Guidelines for cognitive and pilot testing of questions for use in surveys

ESCAP PROJECT ON IMPROVING DISABILITY MEASUREMENT AND STATISTICS IN THE ASIA PACIFIC REGION

December 2010

Statistics Division
Economic and Social Commission for Asia Pacific Region

Table of contents

1. Background to the guidelines

The growing interest in disability statistics globally has led to improved conceptualizations of disability and related measures to provide statistics. These improved measures were required for two major reasons. The first relates to the lack of comparability across countries. The wide range of variation in reported national prevalence estimates for disability[1] suggested that the measures used did not necessarily reflect the same phenomenon across different countries and that the measures themselves were not the same. A second reason was the growing recognition of disability as being about the whole population and not only a minority group of people with disabilities commonly referred to as 'the deaf, blind, crippled or mentally retarded'. This recognition is embodied in the World Health Organization's International Classification of Functioning, Disability and Health (ICF) (WHO, 2001).

Given these reasons the Washington Group on Disability Statistics was formed as a result of the United Nations International Seminar on Measurement of Disability that took place in New York in June 2001, to develop measures of disability for use in censuses and surveys. This was in parallel and with a different purpose to developments on the use of the ICF for clinical purposes. [2]

The development of questions for use in censuses and surveys requires a process of testing and revision to ensure that the questions are in fact measuring the intent of the question. The outcome of testing provides evidence on whether respondents understand the questions and are able to use the response options effectively. The conclusions from the testing process allow the questions to be accepted as is (because they work well), revised (if there are some mis-interpretations or other problems that can be fixed) or rejected (if the questions do not perform well at all).

Disability statistics (or any other statistics based on self-report measures) are really only useful if the performance of the questions is understood. This understanding provides a transparency in the statistics that explain what they reflect and measure.

1.1 The ESCAP Project on improving disability measures and statistics in Asia-Pacific

The UN Development Account project 'Improvement of Disability Measurement and Statistics in Support of Biwako Millennium Framework and Regional Census Programme' (hereafter the ESCAP project) aimed to train people from statistical offices in the region on question performance evaluation and use of disability measures and statistics. This project was successful in providing staff from the six participating countries with a good grounding in methodological issues for survey measurement of disability. This included cognitive testing approaches and pilot testing of questions. In addition, participating countries were integrally involved in the analysis and write up of the results. Staff from project countries expressed the importance of having acquired these skills and

---

[1] See UN's DISTAT at http://unstats.un.org/unsd/demographic/sconcerns/disability/disab2.asp (Accessed 7 Nov 2010)
[2] See for example Stucki, Gerold; Cieza, Alarcos; & Melvin, John, (2007) The International Classification of Functioning, Disability and Health: A unifying model for the conceptual description of the rehabilitation strategy. Journal of Rehabilitation Medicine, Volume 39 (4) pp. 279-285.

being able to apply them in all areas of their work, beyond disability measurement. Given this positive outcome of the project, these guidelines aim to provide a way to increase the number of people who can benefit from the project.

1.2   Aims of the guidelines

The aim of the guidelines, therefore, is to document in a comprehensive and coherent manner the overall experience of the project in a way that can be replicated by other countries wanting to embark on collecting accurate disability statistics. The handbook's target groups include National Statistical Offices wanting to replicate the tests, and other stakeholders wanting to learn from and build on the project's experience, and apply the methodology in other fields.

Specifically the handbook includes information on the following:

1) Why the testing was conducted and why both cognitive and pilot testing was required (section 3)
2) How the cognitive test was organised, conducted and analysed (section 4)
3) How the pilot test was organised, conducted and analysed (section 5)
4) How these combined processes helped to determine the most effective set of disability questions (section 8)
5) How this process can be used in areas other than disability (section 8)
6) Available resources to be consulted (section 9)

2   Current trends in disability measurement and statistics

2.1   Shift of thinking on disability and its influence in measuring it

Disability measures are evolving rapidly and there has been a significant shift in approach over the last 15 years linked to the changing understanding of disability. The main changes that have occurred are as follows:

- Disability is no longer seen only as being blind, deaf, a wheelchair user or intellectually disabled, but it is about the functioning of the entire population. The importance of measuring the population level of functioning is to inform policies for health and health related issues; to aid the development of health interventions required, and to address aspects of the environment that need to be modified to meet all impairment needs of people without consideration of whether they see themselves as disabled or not (see next bullet point).
- The measurement of disability has shifted from asking about 'disability' to asking about 'difficulty' people have in a series of domains. In addition, the response options have changed from using a dichotomous 'yes/no' response format to a scale of four or five response options such as 'no difficulty, some difficulty, a lot of difficulty, cannot do at all' or 'none, mild, moderate, severe, extreme'. These changes have yielded more inclusive measures (Schneider, 2009; Schneider, Dasappa, Khan and Khan, 2009). For example, elderly people and people with chronic health conditions report having difficulties but not to having a disability. Similarly, people are more willing to report a gradation of difficulty (some difficulty or a lot of difficulty) rather than having to choose between only 'yes' or 'no'.

- There has been an increased visibility of disabled people through the disability rights movement and the adoption of the UN Convention on the Rights of Persons with disabilities (UNCRPD). The UNCRPD is being ratified by a growing number of countries which then need to report back on the situation of people with disabilities in their countries. This reporting requires measures of disability that are accurate and reliable and which are transparent in what they are measuring.
- The field of public health is focusing more on the consequences of living with health conditions such as chronic illnesses, injuries and other physical and psychological traumas. This is a shift from only looking at mortality and morbidity. Functional status or presence of disability can be an indicator of the effectiveness of a medical intervention. One example is the provision of antiretroviral therapy to people who are living with AIDS, allowing them to regain their functioning (self care, mobility, concentration, etc.) and being able to be independent and employed. The measure of functional status thus allows a broader socio-economic analysis of effectiveness to be evaluated beyond only laboratory tests.
- The current measures of disability in the Short Set of the Washington Group are focused on basic domains of functioning as these are the simplest and most accessible domains that respondents can report on in a way that is consistent and comparable across different countries and population groups.[3] This provides a measure of people with difficulties in basic domains and further analysis and measures provide more comprehensive information on the other aspects of disability such as education, employment and social inclusion levels of people with and without difficulties.

The shift in approaches to measuring disability reflects the changing models of disability and how it is defined. The individual or medical model of disability was (and at times remains) the predominant model until the 1970s, after which the rise of the disability movement brought to the fore the social model of disability. The medical model highlights the problem as being within the individual and the focus of intervention being medical or rehabilitation intervention. Little attention is paid to the role of external factors in affecting the person's participation in major areas of life. The social model, in contrast, highlights the role of these external factors as the source of the problem and advocates for interventions that change these rather than the individual. Thus the social model advocates for accessible transport and buildings and changing of discriminatory attitudes towards persons with disabilities, for example, rather than interventions that focus on making the individual adapt to his or her environment.

Disability is about both individual and environmental factors; medical and rehabilitation interventions are necessary but so are interventions that make the environment more inclusive and change negative attitudes. The International Classification of Functioning, Disability and Health (ICF) was adopted by the World Health Organization (WHO) in 2001 and highlights these different aspects of human functioning in a comprehensive set of classifications. The model espoused by the ICF is the biopsychosocial model and that is the approach that underlies the work of the Washington Group on Disability Statistics.

---

[3] The work of the Washington Group on Disability Statistics, discussed below, is growing evidence of this.

2.2   Washington Group on Disability Statistics (WG)

The growing interest in developing more standard measurement tools led to the United Nations Statistical Division hosting the United Nations International Seminar on Measurement of Disability in New York in June 2001, to look specifically at the issue of disability measures and statistics. An outcome of the meeting was the establishment of the Washington Group on Disability Statistics (WG), a city group reporting to the UN Statistical Commission. The work of the WG is to develop measures of disability for surveys and censuses that are: a) compatible with the WHO International Classification of Functioning, Disability and Health (ICF); b) reflect the complexity of disability; and c) are comparable across countries and different population groups. [4]

The objectives of the WG were, and continue to be,[5]

1. To guide the development of a small set(s) of general disability measures, suitable for use in censuses, sample-based national surveys, or other data collection instruments, which will provide basic necessary information on disability throughout the world.
2. To recommend one or more extended sets of survey items to measure disability or principles for their design, to be used as components of population surveys or as supplements to specialty surveys. These extended sets of survey items will be related to the general measures.
3. Measures identified in objectives 1 and 2 will be culturally comparable to the extent possible. The ICF model, a useful framework to assist in the development of these measures, will be utilized in developing the measures.
4. To address the methodological issues associated with the measurement of disability considered most pressing by the City Group participants.

The first meeting of the WG was held in Washington DC, USA (hence the name of the group) in February 2002 and was attended by 64 participants from 32 countries, including representatives from disability organisations.[6]

The two main outcomes of the work of the WG so far are a Short Set of six questions for use in censuses since space for questions is highly limited, and the initial components of an extended set of questions. This Short Set was tested in a number of countries, both developed and developing, and indications are that it produces reasonably comparable measures of disability. This is evidenced by the similar trends in findings on how people interpret and answer the questions.[7] Countries are increasingly starting to use this WG Short Set in their censuses (e.g. Viet Nam 2009, Brazil 2010, the Philippines 2010), producing more internationally comparable data. The development and testing of the WG Short Set questions showed that questions are tapping into the construct of disability and functioning as proposed by WHO's ICF-approach and that this is a productive path to follow.

---

[4]See http://unstats.un.org/unsd/methods/citygroup/index.htm - accessed on 13 June 2009 – for more information on the UNSD's various city groups.
[5]WG Website, accessed 13 June at http://www.cdc.gov/nchs/about/otheract/citygroup/objectives.htm
[6]For more information on other meetings refer to www.cdc.gov/nchs/washington_group/wg_meetings.htm
[7]Miller, K, Mont, D, Maitland, A, Altman, B & Madans, J. (2010) Results of a cross-national structured cognitive interviewing protocol to test measures of disability. *Quality and Quantity*; DOI:10.1007/s11135-010-9370-4

The WG Short Set asks six questions each covering a different domain of functioning. The six domains are: seeing, hearing, mobility, cognition, self care, and communication. The set of questions are presented in box 1. The same response options are used for each question:

a. No - no difficulty
b. Yes – some difficulty
c. Yes – a lot of difficulty
d. Cannot do at all

Box 1: Washington Group Short Set

Introductory phrase:
The next questions ask about difficulties you may have doing certain activities because of a HEALTH PROBLEM.

1. Do you have difficulty seeing, even if wearing glasses?

2. Do you have difficulty hearing, even if using a hearing aid?

3. Do you have difficulty walking or climbing steps?

4. Do you have difficulty remembering or concentrating?

5. Do you have difficulty (with self-care such as) washing all over or dressing?

6. Using your usual (customary) language, do you have difficulty communicating, for example understanding or being understood?

The second task of the WG is to develop an extended set of questions to complement the Short Set in order to provide a more comprehensive description of disability at the national level. The cognitive and pilot testing of the extended set took place in 2009 and data were analysed in 2010. These guidelines are the outcome of this testing.

The WG Short Set aims to identify the population at risk of experiencing the disadvantages typical of disability. While this Short Set seems to provide reasonably good measures of population functioning, it is limited in that only one question is asked and only of a limited set of domains of functioning. The population thought to be most excluded from these measures (i.e. not counted in) are people with psychiatric and emotional disabilities (e.g. depression and anxiety) (WG annual meeting discussions, 2007, 2008, 2009).

The aim of the Short Set is to create a demographic variable that can be used in further analysis, in the same way that we have sex and age variables. Disability is not only about difficulties in seeing, hearing, walking, remembering and concentrating, washing and dressing, or communicating, but about the effect of these difficulties when they interact with a person's environment to create disadvantage (e.g. loss of employment, limited education, social exclusion). Thus, a full description of disability in a population comprises measures using the WG Short Set or extended set together with questions on activities such as working, attending school, social interactions, and civic engagement. The level of participation in these life areas is analysed in relation to the degree and

type of difficulty people report. The results of this analysis are the disability related disadvantage people experience.

In summary, the status of questions developed by the WG is (as of end of 2010) that the Short Set is complete and currently being used in a number of censuses and surveys, and the extended set is partially developed with the extended sets for some domains being finalised while for others further testing is required. Reasons for conducting the cognitive and pilot testing

3 Reasons for, and overview of cognitive and pilot testing

3.1 Question evaluation

A good question is one that is relevant to both the research agenda and each potential respondent's experience and knowledge. Question evaluation through cognitive and pilot testing allows poorly performing questions to be fixed to ensure that the questions capture the intended concept (e.g. difficulty in various domains of functioning) and to ensure that data will be comparable across countries and across different sectors of the population in one country.

Question evaluation identifies and documents what questions measure, including errors in the question construction, identifying non-problematic differences (e.g. patterns of correct interpretation, and of calculation, estimation and forming answers), and contribute to the analysis and interpretation of survey data.

3.2 Self reporting in surveys and censuses

Most socio-demographic statistics are obtained using self reported measures with the respondent reporting directly about themselves or about a member of their household (proxy reporting) who is either too young or is too ill to self-report, or has a significant communication or cognitive difficulty that prevents them as an adult to self-report.

