



International
Labour
Organization

2 STATISTICAL METHODOLOGY SERIES

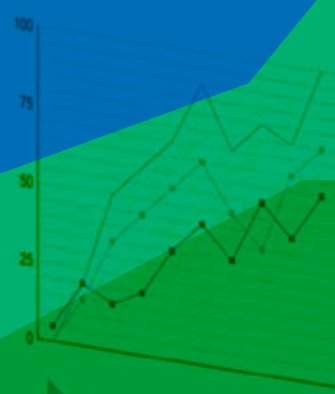
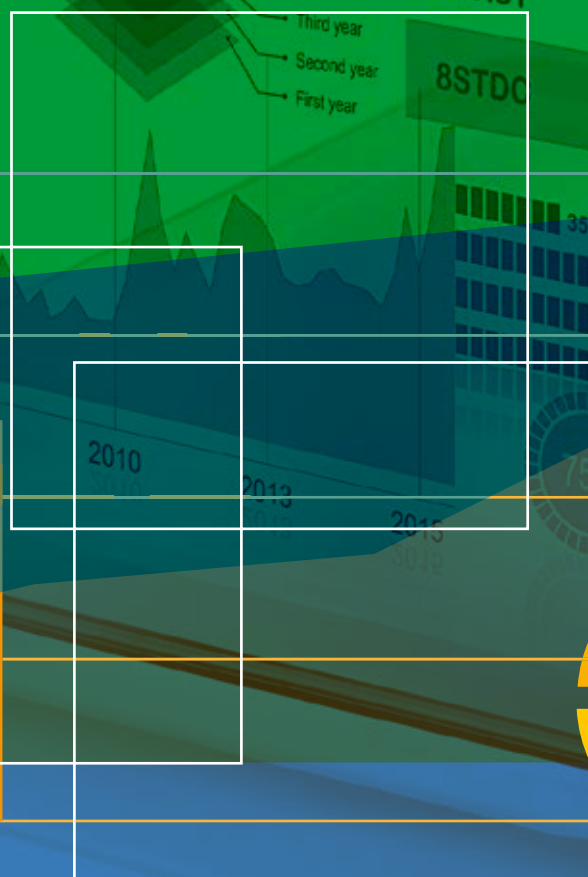
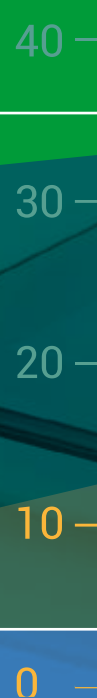
ILO LFS PILOT STUDIES COGNITIVE INTERVIEWING TESTS:

Methodology, process and outcomes

Elisa M. Benes, Kieran Walsh

May 2018

GCMP	-2,32%
STC2K	20,24%
TWFBL	30,01%
SNC5H	16,05%
EVRST	9,5%
8STDC	50,2%



STATISTICS
Department
of Statistics

INTERNATIONAL LABOUR ORGANIZATION

ILO LFS pilot studies –Cognitive interviewing tests:

Methodology, process and outcomes¹

ILO Department of Statistics –Geneva, Switzerland

¹ This document is being reproduced without formal editing.

Copyright © International Labour Organization 2018
First published 2018

Publications of the International Labour Office enjoy copyright under Protocol 2 of the Universal Copyright Convention. Nevertheless, short excerpts from them may be reproduced without authorization, on condition that the source is indicated. For rights of reproduction or translation, application should be made to ILO Publications (Rights and Licensing), International Labour Office, CH-1211 Geneva 22, Switzerland, or by email: rights@ilo.org. The International Labour Office welcomes such applications.

Libraries, institutions and other users registered with a reproduction rights organization may make copies in accordance with the licences issued to them for this purpose. Visit www.ifrro.org to find the reproduction rights organization in your country.

ILO LFS pilot studies –Cognitive Interviewing tests: Methodology, process and outcomes

ISBN: 978-92-2-131681-7 (print)
978-92-2-131682-4 (web pdf)

The designations employed in ILO publications, which are in conformity with United Nations practice, and the presentation of material therein do not imply the expression of any opinion whatsoever on the part of the International Labour Office concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its frontiers.

The responsibility for opinions expressed in signed articles, studies and other contributions rests solely with their authors, and publication does not constitute an endorsement by the International Labour Office of the opinions expressed in them.

Reference to names of firms and commercial products and processes does not imply their endorsement by the International Labour Office, and any failure to mention a particular firm, commercial product or process is not a sign of disapproval.

Information on ILO publications and digital products can be found at: www.ilo.org/publns.

Printed in Switzerland

TABLE OF CONTENTS

I.	Background	7
II.	Cognitive testing theory and practice.....	8
A.	Cognitive Interviewing theory	8
B.	Cognitive Interviewing methods	9
C.	CI testing in official labour statistics	10
III.	ILO Cognitive testing methodology.....	11
A.	ILO CI test design and objectives	11
B.	Capacity building in CI methods	13
C.	ILO CI protocol and supporting materials	13
D.	Target participant profiles	15
E.	Interview recording and summary	15
F.	Analysis and validation strategy	16
IV.	Pilot country implementation practices	19
A.	CI protocol adaptation.....	19
B.	Analysis, identification and reporting main findings.....	26
V.	Key findings and revisions introduced.....	28
VI.	Reflections and recommendations.....	33
A.	Challenges and lessons learned.....	35
B.	Recommendations on CI testing for LFS design	36
VII.	Concluding remarks	38
VIII.	References	39

LIST OF TABLES

Table 1. Stages of the Question-Answer process (adapted from Willis 2005).....	8
Table 2. List of materials developed by the ILO for the cognitive testing phase.....	14
Table 3. Target number of participants by desired work and labour market characteristics	15
Table 4. Team composition, target area, timing and duration of data collection	20
Table 5. Participant selection and recruitment	21
Table 6. Number of participants by country and desired profile.....	23
Table 7. Selected characteristics of the CI	24
Table 8. Illustration of changes to the wording of reference periods to reduce observed variability in interpretation	31

LIST OF FIGURES

Figure 1. Sample questionnaire for cognitive testing.....	14
Figure 2. Interview recording template	16
Figure 3. Interview summary template	16
Figure 4. Analysis framework template developed for the ILO cognitive test	17
Figure 5. Overall structure and order of questionnaire sections by model before CI	29
Figure 6. Dedicated section/question on “For market production” included in models M3, M4 and M5 before the CI tests.....	30
Figure 7. Revised overall structure and order of questionnaire sections by model after CI	31

ACRONYMS

ICLS	International Conference of Labour Statisticians
LFS	Labour Force Survey
CI	Cognitive Interview
NSO	National Statistical Office
UK ONS	Office for National Statistics of the United Kingdom

ACKNOWLEDGEMENT

The completion of the pilot studies and the preparation of the publications has been supported by many colleagues both within and outside the ILO.

Numerous staff in each of the 10 pilot study countries provided technical expertise and logistical support over a period of years to ensure the studies proceeded as planned. The authors would like to express their gratitude for the dedication and expertise they offered at all stages of the process. In addition to the in-kind contribution of staff time and expertise, the National Statistical Offices of Namibia, Peru and Vietnam provided financial support for some or all of the field activities in those countries.

The Pilot Study Programme was organised with the technical leadership of ILO Department of Statistics supported by the ILO's network of regional statisticians including: Africa: Yacouba Diallo, Coffi Agossou and Honoré Djerma; Asia Pacific: Tite Habiyakare; Latin America and the Caribbean: David Glejberman. In the case of the pilot study in Kyrgyzstan, additional technical support was provided by Vladimir Ganta and Igor Chernyshev.

Within the ILO Department of Statistics, a small dedicated team of professional staff and interns were responsible for processing and analysing the pilot study data and providing ongoing support for documentation, planning and implementation of the studies. The authors would like to acknowledge the contributions of Molka Abassi, Estefania Alaminos Aguilera, Anna Belianska, Umberto Cattaneo, Francisco Guerreiro, Jacob Ingnas, Carlos de Porres Ortiz de Urbina and Alina Rodríguez de Vuille, all of whom were involved in different phases of the work. Many thanks are also due to Yves Perardel, who provided substantial technical support during the implementation phase of the studies.

Logistical support was provided by a number of colleagues within the Management Support Unit of ILO Department of Statistics including Ritash Sarna, Catherine Jensen, Michelle de Chaumont, Agnes Kalinga and Virginie Woest.

Funding for the pilot studies and related activities was provided by ILO, Data2x and African Development Bank.

The support of the Office of National Statistics of the United Kingdom was critical during the development stages of the CI process discussed in this report. In particular the authors would like to note the contribution of Genevieve Groom, Peter Betts and Joanne Finney.

Without the support, expertise and commitment of all those listed above this work would not have been possible.

I. BACKGROUND

1. The latest international recommendations on how to measure some of the key headline labour market indicators, including the labour force participation rate, employment-to-population ratio and unemployment rate, are contained in the *Resolution I concerning statistics of work, employment and labour underutilization* adopted in 2013 by the 19th International Conference of Labour Statisticians (ICLS). These standards introduced a number of important revisions that will impact the way work and labour force statistics are collected and disseminated by countries around the world in the years to come. To support their wide implementation, the 19th ICLS called on the ILO to “conduct further conceptual and methodological work including testing” and develop “technical manuals and model data collection instruments” aligned with the latest standards.
2. As follow-up, in 2015, the ILO launched a global project of labour force survey (LFS) pilot studies. The Project had as main aim to develop and test alternative survey questionnaires to collect statistics to produce headline indicators on employment, labour underutilization (comprising time-related underemployment, unemployment and the potential labour force), and own-use production work, in line with the 19th ICLS standards. The findings from the ILO pilot studies will serve to develop evidence-based guidance to support countries in adopting and adapting the new standards in to suit their national context.
3. This report describes the cognitive testing methodology used as first stage of evaluation in the ILO LFS pilot project. The main focus is on country practices in implementing the cognitive tests to assess the model questionnaires. The report is part of the ILO methodological series “*Identifying good practices in labour force survey design*” that describe in detail the Project’s background and methodology and its main findings. This report and others in this series are available in the website of the ILO Department of Statistics.²
4. The report is structured as follows: Section II provides a short introduction to the cognitive interviewing (CI) method. Section III introduces the CI test and protocol designed for the ILO LFS pilot studies. Section IV details the country practices with the implementation of the ILO cognitive tests. Section V highlights the main types of issues identified by topic and changes introduced to the model questionnaires as a result. Section VI discusses the main achievements and challenges of the ILO pilot project’s CI testing phase and recommendations for planning and conducting cognitive tests to inform LFS questionnaire design. A few final remarks are presented in the conclusions.

