not understand what happened to you	***0***	***1***	***2***	***3***
B43 Difficulty performing your work or daily activities	***0***	***1***	***2***	***3***
B45 Feeling guilty for having survived	***0***	***1***	***2***	***3***
B47 Feeling ashamed of the hurtful or traumatic events that happened to you	***0***	***1***	***2***	***3***
B48 Spending time thinking about why these events happened to you	***0***	***1***	***2***	***3***
B49 Feeling as if you are going crazy	***0***	***1***	***2***	***3***
B50 Feeling that you are the only person who has suffered these events	***0***	***1***	***2***	***3***
B51 Feeling that others are hostile to you	***0***	***1***	***2***	***3***
B52 Feeling that you have no-one to rely on	***0***	***1***	***2***	***3***

B53 Finding out or told by other people that have done something that you cannot remember	***0***	***1***	***2***	***3***
B54 Feeling as if you are split into two people, one of you is watching what the other is doing	***0***	***1***	***2***	***3***
B55 Feeling betrayed	***0***	***1***	***2***	***3***

Traumatic Grief

Symptom	Never or No	Sometimes	Often	Always
B56 Hearing the voice of a deceased person speaking to you	***0***	***1***	***2***	***3***
B57 Seeing a deceased person standing in front of you	***0***	***1***	***2***	***3***
B59 Feeling that you have lost your sense of control	***0***	***1***	***2***	***3***
B60 Feeling that the death of someone close to you has changed your view of the world	***0***	***1***	***2***	***3***
B61 (Feeling that you are) having pain the same part of your body or having the same symptoms as people who have died.	***0***	***1***	***2***	***3***
B62 Feeling that moving on with your life (like making new friends, pursuing new interests) would be difficult.	***0***	***1***	***2***	***3***
B63 Feeling envious of others who have not lost someone close.	***0***	***1***	***2***	***3***
B64 and B29 Feeling like you have lost the ability to care about other people.	***0***	***1***	***2***	***3***
B65 Feeling drawn to places and things associated with people who have died.	***0***	***1***	***2***	***3***
B66 Imitating some of the same behaviors or characteristics of people who have died.	***0***	***1***	***2***	***3***

Problems specific to Kurdish survivors of torture in Suleimaniyah governate and their families

Symptom	Never or No	Sometimes	Often	Always
B67 Feeling desperate	***0***	***1***	***2***	***3***
B68 Feeling mentally unstable	***0***	***1***	***2***	***3***
B69 and B25 Feeling inferior to others.	***0***	***1***	***2***	***3***
B70 Wishing you were dead	***0***	***1***	***2***	***3***
B71 and B58 Feeling as if you were already dead	***0***	***1***	***2***	***3***
B72 Waiting for your dead relatives to come back	***0***	***1***	***2***	***3***
B73 The brain is tired.	***0***	***1***	***2***	***3***
B74 Able to express your feelings	***3***	***2***	***1***	***0***
B75 Fighting with others	***0***	***1***	***2***	***3***
B76 Poor relationship with family members	***0***	***1***	***2***	***3***
B77 Able to enjoy feasts or other celebrations	***3***	***2***	***1***	***0***
B78 Drinking too much alcohol	***0***	***1***	***2***	***3***
B79 Thinking too much	***0***	***1***	***2***	***3***

Final Instructions

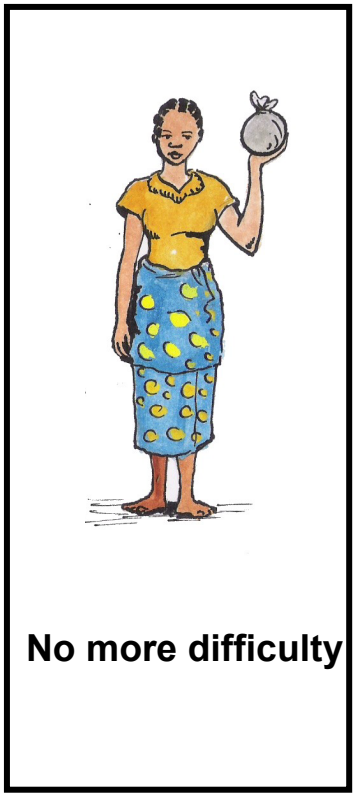
State that this is the end of the questionnaire. Review the questionnaire while still with the interviewee, checking that all questions have been answered and the answers are clear. If not, review the question with the interviewee and insert the missing information. Once you are satisfied that all questions have been answered and the responses are clear, ask the interviewee if it would be OK for you, or your supervisor, to return if they have further questions. If this is OK, ask if there are any restrictions on when you might return and record these in the space below. Finally, ask the interviewee not to discuss the interview with anyone until after the study is over. This is so that their comments do not affect the answers of others who are yet to be interviewed.