The advantages of these measures are that they are easy to apply and do not require complex assessment procedures that can only be done by trained professionals. Thus these are relatively straightforward measures to obtain population level data. In addition, the self-reporting nature of these methods provides a measure of perceptions by the respondents. For example, a measure of health care provider effectiveness by users provides information on how the respondents perceive the service and why they may or may not be using the service. This is not the same as obtaining information through observation of the service provision. These two sources of information complement each other.

The main disadvantage of self reported measures is that they are subjective and reflect individual interpretations, views and life contexts. The response provided is only as good as the respondents' accuracy of understanding, interpretation and selection of responses from the set provided. The question as written and intended by the questionnaire developer takes on a whole other dimension when posed to a respondent who has no real understanding of the survey aim, and may not have experience in responding to survey questions.

The process of responding to a survey question is complex (despite it happening almost instantaneously) and involves a series of cognitive processes. An accurate response only arises if

respondents understand the question, retrieve the necessary information from their memories, review this information and then map their response onto the responses provided for that question. This is the process that informs the reasons for the cognitive testing of survey questions as discussed later in these guidelines in more detail. Of interest here is to note that self reporting is problematic if respondents are not able to carry out these complex processes effectively. The cognitive testing aims to 'check' whether the respondents' understanding and interpretation of the question are correct or not.

A further consideration on potential determinants of what information is provided by respondents are factors such as age, sex, socio-economic context and cultural beliefs. Thus the purpose of cognitive and pilot testing are to understand how questions are interpreted and understood in order to understand whether the questions are performing as intended (cognitive testing), followed by an assessment of the extent of these patterns of interpretation in a larger population (pilot testing).

The cognitive testing tries to determine what stages of cognitive processing yield the major problems, if any. The underlying theoretical framework for question responses includes the following cognitive processing stages as set out in the Cognitive model of Question-Response:

Table 1: Cognitive model of question response

		Cognitive Stage Definition	Response Errors/Question Problems
Stage 1	Comprehension	Respondent interprets the question	Unknown terms, ambiguous concepts, long and overly complex
Stage 2	Retrieval	Respondent searches memory for relevant information	Recall difficulty
Stage 3	Judgment	Respondent evaluates and/or estimates response	Biased or sensitive, estimation difficulty
Stage 4	Response	Respondent provides information in the format requested	Incomplete response options

Adapted from Miller and Willson, 2010 [8]

3.3   Cognitive testing

Cognitive testing is done primarily to collect narratives from respondents to analyse their understanding and interpretations of the questions asked. The pattern of responses across a limited sample of respondents provides insight into the common patterns of understanding and interpretations of people who are typical of the targeted survey respondents.

The methodology is a qualitative one using a limited sample size but with detailed information provided by each respondent elicited in semi-structured interviews. The analysis is primarily thematic in nature and aims to identify common errors in interpretation which will highlight weaknesses in the formulation of the survey questions.

---

[8] The source of this table is the presentation done by Kristen Miller and Stephanie Willson at the July 2010 ESCAP workshop in Bangkok, available on the project website. www.unescap.org/stat/disability/pre-pilot-training/ (accessed 15 November 2010).

Typically a cognitive testing interview will ask the question and elicit a response from the respondent, followed by a discussion on why the respondent gave that answer, what he/she was thinking about when choosing an answer, and what he/she thinks the question is asking about.

Cognitive testing is the first phase of question performance evaluation and is done usually after completion of the first draft of a questionnaire to determine whether there are any significant design problems in the questions. While ideally all questions on a questionnaire should be evaluated, practically this is time consuming and it is difficult to do in a way that maintains a respondent's interest and concentration. Typically measures of new topics or new measures for old topics (e.g. as is the case for disability) should undergo cognitive testing or performance evaluation.

The benefits of cognitive testing are that inherent problems and biases in questions are identified before collecting data from a large sample. The cognitive testing process gives an opportunity to revise the questions and to provide transparency of what is being measured by different questions. Analysts can reach more conclusive decisions as to what the survey data tells them if they have the knowledge of how people understand and interpret the survey questions. This means that, while questions may not always be revised, the range of possible responses is known and can be used in the interpretation of statistical findings from survey data. In other words, cognitive testing provides a better understanding of possible respondent measurement error.

An example from the domain of communication is provided to show the themes that are identified from the narratives of the cognitive testing interviews. The question asked was the following:

> 'Using your usual (customary) language, do you have difficulty communicating, for example understanding or being understood? '

The range of interpretations provided by respondents includes:

1) Physical impairments, whereby respondents described problems with their tongues or mouths that prevent them from being able to speak clearly;
2) Cognition-related problems, in which respondents described difficulties remembering or concentrating such that it is not easy to focus on what others are saying or to speak at length, for example, to tell a story;
3) Hearing-related problems that prevent respondents from being able to clearly hear what others are saying, and
4) Social or interactional difficulties, whereby respondents described having problems interacting or relating to others. These social difficulties could also be broken down into sub-categories, specifically, a) respondents expressing difficulty because they are shy, b) because they talk too fast, c) because of interpersonal problems relating to others such as a spouse or child, or d) because they do not have much education and feel insecure talking to those who do.

The range of interpretations can then be analysed as being 'in scope' or 'out of scope'. In scope interpretations are essentially correct interpretations where the narrative shows that the respondents has understood the question and is responding appropriately.

Table 2: Number of respondents from each ESCAP Project cognitive testing participants for each description of communication difficulties

Country	General-communication skills[9]	Physical	Cognition	Hearing	Social/Interactional				Language
					Shy	Fast-talking	Interpersonal	Education	
Cambodia	7		1		2	1	1	4	1
Canada	1			1		1	3		1
Kazakhstan	6			2					3
Maldives	2	1		2					1
Mongolia	2			1					
Philippines	2	1		2					1
South Africa	1								3
Sri Lanka	1	1					1		
United States	7		1			2	2	1	
Percentage across all countries	40	4	3	11	3	5.3	9.3	6.7	17.3

* Table includes only the cases that included enough detail to categorize within a particular interpretive theme.

While the first three themes (physical, cognitive and hearing) are clearly in scope, this is not entirely true for the social or interactional theme. Indeed, some of these types of difficulties could be learning or affect-related problems. Or it could also be possible that at least some of the respondents are reporting out of scope problems. For example, one of the narratives appears to indicate that the respondent based her answer (some difficulty) on the quality of her marital relationship:

> Respondent : Just trying to get my point across … say if my partner isn't listening …
> Interviewer: Do you feel that you have difficulty describing things to me?
> Respondent: No …well it's not so much delivery of the message…it's being heard.

The questions are revised based on these different interpretations identified in the cognitive testing. The revisions applied to the questions could include a series of probes for use in the pilot test as described in more detail below.

3.4   Pilot testing

Pilot testing is the next stage in the testing or question evaluation process. While the cognitive testing identifies the correct and incorrect interpretations of questions and generates an initial revision of questions, the pilot test can check the extent to which these particular patterns of interpretations are prevalent in a larger sample of respondents.

Pilot testing is not the same as running the survey. The aim of pilot testing is still part of the process of question performance evaluation and is not a substitute for collecting data on a population. The pilot testing results should not be used to provide population estimates on, for example, disability prevalence or the measure being evaluated, but should be used to refine questions and note

---

[9] This refers to difficulties in communication but with no specific mention of factors related to the other categories.

patterns of responses by age groups, males and females, and people from different language, cultural and socio-economic backgrounds.

Based on the patterns of correct and incorrect interpretations identified in the cognitive testing, a series of probes can be developed as question items for the pilot test to generate data on the prevalence of these in a larger sample of respondents.

The example of the probes developed from the cognitive testing narratives on communication difficulties is as follows. The respondent who has communication difficulties is asked: 'Is this difficulty….' followed by the different probes. The response options are 'yes', 'no', 'refused' and 'don't know':

1. Because you sometimes feel shy or have trouble expressing yourself?
2. Because of a physical problem with your mouth or tongue?
3. Because you need to understand other languages or different ways of speaking?
4. Because you sometimes talk too fast?
5. Because you have trouble hearing?

In relation to the example for communication, some of these are in scope, others out of scope and some could be either in or out of scope. Probes 2 and 5 are clearly in scope and indicate a clear problem that could cause a communication difficulty. Probes 3 and 4 are clearly out of scope as they do not describe a cause that is health related. Having difficulty with a second language is not a health related problem but is one of language learning. Probe 2 is not clear in its meaning as it could be signalling some health related underlying problem (e.g. emotional problem) or could merely be a personality trait that is not health related.

The analysis process is a straightforward statistical analyses using frequencies, crosstabulations, correspondence and regression to determine the interactions of question responses. The aim is to understand the implication of certain trends and associations between different responses and associated factors. As stated, above the analysis is not done to obtain prevalence estimates for the sampled population. The benefits of the analysis of the pilot test data is to provide a more extensive review of the question performances and the frequency of occurrence of the different probes identified in the cognitive testing data.

4   The cognitive testing

This and the next section describe the practical steps required for doing a cognitive and pilot test for question evaluation. Details of documents and materials used in the training for the ESCAP cognitive testing training can be found at ESCAP's Statistics Division website.[10]

4.1   Aim and objectives

The objectives of the cognitive test are to get sufficient interviews and narratives within the interviews to provide an insight into how people understand and interpret the questions.

The patterns of mis-interpretation of the questions provide a good source of information on potential error in survey data. However, the aim is to reduce these errors as far as possible and to

---

[10]www.unescap.org/stat/disability/pre-pilot-training/ (accessed 15 November 2010).

understand those that cannot be changed. The outcome of the cognitive testing analysis is to show where and how questions need to be revised prior to pilot testing them.

There are two main components to the cognitive testing – the standard interview where each question is asked as developed and a response requested from the respondent. This is followed by the process of 'getting the story' behind the response. The suggested methodology is to follow up each question response with probes[11] to get at the story of why the person gave that response (see probes to be used below).

Appendix A provides the link to the set of questions used in the ESCAP project cognitive testing of disability measures. A critique of this set of questions is given in section 4.2 below. The questions are the Washington Group extended set of questions which incorporate the Washington Group Short Set.[12]

Ten domains of functioning were included in the extended set including the six included in the short set. The ten domains were vision, hearing, mobility, communication, cognition, upper body mobility and self care, learning, affect (anxiety and depression), pain, and fatigue.

The different components and stages of cognitive testing are the following:

- Developing the questionnaire
- Selecting and training of interviewers
- Practice interviews
- Number and Selection of respondents
- Note taking and audio recording
- Summarising notes
- Capturing the narratives in a software or programme to facilitate the analysis
- Analysis
- Revision of the questions

4.2 Developing the questionnaire

The questionnaire should include enough information to allow for analyses by different socio-demographic variables as well the questions targeted for evaluation. The translation of the questionnaire is a crucial step in the overall process. This is addressed in section 7 below.

The targeted questions should ideally be presented in the order that they will finally appear on the survey questionnaire. However, if there are some questions that are likely to require more discussion they may be better placed at the start of the interview. In addition, it is useful to include a simple question at the start to familiarize the respondent with the approach.

While a question on income does allow for analysis by socio-economic strata, most questions currently available on employment or income are generally cumbersome and tend to be sensitive. If such a question is required it should be placed preferably at the end of the interview.

---

[11] These 'probes' are not the same as those discussed in the previous section. Here we are merely referring to additional questions that encourage respondents to tell the story behind their response.

[12] The results of the ESCAP project testing are available at http://www.unescap.org/stat/disability/analysis/

The interview should have a guide as to the probes to be used for eliciting the narratives. These can either be on the questionnaire (see Appendix A for an example of this) or set out in a separate document.

The example provided in Appendix A has a number of problems listed below. This critique is provided as an example of lessons learnt from doing cognitive testing beyond the question evaluation. The process is important as it can determine the quality of the information obtained.

1. The first is that there are too many questions. The quality of the narratives for the later domains in the ESCAP cognitive test were seriously compromised by the respondents just being too tired to give more than the response to the question.
2. The second problem with the questionnaire is the repetition of questions asking about the age of onset, health vs non-health cause of the respondents' difficulty, and impact of the difficulty. For those respondents with difficulties in more than one domain (as is common), these questions became very repetitive.
3. A third problem with the ESCAP cognitive testing methodology was that the whole questionnaire was administered as a standard interview followed by a repeat of the questions (with responses) to elicit the narratives about the responses. This proved to be too lengthy a process and should not be used.

4.3   Training of interviewers

The qualitative nature of cognitive testing means that few interviewers are required but they should be knowledgeable in doing in-depth interviews in order to elicit sufficiently detailed narratives. The ESCAP project had at least two interviewers per country. This allows for observation of each other while learning to conduct the interviews, as well as providing support to each other. Thus the recommended number of interviewers is at least 2 to provide support and monitoring of each other's interviews.

The length of training is dependent on the experience of the interviewers. For those interviewers experienced in doing in-depth interviews and collecting detailed narratives from respondents, 2 days of training should be sufficient to ensure good cognitive testing data collection. The training would cover the points 1, 2, 3, 4 and 6 as set out below for the training for inexperienced interviewers.