² See: http://www.ilo.org/stat/Areasofwork/Standards/lfs/WCMS_484981/lang--en/index.htm

II. COGNITIVE TESTING THEORY AND PRACTICE

5. Testing of survey questionnaires has long been acknowledged as an important aspect of data quality assurance. While survey testing has focused traditionally on expert reviews, operational and field-related matters, problems related to the wording of survey questions have received much attention in recent years because studies have shown that questions are not always measuring what they are intended to measure. Ambiguity in survey questions, different interpretations made by the participants about the meanings of the questions, and memory recall problems can be an important source of errors in surveys. These, however, can easily go unnoticed if survey questionnaires are not evaluated for comprehension and recall with the targeted population prior to their field implementation (Presser, S. et al, 2004).
6. CI is one of the methods increasingly being used to evaluate potential sources of response error in survey questionnaires. CI is generally used during the questionnaire development phase, after an expert review of the contents, but before its field testing. Developed in the 1980s, the CI method draws from theoretical frameworks based in cognitive and social psychology to assess the cognitive processes participants use to interpret and answer questions in a survey (Collins, 2003; Presser, S. et al, 2004).

A. Cognitive Interviewing theory

7. Following the model developed by Tourangeau (1984), the process of answering survey questions can be thought of as a four-stage process (i.e. comprehension, recall, judgement or estimation and response) as illustrated in [Table 1](#). In each of the four stages, various types of response errors can occur. For example, problems of comprehension (stage 1) can occur when participants are not familiar with some of the key terms used, or if the question is ambiguous, too long or complex. Alternatively, participants might understand the question; however, they might not recall with precision the information requested, thus leading to recall errors (stage 2). Participants might also alter their response based on perceived social expectations (stage 3). Else, participants might select a response option that does not fully fit with their desired answer given the response choices presented to them (stage 4). The CI method attempts to assess the soundness of survey questions by examining their performance vis-à-vis the four stages of the question-answer process.

Table 1. Stages of the Question-Answer process (adapted from Willis 2005)

	Cognitive stage	Cognitive process	Response error types
1	Comprehension	Participants interprets the question	Unknown terms, ambiguous concepts, long or complex formulations, etc.
2	Recall	Recall information from memory	Recall difficulty
3	Judgement or estimation	Arrive at decisions on how to respond	Biased or sensitive information, estimation difficulty
4	Response	Map answer to formal response categories	Incomplete response options, ambiguous options, mismatched options, etc.

8. Additionally, in cross-national surveys, CI can also be used to assess overall instrument equivalence and thus the comparability of the resulting data. For example, CI can help identify errors of comprehension due to translation issues or provide evidence of “construct overlap”, that is the extent to which different linguistic and cultural groups understand key concepts the same way (Fitzgerald, R. et al, 2009; Miller et al, 2011).

9. Overall, the goal of CI is to prompt the participant to share information that provides clues, both verbal and non-verbal, regarding the underlying cognitive process used in answering. That can help us to pinpoint possible issues in the questionnaire design that could lead to response errors (Willis, 1999; Willis, 2005). This information is then used by the survey designers to decide on changes needed to the questionnaire to minimize response errors before field testing.

B. Cognitive Interviewing methods

10. Typically, CI is performed by conducting in-depth, semi-structured interviews with a small, purposive sample of participants. Being a qualitative evaluation technique, participants are selected based on key characteristics of interest. The aim is not to cover all possible groups of the population, but rather those of particular interest for the questions being evaluated (e.g. employed) as well as those without the characteristic but who may be considered as borderline cases (e.g. working without pay). Overall, the number of participants interviewed is low (5-15 per “round”), with the aim to identify the more critical questionnaire problems. The process is repeated, ideally on an iterative basis, after revisions have been introduced to the questionnaire, progressively reducing the sources of potential response error identified.
11. Two major types of interviewing techniques are generally used in CI: think-aloud interviewing and verbal probing techniques (Beatty, P. et al, 2007). In think-aloud approaches, participants are instructed to voice their thoughts as they answer the survey questions. The interviewer interjects little other than to encourage the participants to keep voicing their thoughts. In verbal probing approaches, paraphrasing (e.g. “could you repeat this question in your own words?”) and a variety of probes such as, “what does x mean to you?” or “you hesitated during answering, could you tell me why?” are used to prompt participants on selected aspects of interest to the evaluation.
12. Probes can be targeted to a specific stage of the question-answer process (comprehension, recall, judgement, response) and can be implemented concurrently, that is, right after the participant answers a given survey question, or retrospectively, that is, after a series of survey questions have been completed. The probes may be scripted, i.e. developed prior to the interview and to be used in a more structured interview approach. Probes may also be spontaneous, at the initiative of the interviewer based on the flow of the interview and concerns of interest.
13. In general, these techniques and others tend to be combined in CI taking into consideration the survey questions being evaluated, the characteristics of the participants being targeted, and the level of expertise of the interviewers. In cross-country CI, more structured probes and interview protocols have tended to be used as a means to maintain consistency in the application of the CI method across settings (Willis, 2015a).
14. Analysis of CI tests seeks to assess the likelihood and impact of cognitive issues arising from the questionnaire design and which stage of the question-answer process the issues may arise from. A wide range of qualitative analysis approaches of CI exist. However, important to all is the establishment of a systematic analysis procedure to ensure that no particular case is overemphasized, that findings take into account the full range of responses and that analysts apply a consistent evaluation criteria.

15. Analysis approaches can be top-down or bottom up. In top-down approaches, the CI results are coded based on a list of issues of interest to the researcher (i.e. ambiguous question, missing instructions, technical terms, complex calculation, etc.). Bottom/up approaches rely instead on identification of patterns in the CI results and construction of codes based on the issues identified. The analysis can be further broken down by stage of the question-answer process to identify issues related to comprehension or interpretation, recall, judgement or estimation and response (Willis, 2015b). In general, analysts must examine results within interviews, across interviews (by question) and across target groups with the goal of identifying thematic patterns in question interpretations and response error (Statistical Policy Directive N. 2 Addendum).
16. From this the survey designer can assess the necessity to amend existing question wording and sequences. However, as the process does not typically directly allow comparison of different approaches to wording of questions, the type of updates made rely on expert judgement and further rounds of testing may be warranted.

C. CI testing in official labour statistics

17. In official statistics, CI methods are being increasingly used by national statistical offices (NSO) in more developed countries to inform questionnaire design. CI testing specific to official labour force statistics have been conducted in the United States (Cohanny, S. et al, 1994; Rothgeb, 2007; Bienias, 1988), several European countries and others on an increasing scale over time (Guinea, D. et al, 2003; Fasching, M. et al, 2014). In these contexts, CI testing has been used as an integral evaluation method during the review and update of the countries' national LFS or related household surveys. These tests have almost universally been organised as national tests, i.e. without any cross-country comparisons or co-ordination. However, recent rounds of testing in Europe and Latin America have seen multiple countries use cognitive testing as part of a co-ordinated testing process, albeit with no direct co-ordination of the design in the different countries.
18. Limited information exists on the use of CI testing by NSOs in less developed countries. Existing literature on CI techniques and its usefulness tends to refer to practices tested in contexts with overall higher literacy levels, more structured and urban economies. Recent reviews of cross-cultural CI studies nevertheless suggest existing CI techniques appear to be adequate across different contexts (Willis, 2015a).

III. ILO COGNITIVE TESTING METHODOLOGY

19. From the earliest stages of planning, the ILO recognized the importance of using qualitative assessment methods to inform the design of alternative model LFS questionnaires aligned with the new ICLS standards. The CI method was chosen given the double aim of developing model questionnaires to serve as updated guidance for national LFS data collection in line with the latest standards, and at the same time promote cross-country comparability of the resulting statistics. To support the process and development of the methodology for CI, the ILO engaged the Office for National Statistics of the United Kingdom (UK ONS) during the first half of 2015.

A. ILO CI test design and objectives

20. The overall background of the pilot tests and broad design are described in a separate report in this series. For the purposes of this report it is important to note that the ILO developed five model LFS questionnaires for testing and collaborated with 10 countries from different world regions to conduct the tests. Each pilot country tested two of the five questionnaires to allow for comparisons to be made within the same context.
21. The five questionnaires had a broadly similar scope, each covering the measurement of employment, labour underutilization, own use production work (of goods and services) and main activity, among other subjects chosen to add context for analytical purposes or as interesting topics for study in their own right. Although covering the same topics, the models included a number of design differences that were the focus of the tests. These included differences in the basic structure and order of the questionnaire sections covering the different topics (e.g. employment, own-use production work, job search, etc.), as well as in the wording and type of questions used (e.g. spontaneous versus read aloud response options, single versus multiple response options).
22. A key consideration in developing the methodology for the CI was the range of topics which could be covered. Not all questions nor modules included in the model LFS questionnaires required cognitive evaluation. Furthermore, there was a need to keep manageable the duration of the CIs, first to allow for in-depth assessment of new and priority topics and, second considering the limited experience of partner countries in the use of CI methods.
23. In line with the overall measurement objectives of the pilot studies the decision taken was to focus the cognitive evaluation on the following elements of the alternative questionnaires:
- a. Basic structure and order of sections to capture employment and own-use production work
 - b. Core questions to identify persons employed in the reference period
 - c. Core questions to establish the boundary between employment and own use production work (main intended destination of the products)
 - d. Core questions to identify the unemployed and potential labour force (job search, availability, desire for work)
 - e. Questions to capture main activity based on self-perception
 - f. Questions to identify participation and time spent in own use production work

24. Focus on the above topics, while challenging given the wide scope, was meant to enable comparative assessments of the main design differences introduced in the five model questionnaires developed.
25. Cognitive issues of particular interest included the overall interpretation of the main intention of the questions selected for testing; understanding of key terms considered central in labour force statistics (i.e. pay, profit, job, business); problems of recall and interpretation of reference periods; complexity of selected new questions to identify persons engaged in own-use production work; possible section and question order effects; judgement and decision-making process for perception-type questions (main intended use of products, desire to work, need to work, availability to work); as well as sensitivity and possible social desirability effects for selected questions.
26. A central concern was to assess whether any differences could be observed in the way questions were interpreted and answered by participants with different characteristics, in particular, women and men, persons living in urban and in rural areas, and persons in different labour market situations, including self-employed persons, regular and casual employees, job seekers, and others not engaged in the labour market.
27. In addition, there was interest in examining the equivalence of key survey questions across the pilot countries. This included also evaluation of possible translation issues, as well as potential cultural differences in interpretation and/or question sensitivity.
28. Given the complexity of implementing the CI tests across 10 countries from different world regions, it was decided to conduct a single cognitive testing round in each country and to use a semi-structured interview protocol with standardized materials and procedures (see paras. 35-36). Nevertheless, to the extent possible the cognitive tests were planned in a staggered manner across countries to allow possible revisions of the cognitive protocol based on the initial experiences with its implementation.