Restrictions on Re-visiting?

End the interview

APPENDIX E: EXAMPLES OF NON VERBAL RESPONSE CARDS

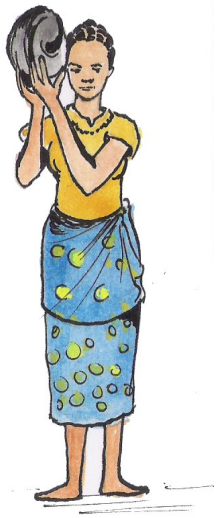
NONVERBAL RESPONSE CARD FOR LOCALLY-DEFINED FUNCTION QUESTIONS*



*Adapted by Chishugi Oswald, IRC/Bukavu, Democratic Republic of Congo

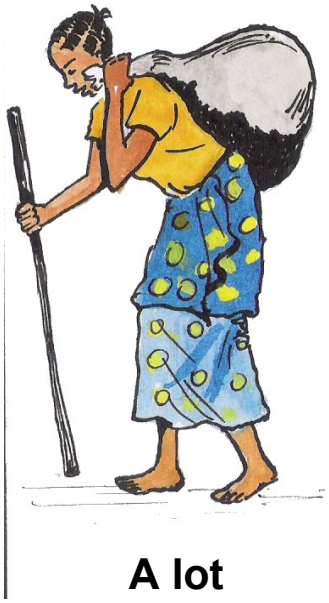
Nonverbal Response Card (symptoms)*

Not at all



A little bit

A moderate amount



A lot

*Adapted by IRC/Bukavu,

Democratic Republic of Congo

APPENDIX F: PILOT STUDY POST-INTERVIEW FORM

These questions are asked at the end of the interview during the pilot interviews only. The responses are recorded on a separate sheet and discussed at the pilot interviews review meeting of researchers, supervisors, and interviewers.

What did you like about the interview? Please explain.

What did you not like about the interview? Please explain.

Were there any questions that you found difficult to understand? Which ones were they? What was it that was difficult to understand?

Were there any questions that you did not like being asked? Which ones were they?

Overall, how did you feel about being interviewed?

APPENDIX G: EXAMPLE OF INTERVIEWEE LISTS (MULTIPLE PROBLEMS)

No	Name	Contact Information	Sex M/F	MH Problem (Y/N)	Depression		Anxiety		Fear	
					Self	Other	Self	Other	Self	Other
1			F	Y	Y	Y	Y	Y	N	N
2			M	N	Y	Y	Y	N	N	N
3			F	Y	N	N	N	N	N	N
4			F	N	Y	Y	Y	Y	N	N
5			F	N	N	N	N	N	N	N
6			F	Y	Y	Y	Y	Y	Y	Y
7			M	Y	Y	Y	N	N	N	N
8			M	N	N	N	N	N	N	N
9			M	N	N	N	N	N	N	N
10			M	N	N	N	N	N	N	N
11			M	Y	Y	Y	Y	Y	N	N
12			M	Y	Y	Y	Y	Y	N	N
13			M	Y	Y	Y	Y	Y	N	N
14			M	Y	Y	Y	N	N	N	N
15			M	Y	Y	Y	Y	Y	Y	Y
16			M	N	N	N	N	N	N	N
17			M	N	N	N	N	N	N	N
18			M	N	N	N	N	N	N	N
19			M	Y	Y	Y	Y	Y	N	N
20			F	Y	Y	Y	Y	Y	Y	Y
21			M	Y	N	N	N	N	N	N
22			M	Y	Y	Y	Y	N	N	N

23			M	Y	Y	Y	Y	Y	N	N
24			F	Y	Y	Y	Y	N	N	N
25			F	Y	Y	Y	N	Y	Y	N
26			F	Y	Y	Y	Y	Y	Y	Y
27			M	Y	Y	N	Y	N	N	N
28			M	Y	Y	Y	Y	Y	Y	N
29			F	Y	Y	Y	NY	N	N	Y
30			M	Y	N	N	Y	Y	N	N
31			F	N	Y	Y	Y	Y	Y	Y
32			M	N	N	Y	N	N	N	N
33			M	N	N	Y	N	N	Y	Y
34			M	Y	Y	Y	Y	Y	Y	Y