For those interviewers not experienced in detailed in-depth interviewing should undergo a 4 day training which covers:

1. What is cognitive testing and why is it necessary for survey questionnaire development
2. The cognitive processes required of respondents in self-reported survey questions
3. Full knowledge of the intention of the questions being tested
4. The probes to be used in cognitive testing interviews
5. Doing in-depth qualitative interviews and the notion of narratives as data
6. Practical exercises in doing cognitive interviews with critical discussion with the whole group of trainees. This would include using the probes and taking notes, and following up if the respondent does not understand with the following strategies:
   o Read the question again
   o Do not tell them what your understanding is, ask them for their understanding of the question

- o Try and get them to explain what they do not understand
- o If they still cannot answer go to the next question.

The materials for training should include:

- Slide shows on cognitive testing and why it is necessary [13]
- The questionnaire to be tested
- A detailed question by question guide for the questions to be evaluated
- A set of possible probes for use in the interviews. These would include questions such as:
  - o 'Why did you answer in that way?'
  - o 'Tell me a little bit more' or indicating to the respondent that you are listening but want more.
  - o 'Can you think of an example of what you are talking about?'

4.4 Number and selection of respondents

Typically, as for other qualitative research, the sample size for cognitive testing is small. For the ESCAP project, each country undertook 20 interviews. The total number of interviews for this testing round was around 150, which included interviews from countries participating in the ESCAP project and others interested in testing the same question set. For an individual country, it may sufficient to do between 20 and 30 interviews to get a good evaluation of the questions. The process of doing the interviews is time consuming as is the analysis process. Thus, it is better to do less interviews but to do them effectively and with detailed narratives than to try and do too many and not generate the required depth of information. Thus it is better to do 10 very good interviews than 20 – 30 poor interviews.

When selecting respondents, it is important to focus on the following features:

- o Mix of men and women
- o Mix of direct and proxy respondents (should the targeted questions be useable in a proxy response interview)
- o Different ages
- o Different socio-economic and educational levels, although problems in understanding the questions seems to be more apparent the lower the educational and socio-economic levels
- o If the questions are about a particular characteristic of life (e.g. disability, employment status) it is important to recruit a mix of people with and without that characteristic (e.g. disabled and non-disabled people, employed and unemployed people)

4.5 Practice interviews

Practice interviews are crucial as part of training. These should be done in front of all trainees as well as in pairs. The experience in pairs should be observed by trainers and discussed in plenary as well. At least two practice interviews should be carried out by each trainee, and possibly a further two done under supervision once training is completed.

---

[13] See the materials prepared for the ESCAP project training and analyses workshops in 2009 and 2010 for examples of slide shows to be used. See www.unescap.org/stat/disability/pre-pilot-training/ (accessed 15 November 2010).

Note taking and audio recording are crucial aspects of the interviewing process. The questionnaire should be printed with sufficient space for notes to be made as the interview progresses. In addition, audio recording of the interview is an important adjunct. While the audio recording is a very useful tool it is not necessary to transcribe the full interview. The audio recording serves as a tool to check any ambiguity or confusion in the notes. Appendix A is presented in this document with little place given for note taking for reasons of space. The document used in the actual cognitive interview should have more space for comments and notes to be made as the interview progresses.

Because the cognitive testing interview involves examining what a respondent is actually thinking or feeling when answering a question, the interview can seem somewhat personal or even strange to respondents, and it is important to make this clear to respondents when they have volunteered for such an interview.

4.6   Summarising notes

Once completed, the interview must be reviewed as soon as possible after the event to ensure good memory of what was said. Additional notes can be entered and summary points made on, for example, the correct or incorrect interpretation of the questions, themes on interpretations, problems with understanding the questions and requirements for repetitions. These summary points form the basis of the combined analysis of all the interviews.

The ESCAP project made use of a new tool for entering the data from cognitive interviews. This web-based programme called Qnotes[14] was developed by the National Center for Health Statistics (NCHS) in Washington DC, USA (see list of resources at the end of the document). This tool allows for the narrative data from the interviews to be entered as notes. The analysis is then simplified as it allows all responses and narratives to one question to be downloaded into a single file. An example of this is shown in Appendix B for responses to different domains of the short question set. This is the raw data filtered by response option selected, by sex, by age, by education level, by socio-economic status or whatever variable is relevant for the purpose of the test.

Capturing the narrative data into Qnotes (or a similar programme) requires as much detail as is possible to allow for as detailed analysis as possible. The essence of the analysis is the stories that come from the respondents.

As part of the training process it is important to include entering notes onto Qnotes. In addition, the first 'real' interview should be entered and comments requested before moving onto doing further interviews so that feedback can be provided on whether the interview has been adequately conducted or not. While the concept of doing a qualitative in-depth interview seems simple, the practical running of such an interview is not so simple. Thus, the trial and comment process is crucial as part of the learning process. The approach to interviewing will change according to comments and critiques of these early interviews.

---

[14] The NCHS Qnotes developers are keen for the tool to be used by any party interested in using this programme and are also willing to provide technical assistance where possible. In the ESCAP project this took the form of comments being inserted into the Qnotes interview data asking, for example, for more detailed explanations (a typical problem noted in this round of testing). In order to use Qnotes permission and a password are required.

Each interview should also include an overview comment about the context of the interview. Examples of this are given in Box 2. These are the actual summaries collected in the ESCAP project.

Box 2: Example of summary comments on interview context

Example 1**:** Interview was with a 62 year old man who had motor issues related to diabetes, heart disease and fairly significant loss in processing capacity. He had a lot of trouble answering these questions exactly and did not understand the idea of day to day activities versus activities outside the day to day realm.

Example 2: The interviewee is a housewife and willing to assist us in this cognitive tests. It took about two hours to complete this interview as I tried to go slow - testing the translated version and the English at the same time. With the limitation of our language I find it easier as well as the respondent to communicate in English and this is because the respondent is literate. The interview was done at her home.

Example 3: The respondent was willing to answer the interview. However, he was not good at talking about most questions due to his problem of speaking and loss of control when speaking – he is a deaf person who struggles with verbal language, and faces difficulty in speaking when the conversation was taking quite long. He said that he experienced difficulty hearing and speaking since he was 7 years old. He said he experienced serious illness of nervous convulsion when was 7. He also mentioned that when he was 14, he again experienced a heavy fever. Now he is 27 and still single.

Example 4: The respondent was very articulate and thinkfull [thoughtful]. She is a junior high specialized teacher for problematic learning kids. She had a ski accident in 2005 and suffers from pain since and had a depression episode during teenage years for which she's still taking medication.

4.7   Analysis

Once all the narratives are entered into Qnotes the analysis can begin.[15]   The section here provides the overall steps and aims of the analysis with illustrative examples from the ESCAP project.

The goals of question evaluation through cognitive testing are to determine the following:

- How do the respondents understand the survey question?
- Do respondents understand the survey question differently?
- Does the question mean the same in all the languages that it is asked?
- Does the question mean the same in all of the cultures that it is asked?

---

[15] A detailed set of slides on the analysis process is available from the ESCAP project website at http://www.unescap.org/stat/disability/ws-July10/cognitive-interviewing.pdf.

- In processing a question, do all respondents recall information and form an answer the same way?

The analysis according to the goals set out above will tell to what extent the data elicited from the questions being evaluated are representative of the phenomena being measured; in other words, do the questions in fact measure what we want them to measure? And in what ways, if any, are the data distorted because the questions are not measuring the intended construct?

The analytical approach is investigative in nature and identifies both patterns of error and patterns of non-error, that is:
1) identifying in-scope and out-of-scope cases, and
2) determining whether there are patterns in those cases.

The analysis tries to determine what stages of the cognitive processing problems arise as described in section 2.4 above on self-reporting in surveys. Are the problems at the comprehension, retrieval, judgement or response selection stages? The analysis looks at why the respondent answered the question as they did and whether their story matches with their response and the intent of the question. This analysis gives insight into potential response errors, patterns of interpretation, and socio-cultural factors that affect the response process.

a) Organizing the data

The analysis can proceed one interview at a time or can be done one question at a time. The discussion below will focus on doing the analysis one question at a time. The strategy is to collect all the narratives for that question from the interviews and organize it as set out in Appendix B. The narratives can be organized by response given (no difficulty, some difficulty, a lot of difficulty and cannot do at all), by sex (males and females) or country (if more than one country is involved). The narratives are read carefully one interview at a time and the different levels of analysis applied as described below.

b) Levels of analysis

The different levels of analysis are set out in the figure below (figure 1) taken from Miller, K et al (in press).

Figure 1: diagrammatic representation of the required levels of analysis for comparative cognitive testing

Required levels of analysis for comparative cognitive testing

Required Levels of Analysis

1. Question response processes (Individual description) — In-interview Analysis
2. Patterns of response process (What the question captures) — Across Interview Analysis
3. Patterned differences in sub-groups (Potential For Bias) — Across Sub-Group Analysis
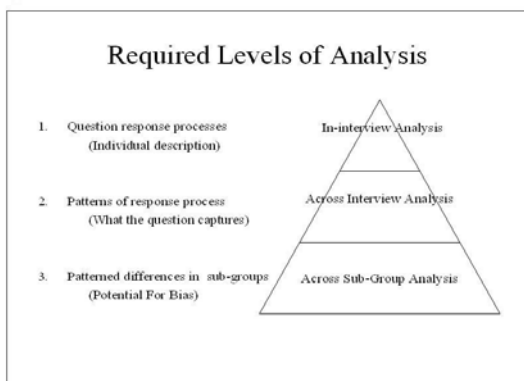
- Level 1: The first interview is reviewed and a comment made about the respondents answer and related narrative. This is the within interview analysis and gives an indication of basic response errors. Typically response errors are noted when the narrative and the answer given do not match because the respondent is unable to remember relevant information, misunderstands the question, is not listening to the whole question, or the response categories are not appropriate.

Box 3: Example of level 1 analysis for cognition domain (remembering or concentrating)(R = respondent; I = interviewer)

<u>Example 1: Respondent with 'some difficulty' remembering or concentrating and with correct interpretation</u>

I: Do you have difficulty remembering or concentrating to do something?

R: I have a problem of remembering. With remembering… if I can get in to the car holding something, if I think of something else, I can forget the object in the car.

I: So you have some difficulty?

R: Maybe as I keep growing older, but now it's not that bad….Like when I'm having something in my hand, when I put it down it happens that I forget it for a long time; like when I'm wearing a blazer - when I take it offer… it happens that when I want to go somewhere I go without the blazer. [i.e. I forget it]

<u>Example 2:</u> Respondent with an unclear response option and incorrect interpretation of question:

I: Do you have difficulty remembering or concentrating to do something?

R: Sometimes I concentrate on doing something such as doing something using my hands. I do concentrate and finish it but something that involves money I struggle to remember and concentrate because I don't have money.

….

I: Can you explain what is it that involves money that you cannot remember and concentrate to do, can you explain what is it like?

R: It's too much. Even here at home many things need money. I just say it in my heart but I cannot implement because I don't have money. This can be like building a house. But going to the bush to cut / fetch grass or wood it is easy to remember and concentrate doing it unlike something that involves money.

- Level 2: The next interview is reviewed and a comment made about the answer and related narrative, and so on until all interviews have been reviewed for a particular question. A tally is made of how many different patterns of interpretation have been identified. This step shows the patterns of interpretations and give information on the ways in which respondents interpret the questions and use the response categories.

The domain of cognition presents a number of challenges for respondents as the intent of the question was not understood by six respondents (out of 10). The remaining four respondents understood and interpreted the short set question correctly. Of these correct interpretations, two reported 'no difficulty' and two reported 'some difficulty'.

Incorrect interpretations: Emotional remembering and reminiscing:

The main incorrect interpretation applied was that of 'emotional remembering' where respondents talked about remembering things that make them sad or unhappy and their 'heart sore'. Respondents 1, 3, 4, 7, 8 and 9 all gave explanations or descriptions of their difficulty as being related to remembering or thinking about difficult aspects of their lives. They described how reminiscing about their difficulties was emotionally painful.

The reasons for this incorrect interpretation are not clear. The fact that four respondents correctly interpreted the questions suggests that the translation is correct. The six incorrect interpretations are more likely to highlight how people understand the question and the notion of remembering. A way to avoid this would be to give clear references to clarify the question intent, such as examples of remembering what needs to be done each day, remembering people's names, etc.

- Level 3: These are then analysed by different sub-groups, such as men vs women, older vs younger respondents, low vs higher educational level, urban vs rural, and so on. This identifies potential biases inherent in the question, where one subgroup seems to be responding differently to other subgroups; e.g. males vs females; younger vs older respondents.

- Level 4 (not on figure): An inventory of interpretations is drawn up and explanations of how different interpretations relate to each other and to the intent of the question are developed.

Specific aspects to look for include correct application of time and distances references. For example, the Washington Group Extended set of questions asked about 'in the past 3 months', 'the last time', 2 litres jug of water, 100 metres, 500m, and so on.

Once the different levels of analysis have been completed, a summary report on the performance of the questions must be drawn up using a table as shown below (table 3) as an example. The table presents a couple of examples from a hypothetical sample of 20 interviews.