Implementing agencies and country CI teams

29. The ILO partnered with the NSO of each pilot country to implement the cognitive tests. The collaborating NSOs all had an established national LFS, and thus strong subject matter experience. In addition, all participating NSOs had expressed interest in becoming familiar with the use of cognitive testing methods for survey development. However, none had prior experience with using CI. Collaboration with the NSOs was deemed crucial given the overarching objective of the pilot project to develop guidance to support the implementation of the new ICLS standards in official work and labour force statistics. Thus, the study design envisioned from the start a comprehensive component of capacity building on cognitive testing theory and methods as well as ongoing technical assistance throughout the planning, implementation and analysis.
30. The NSOs were tasked with establishing the team to be responsible for the planning, conduct and analysis of the cognitive tests and for allocating roles and responsibilities among team members, with guidance from the ILO. As per the focus of the cognitive tests, it was not possible to assume that the CI could be conducted under controlled conditions with technology available to record the interviews. Thus, the study design envisaged the establishment of a small number of teams comprising an interviewer and an observer conducting and recording the interviews using both an audio recorder and note-taking and a supervisor coordinating the daily work, tracking the

participant recruitment and selection, planning for the conduct of interviews in a neutral environment, and reviewing the observer notes and interview summaries. In all countries, separate teams were established and assigned to each model questionnaire tested, to avoid the contamination of findings.

31. While interviewers and observers all received training on CI techniques, observation, note-taking and summarizing, by the end of the training a primary role was generally assigned to each team member based on their performance during the practical exercises. Procedures were also established, as detailed below, to summarize the results of each interview, consolidate the results, analyse and report the main findings.

B. Capacity building in CI methods

32. The ILO provided capacity building on CI for pilot countries throughout the process. This included:
 - a. A three-day Training of Trainers (May 2015) on CI theory, methodology and planning for technical staff from NSOs responsible for implementing the pilot study, delivered by experts in CI from the UK ONS;
 - b. Ongoing remote support to plan the cognitive tests and adapt the cognitive materials to the national context; and
 - c. A hands-on national training, typically of five days duration, for the technical teams assigned to carry out the cognitive tests (NSOs) which comprised:
 - i. Review of survey contents and issues for cognitive testing;
 - ii. Team composition, interviewer and observer roles;
 - iii. Participant selection and recruitment;
 - iv. CI techniques and use of the ILO cognitive protocol;
 - v. Practical exercises, including demonstration interview, mock practice, role playing, and live practice with NSO staff and external volunteers.
 - vi. Note-taking, summarizing and analysis of results.
33. In addition, the ILO participated as observer in the initial CI conducted in all pilot countries and provided further guidance to the teams with a view to improve their CI procedures.

C. ILO CI protocol and supporting materials

34. In preparation for the cognitive tests, the ILO prepared abridged individual survey questionnaires for each of the five models developed for testing. The abridged questionnaires contained only the questions selected for cognitive testing plus a few additional questions needed to facilitate the flow of the interview. Visual aids were used to highlight the questions to be probed, the question blocks to be asked without interruptions, and the timing when the interviewer should switch to the cognitive probes (see [Figure 1](#)).

Figure 1. Sample questionnaire for cognitive testing

35. In addition, an accompanying CI protocol tailored to each of the five models was developed with the support of the UK ONS. The CI protocol provided practical information on the objectives of the tests, identification and selection of participants, instructions on how to conduct the CI, introductory statement to be read out at the start of the interview, as well as the cognitive probes to be used for each survey question being evaluated. Preference was given to the use of concurrent and retrospective probes focusing on the four stages of the question-answer process, as relevant.
36. Additional supporting materials including a sample consent form and tracking forms to monitor the participant selection and recruitment were also provided. Likewise, templates were developed to record and summarize each CI, and to consolidate, analyse and report the results. [Table 2](#) shows the list of materials and templates provided to countries in preparation for the cognitive tests.

Table 2. List of materials developed by the ILO for the cognitive testing phase

	Document	Purpose
1	Abridged Questionnaire for Cognitive tests (M1-M5)	Questions to be tested
2	CI Protocol (M1-M5)	To carry out the CI
3	Interviewer Manual for Model questionnaire (M1-M5)	For training of interviewers
4	Summary of issues for Cognitive Testing (M1-M5)	For training of interviewers
5	Participant tracking form	For participant selection & tracking
6	Interview Recording Template	For note-taking during the interview
7	Interview Summary Template	To summarize individual CI
8	Analysis Framework Template (excel)	To analyse results of CI
9	Analysis Report Template	To report results to ILO
10	Sample Consent form	Sample template if needed
11	List of Common Probes for Cognitive testing	Optional interviewer aid

37. All materials were initially developed in English and translated into French, Russian and Spanish by the ILO. Countries further translated the questionnaires and cognitive protocol to additional national languages where needed (e.g. Arabic, Kyrgyz, Vietnamese). The materials, once translated, were discussed with field staff familiar with the national LFS and further adapted to suit the local context. This included incorporating colloquial words and examples, and in some cases,

restructuring the question formulation to match existing local expressions. In all cases, revisions were discussed and agreed with the ILO to ensure that the intended meaning of the original question was retained.

D. Target participant profiles

38. To enable assessment of how well the questions worked for different types of participants, the ILO CI protocol included a set of target ranges for participants of working age with selected characteristics. The aim was to complete 20 CI for each abridged model questionnaire to be tested, bringing to 40 the total number of CI to be completed per country.
39. Key participant characteristics targeted are shown in [Table 3](#). These included persons in different employment situations, persons producing selected types of goods mainly intended for family use and non-employed persons engaged in various activities. The protocol likewise called for a balanced selection of participants by sex, urban and rural place of residence and broad age group.

Table 3. Target number of participants by desired work and labour market characteristics

Participant profile	Target range	
	per model	overall
Employed (by type of worker)		
• Employee	2-4	4-8
• Casual worker	2-4	4-8
• Contributing family worker	2-4	4-8
• Other self-employed	2-4	4-8
Not employed		
• Job seeker	2-3	4-6
• Young person not in education	2-3	4-6
• Persons looking after family	1-2	2-4
• Other outside the labour force	1-2	2-4
Overlapping characteristics		
Producer of goods mainly for family use		
• Agriculture, animal husbandry	4-6	8-12
• Fishing (as relevant)	1-2	2-4
• Manufacturing	1-2	2-4
Total	20	40

E. Interview recording and summary

40. Taking into consideration the limited experience of NSOs in conducting CI, the ILO in consultation with the UK ONS, developed a practical approach to record and summarize the results of the individual CI. The approach recommended countries to use both note-taking and audio recording to document the interviews. An interview recording template was developed to provide the framework for note-taking (see [Figure 2](#)). Both interviewers and observers were trained on how to use the recording templates and the type of information to capture, focusing in particular on direct quotes and non-verbal behaviours. The audio recordings were meant to be used primarily as supporting evidence during the summary and analysis stages. Thus, no special emphasis was placed on full transcription of the audio recordings.

Figure 2. Interview recording template

ILO Department of Statistics – Pilot Studies 2012

3. Survey answers and probing (see Protocol for summarists)

Module:	
Question:	
Response to Q:	
Verbatim response/ Q&A process/ observations:	
Cognitive probing:	
• Comprehension:	
• Recall, judgement and response:	
• Other probes/ issues:	
[To be repeated for all modules & questions in the Module]	

Figure 3. Interview summary template

4. Survey answers and probing:

(see the Protocol for summarists)

Module M3EMP – Employment in the last week/7 days

Section M3EMP01 to M3EMP03	
Cognitive probing:	
• Comprehension what did the respondent understand:	
Question M3EMP01	
Response to Q:	
Verbatim response/ Q&A process:	
Cognitive probing:	
• Comprehension – what did respondent understand:	
o Work for pay or profit	
o in cash or in kind	
o Consider other types	
o If not, why?	
• Recall, judgement – how did respondent decide answer:	
o Yes/No	
• Other probes/ issues:	
o Easy or difficult to understand	
Other comments:	
• Respondent	
• Interviewer	

41. The ILO also developed an interview summary template to be completed after each interview to summarise the main conclusions on each question probed, in this case, following the four stages of the question-answer process (see [Figure 3](#)). The observers had the task of preparing the interview summary, however, they were required to discuss and confirm the summary findings with the interviewer.

F. Analysis and validation strategy

42. The analysis strategy developed for the cognitive phase relied on successive aggregation of results, at national level by the NSO team assigned to carry out the CI, and globally, by the ILO team responsible for the pilot study. Emphasis was placed on textual summaries containing direct quotes to illustrate the main findings. No coding system was used; however, the interview summary and analysis strategies required structuring the results with particular reference to the four stages of the question-answer process (comprehension, recall, judgement and response) to facilitate identification of particular cognitive issues that could lead to response errors and inform possible revisions needed. Emphasis was placed on the recording of verbatim quotes which illustrated the main findings.
43. To support the country level analysis, the NSOs were instructed to transcribe the main findings from each individual interview into an analysis framework template prepared by the ILO for this purpose. The analysis framework template was prepared in excel to ensure that all NSOs could carry out the analysis without the need to acquire specialized software and training for qualitative analysis. The framework template served to facilitate the process of identifying common themes or patterns for each question evaluated across the four stages of the question-answer process (see [Figure 4](#)).

