

## **A new approach to cross-cultural function assessment**

Paul Bolton, M.B.B.S., M.P.H., M.Sc.<sup>1</sup>

Alice M. Tang, Ph.D.<sup>2</sup>

<sup>1</sup>Center for International Emergency, Disaster and Refugee Studies, Department of International Health, The Johns Hopkins University Bloomberg School of Public Health

<sup>2</sup>Department of Family Medicine and Community Health, Tufts University School of Medicine.

Address for Correspondence:

Alice M. Tang, Ph.D.  
Department of Family Medicine and Community Health  
Tufts University School of Medicine  
136 Harrison Avenue, Posner 4  
Boston, MA 02111

## **Abstract**

### *Objective*

To create and validate a new method for cross-cultural function assessment that avoids the major problems with existing methods.

### *Methods*

We used free listing to learn about tasks important to local people. Community-specific function questionnaires based on these tasks were then created and used in community-based surveys.

The survey results were used to assess the questionnaires' internal reliability (Cronbach's alpha), combined test-retest and across-interviewer reliability using repeat interviews, and (in Uganda) criterion validity by comparing assessment by self to assessments by cohabiting adults.

### *Results*

Field trials of this approach were conducted in rural Rwanda and Uganda. Differences between tasks identified by free listing were greater between sexes than sites. Cronbach's alphas for male and female questionnaires were respectively .815 and .822 in Rwanda and .886 and .881 in Uganda. Pearson correlations for combined test-retest and across-interviewer reliability were respectively .469 and .640 for Rwandan men and women and .797 and .871 in Uganda.

Correlation between self-assessment and co-habiting adults was .904.

### *Conclusions*

We have developed an alternative to the existing approach of adapting western function instruments to other cultures and situations. The field trials have demonstrated that this approach is rapid, feasible and can yield valid and reliable instruments. Developing instruments locally

Cross-cultural function assessment

avoids the problems of limited local relevance and appropriateness associated with adapting western instruments. Although each instrument created in this way is culturally bound, they are 'cross-cultural' in the sense that each refers to the tasks most important to local people. This approach should prove useful for both researchers and aid agencies working in non-western countries.

## **Introduction**

This work is part of a larger project to develop methods for assessing the prevalence of specific mental health problems in developing countries, and their impact on function. The overall aim is to develop an approach to cross-cultural assessment suitable for use with any population. 'Suitable' means that the approach must be within the resources usually available for this type of research. This requires the use of self- or interviewer-administered instruments rather than clinician-administered, since highly trained clinical personnel familiar with the local culture are not available for many non-western cultures (while it is true that some developing countries in Africa do have well trained mental health workers, these usually represent only a few of the many cultural groups existing in those countries and are limited to the larger cities). Suitability also refers to the need for adaptability across many types of cultures, demonstrable scientific validity, and for generating data of practical significance both for researchers and aid agencies that would want to make use of the resulting data.

The approach we developed for assessing the prevalence of mental illness has been described elsewhere [1]. With regard to function, we reviewed the literature for standardized

#### Cross-cultural function assessment

self- or interviewer-administered assessment methods. The most widely used appears to be the SF-36 [2-4] which has been adapted for many developed countries [5]. The desire for brevity has also spawned shorter versions of this instrument, including the SF-20 [6] and SF-8 [7]. Other standard instruments include the Work and Social Disability Scale (WSDS) [8], the Disability Scale (DISS) to assess work, social and family problems [9], and the World Health Organization's (WHO) Brief Disability Questionnaire (BDQ) [10]. With the exception of the BDQ, all were created in developed countries. There has been some work to develop standard instruments less specific to advanced western nations. Attempts have been made to adapt the SF-36, but the major effort has been WHO's Disability Assessment Schedule, versions I and II [11-12].