Table 3: Examples of the content of a summary report on cognitive analysis using hypothetical information

Question	Number of respondents with correct in scope interpretations	Number of respondents with incorrect out of scope interpretations	Nature of incorrect interpretations	Subgroup Bias
Do you have difficulty seeing even when wearing your glasses? (VIS_SS) (N=20)	16	4	Comprehension problems because of glasses clause	No evident subgroup biases
Do you have frequent pain? (PAIN_1)	12	8	• Judgment error linked to difficulty in understanding 'frequent' (5 interviews) • Recall difficulty of not being able to remember all instances (2 interviews) • Response error where respondent does not give any clear response even when prompted (1 interview)	Effect of age is seen in response error; lower education effect seen in comprehension errors. Women tend to report pain more often than men.

4.8 Revision of questions for pilot testing

Once all questions have been analysed for all interviews, a decision is made on what revisions are required. These would include the following:

- The number of different interpretations noted and how many were 'in scope' and how many 'out of scope';
- The number of respondents with in and out of scope interpretations. If the majority of respondents have in scope interpretations and narratives that match the answer given, the question can be said to be performing well. The few out of scope interpretations provide some indications of possible response error in a full survey. The pilot test will then test how many respondents generally have these different interpretations and a final decision is made only after those data are available.

In order to prepare for the pilot test, a decision is made to a) keep the question as is; b) revise the question in a minor way; or c) throw out the question and rethink it completely. For questions that require substantial revisions or to be redrafted completely, a new round of cognitive testing of just those questions is recommended.

4.9 Analysis examples

This section gives a range of examples of analysis of cognitive testing information and the conclusions reached. The text is taken as is out of the final report of the ESCAP project. The examples do not report on all the steps undertaken in the analysis as set out in the preceding section.

4.9.1 A domain where questions perform as intended: Hearing[16]

The questions asked about hearing difficulties (WG Short Set), use of a hearing aid and frequency of use, hearing in noise followed by hearing in quiet. People with no difficulty hearing in noise skipped the question on hearing in quiet. All respondents who indicated having hearing difficulties were asked the age when the difficulty started, and whether it had any impact on their ability to carry out daily activities.

Box 5: Hearing questions asked in the cognitive testing interviews

Questions	Response Options
SS2: Do you have difficulty hearing, even if using a hearing aid?	1) no difficulty 2) some difficulty 3) a lot of difficulty 4) Cannot do at all/ unable to do
2.1 Do you use a hearing aid? 2.2 **If yes:** How often do you use your hearing aid(s)?	1) Yes 2) No *If **Yes**, read **hearing aid** in **noisy room** and **quiet room**.* 1) All of the time 2) Some of the time 3) Rarely 4) Never
2.3 Do you have difficulty hearing what is said in a conversation with one other person in a noisy room [***even when wearing your hearing aid(s)***]?	1) no difficulty 2) some difficulty 3) a lot of difficulty 4) Cannot do at all/ unable to do *If **No difficulty**, go to <u>next section</u>.*
2.4 Do you have difficulty hearing what is said in a conversation with one other person in a quiet room [***even when wearing your hearing aid(s)***]?	1) no difficulty 2) some difficulty 3) a lot of difficulty 4) Cannot do at all/ unable to do
11.1i How old were you when the difficulty hearing began?	_____ age in years
12.1i Is your difficulty hearing due to a health problem or something else?	1) Due to a health problem 2) Something else: _____
13.1i Does your difficulty hearing limit your ability to carry out daily activities?	1) Yes 2) No
13.2bi Does your difficulty hearing limit your ability to carry out other activities that are not part of your day-to-day life?	1) Yes 2) No

The analysis of the cognitive testing responses looks at whether the intent of the questions was understood and what confusions, if any, arose from the response options. In addition, the interpretation of the hearing aid clause was analysed.

The intent of the question SS2 (see table 4) seems to have been clear to most respondents. Of the 92 respondents who provided comments on their responses 67 indicated that they had understood the question intent and provided a response that met their description of 'no difficulty' or 'difficulty'. This is noted by the number of examples provided by the respondents highlighting the activity of hearing in different contexts. These included examples of listening to a range of sounds, loud and soft, far and near, playing a musical instrument, hearing birds, diseases of and trauma to the ear. One teacher, who responded with 'no difficulty', commented 'I really don't have any difficulty and

---

[16] The analysis for the first question only is presented but the conclusions refer to all the tested questions. The reader is referred to the full chapter in the report for more details.

my students know that I can even hear them whispering'. Some referred to having had a hearing test which indicated normal hearing even if there was some loss in some of the higher frequencies, or being in 'fine physical form'.

	'Do you have difficulty hearing, even if using a hearing aid?'					
	No	Some	A lot	Unable to do	Skipped; not asked	Total persons
Cambodia	11	3	3	<1	4	21
Canada	14	1	1	<1	1	17
Kazakhstan	14	2	1	1	2	20
Maldives	15	2	3	<1	<1	20
Mongolia	15	3	2	<1	<1	20
Philippines	12	5	2	1	<1	20
South Africa	5	<1	<1	<1	2	7
Sri Lanka	2	1	1	1	10	15
United State	13	3	<1	<1	3	19
Total Persons	101	20	13	3	22	159
Percentage (excl. skipped)	74	15	9	2	-	100

The respondents who reported having difficulty tended to report examples such as difficulty on the phone, in noise, being completely deaf, having a recognized unilateral hearing loss, having tinnitus[17], having problems even when wearing a hearing aid, ageing and reporting various forms of illness or trauma (e.g. noise damage, being kicked on the side of the head).

The 25 respondents (out of 92) who provided ambiguous responses varied in the reasons for this ambiguity. One example was the confusion with the hearing aid clause. Nine respondents responded to the hearing aid clause rather than about hearing. Most were able to respond appropriately once the confusion was explained. The confusion occurred only with respondents who reported 'no' or 'some difficulty'. None of the respondents reporting 'a lot of difficulty' or 'cannot do' were confused by the clause. Some of the confusions lead to respondents saying 'no difficulty' because they interpreted the question to be about a hearing loss that is severe enough to warrant the use of a hearing aid. For example, one USA respondent said 'Yes I do have a problem hearing but I don't wear a hearing aid'. He reported having 'no difficulty' hearing, when in fact he should have responded as having 'some' or 'a lot of difficulty'. Another USA respondent described his confusion well:

> 'You asked about a hearing aid and I'm thinking I don't have no hearing aid but I can still hear pretty good. So that's what threw me off, when it said *with a hearing aid*, I'm like I don't even have one of them, so why is that...how's that going to help me'.

A second USA respondent reported 'no difficulty' and clarified this by saying 'Because I don't wear a hearing aid. Yes I do have a problem hearing but I don't wear a hearing aid.'

---

[17] Tinnitus is a constant humming, rushing/roaring or high pitched sound in the ears often but not always associated with a hearing loss.

A number of respondents in the middle and low income countries did not know what a hearing aid was. Unlike eye glasses, hearing aids are not common and people may have ignored the hearing aid clause because of this unfamiliarity.

A number of problems experienced in relation to hearing were reported as being listening in noise and having a hearing loss in one ear only. Responses were reported as either 'no' or 'some' difficulty for the similar description by different respondents. These are good examples of borderline cases. The respondents who reported 'no difficulty' would give examples such as too much noise at a concert, ceremony or party, or other similar contexts as the only time when they have some difficulty hearing. A few respondents reported a unilateral hearing loss but having no difficulty, with one South African respondent considering the setting in which she finds herself to decide whether she has a hearing loss or not. She replied 'no difficulty' because the setting of the interview was a quiet one. A respondent from the Maldives reported having 'some difficulty' because of a hearing loss in one ear and problems hearing at a distance of about 10 feet.

A further ambiguity arose from people conflating concentrating with hearing. One Canadian respondent described how her family have learnt to get her attention before talking to her. When asked the questions (SS2) again, she responded 'no, what I have is not a hearing problem.' While this was not a common response, it does reflect the close relationship between hearing and concentrating.

Only 5 respondents reported using a hearing aid and, of these, three reported using it 'all the time'. The other two used them rarely or never. In general, some of the reasons for not using a hearing aid other than not needing one, included:

- Not knowing about a hearing aid with the added response that if they were given one they would like to use it. One such respondent indicated having 'a lot of difficulty' hearing while another reported 'no difficulty'. This was the most common reason for not using a hearing aid after the reason of not needing one

- Being told that use of a hearing aid is not indicated for respondents who cannot hear at all or who have a unilateral hearing loss.

- Getting no benefit from using a hearing aid from respondents reporting 'a lot of difficulty' hearing or 'unable to hear at all'

- Running out of batteries (when hearing aid was provided for free) and so giving up using it by a respondent reporting 'some difficulty' hearing

- Unable to afford it

Impact of hearing difficulties on activities

When asked whether the hearing difficulties had an impact on their daily and non-daily activities, the examples given included working outside of the normal workplace, attending religious ceremonies, social situations, visiting cultural establishments, receiving visitors at home, shopping, talking to a bank teller, hearing approaching traffic (for more severe difficulties) and negotiating

airports. Non-daily activities were seen as being infrequent in occurrence and hence some respondents reported no impact on these.

Conclusions from cognitive testing of hearing questions

In view of the trends in the hearing questions from the cognitive testing interviews, the questions were revised only minimally for the pilot testing. The first question was kept unchanged as it forms part of the Washington Group Short Set of questions. The questions on use and frequency of use of a hearing aid were left unchanged. The two questions on hearing in quiet and in noise were reversed starting with 'hearing in a quiet room'. If respondents reported 'cannot hear at all' in a quiet room, they were not asked about difficulties hearing in a noisier room.

4.9.2 A domain where questions require further testing: Pain

The following question set on pain (Box 6) was included in the interview protocol for the 2009 round of cognitive testing in the ESCAP region. The set asks about having frequent pain, use of medication, duration, intensity, consistency, age at onset and whether the pain has any impact on daily and other activities. Note that unlike in other domains, there is no single "short set" question for pain as the multiple rounds of testing for pain have demonstrated that a single question is not feasible.

Box 6. Pain questions administered in the cognitive test interviews.

Questions	Response Options
9.1 Do you have frequent pain?	1) Yes 2) No
9.2 Do you use medication for pain?	1) Yes 2) No *If "No" to both 9.1 and 9.2, skip to next section.* *If "Yes" to 9.1 continue with 9.3.*
9.3 In the past 3 months, how often did you have pain? Some days, most days or every day?	1) Some days 2) Most days 3) Every day
9.4 Thinking about the last time you had pain, how long did the pain last? Some of the day, most of the day or all of the day?	1) Some of the day 2) Most of the day 3) All of the day
9.5a Thinking about the last time you had pain, how much pain did you have, a little, a lot, or somewhere in between a little and a lot?	1) A little 2) A lot 3) Somewhere in between a little and a lot *If "Somewhere in between" to 9.5a, continue with 9.5b. Otherwise, skip to 9.6.*
9.5b Would you say the amount of pain was closer to a little, closer to a lot, or exactly in the middle?	1) Closer to a little 2) Closer to a lot 3) Exactly in the middle
9.6 Thinking about the last time you had pain, was the pain worse than usual, better than usual, or about the same as usual?	1) Worse than usual 2) About the same as usual 3) Better than usual
9.7 How would you describe your pain?	

9.8 How old were you when the pain began?	_____ age in years
9.9 Is your pain due to a health problem of something else?	1) Due to a health problem 2) Something else: _____
9.10 Does your pain limit your ability to carry out daily activities?	1) Yes 2) No
9.11 Does your pain limit your ability to carry out other activities that are not part or your day-to-day life?	1) Yes 2) No

Reporting of Pain

From the outset of testing pain questions, it has been clear that whether respondents report pain and what they are reporting as pain varies. The data from this round of cognitive testing also illustrates the variation among respondents in the reporting of pain. These variations appear to be related to a number of factors, including how the respondent interprets 'frequent'. More information on this is provided below. Respondents' reports of pain vary by whether or not the cause of the pain is believed to warrant report. For example, firstly, self-inflicted pain or pain that results from overwork is discounted by some, but not all, respondents; secondly, depending on whether their pain is a frequent experience or 'usual' or typical experience for them, as well as whether they believe the question is asking about 'usual pain'; lastly, the results show differences in reports associated with respondents' beliefs that their pain is 'intense' enough to report. For example, the experience of discomfort is reported by some respondents as pain, but not by others.

Interpretation of 'Frequent' and Frequency of Pain

Previous versions of the initial pain question have demonstrated that asking, 'Do you have pain?' captures a wide range of experiences, including discomfort and fatigue for example, which are out of scope for our purposes. The current version of the question inserted the word 'frequent' as an attempt to capture pain experienced above a relatively low or common threshold. The word 'chronic' was considered and seen to be a medical term not universally understood. Hence 'frequent' was chosen. Ultimately, the goal was to try to avoid capturing the occasional, routine experience of pain that lasts only for a short period of time and is easily resolved by medication.

The cognitive test provided some evidence that respondents vary in their interpretations of 'frequent' in the initial pain question. Some respondents asked immediately what was meant by this word. Others asked if 'frequent' was different from 'chronic' or 'constant pain'. When respondents were asked by interviewers how they interpreted 'frequent', various interpretations were reported including: constant, every day, every week, and every time it rains.

Type of Pain

Many different types of pain were reported in the cognitive test. The majority of reports of pain were based on physical pain. Among the responses recorded were long-term injury; injuries without specification of duration; disease-related pain such as liver disease, osteochondrosis and scoliosis; muscular pain and soreness. Only two respondents mentioned emotional pain when probed about their pain. Furthermore, the sites of the pain experienced covered nearly every part of the body, including the head, neck, eyes, teeth, back, arms, knees, feet, etc.