Figure 4. Analysis framework template developed for the ILO cognitive test

FRAMEWORK CHART		MODULE: MCEMP - Employment													
Interview ID number	Respondent characteristics	Question MCEMP01: Work for pay or profit					Question MCEMP02: Work for pay or profit recovery list					Question MCEMP03: Work as contributing or assisting family			
		Answer to survey question	Comprehension	Recall	Judgement	Other	Answer to survey question	Comprehension	Recall	Judgement	Other	Answer to survey question	Comprehension	Recall	Judgement
18	Age group: Sex: Level of education: Place of residence: Type of employment: Type of over-use production work: Type of transaction/employment:														
19	Age group: Sex: Level of education: Place of residence: Type of employment: Type of over-use production work: Type of transaction/employment:														
20	Age group: Sex: Level of education: Place of residence: Type of employment: Type of over-use production work: Type of transaction/employment:														
Findings															
Main findings	Insert main conclusions/summary of issues for each question														
Likelihood	High Low														
Impact	High Low														
Recommendations															
Recommendation 1															
Recommendation 2															
Recommendation 3															

44. The analysis framework template also included a final section for each question where analysts were asked to provide their overall assessment of how the question had performed based on two criteria:
- Likelihood:** The likelihood a participant would not understand the question as intended and be able to accurately respond.
 - Impact:** The impact of any errors in response. Impact was considered high if the result of misunderstandings by participants was likely to lead to their misclassification.
45. For each criteria a three-point scale (high, medium low) was used. Based on the ratings, countries were asked to provide recommendations for changes to the questions or indicate if they felt no change was needed.
46. The NSOs were required to prepare a report describing their practices in planning and conducting the CI, the main findings per model and question evaluated, and their conclusions and recommendations. The ILO provided a template for the preparation of the final reports that included an annotated table of contents and examples of how to report cognitive findings by stage of the question-answer process based on textual analysis. Countries were required to submit to the ILO the final reports along with the summary analysis templates used to aggregate and conduct the analysis of the CI. These documents could be submitted to the ILO in English, French, Russian or Spanish.
47. The ILO reviewed the individual interview summaries transcribed in the analysis framework, the main findings reported by the countries and conclusions arrived at. Any discrepancies, ambiguous or unsupported findings identified were referred back to the pilot country for further clarification. Furthermore, the ILO conducted a review of the summary analysis templates and country reports to identify the main findings and any common patterns across models and countries. The cross-model and cross-country analysis conducted by the ILO relied on further aggregation of the results

by model and survey question evaluated. Insights gained through observation of selected CI across pilot countries by the ILO technical team were also instrumental in the interpretation and overall assessment of the CI findings.

48. The findings were used by the ILO to draw broader conclusions on the suitability of each of the five model questionnaires and to identify necessary changes to the model questionnaires to be introduced prior to the start of the field testing stage.
49. The main findings from the cognitive testing phase were discussed with the pilot countries during an analysis workshop held in Geneva at the end of the pilot study project in November 2016.

IV. PILOT COUNTRY IMPLEMENTATION PRACTICES

A. CI protocol adaptation

50. Implementation of the ILO cognitive tests took place between July and October 2015. The first country to implement the cognitive tests was Moldova. This first experience was used by the ILO to assess the overall performance of the CI protocol prior to its rollout in other pilot countries. A number of issues were identified with the overall length of the interview and the types of probes that had been selected for use. In particular, comprehension probes such as “what does [x] mean to you?” and paraphrasing “can you repeat this question in your own words?” were found to be somewhat difficult to answer especially for participants in rural areas, as was the use of think-aloud approaches.
51. Based on this experience, the CI protocol was revised to include probes aimed at capturing similar information but that were less abstract and more relatable to participants given their immediate experiences, such as “when I said [x] what did you think about?; “could you give me examples of “x”?”; “when I asked [x], you said [y], can you tell me a bit more about that? These revisions proved to be very important for the implementation of the cognitive tests across the remaining pilot countries.
52. The revised CI protocol was further assessed in the second country to conduct the cognitive tests (Namibia) and few additional refinements were introduced to facilitate its use by interviewers and to shorten the overall duration of the interviews. After these revisions, the CI protocol was maintained without further changes in the remaining pilot countries. Any adaptations beyond this point were limited to address translation issues or include local examples as relevant.

CI team composition

53. The cognitive tests were for the most part implemented by small teams of technical staff and experienced field interviewers from the NSO (see [Table 4](#)). The technical staff assigned to conduct the cognitive tests all had prior experience with the national LFS, thus they had strong subject matter knowledge. The field interviewers were generally selected from the existing pool of household survey interviewers, and in many cases, were field supervisors responsible for quality assurance. Only one pilot country, Tunisia, engaged an external market research firm with expertise in conducting qualitative studies to carry out the cognitive tests.
54. In all cases, the team members participated in the training on CI methods organized by the ILO in collaboration with the NSO. The duration of this training was typically 5 days. Most countries also provided additional training on labour force concepts and definitions, usually a few days in advance of the ILO trainings, to ensure a common understanding of the underlying concepts and issues being assessed.

Table 4. Team composition, target area, timing and duration of data collection

Country	Interview Team		Data collection		
	Composition	Size	Target area(s)	Month	Duration
CMR	NSO technical staff & field interviewers	9	Yaoundé, Nkometou, Bachenga	OCT 2015	5 days
ECU	NSO technical staff	8	Sierra (Quito, Cotopaxi) & Costa (Guayas) regions	SEP 2015	4 days
CIV	NSO technical staff & field interviewers	9	Abidjan & Anyama regions	AUG 2015	5 days
KGZ	NSO technical staff & field interviewers	4	Chuisakaya Region	AUG 2015	6 days
MDA	NSO technical staff & field interviewers	4	Multiple regions	JULY 2015	3 weeks
NAM	NSO technical staff	6	Ohangwena region	AUG 2015	3 weeks
PER	NSO technical staff & field interviewers	5	Lima, Huaura & Cañete provinces	SEP-OCT 2015	5 days
PHL	NSO technical staff & field interviewers	12	Cebu area	SEP 2015	6 days
TUN	External company	5	Tunis area	SEP 2015	5 days
VNM	NSO technical staff & field interviewers	15	Hanoi area	SEP 2015	5 days

55. The approach taken to establish and train the cognitive team served to ensure that both interviewers and observers had a generally good understanding of the intended objectives of the questions being tested and the issues of concern. They also had overall good skills with establishing rapport with participants, maintaining control of the interview flow and interview setting.
56. Given their background in quantitative surveys, interviewers generally had solid skills in applying the cognitive protocol and structured probes provided. Nevertheless, some struggled with marking the transitions between the questionnaire being tested and the retrospective probes. Additionally, few were comfortable with using more spontaneous probing when insufficient detail was obtained through the pre-set probes or with improvising when participants showed difficulty with the established cognitive probes. This overall, limited the depth of the information captured during CI and, at times, also introduced noise or error, when participants felt confused or stressed by the types of probes used.

CI study sites

57. The cognitive tests took place in specified target areas selected by the pilot countries to facilitate identification of participants with the desired profiles. This for the most part involved a combination of urban and rural areas with different levels of integration to local or regional markets where persons involved in mixed and subsistence agriculture, animal husbandry and/or fishing could be found. Based on these requirements, most pilot countries selected a target urban centre and a few villages in adjacent rural areas. In some cases, countries targeted two distinct areas to capture different primary industries, such as fishing in a coastal area and agriculture in the highlands. Only Moldova selected participants from multiple areas of the country (see [Table 4](#)).
58. Most pilot countries completed the CI within a concentrated one-week period. This was generally necessary as in most pilot countries the CI team had to travel to the target area to conduct the interviews. In some cases, countries reported that the time allocated to complete the targeted number

interviews was rather limited. A main issue faced was the need to identify and recruit new participants in cases where those selected initially were not located, declined to participate, or had a new labour market situation and thus, no longer fit the desired profile. Only two countries spread the CI over a 3-week period. In Moldova, this was necessary as the interviews were combined with other work assignments. In Namibia, local approvals, recruitment and selection of participants had to be organized upon arrival to the target area, thus extending the time needed to complete the cognitive test.

Participant selection and recruitment

59. The selection and recruitment of participants with the desired profiles was particularly difficult for the NSOs. Pilot countries used a variety of approaches to identify and select participants (see [Table 5](#)). Half of the pilot countries used as basis the results from a previous survey that collected labour-related information as thus could help in identifying potential candidates with the desired profiles. This included a recent LFS, living standards survey, child labour survey and similar. In these cases, NSOs were advised to select only participants that did not directly participate in the survey, but whose information had been provided by a family member, to reduce potential contamination. Although this strategy appeared to be the most accessible to NSOs, several reported that over half of the participants identified using this approach had changed their labour market situation by the time they were contacted to participate in the cognitive test and thus did not always fit the desired profiles for the study.

Table 5. Participant selection and recruitment

Country	Participant selection method	Incentive
CMR	Previous survey	Yes
ECU	Previous LFS	Yes
CIV	Previous survey	Yes
KGZ	Previous survey	No
MDA	Convenience sampling	Yes
NAM	Radio Ad & targeted area	No
PER	Previous LFS	Yes
PHL	Purposive sampling	No
TUN	Quota sampling	Yes
VNM	Convenience sampling	No

60. Four additional countries used a variety of convenience, purposive and quota sampling techniques. Moldova and Vietnam used convenience sampling where the networks to identify potential participants were created through NSO staff. Moldova, for example, requested LFS interviewers to identify potential family members, friends or acquaintances that fit the desired profiles. A list of more than 50 potential participants was created and from these 40 were selected. To limit possible contamination effects, interviewers were instructed not to disclose detailed information about the study with the selected participants, nor allowed to interview them. Rather appointments were arranged only with non-related interviewers. Vietnam used a very similar approach creating a list of 60 potential participants identified through NSO staff from which 40 were randomly selected taking into consideration their sex, age group and place of residence. The Philippines, by contrast, created a network to identify participants based on local area leaders who provided a list of potential participants from which participants were selected.

61. Finally, Tunisia and Namibia, used various strategies to raise awareness about the study among the target populations and to recruit participants. Namibia, for example, implemented a strategy that involved identifying target neighbourhoods and villages, requesting the approval of local leaders, then making an appeal by radio to the local population, and finally randomly visiting households in the targeted areas to recruit potential participants.
62. In addition, a number of procedures were implemented by countries to promote participation. For countries identifying participants through a prior household survey this included initial calls and/or visits to request participation, explain the overall purpose of the study and set-up appointments. Most countries also prepared official letters providing general information about the cognitive study that were handed over to potential participants. In addition, six out of the ten pilot countries used incentives as a way to reduce barriers to participation. In particular, cash incentives were provided to cover transport costs when conducting the interviews in the NSO offices or in recognition for the time spent participating in the interview. In-kind incentives, used in Peru only, were offered as a token of appreciation. In all cases, the incentives were provided prior to the start of the interview and participants were informed of the voluntary nature of their participation. Although not used to promote participation, the Philippines provided lunch to participants upon arrival to their appointment for the interview.
63. Nevertheless, despite these various strategies, countries indicated having found particularly challenging the participant selection and recruitment process. Among the challenges cited were difficulty in identifying participants with selected labour-market characteristics, limited flexibility to track the progress achieved in meeting the target number of interviews per desired profile, problems to deal with last minute cancellations or no-shows, and lack of mechanisms to identify additional participants from underrepresented profiles. In addition, some of the NSOs indicated that their procedures did not allow for the use of incentives, as this could impact the general procedures established by the office in the conduct of official household surveys.