Yet none of these standard instruments fulfilled our suitability criteria for cross-cultural work. Those developed in western countries contain too many culture-bound questions that are difficult to adapt to other situations; for example, questions on the ability to climb stairs and go 'shopping' are difficult to adapt to situations where these things do not exist. Even responses to less culture-specific questions can be difficult to interpret across populations with varying physical requirements. For example, the ability to walk a mile may represent an acceptable level of function for an American but would be a grossly inadequate standard for an African nomad. Even more problematic is the reliance of both western and non culture-specific instruments on questions referring to 'component' activities such as lifting or walking. This is done in an effort to improve generalizeability. It is useful in some types of research but does not translate into direct assessment of the ability to complete the tasks important to daily existence. Nor do any of

the instruments we reviewed acknowledge the major differences in the roles of men and women that still predominate in most developing countries.

This paper describes a new functional assessment approach developed to address these problems and meet our suitability criteria for cross-cultural research. We also describe our experiences using this approach in field trials in rural areas of Rwanda and Uganda.

## **Methods**

### *Overview*

There are three stages to creating the population-specific function assessment instrument: 1) creation of the template (one time only); 2) free-listing and data analysis to identify tasks for the template (for each new population); and 3) use in a survey, including validity and reliability testing (for each new population). We conducted stages 2 and 3 among two African populations. In 1999 we used this approach to develop an assessment instrument for people living in Kanzenze commune in Rwanda, a rural area close to the capital Kigali that was severely affected by the 1994 genocide. The instrument then formed part of a community-based survey of depression. In 2000 we repeated this process in the Masaka and Rakai districts of southwest Uganda, again creating a local function assessment instrument that was also used as part of a community depression survey.

### *Creation of the function assessment template*

Prior to conducting any assessments, we developed the template shown in figure 1. As

#### Cross-cultural function assessment

the term implies, its form stays the same across assessments while the actual tasks included change with each new community. The template has space for nine tasks and an open-ended ‘other’ category. We chose this number in order to provide sufficient coverage of important tasks while keeping the questionnaire short. For each task listed, respondents are asked to assess how much difficulty they currently experience in completing that task, compared with most other people of their age and sex.<sup>1</sup> The response categories are then read to the respondent. They are also shown a card with a series of drawings, each representing one of these response categories in the form of a person carrying an increasingly heavy weight. The corresponding response category is written under each drawing and the interviewer points to the drawing as they say that category. This card is used because of the large number of possible responses (five) and to assist respondents who are illiterate.

The template also includes a section on the causes of difficulty for each task.

Interviewers ask this question if the respondent reports at least mild difficulty in completing the task. It is included as a way of keeping responses focused on health. If the respondent lists the cause(s) as a lack of some external physical resource not related to health (such as assistance or money) the interviewer asks the respondent to consider how much difficulty they would have if they had those resources. Interviewers refrain from recording the level of difficulty for a task until the respondent gives causes that are related to physical, mental or emotional problems. If no such causes are given, the level of difficulty is recorded as ‘none.’

---

<sup>1</sup> In Rwanda, our first field trial site, respondents were asked to assess their level of difficulty compared to a ‘normal person’ of their age and sex. However, this was changed in Uganda, the second trial site because of concerns that the use of a hypothetical ‘normal’ person was too abstract a concept for the respondents.

#### Cross-cultural function assessment

A simpler method would be to directly ask the respondent to consider only dysfunction due to poor health, as is done in most existing function assessment instruments. However, across (and even within) cultures, the meaning of a complex term like 'health' varies widely and it can be difficult to determine its precise meaning. Even more problematic is that there is often no equivalent term for the western concept of 'health', so translators are forced to use the closest term or phrase available. Where the translation is not (or cannot be) accurate, asking respondents to include only health'-related issues may result in inclusion of non-health-related causes (from the researcher's viewpoint) or exclusion of health-related causes. For example, in many parts of Africa, mental health issues are thought of in terms of witchcraft or spirits rather than as health problems. Therefore, asking for only health related causes would result in the exclusion of most mental health issues. Since what falls under the health rubric varies by culture and even by individual communities, a comprehensive knowledge of the local meaning of 'health' would be required in order to use the term.