Pain Medication

The question on pain medication was included in the set to provide some information on the degree of pain experienced, as well as to assess accommodation (and functioning with or without the accommodation). This is based on the assumption that, in most cases, the greater the pain experienced the greater the likelihood an individual will use pain medication. It was also included as a way to interpret (during data analysis) the information provided in the frequency, duration and intensity questions. It was not included as a way to filter out those respondents who report experiencing pain, but for whom medication alleviates the burden of that experience. Everyone who reports pain in the initial question receives the follow up questions, regardless of their answer to the pain medication question.

Responses to 'Do you use medication for pain?' depended greatly on the interpretations of, and emphasis placed on, the term 'use' and 'for pain'. For some respondents, some medicines did not qualify as pain-relievers, for example those items typically associated with complimentary or alternative medicines. For others, medicine included water therapy, supplements, patches, exercise, and calcium, to name a few. For some respondents, it is the form of the medicine that dictated its report. Ointments and non-prescription drugs did not qualify for some respondents; for others these items along with tablets, prescription drugs, and other more traditional type drugs were counted as medicines. Many respondents were unsure of what medicines should be included and asked the interviewer for clarification.

Duration, Intensity and Consistency of Pain

The cognitive test did not capture as much information as desired about these important dimensions of pain. In general, respondents answered the test questions, but the majority were either unable to answer, or due to time constraints were not asked, the probe questions which provide valuable interpretative information.

Some information was provided that was used to inform the pilot test, however. First, there is some evidence that respondents have difficulty estimating how long their pain lasts. Part of the evidence highlights the difficulties in accurately pinpointing the onset of the pain and the end of the pain experience. Some respondents chose varying metrics by which to estimate the length of time of their pain. For example, if the pain lasted a couple of hours or for an afternoon, they asked if that means 'some of the day' or 'most of the day'? Others were unsure of what metric to use at all. Clearly the response categories were problematic and did not correspond well to how most respondents measured the duration of their pain. Further, the phrase 'the last time' was not always included as part of the response process. When probed, some respondents had not limited their answer to just their last experience of pain.

In contrast, the response categories 'a little', 'a lot' and 'somewhere in between' seemed easy for most respondents, and respondents appear to have little difficulty with the follow up question for those that answered 'somewhere in between a little and a lot', although most responses still fall at the extremes or exactly in the middle.

Question 9.6 asks respondents to rate their last episode of pain as 'worse', 'better' or 'about the same as usual'. While little information was obtained during the probes, it is evident from the data

collected that this question was especially difficult for those who do not experience pain in discrete periods. For these individuals common verbatim responses to probes included, the pain is 'always similar', 'always there', 'constant' and 'consistent'.

Cognitive Test Conclusions for Pain domain

Several important findings emerged from the cognitive test of the pain question set.

1. Whether pain is reported or not by respondents varies. The variation occurs by respondents' interpretation of 'frequent' as a qualifier of the pain, by cause, by frequency, and by intensity of the pain experience. Whether these variations occur as a result of socio-cultural differences, or are influenced by age, sex, education and other demographic factors is unknown. Clearly, the finding strongly supports the idea that pain must be measured along multiple dimensions in order to adequately and accurately capture the full experience of pain.

2. There is some evidence that the meaning of the word 'frequent' is not consistently interpreted by respondents. Thus, the initial pain question alone may not serve as a reliable screening question for the remainder of the set.

3. One consistent finding concerns the type of pain. When pain is reported, it is predominantly physical pain associated with a specific part of the body and the result of an injury or acute or chronic condition.

4. The use and types of medicines reported vary in ways that do not provide clear evidence of how the data should be interpreted, although medication remains an important accommodation for pain and should not necessarily be excluded based on differences in type or frequency of use. Moreover, without asking about pain with and then again without medication, it is unclear whether we are ascertaining pain with or without accommodation.

5. Finally, information about the frequency, duration and intensity of pain is important but is also highly subjective and heavily influenced by whether the pain experience is episodic or continuous. Thus, many of the findings suggest quite a bit of interpretative variability.

4.9.3   A domain where questions are rejected: Learning

The cognitive and pilot testing of the learning questions led to these questions being rejected and a recommendation being made to either not include this domain in the WG Extended Set or to redraft the questions completely.

1. **Do you have difficulty understanding and using information like following directions to get to a new place?**

Response options:

1. No difficulty [If no difficulty, the respondent was directed to the next section/domain in the cognitive interview]
2. Some difficulty
3. A lot of difficulty
4. Cannot do at all/ unable to do

2. Do you have difficulty learning new things such as the rules for a new game?

Response options:

1. No difficulty
2. Some difficulty
3. A lot of difficulty
4. Cannot do at all/ unable to do

3. How much difficulty did you have in analyzing and finding solutions to problems in day to day life?

Response options:

1. None
2. Mild
3. Moderate
4. Severe
5. Extreme/Cannot do

Learning findings:

An assessment of the findings that resulted from the cognitive testing of the learning questions includes an interpretation of the process of respondent comprehension, retrieval, judgment and response. We are interested in understanding the respondents' judgment processes and response patterns to the questions specifically through their interpretation of the question (comprehension), and their determination of what they deems relevant information (retrieval).

1. The first question involves some interpretation by the respondent.

Understanding and using information like following directions to get to a new place is a complex question that includes several actions: understanding and using information, following directions, and getting to a new place.

A review of responses to the question revealed that responses focused on the third (last) action: getting to or finding a new place. This, in turn, involves varied interpretations, such as following directions, using a map, reading street signs, and needing assistance to walk or use various modes of transportation. Of the 124 respondents , the majority (73 percent) replied *no difficulty* and some went on to explain that they used aids (maps, GPS, or MapQuest) to assist them. Some respondents, however, never go to new places and responded to the question either *no difficulty* or *can't do at all*. Twenty three percent of respondents reported at least some difficulty and most often referred to unfamiliar places, the fear of getting lost and difficulty concentrating on instructions. In a few instances respondents indicated other difficulties like vision (blindness or difficulty reading street

signs) or mobility (needing assistance to get around) that affected their ability to learn and follow directions.

Among the valid interpretations that respondents offered to the question on *using information* were: thinking & logic skills at work, school work, directions for household chores, and following directions in general.

2. *Learning new things such as the rules for a new game* was primarily seen as asking about learning a new game, which involves the level of difficulty of the game, and the ability to understand directions written or spoken (in the respondents own language or a foreign language). Among those who responded to the question (n=52), 50 percent claimed *no difficulty* and 42 percent had at least *some difficulty*. A few respondents (5) claimed that they never play games, and their answers ranged from *no difficulty*, *don't know*, *can't do at all* or they provided no answer at all. A respondent who claimed not to have time to play games offered the example of cooking meals as an alternative. In a few instances (4), respondents mentioned other difficulties like vision ("because of sight") that impeded their ability to learn a new game; and two respondents mentioned their age ("she is old and no longer able…") as the reason for their difficulty (one responded *cannot do at all* and the other chose not to respond).

Interpretations of this question included putting furniture together, learning how to feed livestock, cooking (following a recipe), dancing, school work, and life in general.

3. Responses to the question on *analyzing and finding solutions to problems in day to day life* (n=41) elicited interpretations that included examples of daily problems like family issues, work issues, money problems, interpersonal relations, neighbourhood crime and even being able to do puzzles. The myriad responses to this question reflect the imprecision of the 'problems in day to day life' clause; and they do not necessarily capture the aspects of learning that we would want or expect through 'analyzing and finding solutions'.

Of the 41 responses to this question, 46 percent reported *no difficulty* and 46 percent reported at least *mild difficulty*. Some respondents (3) had difficulty understanding the question; in particular some had difficulty with the word *analyze.* This raises the issue of potential problems related to the effects of socio-economic status on the ability to interpret/ understand the question.

In summary, cognitive findings on the learning questions indicated that these were not getting at general learning but were being interpreted as asking about the specific example – issues of playing games and or being able to follow instructions.

Revisions for Pilot test Questionnaire

Based on these findings it was decided to:
- Simplify the child focus question deleting *new things like* and focusing only on *the rules for a new game*;
- Add *a new cell phone* to adult question in an attempt to get the respondents to focus on learning rather than the specific example. In addition, a probe question was added as a follow up to this question to learn more about how respondents were responding; and
- Drop the third question on *analyzing and finding solutions to problems in day to day life.*

5   The pilot testing

With the completed cognitive testing results, a pilot test is prepared. The stages in this process are:

- Developing the questionnaire
- Training of interviewers
- Sample size and selection
- Analysis
- Revision of the questions

This section sketches out the process of pilot testing. It should be read in conjunction with the final report on the cognitive and pilot testing in the ESCAP project to get a fuller understanding of how to proceed and to get examples of the analysis undertaken.

5.1   Aims and objectives of the pilot test

The pilot test follows on from the completion of the cognitive testing and any further iterations of the cognitive testing. The results of the cognitive testing analysis are used to inform the design of the pilot testing questionnaire.

The aim of this phase is to continue the evaluation of question performance. It is not the final survey. The cognitive testing results would have thrown up several issues with the question design and format. The example of pain in table 3 in the above section shows that there are three possible interpretations of the question 'Do you have frequent pain?' The reader should accept as given that further analysis of the other pain domain questions showed that there were different interpretations of what frequent pain meant and how the answer to this first question was supported or not by answers to the other pain questions.

This information is used to set up some hypotheses to be tested in the pilot testing. These would include examples such as the following:

- Reading out a number of possible definitions of frequent pain from which the respondent has to select one that best describes his or her understanding. The pilot test analysis would determine the prevalence of each of these definitions. A high prevalence of problematic definitions would signal the need to revise the question, while a high prevalence of correct definitions would indicate that the term 'frequent pain' should remain.
- A hypothesis that states that not having frequent pain is associated with infrequent pain in the last three months, and low intensity and duration of pain the last time the person experienced pain. Statistical analysis of the pattern of responses for the set of pain questions will provide results that accept or reject this hypothesis.

The objectives of the pilot test is to complement the cognitive test results in a way that allows for a final decision to be made on the question evaluation, and thus remains an evaluation exercise and not a measure of prevalence of various difficulties in the given population. However, the pilot test will look at different demographic factors in the analysis to determine whether the biases noted in the cognitive test remain in a larger and more representative sample, such as the trend of women reporting pain more often than men.

Appendix A presents the questionnaire used in the pilot test of the ESCAP project. This form requires much less space than the cognitive testing form as note taking is not a central feature of this part of the question evaluation process.

5.2   Developing the questionnaire

The questionnaire is set out as it would be in the full survey with the questions in the same order and with the relevant instructions for skip patterns and instructions to be read out to the respondent. The questionnaire uses the revised questions arising out of the cognitive test. Basic demographic questions are included to allow for a detailed statistical analysis.

The translation of the questionnaire is a crucial step in the overall process. This is addressed in section 7 below.

Some of the features of the ESCAP pilot test questionnaire (Appendix A) should be noted to allow the lessons learnt to be effective:

- The length of the questionnaire remains an issue but was much less of an issue than for the cognitive testing questionnaire;
- The repetitive use of the questions at the end of each domain section on the impact of any difficulties remains a problematic issue and should be revised. The best way to do this has not been determined as yet.
- The probes used in some of the domain sections were developed based on the cognitive testing analysis. These probes, however, remain 'untested'.
- The list of different life domains affected by difficulties was problematic as respondents struggled to understand the difference between these, especially between the first two:
  a) Working to support you or your family?
  b) Working outside the home to earn an income?
- The list of impairments and health conditions at the end of the questionnaire were used to collect data for comparing different impairments or health conditions in relation to different profiles of responses on the WG Extended Set questions. This is not a necessary component of a pilot testing for disability measures.

5.3   Sample size and spread

The sample size for the pilot test is much larger than the one for cognitive testing as the interview is a standard application of a questionnaire without follow up questions on the respondents' understanding and interpretation of the questions.

The sample size for the ESCAP project was 1000 respondents per country giving a total sample of 6000 completed questionnaires. There is no prescriptive sample size but it should be large enough to allow for sufficiently detailed statistical procedures.

a)   Selection of sampling areas

The samples should cover a few areas that represent the typical regions of the country. For example, in Sri Lanka there are three areas denoted in the census enumeration areas: urban areas, rural areas, and tea plantations. In South Africa the types of enumeration areas include urban informal, urban formal, rural traditional areas, and rural farms. Other countries may only have urban and rural areas.

b) Selection of households

One of each of these areas is selected and a random sampling strategy applied for selecting households to ensure the required sample is realised. This is no different to applying sampling techniques for a full survey except that the size of the sample is greatly reduced. The final sample size for the pilot test will depend largely on the funding available and the costs of travel and accommodation of fieldworkers.

The households were randomly selected in different ways across the countries. Some listed all the households and selected the required number of households, while others selected every n-th household without listing using a systematic sampling technique.

c) Selection of individuals within the household

In each selected household, two adults and, in every alternate household, one adult and one child were randomly selected for an interview. The selection of these individuals was done by the birthday method. The household respondent is asked for the name of the adults or one adult and one child whose birthday is next from the date of the interview. If that person is not available for the duration of the data collection, the person with the next birthday is selected.