Participant characteristics

64. Despite the difficulties reported in identifying and recruiting participants, for the most part, countries were able to meet the target number of participants with the desired profiles as specified by the ILO cognitive protocol.
65. [Table 6](#) shows the number of participants that completed the CI by country and desired profile. In terms of age group, some countries reported problems recruiting a sufficient number of young adults (15-24 years). This was particularly the case for Kyrgyzstan, Moldova, Peru, Philippines and Vietnam. In terms of sex distribution, the study called for a balanced number of women and men to participate. This was achieved in all countries except Moldova and the Philippines. A balanced representation of participants from urban and rural areas was also mostly achieved, except in Moldova, the Philippines and Vietnam, likely due to the starting point of the convenience sample selection process implemented (NSO staff members). However, as described below, even in those three countries the targeted number of individuals engaged in agriculture or fishing work was still achieved thus enabling the assessment desired of questions relevant to those types of activities such as the boundary between employment and own use production of goods.

Table 6. Number of participants by country and desired profile

	Target range	Actual Interviewed									
		CMR	ECU	CIV	KGZ	MDA	NAM	PER	PHL	TUN	VNM
<i>Age group</i>											
15-24 years	[10-14]	12	14	9	6	7	10	8	7	13	6
25+ years	[26-30]	32	26	31	34	33	30	32	32	27	34
<i>Sex</i>											
Men	[18-22]	22	22	20	19	9	20	19	16	22	19
Women	[18-22]	22	18	20	21	30	20	21	23	18	21
<i>Place of residence</i>											
Urban	[18-22]	24	20	20	20	26	25	23	37	20	30
Rural	[18-22]	20	20	20	20	13	15	17	3	20	10
<i>Employed</i>											
Employee	[4-8]	7	9	7	17	21	6	7	5	5	18
Self-employed	[8-16]	24	9	8	8	7	21	21	12	12	11
Family helper	[4-8]	2	3	5	3	0	2	3	7	5	3
<i>Not employed</i>											
Job seeker	[4-6]	5	3	1	3	7	6	5	3	6	2
Household care	[2-4]	2	4	3	1	0	0	2	2	3	2
Other	[6-10]	1	3	12	8	3	5	2	1	9	4
<i>Producing for family use</i>											
Agriculture/fishing	[12-16]	19	9	18	7	16	37	19	14	13	14
Total	40	44	40	40	40	39	40	40	39	40	40

66. In terms of labour market characteristics, relatively low numbers of participants that worked as contributing family workers were interviewed by most pilot countries. Also, particularly difficult was the identification and recruitment of potential discouraged job seekers. In this regard, the NSOs noted not having the tools to identify persons in this situation ahead of the cognitive test. Nevertheless, a high number of self-employed participants and participants engaged in production of goods from agriculture and/or fishing mainly for family use were successfully recruited and interviewed.
67. The overall targets for employed, not employed persons and persons producing goods intended mainly for family use were set to provide sufficient cases to identify the main possible sources of response error during the cognitive tests. Nevertheless, the protocol did not include procedures to monitor the achievement of sample saturation. This was in part due to limited experience in setting parameters to establish when saturation point had been reached, as well as to the complexities of coordinating a multi-country pilot study with multiple testing phases. In the case of selected profiles, contributing family workers for example, it is likely that saturation point was not reached.
68. While it was not possible for all countries to fully match the target sample in all cases, it is nonetheless believed that the distributions of the samples were broadly sufficient to enable meaningful assessment of cognitive issues with the model questionnaires. Additionally, comparison of results across pilot countries further served to assess the potential diversity of issues that could arise with the model questionnaires.

Venue of CI

69. For the most part, the CI were held in the home of the participant [Table 7](#). In some countries, particularly for participants living in an urban centre, the interview took place at the NSO offices. Only in Tunisia, the interviews were conducted in a room set-up especially for conducting and observing qualitative interviews. This was possible because Tunisia hired a market research firm that had facilities to carry out qualitative studies. In the Philippines, a conference room at a local hotel in the selected area was booked for this purpose and participants were requested to arrive for the interview at a scheduled time.

Table 7. Selected characteristics of the CI

Country	Venue	Language	Direct interpretation	Recording
CMR	Household	French, Eton	Yes	Audio & Notes
ECU	NSO office & household	Spanish	No	Audio & Notes
CIV	NSO office & household	French, local languages	Yes	Audio & Notes
KGZ	NSO office & household	Kyrgyz, Russian	No	Audio & Notes
MDA	NSO office	Romanian, Russian	No	Audio only
NAM	Household	Oshiwambo, English	Yes	Audio & Notes
PER	Household	Spanish	No	Audio & Notes
PHL	Meeting room & household	Cebuano, Tagalog, English	Yes	Audio & Notes
TUN	Recording room	Tunisian Arabic	No	Audio & Notes
VNM	NSO office	Vietnamese	No	Audio & Notes

70. Conduct of the interviews in the NSO offices generally provided a quiet, neutral environment for the interview where participants could focus without distractions. Nevertheless, some problems were faced with no-shows, with some participants arriving at times other than scheduled, and in some cases, participants having to wait before being interviewed. In the case of interviews taking place in households, these tended to face similar constraints as typical household survey interviews; that is, the interviews were not always conducted in privacy as family members tended to appear at times, and participants tended not to be as focused given other responsibilities. In addition, sometimes finding a surface to place the materials needed for the interview was also a challenge. Interviewers and observers had to manage two separate instruments as well as the audio recorder. This sometimes posed problems, especially when the interviews had to be conducted at the dwelling's entrance. For the most part, however, interviewers were well familiar with techniques to manage the household environment to ensure a good quality interview.

Language of CI

71. Given the global distribution of the pilot countries, from the start it was expected that the CI would need to take place in multiple languages. In preparation for the cognitive tests, the ILO worked with the NSOs to translate and adapt the cognitive protocol and supporting materials to the country's official language(s). In addition, the cognitive training workshops included an initial session where the materials were further reviewed and adapted to include local examples and expressions as relevant, in collaboration with the team established to conduct the CI.
72. In six out of the ten pilot countries, it was established that additional local languages would be needed, given the areas targeted for the study. In those countries, interviewers were required to be fluent in the local target language. Where the local languages had a written form (e.g. Tagalog), agreed translations for the questions and probes were further included in the cognitive protocol. Where the local languages were only spoken (e.g. Oshiwambo), special sessions to agree on key local terms and translations to use were organized as part of the training workshop.
73. Although the project included plans for handling translation and adaptation of the materials, the time allocated to these activities was overall found to be insufficient. Translations had to go through multiple revisions to make sure the language used was appropriate for conducting interviews with the general population in a given country. For example, even when official Arabic, French, Russian or Spanish translations were prepared, there was a need to review the translations and adapt them to national versions of the language (i.e. Tunisian Arabic, Ecuadorian Spanish, etc.). In addition, another round of revisions was needed to ensure the language was appropriate for survey implementation.
74. Crucially important was also the final review and local adaptation of the materials with experienced field interviewers, who contributed common expressions and examples appropriate to the target participants. This final stage was integrated as part of the CI training workshop conducted in each country. A consequence was that materials had to be revised during the workshop and reprinted just prior to the start of the CIs, which were generally scheduled to start just after the completion of the training workshop.
75. Promoting agreement among interviewers in the terms and wordings to be used when direct interpretation was necessary was an additional step not initially foreseen by the project, but introduced as a general procedure after this need became evident in the second country to implement the cognitive tests (Namibia). While in no case did these extra steps prevent the process from proceeding, it would have been preferable to allow additional time for the finalisation of the materials before the CIs.

Duration of the CI

76. The average duration of the CI varied by model tested and participant profile. For the shortest model, M2, pilot countries reported average durations of around 15 minutes. For the longest models, M1 and M4 however, the interviews could last up to 1 hour, particularly with participants who were employed and also producing goods intended mainly for family use.
77. Based on the observation of CI in various pilot countries, it was clear that participants struggled with focus when the interviews lasted longer than 30 to 40 minutes. Longer duration of the interviews in certain scenarios were observed particularly in the first two countries to implement the cognitive tests (Moldova and Namibia). As indicated earlier, the CI protocol was revised to

include a reduced number of probes per question, as well as simpler probes that were more relatable for the participants. These revisions improved the flow of the interview and served to maintain the overall duration of the interviews within acceptable limits. Nevertheless, it was clear that the cognitive tests included more topics than would have been optimal for evaluation in a single round of cognitive testing.

Interview recording and summary

78. As recommended in the CI protocol, most pilot countries used both audio recording and note-taking to document the interviews. Moldova was the only country to rely alone on audio recordings. The use of note-taking and audio recording proved to be crucial in supporting the timely processing and identification of main findings for each completed interview. In the case of Moldova, the use of audio recording alone implied that the summary for each of the interviews completed had to be prepared based on a review of the audio tapes. This increased the work burden of the country CI team leader compared to other pilot countries. Nevertheless, deadlines to submit the cognitive test results to the ILO were not impacted.
79. The procedure established by the CI protocol to record and summarize the main results from each interview worked well overall across countries. In particular, requiring observers to summarize the main results shortly after completing the interview and to discuss the findings with the interviews were crucial to the overall analysis process. During training, it was necessary to provide observers with strategies and short-cuts to improve note-taking. Once observers had practiced with a few mock interviews, they were generally able to adequately capture the relevant information and to produce summary results taking into account the four stages of the question-answer process.
80. An ongoing issue, however, was the limited inclusion of direct quotes to illustrate key findings. There was a general need to instruct CI teams to review the recordings with a view to recover full quotes of statements made by participants during the interview for inclusion in the interview summaries. Despite this, given the limited experience, resources and time available to conduct the CI, the procedure established by the CI protocol to prepare interview summaries was considered appropriate.

B. Analysis, identification and reporting main findings

81. Countries were given flexibility in deciding who would conduct the aggregation of the interview summaries, their analysis and identification of main findings and drafting of the country report. In some cases, selected interviewers were assigned to carry out these tasks. In other cases, it was the CI country leader that undertook this work. The analysis framework template developed to guide the analysis worked well in providing a basic structure to organize the findings by question and identify common themes across the four stages of the question-answer process. While not designed to allow comprehensive textual evaluation nor coding of recurrent themes, the approach was found to be practical and accessible for basic qualitative analysis by persons new to CI methods.
82. A number of issues were nevertheless observed with the analysis and reporting process adopted in the project. Most important was the uneven interpretation and evaluation of findings by different analysts within countries. In particular, differences could be observed in the way that common issues had been interpreted and reported in countries where two staff had been assigned to separately conduct the analysis and write-up for each model tested. This was evident not only in the way the issues were described, but also in the overall assessments of likelihood and impact reported by individual analysts. When these situations were evident, the ILO requested CI team

leaders to review the aggregation of results included in the analysis frameworks, their interpretation and evaluations arrived at with a view to improve consistency in the assessments.