By avoiding the term 'health' and instead having the respondent explicitly state all the causes the trained interviewer, rather than the respondent, can judge what should be excluded. In cases where the respondent listed a cause for which it was not clear whether this was a health problem or a lack of resources (such as spirits or witchcraft) the interviewer was trained to ask how this cause led to the dysfunction. For example, if witchcraft were listed as the cause of inability to cook, the interviewer would try to establish whether witchcraft was thought to be acting by affecting health or innate ability to do the task, or whether the respondent thought that witchcraft had reduced the supply of outside help or the things needed for cooking.

Cross-cultural function assessment

Having the interviewer record all the responses that they accept also provides a check on the interviewer's judgement: both supervisors and the project director review each completed template during the preliminary pilot study and the survey. Any problems are referred back to the interviewer and remedial training is provided when necessary.

### *Free listing*

In each community the tasks to be inserted into the template are determined by the results of a free-listing exercise. This is a qualitative technique in which respondents are asked to generate a list of items in response to a specific question [13]. Each respondent is asked to provide three free lists by answering three questions in turn:

*'What are the tasks that men/women must do regularly to care for themselves?'*

*'What are the tasks that men/women must do regularly to care for their family?'*

*'What are the tasks that men/women must do regularly to care for their community?'*

For each question the interviewer probes the respondent to give as many tasks as they can think of. Once this is done, the interviewer then revisits each task and asks for a short description of each. It is best to state all three free-list questions before asking for responses to the first question. Otherwise we have found that respondents tend to include family and community tasks in their answers about self-care.

In each community 20-40 respondents are interviewed. Convenience samples are used,

Cross-cultural function assessment

the only restrictions being that all must be adults (at least 18 years of age), half must be women, and the sample must reflect ages from young adult to the elderly and a range of villages across the study site. Men and women are asked only about tasks specific to their gender. Both tasks and their descriptions are recorded in the local language, using the exact words of the respondent.

The completed interviews are then combined into six composite free lists, referring separately to self, family and community tasks for each gender. We then remove from each list those tasks for which dysfunction does not clearly affect other persons. This is done in order to focus on those tasks that most affect the family and community. In Rwanda this caused us to remove a single task - 'prayer' - as we could not see a clear way in which inability to pray would significantly affect others (although this is certainly possible). On the other hand, washing oneself was retained since inability or difficulty doing this would require assistance from others and therefore affect them. Tasks that are listed twice in different forms are also removed. In Rwanda many of the men listed 'sending children to school' as an important task in caring for the family. This was removed from the list since most of the men interviewed stated that their contribution to schooling was limited to earning enough money to pay the fees, and earning money was already listed as a separate task. Hence only two tasks were removed from the free list in Rwanda. In Uganda all the frequently mentioned tasks clearly affected others and were not repetitions of other tasks. Therefore, none were removed.

After reviewing the list of tasks, as above, we then choose the nine most frequently mentioned, including at least one task from each of the categories of self care, family and community tasks. These tasks are then inserted into the template (figure 1), under the heading of

#### Cross-cultural function assessment

‘task or activity’. In the same way we choose nine tasks for women and insert them into an identical template. We chose nine tasks in order to have ten categories in the final instrument (including 'other' tasks). Ten is used to provide sufficient variety of function but not to make the questionnaire too long (otherwise our decision to have ten categories was arbitrary). This completes the draft function assessment instruments for men and women. For both sexes an ‘other’ task category is included as a permanent part of the template. This provides a check on the results of the free-list: if an important task were missed, we would expect it to be mentioned frequently under ‘other’ tasks in the pilot study and the main survey. We could then decide whether to add it to the instrument (if it became apparent in the pilot study) or use the results in the analysis (if it emerged only during the main survey).

#### *Reliability and validity testing*

Once the pilot testing is completed, the functional assessments are further tested in community-based surveys. The survey results are used to test the reliability and validity of the function instrument. Unless the majority of interviewees respond to the ‘other’ category it is not used in the testing or data analysis.