This approach to random selection is easy to apply but is not as rigorous as other methods such as listing and Kish methods of sampling. However, it was felt to be sufficiently rigorous for the pilot testing purposes as the aim was not to obtain prevalence estimates.

d) Examples of sampling plans were extracted from participating country reports[18]:

    1. Cambodia:

"In each selected village, sample households were chosen by supervisors using systematic sampling. Enumerators did the selection of respondents from each household. Based on the selected sample households, the final selection of respondents from every first household were two adults and for the following household one adult and one child, alternatively following the list of samples."

    2. Kazakhstan:

"Since Kazakhstan is a very large country, it was decided that in order to account for all regional differences the survey should cover representatives of different regions, northern, southern, western and eastern parts of the Republic. In these regions, the current survey network, i.e. the census area principle, was used to select households. The advantage of such household selection was that the current network has a database covering all members of households: number of members, their age, sex, educational attainment, employment status, etc. Respondents were selected as follows: 2 persons per household (2 adults or 1 adult and 1 child aged 5 or above), with the random sampling principle ensured through "next birthday" rule. Urban/rural and male/female ratios were secured."

---

[18] See www.unescap.org/stat/disability/task-team

3.   Maldives:

"The pilot test has been carried out targeting coverage of a 1000 people – 500 people from urban areas and 500 from rural areas. The 2006 Population and Housing Census recorded Malé's total urban population as 103,693 persons. The island Mulah of Mulakatholhu (rural) had a total population of 1,129. The sample selection from the island Mulah was based on 2006 Census data which showed a total of 213 households. The island's average household size was five persons per household. In Male', the total number of households was 14,107 with an average household size of seven persons. Within each selected household the questionnaire was administered to two adults (aged 17 or older – personal interview wherever possible) and in every second household the questionnaire was administered to one child through a proxy (ages 5-16 years). Where more than one adult and/or child lived in a selected household, interviewees were selected based on whose birthday would be reached first.

Problems and Selection errors included:
    1. Unexpected number of vacant households
    2. Single person household
    3. Household refusing
    4. Individual person refusing"

4.   Mongolia:

"The pilot test was conducted using a two-stage sampling method, which gives an equal probability of the selection of households. The sample frame comprised the listings of households prepared annually in khoroo (a low level administrative unit of districts within the Capital city) across the Capital city and bags at soum level. The actual sampling frame was based on the 2008 end-of-year population data. According to the organizers' recommendation, 60 percent of total respondents selected lived in the urban area, (this time selected the capital city) and 40 percent lived in rural areas. For the sampling frame, Songinokhairkhan district, which is a highly populated district, was selected in Ulaanbaatar city due to its representativeness. From the rural areas the Selengeaimag was selected."

5.   Philippines:

"The sample provinces, municipalities, barangays and enumeration areas (EAs) for the Field Test on Functional Difficulty were selected using the following criteria:
    1. accessible to Metro Manila and from the provincial office
    2. generally with Filipino or Tagalog speaking households
    3. with updated listing of households
    4. with high percentage of persons with disability based on the 2000 Census.

The updated listing of household based on the 2007 Census of Population was used in the selection of sample households. A total 520 respondents in about 260 households was interviewed in each sampled province. To fulfill the requisite to select an urban and rural area for the field test, sample areas were selected in two provinces: Cavite for rural area and Rizal for urban area.

The general rule was to complete the Household Roster of the sample household and interview the selected respondents at the time of visit. Within each selected household, the Field Test Questionnaire was applied through personal interview to two adults (17 years old and older) and in every second household, to one adult and one child (5 to 16 years old), through proxy. The adult and/or child in the household were selected by taking the one with the next birthday. It must be noted that the respondents need not be literate. "

6. <u>Sri Lanka</u>:

"Geographically, Sri Lanka could be divided into 3 major sectors, namely Urban, Rural and Estate sectors. Due to the various characteristics and different dialects used in these locations, an attempt was made to capture to what extent the questionnaire could be used for future health surveys. Therefore a number of clusters were selected to represent the proportion of population in those sectors at initial stage and Urban and Estate sectors were over-weighted by one census block each to cover distinguishing variations on characteristics of the people living in those areas compared to rural population."

5.4 Selection and training of interviewers

The sample size and range of language groups to be included in the pilot test will determine the number of interviewers. The basic principle of not allowing interviewers to do too few or too many interviews will apply as it does in any survey implementation. The bases for determining the number of interviewers to be hired were determined by the countries based on the total workload, the expected number of completed interviews per person day, expected percentage of callback and wrap-up activities.

Since the sample size was not large, small teams were set up including 5–6 interviewers per team with a supervisor. It is important that interviewers have the possibility to share experiences at the end of each day and summarize these as a useful part of the information collected throughout the pilot test.

The training should be scheduled for 4 days including a number of practice interviews with each other and an outing to do 'real' interviews in an area close to the training venue. These 'real' interviews must be observed and followed up with a discussion in plenary. Experience from the ESCAP project is that these 'real' interviews were invaluable as they generated much discussion and raised a range of questions on the survey questionnaire and its implementation. Some of the points raised from these 'real' interviews are provided below from the Sri Lanka training as this was the first country to undertake their training and field test.

General and question specific comments from practice interviews:

- Timing - took between 40 and 70 minutes.
- People were willing to be interviewed
- Most people had one or more difficulties
- Skips were difficult
- Need more examples at hand (*possibly put some in questionnaire with italics*)
- Background questions:
  - o 'single' never used – only never married
  - o 'Unemployed and not looking for work' is not a category used in Sri Lanka
- The activities questions at end of each domain are not all relevant for Sri Lanka. A and F are seen as being too similar. Discussed that A was about economic activities. Suggestion is that A and B be merged into one category.
- The main activities that are relevant for Sri Lanka are
  - o Economic activity to take care of self and household
  - o Social activities
  - o Household work
  - o Transportation

  Final decision on the activities/impact Qs – keep as is except for possible merger of A and B.

  There will be a number of not applicable, but that's OK.
- behavioural coding questions (in grey shaded boxes) are asked of the proxy respondent's understanding and asking for clarification. The point is that it is about how the proxy has understood the question.
- Age of onset: An example was given of a respondent who had difficulty seeing since 30 years of age, got glasses and it was rectified (no difficulty) and then came back again at 35 yrs of age because glasses were no longer effective. What is the age of onset? *The first age when it started is the age of onset*
- Issue of the 'even when wearing glasses' clause was raised again as being problematic. The question was asked without it and prefaced by asking if they wear glasses. *It was stressed that they should ask it as is and then follow up with:*
  - o Do you wear glasses?
  - o If yes, re-ask the question 'even when wearing them
  - o If no, ask do you have difficulty seeing.
- For mobility, a pregnant respondent responded as having difficulty walking more than 100m. This should be captured as a valid response even though condition is temporary, but note pregnancy in the COND_2 question under other illnesses.
- For communication, one respondent has a hearing loss and attends a special school where he learns using sign language. At school has no difficulty in usual language but has a lot of difficulty with spoken language at home. Interview was a proxy by mother and since she was describing the home context, it was decided to use that as the reference – i.e. code as 'a lot of difficulty'. The impact questions were the also asked in relation to the home context rather than school. Same child for cognition, point was made that he clearly does not have a cognition problem as such but rather a way to express it. Cognition is about thinking and not about communication. Thus the response was marked as 'no difficulty' since he can do it in sign language.
- For learning, examples to use include learning a new recipe, helping children with homework, switching on VCR and radio (although when given this last example, a respondent said she just gets her *daughter to do* it.

The training also provides a final opportunity to check on translations and on any errors in the question wording, skip patterns, and other formatting errors on the questionnaire. Thus the full set of questionnaires should not be printed until the end of the training process to ensure that these errors are changed for the final field version.

The materials for the training include:

- Practice copies of the questionnaire
- One copy of the question by question specification guide detailing the intention of each question and how it should be administered. This document is for reference in field when faced by a complex situation[19]
- Pens and pencils
- Some materials on definitions of complex concepts, such as the difference between anxiety and depression on the ESCAP project[20]
- Note books to make notes on issues arising in field, such as questions that are always difficult for people to understand, complaints about sensitive questions, comments on the topics being covered in the pilot test questionnaire, etc.

The interviewers should be experienced in survey questionnaire administration and be fluent in the languages in which they will be administering the questionnaire. Typically field workers have completed full basic education and may or may not have post schooling qualifications. The outcome of training should be good familiarity with and accurate administration of the questionnaire. In addition, the fieldworkers should have skills on how to conduct a survey interview in a manner that retains the interest of the respondent while still asking the questions in the required format.

The interviewers should be told what criteria to apply in deciding when to do a proxy interview rather than a direct one. The main criteria include:

- **Age:** most children under 14 years of age will not respond for themselves. Consent from the parents or guardians of the child 14 – 17 years of age and assent from the child must be obtained before doing a direct interview with this age group of children.
- **Adults who are not able to respond for themselves**: These adults include those a) too sick to be interviewed (or they should be interviewed over a couple of sessions); b) not cognitively able to understand the interview process or give legal consent; c) not able to communicate effectively in the spoken language used for the interview (e.g. a Deaf sign language user, a foreigner). These people, should, however, be interviewed using an interpreter rather than as a proxy interview; or d) not able to communicate effectively because of a stroke or head injury.

5.5 Data capturing

The data from the questionnaires should be entered in the system currently used for other surveys by the implementing agency and thoroughly revised for data-entry errors.

---

[19]See http://www.unescap.org/stat/disability/task-team/Guide-field-testing-E.doc
[20]See http://www.unescap.org/stat/disability/task-team/WHO-anxiety-depression-symptoms.doc

5.6   Analysis

The pilot test data analysis comprises statistical procedures ranging from simple frequencies and cross tabulations through correlations and regression analyses to identify significant patterns of interpretations and to test hypotheses.

The questions asked of the analysis will include:

- The extent of the different interpretations (in and out of scope) to determine whether the out of scope ones are cause for concern or so infrequent that they are not likely to a significant impact on the final data set
- Crosstabulation and regression analysis of responses to determine relationship between the patterns of responses on extended set of questions, impact of the difficulty on daily activities, and age and sex factors. The example of pain given above would be such an example
- Significant statistical differences between subgroups to confirm or reject differences or biases noted in the cognitive testing

The culmination of the analysis is a decision making process on the performance of the questions. The performance will be rated according to the following:

a) Works well and the performance is well understood, consistent and transparent as to potential sources of response errors
b) Problems are clear but questions needs revision and possibly further testing
c) Can't decide and needs further testing in order to make and informed evidence based decision
d) Does not perform well at all and should be omitted or fully revised

6   Information dissemination

An important part of the question testing process is to inform relevant role players within government and civil society about the testing exercise in order to elicit their cooperation and input as far as possible. The more involvement these role players have in the overall process the more likely they are to understand and use the data collected. They should be involved at all stages of the process.

In addition, information may need to be disseminated to the areas selected for the pilot testing to prepare those areas for the presence of interviewers and, therefore, reduce non-response and refusals. This component of information dissemination was small compared to what is usually required in a full survey. But it does remain necessary and important. Typically, statistical offices have well developed strategies for undertaking this information dissemination and the same strategies should be used for the pilot testing.

6.1   Developing the questions for testing

The use of data is dependent on the need for such data from government ministries and civil society organizations. The data requirements from government ministries are typically for developing, implementing, and monitoring policy and service delivery, while for civil society the needs are often for advocacy purposes (e.g. in the field of disability). These needs for data should inform the

questions developed for surveys to collect relevant data. Thus, it is important to involve these role players from the onset of the project to ensure that relevant questions are developed.

6.2 Reporting on cognitive test results and preparing for pilot testing

It is useful to report back the results of the cognitive testing and how it informs revisions of the questions for pilot testing. The input from the role players can assist in assessing the findings of the cognitive test and how these inform the revisions for the pilot testing.

The involvement of the role players in the training of interviewers and observations in the field during the pilot test can give the process transparency and credibility. In addition, specialist knowledge from the different ministries and sectors can provide useful assistance to the interviewers and supervisors while in field.

6.3 Analysis and reporting of pilot test results

The final stage for involving the role players is in the interpretation and finalization of the pilot test and overall results that will lead to the final questions to be used on the full survey. This continues the process of ensuring correct interpretations from people working in the topic area of the questions as well as ensuring their understanding and hence effective use of the final survey results.

6.4 ESCAP project experiences

All six countries used various forums to inform role players of the nature of the testing exercise and feedback the results of the cognitive testing. Typically the ministries involved were those of Health, Labour, Social Welfare and Development, and Gender and Disability issues, and the civil society organizations were the disability sector in each country and non-governmental organizations working in this sector.

7 Translation of cognitive and pilot questionnaires

The success of comparative measures of disability or of any other phenomenon is predicated on the questions being the same in all languages in which they are asked. Translation of the questionnaire is, therefore, a crucial component of the question evaluation process but a difficult one. Statistical offices will already be applying translation approaches in their ongoing development of questionnaires as few countries have a single language for the whole population.