83. Furthermore, as described earlier, prior to conducting the cross-model and cross-country analysis of the results, the ILO reviewed each individual interview summary submitted by countries in the analysis framework files, as well as the main findings and conclusions arrived at. All discrepancies, ambiguous or unsupported findings identified were referred back to the pilot country for further review and clarification. A critical part of this review process was to ensure that all findings were substantiated with direct quotes or descriptions of behaviours and observed interview dynamics; and that interpretation and reporting of patterns identified were consistent across models and countries, to the extent possible.
84. Other problems related to insufficient time allocated to support the analysis of the cognitive results at country level as well as insufficient time to review the cross-country results prior to the start of the field tests. Effectively, only one month was allocated between the end of the cognitive testing phase and the start of the field tests. This meant that the time allocated to revise the model questionnaires and accompanying materials in preparation for the first field test by a pilot country was very compressed. In this regard, the analysis framework files played a crucial role in facilitating the prompt submission of basic CI results by pilot countries to the ILO. This allowed ILO to identify early on the main issues requiring changes in the model questionnaires. Likewise, the staggered implementation of the CI in pilot countries also enabled ILO to begin identifying patterns prior to the full completion of the cognitive tests in all 10 pilot countries.
85. Another important challenge faced by countries was the drafting of the country reports with the main findings from the CI. While the ILO provided an annotated template for the report, NSO staff struggled to allocate time for this activity, as it overlapped with other responsibilities as well as with the start of plans for the field tests. In one case, it was necessary to hire a consultant to support the NSO to complete this stage of reporting.

V. KEY FINDINGS AND REVISIONS INTRODUCED

86. The cognitive tests provided very useful insights about the functioning of the alternative model questions and sequences tested. This section provides a general description of the types of issues identified through the cognitive tests and the type of revisions introduced in the model questionnaires prior to the field tests. Detailed descriptions of the findings and revisions introduced are included in the corresponding reports from this series, accessible in the website of the ILO Department of Statistics.

Main types of findings

87. Overall, the results of the CI were encouraging in that no major differences were observed in how participants across countries understood the overall intention and underlying concepts for key questions needed to identify the employed, unemployed, time-related underemployed and the potential labour force. Many areas of the questionnaires seemed to be well understood and easy to answer across pilot countries. The cognitive tests provided general evidence of consistency of interpretation of underlying concepts across countries for concepts, such as “work for pay or profit”, “job search”, “desire to work”, “availability to work”, and “main intended destination.”

88. At the same time, the tests served to document a number of issues in LFS questionnaire design that require attention to reduce potential problems with the identification of the employed, unemployed, potential labour force or own-use producers of goods due to response errors. For the most part the issues documented through the cognitive tests were consistent across pilot countries and, thus, pointed to issues with the original design of the model questionnaires or to complex concepts inherent to labour force and work statistics, rather than to culture or language specific issues. This included, in particular, problems related to the overall structure and approach taken to establish the boundary between employment and own-use production work, particularly in model M1 which led to some double reporting of the same activity in different parts of the questionnaire.

89. Other issues documented included problems with uneven interpretation of reference periods; problems with recall and reporting of hours worked, especially in the case of activities performed intermittently or simultaneously; some problems with understanding of specific terms and phrases commonly used in LFSs; as well as issues with identification of work performed by participants who help either in family businesses or in paid jobs held by family members.

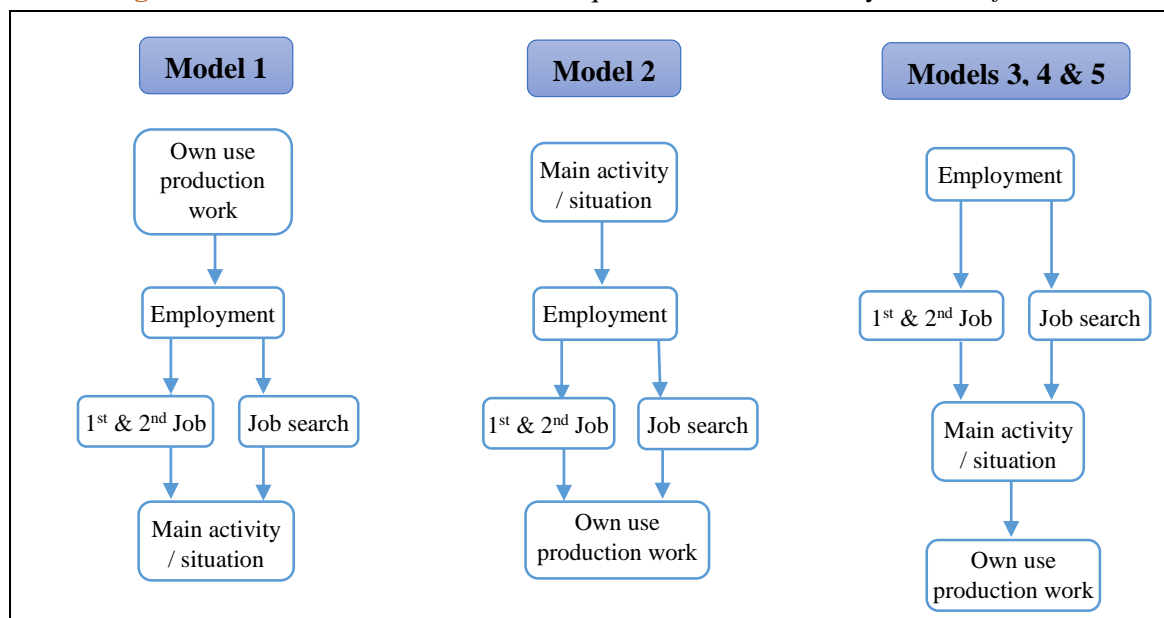
90. The cognitive tests further indicated that the problems identified tended to be more pronounced among participants from rural areas or with lower education levels. This is particularly important given that many of the more complex issues to be captured in a LFS (e.g. work in agriculture, the boundary between employment and own-use production work) are most relevant to persons living in rural areas.

91. While no specific evidence was found regarding differential interpretation of questions by sex, some of the problems observed tended to affect women more compared to men given gendered expectations that result in women often combining paid and unpaid work. Furthermore, some issues with gender sensitivity were documented in particular with the questions on own-use provision of services. When asked about performing “housework” for example, some male participants tended to say “that is women’s work” and show some discomfort when answering. This was observed only in some countries, pointing to culture specific factors that need to be taken into consideration when developing stylized survey questions to capture this topic.

Main types of revisions introduced

92. Based on the findings from the CI tests, a number of revisions were made to the model LFS questionnaires prior to the start of the field testing stage. These included: (a) revisions to the basic structure and order of the questionnaire sections; (b) clarification of reference periods; and (c) revisions to the format and wording questions.
93. One of the more substantial revisions stemming from the CI tests was a change to the basic structure used in model M1. As described in Report I of this series, M1 was designed specifically for use in contexts with high levels of subsistence activity. M1 was designed to start with a section on “own-use production work” followed by a section on “employment” with a view to promote measurement of both activities fully (see Figure 5). Nevertheless, the CI tests indicated that, as originally designed, this approach was not optimal to distinguish between persons employed in agriculture and/or fishing and those producing mainly for own final use (i.e. own-use producers). The structure had resulted in questions on subsistence activities that were too long and complex to understand, leading to double-reporting of the same activity in the first two sections of the questionnaire.³

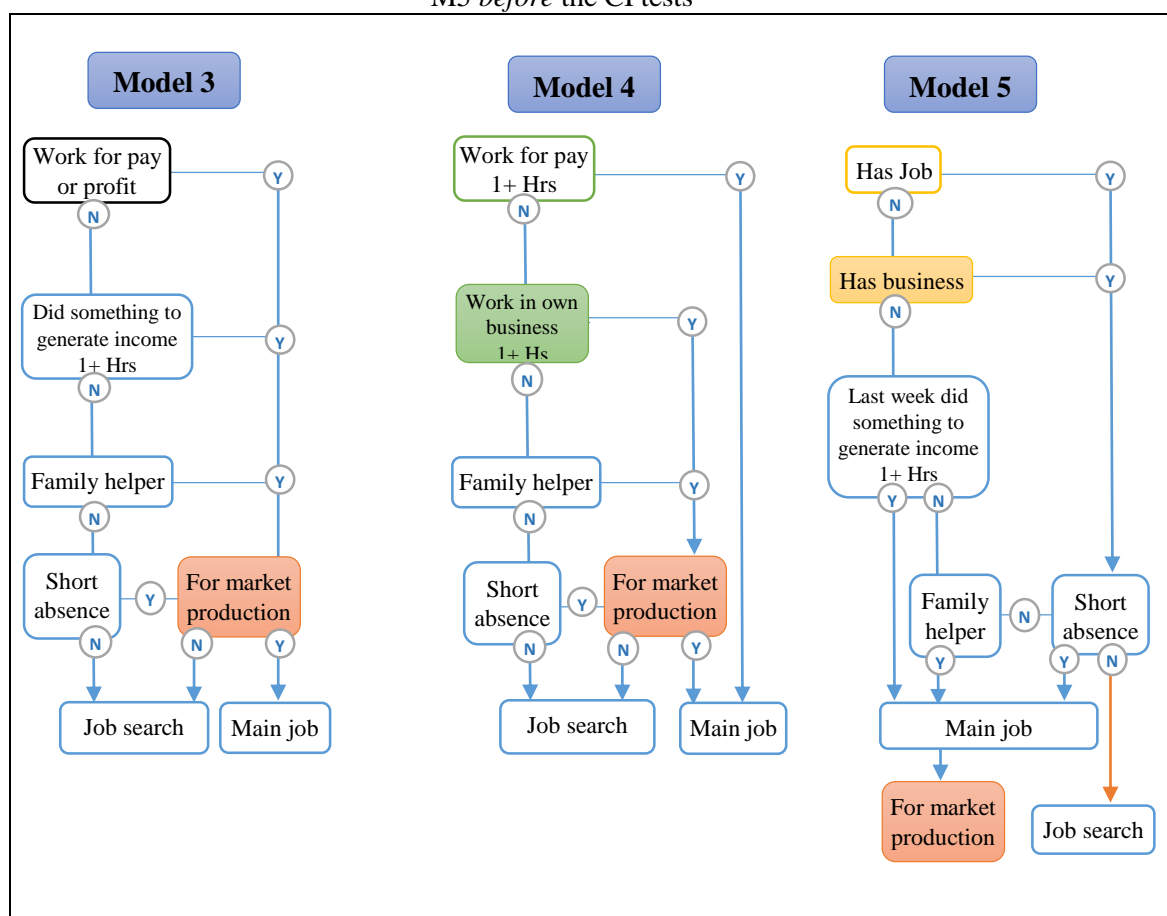
Figure 5. Overall structure and order of questionnaire sections by model *before* CI



94. At the same time, the CI tests indicated that the approach taken in models M3 and M4 worked much better to separately identify the employed (producing mainly to sell) from own-use producers (producing mainly for own use). These models included a dedicated section that asked persons working in agriculture and/or fishing to indicate the “main intended destination of the products” (see Figure 6). Findings from the CI tests indicated that this approach more closely matched how participants thought about their work, first focusing on the activity (farming, raising animals, fishing), and second on the intended use of the products or animals (for sale/for own use).