Internal reliability is assessed by calculating Cronbach’s alpha and conducting item analyses. Additional data to study test-retest reliability is provided by having each interviewer’s supervisor repeat at least 10% of their survey interviews, chosen at random from among all the respondents and repeated one to three days after the initial interview (in Rwanda, interviews were repeated up to 12 days later due to logistical problems). We chose to reinterview 10% of the

#### Cross-cultural function assessment

sample because this was the highest figure that would not interfere with the supervisors' other duties. The scores for each task are added to give two summary measures of dysfunction severity for each respondent – one for the initial and one for the repeat interviews. On this scale, '0' represents minimal dysfunction whereas '36' represents the maximum dysfunction on all tasks (a score of 4 on all nine tasks). If respondents don't answer some questions – usually because they are not responsible for that task, for example child-care for a woman with no children – the average of their other responses is used in place of the missing data. Then, the overall correlation between initial and repeat interview scores is calculated as a measure of test-retest reliability. Since the repeat interviews are done by supervisors, this is also a test of the reliability of the instrument across different interviewers.

In the Uganda survey we also attempted to validate the accuracy of self-assessment of function. Whenever supervisors re-interviewed respondents they also asked if there were any other adults living with the respondent who were available for interview at that time. If so, this adult was also interviewed about the respondent's functional ability, using the same instrument. Neither respondent nor cohabiting adult witnessed the other's interview. As with the test-retest analysis, function scores were calculated based on each interview and then compared by measuring their correlation. Each respondent's score was then plotted against their score according to the cohabiting adult. This was done to determine whether disagreement was due to greater estimation of dysfunction by the respondent or by the other adult.

#### **Results of Field Trials**

*Identifying tasks for the function template*

In both Rwanda and Uganda, interviewers chose a convenience sample of 40 respondents for the free-listing exercise. Table 1 shows the tasks resulting from the free lists that were included in the function instruments in Rwanda and Uganda. In Rwanda, the men mentioned only seven tasks frequently enough to be included in the instrument (we set an arbitrary minimum for inclusion of mention by at least five respondents). Otherwise, there were nine tasks in each gender-specific questionnaire.

Table 1 also shows that differences in tasks were much greater between the sexes than between sites: six of the tasks were identical or very similar for Rwandan and Ugandan men and the same number for women. Some tasks appeared to be specific to occupation (such as farming), but were mentioned so frequently that we included them in the instrument.

*Reliability and validity testing*

Results of the item analyses and internal reliability testing were based on the surveys we conducted in Rwanda (n=368) and in Uganda (n=587). Cronbach's alphas for male and female questionnaires were respectively .815 and .822 in Rwanda and .886 and .881 in Uganda. In item analyses, all items in Rwanda and Uganda performed well: Cronbach's alpha was not substantially increased by the removal of any item in any of the four questionnaires (results not shown).

The Pearson correlation coefficients for test-retest reliability are given in table 2. As noted in the Methods section, in both Rwanda and Uganda the respondents were a random

#### Cross-cultural function assessment

sample of at least 10% of the survey respondents. At both sites all those persons randomly selected were found and reinterviewed (response rate = 100%). The results show poor test-retest reliability for male function assessments in Rwanda and only adequate female agreement. In contrast, agreement in Uganda was very good (and likewise higher among women than men), confirming good reliability for both male and female instruments in Uganda across interviewers and time.

During the main survey in Uganda, thirty-seven interviews were conducted in which a cohabiting adult was also interviewed about the function of the respondent. A scatter plot comparing the function scores based on the respondent's own responses versus those of the cohabiting adult (with the corresponding line of best fit) is shown in figure 2. The plot suggests good agreement between self-assessments of function and assessment by others - the slope is approximately one, with a point of origin near zero. The relationship between the scores explains >80% of the variation in the plotted values ( $r^2=.82$ ). The Pearson correlation coefficient ( $r$ ) for the respondents' versus the cohabiting adults' responses was correspondingly high - .904.

#### **Discussion**

When we originally began this project our goal was to find or develop a single instrument that would effectively measure the ability to do tasks important to daily existence, and that such an instrument should be useable across cultures. We abandoned trying to adapt existing instruments, for the reasons described in the introduction. When we tried to develop our own instrument we quickly realized how much these tasks vary according to sex, culture, and

Cross-cultural function assessment

environment. So instead we opted to maintain local relevance by creating a new instrument for each community, based on tasks that are important locally. We decided that the decision of which tasks to include in the instrument (i.e., which were ‘important’) would best be made by local people, rather than by outsiders like ourselves. From the research point of view, having the community choose the tasks greatly enhances content validity.