The current approaches to translation highlight the following:

- The importance of a conceptual and not a literal translation.
- The use of a phrase if a word is not available to ensure the concept meaning is retained correctly.
- Avoidance of emotive terms and especially negative terms, even more so in the field of disability.
- The translation must be checked prior to the cognitive testing and revised after it based on the findings. A further revision can be done after the pilot test has been completed and analysed if indications are that this is required.
- The language used should be colloquial and clear and not use possibly correct but unfamiliar terms for low literacy populations.

Two main approaches to translation are currently referred to in the literature: a) forward and backward translations; and b) translation by a committee.

7.1 Forward and backward translations

This approach requires that an expert in the topic area and in the required language translates the whole questionnaire. The translated questionnaire or some key phrases (without the original language, such as English in the ESCAP project) is given to second person who translates it from the specific country language back to English. Any major discrepancies are then discussed by the project team and a decision made as to the final translation to be used.

7.2 Translation by committee

This approach seems to be used more and more as it allows for additional discussions from the start. A group of two to four people is formed combining expertise in the topic area and the languages into which the questions will be translated. This group meets to discuss the questions and their intent, and then agrees on the best translations to be used. The outcome of this process is then checked by a couple of other experts in the field. This approach is particularly useful when more than one language will be used as the teams can meet altogether to discuss the intent of the questions to ensure uniformity of conceptual translations across the different languages.

7.3 The translation process in the ESCAP Project

In the ESCAP project the countries were encouraged to use the translation by a committee approach. Each country established its own group of people to undertake the translation. Some used a group of people from the Ministry of Health and the Statistical Office. Others used a group of interviewers. The exact structure of the committee is not important as long as it does reflect knowledge of the original and target languages, and more importantly, of the content and purpose of the questions. It should not be too large either, as this may make reaching a consensus translation difficult.

The training of the interviewers for the pilot test was found to be a very useful platform for checking translations. In all six countries problems in the translation were noted and addressed during this training. This is an important step in the translation process which allows not only for a final revision of the translation, but also ensures that the trainee interviewers develop a good understanding of the questionnaire. Final revisions can also be made after completing the practice interviews at the end of the training period.

This makes it important not to print the full set of pilot test questionnaires prior to the completion of the training process.

The extract from the Cambodia country report[21] is provided as an example of a translation process. They used a combination of the two methods described above. Most of the six countries used some combination of these two approaches. The main aim is not as much the approach used as the final product of a translation that is semantically congruent with the original English text and is easily understood by typical respondents.

---

[21] See http://www.unescap.org/stat/disability/task-team/Cambodia-report.pdf for the full report and other country reports.

Translation of questionnaire (Cambodia example)

The translation of the questionnaire was led by NIS [statistical office]. The completed translation of the questionnaire into the local (Khmer) language was sent to concerned ministries/agencies and specialized organizations working with disabled people, for comments. Moreover, before starting the implementation of cognitive test, the NIS also hired a local consultant to translate the questionnaire back into English in order to verify translation accuracy.

Who was involved?
The following organizations contributed to the translation process: Ministry of Social Affairs, Veterans and Youth Rehabilitation, Disabled People Action Council, Handicap International of Belgium and France and the Trans-cultural Psychological Organization. In addition, the completed translated questionnaire (both English and Khmer versions) were comprehensively circulated through meetings on disability classification working group led by Disabled People Action Council, for review and comment before finalization.

What were the main issues/observations?
The main issue in translating the questionnaire was the use of some terminology in English, which does not have a direct translation in Khmer. For example, "anxiety" and "depression" – both words needed to be explained further in Khmer to avoid confusion and misinterpretation from respondents. Another issue was the length of some questions when translated into Khmer – asking lengthy questions to respondents was sometimes complicated.

Additions or changes to question set
The format of the questionnaire translated into local language was kept in its original form and structure as it was in English. For example, number of questions, response codes, response categories and instructions in *italic* and **bold**. For ease of use by enumerators and to avoid mistakes during the interview, instructions were provided for every question. An asterisk (*) was added so that Enumerators could determine skip patterns and be directed to the next sections/domains.

Some questions were revised to better fit in the Cambodian context. For example, instead of *"walking 100 meters or about the length of one football field or city block",* it was changed to "*the length of one hectare rice field*"; or from *"raising of 2 liter bottle of water or soda from waist to eye level"* to *"raising of 2 liter bottle of water or orange from waist to eye level"*; or *"walking up or down 12 steps"* instead of "*walking up or down a small hill*". Questions on income were changed to monthly income quintiles, ranging from lowest to highest quintile in accordance to Cambodian Household Survey and as for questions on highest level of education, additional response categories were added accordingly.

8 Concluding remarks and lessons learnt

This section brings together some concluding remarks on the approaches of doing cognitive and pilot testing and, specifically, of combining the two approaches.

8.1 Lessons learnt

While each section above on the cognitive and pilot testing gives some points on lessons learnt on the individual questionnaires, this section looks at the overall process and what issues are important to consider when repeating this type of exercise.

a) The number of questions to be tested should be kept to a minimum. The number of questions included in the ESCAP project were too many. If a country undertakes a cognitive and pilot testing project, the focus should be on those questions that: a) remain problematic, (e.g. affect, fatigue, pain), together with one domain that is known to be working well across a number of countries (e.g. hearing), and b) new questions not tested before (e.g. questions on environmental barriers).

b) The questionnaire translation is a process where an initial translation is done in discussion with three or four people knowledgeable on the topic and language. This is then followed up by further checking, and revisions during the training of interviewers, and a final revision after a few practice interviews done at the end of training.

c) While the time lapse between the cognitive and pilot testing should not be too long, it should allow for a thorough analysis of the cognitive interviews to inform the revisions required for the pilot test. The time lapse will depend very much on the individual organisation and the amount of time allocated to the project. The recording of the cognitive interview notes and analysis should be done as soon as possible after the completion of these, to ensure sufficient recall.

d) Practice in doing cognitive interviews is essential and should be done in pairs with an overall mentor such as one of the resource people listed at the end of these guidelines. This is especially important for statistical office personnel who are more used to administering survey interviews than doing in depth interviews.

e) The combination of the cognitive and pilot testing processes provides a very effective testing process for new survey questions.

8.2 Beneficial outcomes of combining cognitive and pilot testing

The discussion above of cognitive and pilot testing and the example on communication shows that the progression from cognitive testing to pilot testing provides the following:

- Intensive analysis of potential problems in question performance (small sample of cognitive interviews)
- Extent of these problems in a larger sample (pilot test)

Cognitive testing helps us understand the ways in which a question performs across different respondents to highlight any question design problems. The pilot testing helps us to understand the extent to which the performance differs across respondents to highlight the extent of a problem identified in the cognitive testing.

Complementary information from the cognitive and pilot tests gives a comprehensive picture of the question performance. Doing only a cognitive test gives information on possible incorrect interpretations of questions and suggests possible revisions to the questions. The pilot test gives further information on the revised questions and whether the incorrect interpretations are significant or not, or whether very few respondents in fact show these misinterpretations. This is important information for understanding measurement error related to respondent variables in the final survey data set.

The outcome of the cognitive and pilot testing process may need to be repeated, although it is better to repeat the cognitive testing stage and only do the pilot testing when the cognitive testing results suggest few problems.

8.3   Using this approach in areas other than disability

The guidelines present the process of undertaking a combined cognitive and pilot test with the focus on the experiences from the ESCAP project testing disability measures.

As described above disability measurement is a difficult and rapidly changing field and hence requires careful testing to ensure we are using accurate and comparable measures. However, the field of statistical data is wide and many other areas of enquiry would also benefit from such testing. Two particular instances come to mind:

a) **Testing of new topic areas**: The area of income and expenditure statistics has undergone much development over recent years moving from asking about income directly to approaches that ask about assets, sources and stability of income, as well as expenditure. The impact of these changes are important to document and would benefit from being thoroughly tested using a combined cognitive and pilot testing strategy.

b) **Testing of areas which typically shows odd results or important measurement error:** Ongoing collection of data on topics such as employment, educational achievement or access to services may benefit from cognitive and pilot testing to check whether measures are performing consistently.

Use of this approach in other areas merely requires that the correct steps be followed. Once this is done a few times more and more statistical office staff will become familiar with the techniques and the benefits of collecting such information will be highlighted.

8.4   Conclusion

These guidelines have focused on testing of disability measures. Thus these guidelines are for:

- Testing existing questions that typically yield confusing results in surveys and that would benefit from a full cognitive testing to try and find out what the potential sources of these errors are. Cognitive testing of these existing questions will provide transparency as to their performance and highlight the different interpretations that may be causing problems in data analysis.
- Developing and testing new questions for areas such as disability, wellbeing, poverty and employment, which are all complex phenomena and difficult to measure in self report surveys.

- Linking up with people working on similar areas of measurement and to be able to compare results in a meaningful manner, such as comparing the interpretation across different countries and populations. This provides a growing body of evidence that enhances our understanding of these different measures and how to compare them.

9 Resources

This section is divided into websites, resource persons and readings.

9.1 Websites and programmes

1. ESCAP disability statistics project website:
   - Overall and country reports on the cognitive and pilot testing results
   - Materials from training and other workshops held during the project

http://www.unescap.org/stat/disability/index.asp

2. Washington Group on Disability Statistics
   - Reports on the WG meetings since February 2002
   - Various documents on the development and testing process

http://cdc.gov/nchs/washington_group.htm

3. Training manual on disability statistics published jointly by ESCAP and WHO, 2008
http://www.unescap.org/stat/disability/manual/index.asp

9.2 Resource persons

Mr Chhan Lay
Vice Bureau Chief
National Institute of Statistics (NIS)
386 Monivong Boulevard
Boeung Keng Kang I
Chamkarmon
Phnom Penh, Cambodia
Tel: +855 12 883 419
Fax: +855 23 213 650
Email: lay.chhan@gmail.com

Ms Tserenkhand Bideriya
Head, Data Processing and Technology
Department
National Statistical Office
Government Bldg No.3
room 303, Baga Toiruu 44
Ulaanbaatar 20ª, Mongolia
Tel: +976 51 267 885
Fax: +976 51 327 885
Email: bideriya_tserenkhand@yahoo.com

Ms Loona Abdul Hakeem
Statistical Officer
Department of National Planning
Ghaazee Building
Male, Maldives
Tel: +960-334 8383
Fax: +960-332 7351
Email: loona@planning.gov.mv

Ms Paula Monina Collado
Deputy Administrator
National Statistics Office
P.O. Box 779
Manila, Philippines
Tel: +632 716 0369/715 7758
Fax: +632 715 7758
Email: M.Collado@census.gov.ph

Ms Indumathie Ranjanadevi Bandara
Deputy Director
Department of Census & Statistics
15/12 Maitland Crescent
Colombo, Sri Lanka
Tel: +94 11 255 2538
Fax: +94 11 255 2538
Email: indu.bandara@statistics.gov.lk

Ms Marguerite Schneider
Project consultant
Tel: +27-116467639
Fax: +27-86 6840064
Email: margie_who@mweb.co.za

Ms Kristen Miller
Survey Methodologist
National Center for Health Statistics
3311 Toledo Road
Maryland 20782, USA
Tel: +1-3014 584625
Fax: +1-3014 584031
Email: ksmiller@cdc.gov

Andres Montes
ESCAP Statistics Division
United Nations Building
Rajadamnern Nok Avenue
Bangkok 10200, Thailand
Tel: + 66-(0)2288-1655
Fax: + 66-(0)2288-1082
Email: montesa@un.org

9.3   Articles

1. Miller, Kristen. 2003. "Conducting Cognitive Interviews to Understand Question-Response Limitations among Poorer and Less Educated Respondents. *American Journal of Health Behavior*. 27(S3):264-272.

2. Miller, Kristen. 2006. Results of Cognitive Testing for Oral Health Questions. Web site: http://wwwn.cdc.gov/QBANK/Home.aspx

3. Miller, Kristen. 2007. Design and Analysis of Cognitive Interviews for Cross-National Testing. 2007 European Survey Research Association Annual Meeting. Prague, Czechoslovakia.

4. Miller, Kristen -. 2008. Results of the Comparative Cognitive Test Workgroup Budapest Initiative Module. Web site: http://wwwn.cdc.gov/QBANK/Home.aspx

5. Miller, Kristen, Rory Fitzgerald, José-Luis Padilla, Stephanie Willson, Sally Widdop, Rachel Caspar, Martin Dimov, Michelle Gray, Cátia Nunes, Peter Pruefer, Nicole Schoebi, and Alisú Schoua-Glusberg. 2009. "Design and Analysis of Cognitive Interviews for Comparative Multi-National Testing." Forthcoming, *Field Methods*.

6. Miller, Kristen and Stephanie Willson. 2002. Results of Cognitive Testing for the NCHS Best Race Question. Web site: http://wwwn.cdc.gov/QBANK/Home.aspx

7. Miller, Kristen (in press – chapter in a book) Cognitive Interviewing.

8. DeMaio, Terry and Ashley Landreth. 2004. Pp. 45-67 in *Methods for Testing and Evaluating Survey Questionnaires*, edited by Stanley Presser, Jennifer Rothgeb, Mick Couper, Judith Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer. Hoboken, New Jersey: John Wiley and Sons, Inc.

9. Schneider, M., Dasappa, P., Khan, N., & Khan, A. (2009) Measuring disability in censuses: The case of South Africa. ALTER, European Journal of Disability Research. 3: 245 – 265.