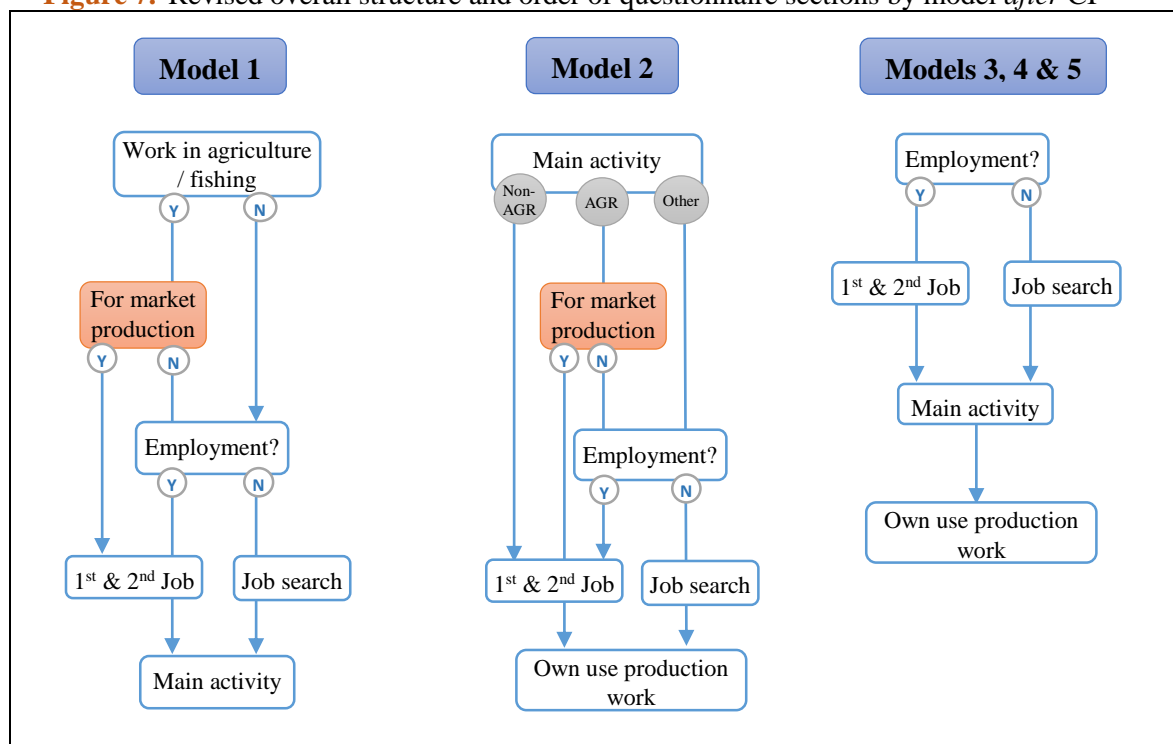
³ For more details, see reports IV on “Employment”, VIII on “Own-use production work” and X on “Work in Agriculture” in this series.

Figure 6. Dedicated section/question on “For market production” included in models M3, M4 and M5 before the CI tests



95. These findings were particularly useful not only in documenting the impact that a structure such as that used in M1 could have on reporting but also on the type of revisions that could be introduced to reduce the potential for response errors. Following the findings from the CI tests, the basic structure of M1 was significantly revised by changing the starting section to capture all work in agriculture and/or fishing, regardless of the intended use of the products, followed by a new dedicated section where persons engaged in agriculture or fishing were asked a series of questions to assess the main intended destination of their production, and on that basis, distinguish between persons to be treated as employed (producing mainly to sell) and those to be treated as own-use producers (producing mainly for own final use). A similar revision was also introduced into M2, with a smaller impact on the basic structure of the M2 approach (see [Figure 7](#)).

Figure 7. Revised overall structure and order of questionnaire sections by model *after* CI



96. A second type of revision introduced across all five model questionnaires was a clarification of the various reference periods used such as “last month”, “last 30 days”, “last week”, “last 7 days” which, as per the CI findings, tended to be interpreted inconsistently by participants. While it was not possible to assess whether this led to actual misreporting of activities, it was decided that simple clarifications could be introduced to all model questionnaires to reduce the observed variability in interpretation. This was accomplished by systematically specifying the start and end of the reference period in the wording of the questions. Table 8 provides examples of the type of changes made to the wording of reference periods in all five model questionnaires. To limit respondent burden, these changes were applied for the most part only to the first question that established the reference period for a given set of questions.

Table 8. Illustration of changes to the wording of reference periods to reduce observed variability in interpretation

Wording of reference periods <i>before</i> CI tests	Wording of reference periods <i>after</i> CI tests
During last week, ...	During last week, that is from [DAY] up to [DAY] last week, ...
During the last 7 days, ...	During the last 7 days, that is from [DAY] last week up to yesterday, ...
In the last 4 weeks, ...	In the last 4 weeks, that is from [DATE] up to last [DAY], ...
In the last 30 days, ...	In the last 30 days, that is from [DATE] up to yesterday, ...
In the last 12 months, ...	In the last 12 months, that is from [MONTH] last year up to last [MONTH], ...

97. Finally, the most common type of revisions introduced based on the CI findings were changes to the wording and format of selected questions. These for the most part involved simplifications to reduce question length or complexity, replacement of specific terms found to be too technical or

abstract, rewording of questions to place emphasis on key terms and inclusion of common expressions or examples as illustration. In some cases, these revisions involved splitting questions to reduce complexity; combining separate questions that were too similar in interpretation; or inverting the order of selected questions to reduce misreporting in closely related activities, such as cooking for children and childcare, or making renovations to the family home and small repairs.

98. Changes to the question format were made only in the case of questions aimed at capturing main activity based on self-perception. In this case, the CI tests made it evident that response options needed to be read out by interviewers to improve interpretation of the question, set the context for respondents to decide on their answer, and facilitate reporting.

VI. REFLECTIONS AND RECOMMENDATIONS

99. In reflecting about the implementation of the CI testing phase of the ILO LFS pilot studies, it is useful to highlight some of the key achievements and challenges identified.

Evidence of overall construct overlap in key criteria

100. Overall, CI phase of the ILO LFS pilot studies proved to be a useful and cost-effective method to assess the design of LFS questionnaires and inform their revision prior to field implementation. The CI tests provided evidence of good overall construct overlap across countries in some of the core criteria needed to classify persons as employed or unemployed, as well as in the new criteria needed to identify potential labour force and own-use producers.
101. While some issues emerged with interpretation of selected key terms, the consistency in findings across countries suggests that respondents generally interpret in a similar way the main intention of the questions used in household surveys to produce some of the core and new headline indicators of the labour force. This was the case for behaviour-type questions, such as “did you do any work for pay?” or “did you do anything to try to find a job?”, as well as for perception-type questions, such as “do you want to work at present?” or “are you available to start working in the next two weeks?”. Furthermore, in the case of perception-type questions, the CI tests revealed consistency in the factors that participants, across contexts, took into account in deciding their answers.

Identification of LFS questionnaire design issues

102. The CI tests further provided a wealth of information on potential sources of response error specific to LFS questionnaire design. Some of the main design issues identified related to the structure of the questionnaires, for example, in the case of the approaches developed to establish the new boundary between employment and own-use production work. This is not surprising given the limited accumulated experience with measuring employment as more narrowly defined and with fully capturing participation in own-use production work in LFS.
103. Other cognitive issues identified related to existing practices that have been commonplace in national LFS design, such as issues with inconsistent understanding of some key terms and phrases or uneven interpretation of commonly used reference periods. Some of these findings, for example, in the case of the questions to identify persons employed, to capture hours worked, or to capture job search, coincide with results reported in CI tests conducted by national statistical offices in high income countries (Bienias 1988; Guinea et.al. 2003; Rothgeb 2007), thus pointing to common design challenges that need to be taken into account across contexts when developing LFS questionnaires.
104. In addition, CI testing of the five alternative models also served to shed light on design issues specific to each approach. For example, in the case of M2, the CI tests made evident that the interpretation of the starting question on “main activity at present” is highly sensitive to the question wording and formulation as well as its mode of application.

Identification of good practices in LFS questionnaire design

105. Beyond making visible the type of issues in LFS questionnaire design that can impact responses, the CI testing phase also provided invaluable information about their root-causes and thus, the types of revisions needed. For example, in the case of agricultural work, the CI results showed that

respondents tend to think first about the activity, and separately about its main intended use. Thus, to address the structural problem of separating own-use production work and employment, the solution pointed to the need for a common questionnaire section where all work in agriculture or fishing is routed and where separate follow up questions are asked to establish the “main intended destination” of the products. This was the approach originally used in models M3 and M4. As described earlier, Models M1 and M2, while retaining their own logic, were revised to use a similar structure for establishing the boundary between employment and own-use production work.

106. Overall the insights learned through the CI tests will be used by the ILO to develop updated practical guidelines on LFS questionnaire design that are aligned with the latest ICLS standards. Nevertheless, the experience of the CI tests also made evident the importance of national adaptation, translation and testing of LFS survey questionnaires to minimize potential sources of response error. Model survey questionnaires and guidelines can serve the important role of providing general guidance on viable approaches to apply internationally agreed concepts and definitions. National adaptation and testing, would remain essential to validate the survey questionnaire in the local context.

Relevance of CI testing methods across contexts

107. The ILO pilot studies experience confirmed the usefulness and relevance of CI testing methods to assess issues of questionnaire design across different contexts. No specific problems with the CI methodology were identified that would question its applicability in the different settings where the tests took place.
108. Nevertheless, the project’s experience coincides with issues reported in the literature on cross-cultural CI (Willis, 2015b). In particular, attempts to use think-aloud and paraphrasing during the CI tests generally increased the stress of participants and were seldom useful in providing insights about the question-answer process. Abstract probes on cognition (i.e. to you what does “X” mean?) tended also to be difficult, especially for participants with lower education levels or from rural areas. More conversational or factual probes tended to work best. This included, for example, probes such as “What did you think about when I said “X”?” “Can you give me examples of ...?” “Can you tell me more about that?”