Our field trials in Rwanda and Uganda have provided preliminary evidence that this type of approach is feasible. All staff at each site were local people with a high school education who were trained by one of the authors. The total time required to create and test the instruments at each site (including staff training in qualitative methods but not including the quantitative survey) was 6-8 working days.

There remain issues with this approach. In Rwanda test-retest reliability was poor for men and barely adequate for women. The most likely explanation for this was that the interviewers were affecting the responses. This was most likely due to a lack of interviewing skills on the part of the interviewers, supervisors or both. Lack of experience did not appear to be a factor as agreement was no better at the end of the study period than at the beginning. Although functioning may have changed over time we did not believe the changes would be great even in the cases where 12 days elapsed between the first and repeat interviews.

We addressed the perceived lack of skills by more intensive training in the Uganda study. Specifically the interviewers and supervisors spent much more time practicing interviewing under the direction of the trainers before proceeding to the field. Feedback on these practice sessions emphasized a standardized approach to probing and not influencing responses through

#### Cross-cultural function assessment

word or action. We also reduced the time between interview and re-interview (1-3 days), although this was done to reduce the time required for data gathering. Reliability was much improved in Uganda, but it will take further studies to determine if this was a local phenomenon or due to our changes in training.

A second issue is that this approach asks the respondent to think hypothetically. Respondents were asked not for the level of difficulty actually experienced but for what *would* be experienced if environment and resources were not a problem. The impression of interviewers was that this was not difficult for respondents in either Rwanda and Uganda, but we have little data to suggest how accurate this approach is. These data are limited to Uganda where we asked both respondents and their cohabitants about their function, using this approach. We found a very high degree of agreement which suggests good accuracy. However, multiple similar assessments will be required before we can feel confidence in this approach.

From a research viewpoint it may be argued that this approach is not 'cross-cultural' since different instruments are created for different communities which prevents comparisons between communities. On one level this is true: the tasks of men and women in the same community, and the tasks in different communities, do vary. However, we believe that comparisons can be made because the basis of each questionnaire is the same: the tasks considered to be most important by local people in terms of caring for themselves, their families and their communities. Therefore comparisons would consist of asking the question 'how does function vary between communities on the tasks that local people feel are most important?'

#### Cross-cultural function assessment

We also have another reason for contending that this approach is 'cross-cultural.' So far we have done only two assessments using this approach in similar types of communities (rural and impoverished) but in different countries. Yet we found that there were many similarities between the tasks chosen by the same sex at the two sites. As more assessments are done in similar sites we anticipate that certain common key tasks will continue to emerge. These recurring tasks could then form the bases for a group of standard instruments specific to particular *types* of communities (for example, African rural communities). This approach may ultimately lead to instruments that are not only more directly relevant to individual communities but are also generalizable. In the meantime, by generating instruments within the communities to be studied, we believe that this approach offers a feasible solution to the problems associated with current methods of cross-cultural function assessment.

## References

1. Bolton P (2001) Cross-cultural validity and reliability testing of a standard psychiatric assessment instrument without a gold standard. *J.Nerv.Ment.Dis.* 189:238-42.
2. Ware JE Jr. (1993) SF-36 health survey manual and interpretation guide. The Health Institute, New England Medical Center, Boston, MA.
3. Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A (1995) Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Medical Care* 33:AS264-AS279.
4. McHorney CA, Ware JE Jr. (1995) Construction and validation of an alternate form general mental health scale for the Medical Outcomes Study Short-Form 36-Item Health Survey. *Medical Care* 33:15-28.
5. McDowell I, Newell C (1996) The short-form-36 health survey. In: *Measuring Health: A guide to rating scales and questionnaires*. 2<sup>nd</sup> ed. New York and Oxford: Oxford University Press 446-60.
6. Stewart AL, Hays RD, Ware JE Jr. (1988) The MOS short-form general health survey. *Medical Care* 26:724-35.
7. SF-8 - a new short form for monitoring health outcomes. Available at

Cross-cultural function assessment

<http://qmetric.com/innohome/insf8.shtml> Accessed March 15, 2001.