10. Schneider, M. (2009) The difference a word makes: Responding to questions on 'disability' and 'difficulty' in South Africa. Disability and Rehabilitation. 31(1), 42 — 50

11. Tourangeau, Roger, Lance Rips and Kenneth Rasinki. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press

12. Willis, Gordon. 2005. *Cognitive Interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications, Inc.

10 Appendices

10.1 Appendix A: ESCAP project cognitive and pilot test questionnaires

Cognitive test

www.unescap.org/stat/disability/pre-pilot-training/cognitive-interviewing-protocol.pdf

Pilot test

www.unescap.org/stat/disability/task-team/ESCAP-WG-extended-question-set.doc

10.2 Appendix B: Examples of narratives from three domains and comments on the interpretations

10.2.1 Points about the narratives

The narratives are presented as they were recorded on QNotes with a light edit of the spelling and grammar to ensure that the reader can understand them. The narratives given are only a few examples from some of the countries. When drafting the notes from interviews the main focus should be on getting the information down and, if required, formatting and editing can happen later. The original notes should be done in the language of the interview and translated from the written notes.

The narratives vary greatly in length. The longer narratives are the most useful as they help understand the explanations behind the response given. The number of blanks or 'did not probe' should be kept to a minimum. The comments on the nature of the interpretation and the correctness of this in relation to the question intent are not obtained from Qnotes but added separately.

Examples of the first question only of the three domains are presented here and only a few examples since the total print out for all the questions is a over a few hundred pages in length.

10.2.2 Hearing – Do you have difficulty hearing, even if wearing a hearing aid? (no difficulty, some difficulty, a lot of difficulty, cannot do at all)

Country and interview number	Degree of difficulty	Narratives	Comments for analysis process
Cambodia 1	No difficulty	I thing the question ask me do I have problem with hearing or not. But I'm fine I don't have any problem with my ears.	The response given matches the narrative and, while not much explanation is given, it seems that the intent of the question has been understood.
Canada 10	No difficulty	my wife wonders sometimes, but no I don't seem to have any difficulty Probed: Well, I was thinking that…but I'll say the same thing about her…now I'm sure my hearing isn't as acute as it once was, but I don't seem to have any difficulty in my everyday work, or I have no problems	Correct interpretation of the question intent. The issue with this response is what is considered normal hearing (no difficulty) vs 'some difficulty'. This person is reporting his hearing to be fine and suggests

		hearing things in my day-to-day life	('as it once was') that it is age related and does not affect his overall functioning and, hence, is not an issue.
Canada 2	No difficulty	I don't have difficulty but if I'm not attentive to someone and the person talks to me, I might not understand or answer because I don't realize someone is talking to me. My family knows now they have to get my attention before talking to me" Retrospectively the respondent was re-asked the question to see if she would change her mind, but she said "no, what I have is not a hearing problem."	The response suggests a correct interpretation but also highlights the close relationship between paying attention and hearing. It may well be that having to get her attention is because of a hearing problem. Her response is not about misunderstanding the question.
Kazakhstan 11	No difficulty	The respondent has presented himself to the doctor to check hearing and the doctor said separate phrases very quietly and asked if he hears them. There are no problems with hearing.	The respondent gives an example of a hearing test (of sorts) and having passed that and so judges his hearing as being 'no difficulty'. The interpretation is correct and the explanation fits the response.
Maldives 16	No difficulty	She said she can hear clearly without using any kind of hearing aid, different kind of sounds at a distance of 15feet	The respondent interprets the question correctly and gives an example to explain this. This matches her response of 'no difficulty'.
USA 10	No difficulty	Upon hearing this question for the first time, this respondent made a puzzled face, and stated "I'd say that was kind of strange. Because you're talking about a hearing aid? Right?" I read the question again and he stated, "No." When I went back and probed on this question I asked him about what he was thinking about and he replied, "Yeah, because you asked about a hearing aid and I'm thinking I don't have no hearing aid but I can still hear pretty good. So that's what threw me off, when it said 'with a hearing aid' I'm like I don't even have one of them, so why is that...how's that going to help me? I can answer the question but I don't need no hearing aid." He was able to respond "no" because he doesn't have any problems hearing. He reported that he can hear some things that no one else can hear. Gave an example of often being able to hear an ambulance or fire truck coming before others do. However, then he went on to say that in one ear he can hear excellent but in the other ear he can't hear so great because he had frost bite in that ear. He said the frost bite affected his hearing somewhat in that ear. He had his hearing tested and his hearing was worse in the one ear that had frost bite than it	This respondent shows a misinterpretation of the question as being about wearing a hearing aid and not about hearing. The extensive probing leads to a correct response and interpretation. This example brings to the fore a potential problem with the 'if wearing a hearing aid' clause in the question.

		was in his other ear. He said that he can still hear somewhat out of the frost bitten ear but "it's not as clear and sharp as I would like it to be." However, it doesn't bother him enough to affect his hearing. He said he had no problem hearing me in the interview. He doesn't experience problems hearing in noisy room. He can still hear people who are sitting next to his frost bitten ear	
USA 4	No difficulty	He reports no difficulty here, but later says he has some trouble with background noises that may be beyond what's normal. He said background noise seems to take precedent over voices. He gave the example of watching a movie at home. The music in the movie makes it very hard for him to hear what the characters are saying. He's wondered if a better sound system would allow him to minimize the music and maximize the dialogue. He got his hearing tested and they told him he has some degree of hearing loss for higher ranges of pitch. But they didn't suggest a hearing aid and said some loss is normal as you age. It didn't seem to be defined as a problem by them.	The interpretation of the question is correct but this is another example of the cut off point between a 'no difficulty' vs 'a lot of difficulty'. The reliance on the professional opinion seems key in deciding this cutoff point (i.e. no difficulty) even though he seems to be experiencing some problems in noise.
Canada 4	Some difficulty	On probe: Where I have the greatest difficulty is my experience when I'm on a tractor and such…for me it's just background noise…if there's background noise and someone is speaking I may have trouble understanding what they're saying	The question is correctly interpreted and the explanations and examples given match the response of 'some difficulty'.
Mongolia 19	Some difficulty	My left ear is buzzing and I have tinnitus.	The question is correctly interpreted and the examples given match the response of 'some difficulty'.
Cambodia 2	A lot of difficulty	She said 'a lot of difficulty'. She confirmed that she never used hearing aid. She continued that 'now my hearing function is reduced a lot if compare to when I was young', by 50% to 60% reduced. She gave an example: 'More often when I'm listening to my son or daughter they talk with me, one time I can't understand anything at all. I don't remember what they are talking about. So that sometimes I must to ask people tell me two or more than this when I communicate with them.	The interpretation of the question is correct as shown by the examples and explanations given. Her response matches these examples.
Kazakhstan 12	A lot of difficulty	The respondent since childhood uses a hearing aid, but all the same he should see lips of the speaking person to understand about what is being said, and it is difficult to it to understand unfamiliar words	The interpretation of the question is correct as shown by the examples and explanations given. His response matches these examples.
Maldives 16	A lot of	She said lot of difficulty. Even with the	The interpretation of the

	difficulty	hearing aid she had lot of problem in hearing. She cannot hear clearly what people say.	question is correct as shown by the examples and explanations given. Her response matches these examples.

10.2.3 Learning – Do you have difficulty understanding and using information like following directions to get to a new place? (no, some difficulty, a lot of difficulty, cannot do at all)

Country and interview number	Degree of difficulty	Narratives	Comments for analysis process
Cambodia 7	No difficulty	She mentioned that the reason of difficulty for her in using of new information is only in cases of complicated matters such as the way to resolve mathematical equations or regression or matters related computer programs. She said she don't think understanding or using information, for example following the instruction or advised by someone with simple way, such find the new place, new game in computer and mobile phone are difficulties. Further, she added, follow the map are easy to do it. She said "no difficulty".	Seems to understand notion of learning and gives useful examples.
Canada 11	No difficulty	Probed: I was thinking about following directions and turning left, right, whatever…I was also thinking about problem-solving questions when you have to read through it and read each step to get to the next part of the question	Seems to understand notion of learning and using information to solve problems and gives useful examples.
Canada 12	No difficulty	Probed: I thought about my GPS and how I use that. So no problem.	Incorrect interpretation as respondent focuses on following directions and not on learning.
Maldives 3	No difficulty	She said no difficulty. She was considering going to places using information. She said I don't usually go places, only if i have to go to the doctor or to my mother's friend's place. I have never tried going to a places using information. I don't know haven't tried, I don't know.	This respondent focuses on the example rather than the notion of learning.
Kazakhstan 10	Some difficulty	In familiar district - problems do not arise, and in unfamiliar places - there can be problems	It is not clear from this narrative whether the respondent is referring to learning as in using information or rather just responding about problems getting around in unfamiliar places.
Kazakhstan 13	Some difficulty	The respondent thought of a situation when instructions are not exact	This interpretation is limited to the example rather than a broader notion of learning.
Canada 14	Some difficulty	Some difficulty but Mapquest helps [type of GPS device]	This interpretation is limited to the example rather than a broader notion of learning.
Canada 15	Some	some difficulty … that has to do with focus	The interpretation is incorrect

Country and interview number	Degree of difficulty	Narratives	Comments for analysis process
	difficulty	and the last question you asked (hand dexterity) the answer to that is complicated by the fact that I don't have depth perception anymore and one eye is shot so putting the cap back on a pen...I miss - I have marks all over my hands and since I take insulin, putting the cap on needles...I have to be very careful as well...the physical part is not difficult...it's lighting it up visually Probed: focus...yes...in the past, I used to drive and was quite good at navigating and knowing my way around ... now all this has been complicated by health issues ...focussing on what I should do next or where to go from here...it becomes more confusing for me	and seems to be confused as well as what is being talked about – wandering off the point.
Maldives 12	A lot of difficulty	She said she had lot of difficulty in learning. She said here that she cannot concentrate or learn that was taught. She always enjoyed playing. She also said to her teacher that she has a lot of difficulty in understanding what was being taught in the class.	This seems to be a correct interpretation with examples that look at learning rather than focusing on the example of directions only.
Maldives 15	A lot of difficulty	He said lot of difficulty, because he is not used to walk so much and now he doesn't know some new places. There are many new places. The island was not like when he was young. Even if he is given instructions he will not be able to follow.	The narrative is not detailed enough to clarify whether this is a correct or incorrect interpretation.
Mongolia 14	A lot of difficulty	Can't use the mass media. Can't receive the information owing to the blindness.	Incorrect interpretation.
Philippines 4	Cannot do at all	The respondent mentioned that she cannot follow directions going to a new place. She does not go outside the barangay (village) alone.	Incorrect interpretation.

10.2.4 Pain – Do you have frequent pain (yes/no)

Country and interview number	Degree of difficulty	Narratives	Comments for analysis process
Cambodia 1	No	The respondent said that it happened rarely. So, it is no problem for her and therefore she answers "No". and she said that if it happened every day or every week, she will answer "yes".	Seems to be correct interpretation of 'frequent' and examples fit the response.
Maldives 19	No	She said no frequent pain but she said when she gets fever then she gets headaches and body aches. also said that it happens very rarely.	Seems to be correct interpretation of 'frequent' and example fits the response.
Philippines 6	No	Although she has arthritis, she does not experience frequent pain.	Seems to be correct interpretation of 'frequent'
Philippines 14	No	The respondent, however, mentioned that if he rides the jeepney [form of transport] and he is pushed to give others some space, he feels the pain in his hips.	Seems to be correct interpretation of 'frequent' and example fits the response.

Philippines 10	Yes	Sometimes she feels pain especially when overworked with household chores, walking long distances, etc	This may be a borderline case of frequent pain and maybe an overestimation of the frequency of that pain.
Philippines 8	Yes	The respondent asked if this refers to physical or emotional pain. She mentioned, however, that physical pain first comes to her mind when she was asked with the question. The respondent related that she had a bone screening in the previous month. She had to go to the town center for the examination and she felt tired and her body was painful from walking. However, she added that she was glad that she underwent the screening and she found out the she is already diagnosed as osteoporosis. She is thankful for the diagnosis and that she already started to take calcium supplement. The local Social Welfare office provides her with calcium supplement.	This may be a borderline case of frequent pain and maybe an overestimation of the frequency of that pain.
South Africa 4	Yes	She then said 'sometimes I feel them' - when pushed about whether these were frequent, she said: no, it's not frequent but I just feel it sometimes' response should have been no. She indicated that she has not yet felt big pains	This may be a borderline case of frequent pain and maybe an overestimation of the frequency of that pain.
Mongolia 8	Yes	Backache and pain in the leg. It is difficult when I sit longer period or stands up slowly.	Seems to be an accurate interpretation of 'frequent pain' and the example given fits with the response.
USA 10	yes	For this question I stated, "you said you had frequent pain, tell me about that." He stated that he has pain in his knee and his ankles that will act up because of the weather or if has been pushing himself too much. He has no other pain other than these things. This was all he was thinking about when answering.	This may be a borderline case of frequent pain and maybe an overestimation of the frequency of that pain, but possibly some overreporting.
USA 13	Yes	He said he has daily pain. No time to probe.	Daily pain suggests accurate interpretation of 'frequent pain'.