Interest among NSOs in CI methods for survey development

109. The ILO pilot studies experience made evident the strong interest by NSOs in becoming familiar with CI to inform questionnaire design and data quality assurance. Pilot countries clearly saw the value for improving their survey design and were often surprised to see standard question formulations causing misunderstandings or recall problems.
110. A strong collaboration was established with the countries NSOs during the implementation of the CI tests. Technical and field staff involved in the tests gained sound skills in the CI method as well as a deeper understanding of the new changes introduced in the international statistical standards by the 19th ICLS.

A. Challenges and lessons learned

111. An important challenge for the project was limiting the scope of topics to be evaluated during the CI testing phase. The choice of topics was made based on the need to develop evidence-based guidelines for countries to implement the new ICLS standards in their national LFS. This implied that questions required to measure core headline indicators, such as employment and labour underutilization needed to be evaluated in the tests.
112. The experience of the ILO CI tests, however, indicated that the CI were generally long, lasting over 40 minutes, and at times burdensome for participants. The inclusion of too many topics also meant that only a few probes could be included for any survey question being assessed. This overall limited the depth of qualitative information that could be captured. Likewise, given the scale of the multi-country pilot project only one CI test per country could be planned. Ideally it would have been useful to have a second iteration of the CI test, either to split the topics covered, or to re-evaluate topics for which further evidence was needed.
113. The ILO pilot studies evidenced the usefulness of developing model materials establishing clear instructions for each step of the CI process. The CI protocol developed was particularly crucial to ensure a consistent application of the CI across contexts. However, the project did not envisage an initial testing of the CI protocol prior to its roll-out. During the first implementation, it became clear that some of the probes used were not adequate for some participants. As described earlier, the CI protocol had to be revised to include simpler probes and to shorten its duration prior to its use in subsequent country CI tests.
114. Given the focus of the project in developing guidance on LFS survey design for use by NSOs, it was imperative to work with NSOs in the implementation of the CI tests. Participating NSOs had strong interest in the CI method but no prior experience. Thus, capacity building of NSO staff in CI techniques was critical to the project's success. In most pilot countries, interviewers were mostly either technical staff from the NSOs, or experienced LFS supervisors or interviewers. Each profile had its pros and cons. Technical staff usually understood the substantive aim but struggled sometimes with interview flow. LFS field personnel sometimes could adapt well to semi-structured interviewing, but in many cases, expected to have a structured set of probes to ask. The challenge for the project was developing a CI protocol that provided sufficient structure to elicit the necessary insights from participants without requiring strong improvisational skills by interviewers. Regardless of the profile of the interviewers, practice was critical to ensure the interviewer could establish a good flow of communication with the respondents.
115. The project envisaged handling translations into multiple languages as well as country adaptation. Nevertheless, the translation work took more time and resources than envisaged. Materials were translated either directly by ILO technical staff or through consultants and then checked by the technical staff of the NSO in the country. Further revisions and adaptations were also introduced after consultation with NSO field staff with experience in implementing the national LFS. The experience from the project made it clear that the questions needed to be tailored at each country language context instead of being literally translated. Often there was a different sentence structure or phrase that would capture best the intended aim and meaning of the questions. While this level of translation and adaptation can represent an important challenge for cross-country surveys, it was clearly necessary to ensure the comparability of the model instruments. This points to limitations in processes of 'input harmonisation' across countries.
116. Beyond translation and adaptation of the materials, during the conduct of the CI tests it also became clear that there would be a need to handle direct interpretation to local languages in some contexts.

This had not been initially envisaged by the project and had to be addressed during the training workshops with the country CI teams. The approach established involved running special sessions focused on identifying equivalent phrases in the local language and having all interviewers to discuss possible alternatives and agree on a single interpretation to be used during the CI.

117. The process established to record, summarize and identify the key findings from the CI tests worked very well. However, an ongoing challenge was to obtain clear illustrative evidence from countries and consistent interpretation of the findings by country analysts assigned to this task. Several countries did not include direct quotes to illustrate the main findings described. Usually they provided paraphrase or the writers' evaluation. While the project established a simplified analysis approach and provided training on the basics of CI analysis and reporting, it was clearly not sufficient. Based on the project's experience, there was a need for more substantial on-going support during the analysis and write-up phase.
118. A final challenge was the limited time allocated between the end of the CI phase and the start of the field tests. Effectively, only one month was left for turning around revised model questionnaires. Sufficient time needs to be allocated after the CI tests for review and evaluation before any further testing is defined.

B. Recommendations on CI testing for LFS design

119. Being a new methodology, introducing the CI method in NSO practices will require extensive training and capacity building as well as the development of new and different procedures than those usually applied for survey implementation.
120. Particularly important will be introducing procedures that facilitate the recruitment and selection of participants through a variety of non-sampling methods. Reliance on information from previous surveys is often not sufficient. Rather a combination of selection methods appears to work best. This includes use of network and snow-ball sampling techniques as well as outreach to various groups as per the requirements of the studies.
121. While challenging for many NSOs, the use of incentives to facilitate and promote participation in CI tests may need to be considered. This is particularly important if participants are required to travel or allocate a significant amount of time to participate in the CI test.
122. CI testing also requires procedures that allow the flexibility to plan and conduct small-scale iterative tests. This includes availability of neutral facilities for conducting the interviews not only in the central office but also in regional or local offices; ability to develop and print materials on a small scale within short timelines; and working methods that facilitate close communication with the survey developers and integration of CI results into the survey design process.
123. Another essential component is the development of CI protocols and reporting templates. NSOs seeking to establish the use of CI methods should develop clear and simple CI guidelines and templates, based on previous experiences, but that can be readily adapted as needed. There is a need to test different probing strategies and identify those that work best within the national context. Probes that appear to work well with urban and more educated groups may not necessarily work with persons with lower levels of education or living in rural communities. The CI protocols should highlight these considerations and provide appropriate guidance.

124. The skills of the interviewers are essential to ensure the usefulness of the CI method. NSOs interested in establishing CI teams will need to identify a few staff with both ability to grasp conceptual issues and to improvise during interviews. Practical training and ongoing experience is necessary to develop and strengthen those essential skills. Likewise, important is the development of skills in qualitative analysis and establishment of procedures for consistent interpretation of results. Observation of CI interviews by the survey developers is also particularly important to inform the interpretation of the results and to the overall quality of the CI process.

VII. CONCLUDING REMARKS

125. Overall, over 400 individual interviews were completed during the CI testing phase of the ILO LFS pilot studies project. The CI phase was successful in achieving the objectives of identifying issues specific to LFS questionnaire design that can impact responses. They also provided evidence of good overall construct overlap across settings in the core operational criteria being used to identify persons employed, unemployed, and the new groups of potential labour force and own-use producers. Furthermore, the CI tests yielded useful insights on approaches to introduce the needed changes to align national LFSs, following different overall structures, with the new international statistical standards on work, employment and labour underutilization.
126. The CI method was found to work generally well across context, albeit some challenges faced, in particular with the selection and recruitment of participants by NSOs and initial difficulties with the use of paraphrasing and more abstract cognitive probes. CI techniques have a strong potential of being incorporated by NSOs in developing countries as part of their survey development practices. More developments in this area are needed to facilitate its adoption as part of the regular production of official statistics.

VIII. REFERENCES

- Beatty, P. et al. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2), pp. 287-311. doi:<https://doi.org/10.1093/poq/nfm006>
- Bienias, J. (1988). *Preliminary laboratory research on current Population Survey Questionnaires with employed persons*. Bureau of Labor Statistics, Washington DC. Retrieved 2017, from https://ww2.amstat.org/sections/srms/Proceedings/papers/1988_032.pdf
- Cohanny, S. et al. (1994). Employment and Earnings. *Revisions in the Current Population Survey Effective January 1994*. Retrieved from <https://www.bls.gov/cps/revisions1994.pdf>
- Collins, D. (2003). Quality of Life Research. *Pretesting survey instruments: an overview of cognitive methods*, 12. Kluwer Academic Publishers. Retrieved from <https://doi.org/10.1023/A:1023254226592>
- Fasching, M. et al. (2014). *Cognitive testing of proposed LFS-Questions measuring the employment status*. Statistics Austria, Vienna.
- Fitzgerald, R. et al. (2009). *Testing for equivalence using cross-national cognitive interviewing*. Working paper series, Centre for Comparative Social Surveys, City University, London. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.3862&rep=rep1&type=pdf>
- Guinea, D. et al. (2003). *How people answer Labour Force Survey questions about economic inactivity*. Technical report, Office for National Statistics, Quality and Risk Management Division, London. Retrieved from www.ons.gov.uk/ons/
- ILO. (2013). Report of the Conference. *19th International Conference of Labour Statisticians*. Geneva: ILO.
- ILO. (2018). *LFS Pilot Study Programme*. (ILO, Producer) Retrieved from ILO: http://www.ilo.org/stat/Areasofwork/Standards/lfs/WCMS_484981/lang--en/index.htm
- Miller et al. (2011). SAGE Journals. *Design and analysis of cognitive interviews for comparative multinational testing*. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1177/1525822X11414802>
- Presser, S. et al. (2004). *Public Opinion Quarterly. Methods for testing and evaluating survey questions*, 68. Retrieved from <https://doi.org/10.1093/poq/nfh008>
- Rothgeb, J. (2007). *ACS Labor Force Questions: Results from Cognitive Interview Testing*. Research Report Series (Survey Methodology #2007-16), U.S. Census Bureau, Statistical Research Division, Washington DC. Retrieved from <https://www.census.gov/srd/www/abstract/rsm2007-16.html>
- Statistical Policy Directive N. 2 Addendum*. (n.d.). Retrieved November 2017, from The White House President Barack Obama : https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/directive2/final_addendum_to_stat_policy_dir_2.pdf
- Tourangeau, R. (1984). Cognitive science and survey methods: a cognitive perspective. In T. Jabine, & J. Tanur (Ed.), *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines* (pp. 73-100). Washington D.C.: National Academy Press.

- Willis, G. (1999). *Cognitive Interviewing. A "How to" Guide*. Research Triangle Institute. Retrieved from <http://www.chime.ucla.edu/publications/docs/cognitive%20interviewing%20guide.pdf>
- Willis, G. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks: SAGE publications.
- Willis, G. (2015a). Oxford University Press. *Analysis of the cognitive interview in questionnaire design*. Retrieved from <https://global.oup.com/academic/product/analysis-of-the-cognitive-interview-in-questionnaire-design-9780199957750?cc=ch&lang=en&>
- Willis, G. (2015b). Public Opinion Quarterly. *The practice of cross-cultural cognitive interviewing*, 79. Retrieved from <https://doi.org/10.1093/poq/nfu092>