8. Mauskopf JA, Simeon GP, Miles MA, Westlund RE, Davidson JR (1996) Functional status in depressed patients: the relationship to disease severity and disease resolution. *J.Clin.Psychiatry* 57:588-92.
9. Sheehan DV, Harnett-Sheehan K, Raj BA (1996) The measurement of disability. *Int.Clin.Psychopharmacol.* 11 Suppl 3:89-95.
10. Ormel J, VonKorff M, Oldehinkel AJ, Simon G, Tiemens BG, Ustun TB (1999) Onset of disability in depressed and non-depressed primary care patients. *Psychol.Med.* 29:847-53.
11. Janca A, Kastrup M, Katschnig H, Lopez-Ibor JJ, Jr., Mezzich JE, Sartorius N (1996) The World Health Organization Short Disability Assessment Schedule (WHO DAS-S): a tool for the assessment of difficulties in selected areas of functioning of patients with mental disorders. *Soc.Psychiatry Psychiatr.Epidemiol.* 31:349-54.
12. World Health Organization. (1998) WHO-DAS II Disablements Assessment Schedule Intermediate Version. Revised Draft for Field Trials Phase 2. World Health Organization.
13. Weiss W and Bolton P (2000). Training in qualitative research methods for PVOs and NGOs. Sessions 4.1-4.3. Available at: [http://www.jhsph.edu/refugee/tqr\\_a\\_desc.html](http://www.jhsph.edu/refugee/tqr_a_desc.html). Accessed March 15, 2001.

**Table 1: Tasks frequently mentioned by respondents in Rwanda and Uganda and included in function instruments (number of respondents).**

<u>Men</u>		<u>Women</u>	
<b>Rwanda (20)</b>	<b>Uganda (20)</b>	<b>Rwanda (20)</b>	<b>Uganda (20)</b>
Wash self	Personal hygiene	Wash self	Personal hygiene
Earn money	Farm	Care for children	Care for children
Advise the family	Head the home	Cook	Cook
Manual labor	Manual labor	Wash clothes	Wash clothes/utensils
Socialize	Plan for the family	Clean house	Clean house
Dress self	Participate in community development activities	Dress self	Participate in community development activities
Attend meetings	Attend meetings	Attend meetings	Attend meetings
	Socialize	Socialize	Grow food
	Participate in burial ceremonies	Transmit culture	Console and assist the bereaved

Cross-cultural function assessment

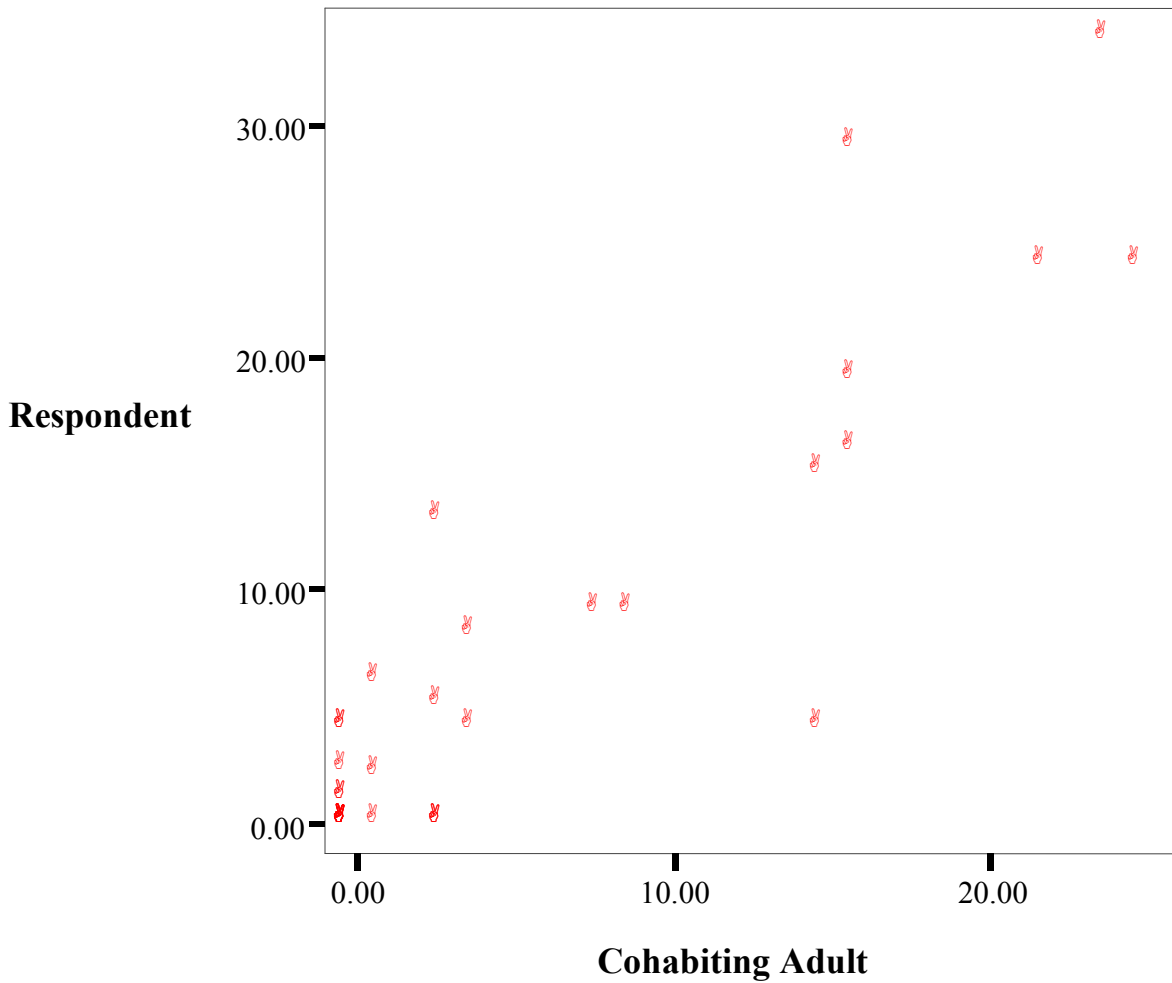
**Table 2: Pearson correlation coefficients for test-retest reliability (number of repeat interviews)**

	<b>Men</b>	<b>Women</b>	<b>Overall</b>
<b>Rwanda (37)</b>	.469	.640	.571
<b>Uganda (72)</b>	.797	.871	.833

**Figure 1: Function assessment template**

Task or activity	Degree of Difficulty completing task or activity:					Cause of difficulty
	None	Little	Moderate	A lot	Often cannot	
<b>1.</b>	0	1	2	3	4	
<b>2.</b>	0	1	2	3	4	
<b>3.</b>	0	1	2	3	4	
<b>4.</b>	0	1	2	3	4	
<b>5.</b>	0	1	2	3	4	
<b>6.</b>	0	1	2	3	4	
<b>7.</b>	0	1	2	3	4	
<b>8.</b>	0	1	2	3	4	
<b>9.</b>	0	1	2	3	4	
<b>10. Other</b>	0	1	2	3	4	

**Figure 2: Correlation of function scores – respondents' own assessments vs. assessments by cohabiting adults (N=37)**



**Respondent = 0.70 + 1.08 \* Cohabiting Adult**

**R-Square = 0.82**

## **Acknowledgements**

This work was part of the Complex Emergency Response and Transition Initiative (CERTI), a project of the Africa Bureau of USAID. The study was conducted in collaboration with World Vision International. Informed consent was obtained from all participants prior to interview and the study was approved by the Committee on Human Research at the Johns Hopkins Bloomberg School of Hygiene and Public Health.

Xx insert figure 2 (non-verbal card).

Xx add table and describe/discuss it in the body.

