



**Are you thinking what I'm meaning?  
Experiences from cognitive interviews in multi-country survey tests**

Elisa Munoz F. Benes\*

International Labour Organization, Geneva, Switzerland – [benes@ilo.org](mailto:benes@ilo.org)

Kieran Walsh

International Labour Organization, Geneva, Switzerland – [walshk@ilo.org](mailto:walshk@ilo.org)

**Abstract**

In recent decades, cognitive interviewing has gained recognition as a practical technique to identify and reduce potential sources of response error in surveys. Extensions to cross-national surveys are further expanding the application of cognitive interviewing to assess instrument equivalence and thus assure the comparability of survey results across countries. While the technique has become part of the standard validation tools used during survey development by many organizations, in official statistics, this practice is for the most part restricted to a few national statistical offices (NSO) in more developed countries. Nevertheless, strong interest exists among NSO in less developed countries. This paper describes the experience of the ILO in cross-national cognitive interviewing to support the development of model survey questionnaires meant to serve as guidance to countries in applying the latest international standards on the measurement of work, employment and labour underutilization. The findings attest to the usefulness of cognitive interviewing in reducing potential sources of error during survey development that may impact the overall validity and comparability of the results. They also shed light on the applicability of different cognitive interviewing techniques in different cultural, linguistic and socioeconomic settings. Particular issues requiring further attention to support the adoption of cognitive interviewing by national statistical offices are also highlighted.

**Keywords:** cognitive interviewing; labour force surveys; labour statistics; ICLS standards.

**1. Introduction**

Testing of survey questionnaires has long been acknowledged as an important aspect of data quality assurance. However, only in recent decades have survey questions and instruments been evaluated in a more systematic manner using a variety of qualitative approaches. Among these, cognitive interviewing (CI) has gained particular recognition as a practical technique to identify and address potential sources of response error due to questionnaire design. By focusing on how respondents understand, recall and answer survey questions, CI can help not only minimize response errors but also serve as a tool to assess the overall validity of the data collected (Collins 2003).

Although less widespread, CI is also gaining popularity in cross-national surveys. In this context, a crucial concern has been to assess instrument equivalence as a way to improve the comparability of the results (Fitzgerald et al. 2009) Main areas of focus in CI for cross-country surveys include identifying errors due to translation issues; providing evidence of “construct overlap”, that is the extent to which different linguistic and cultural groups understand key concepts the same way; and more generally, assessing potential differences in interpretation and response across different groups (Miller et al. 2011).

This paper discusses the experience of the International Labour Organization (ILO) in applying CI to test and develop model Labour Force Survey (LFS) questionnaires for use in different contexts. It describes the process introducing cognitive interviewing methods in countries from different regions of the world and illustrates the usefulness of the technique to identify potential sources of response error and assess the equivalence of the instruments across contexts. To place this discussion in context, the paper begins with a brief overview of the survey development work being carried out by the ILO. The paper concludes by highlighting some of the key lessons learned in conducting cross-country CI.



## 2. ILO labour force survey pilot studies

As the United Nations Agency responsible for labour issues, the ILO has as one of its core mandates to promote the setting of international standards on labour statistics. For this, every 5-years, since 1923, the ILO holds the International Conference of Labour Statisticians (ICLS), where representatives from Governments, Workers and Employers discuss and adopt statistical standards on a variety of labour-related topics. A main aim of these standards is to provide guidance to countries on how to establish or improve their labour data collection programmes to address national needs and at the same time facilitate cross-country comparisons.

In 2013, the 19th ICLS adopted new standards on statistics of work, employment and labour underutilization (ILO 2013).<sup>1</sup> These standards introduced a number of changes that will impact the way countries conduct household surveys to collect statistics on key headline indicators on employment and unemployment. In addition, the standards provide definitions for new indicators of labour underutilization and new forms of work such as own-use production of goods and services that will need to be integrated into national survey programmes. To support countries during the transition, the 19<sup>th</sup> ICLS further called on the ILO to conduct pilot tests and develop guidance, including model labour force survey materials aligned with the latest standards.

As follow-up, in 2015 the ILO launched a global pilot labour force survey (LFS) programme<sup>2</sup> focused on developing and testing survey questions aligned with the new standards with a view to measure:

- Employment defined more narrowly as work for pay or profit.
- Labour underutilization, including unemployment, time-related underemployment and the newly introduced potential labour force.
- Participation in own-use production work, including production of goods and services.
- Main activity as self-declared.

Based on a review of existing national LFS practices, the ILO developed five alternative questionnaires for testing. While all questionnaires were designed to measure the same topics, they included differences in question wording and order, the way response options were meant to be implemented and, in some cases, the structure and order of different sections. Overall, the biggest differences related to the section structure and wording of questions to separately identify employed persons and own-use producers of goods. For topics such as time-related underemployment, unemployment and the potential labour force the differences focused only on selected new elements introduced by the standards, such as the measurement of “desire to work”.

A main objective of the pilot was to assess how well the alternative questionnaires performed in different languages, cultures and socioeconomic contexts, and whether they were equivalent in terms of classifying persons as employed, underemployed, unemployed, potential labour force and/or own-use producers. At this early stage of the testing it was decided not to focus on whether the questionnaires yielded comparable estimates or on assessing the potential impact of the changes on key labour market indicators. Rather emphasis was placed on questionnaire design, adaptation and equivalence issues.

The project developed a methodology that combined both CI and small-scale field tests using a split sample design. Each of the 10 pilot countries was assigned two questionnaires for testing. One was chosen to be closest to their current practice while the second was to serve as contrast and to maximise testing across regions as deemed relevant (see table 1). To support consistent implementation, the ILO developed a full set of testing materials and instructions. All materials were initially developed in

<sup>1</sup> See: [http://ilo.org/global/statistics-and-databases/meetings-and-events/international-conference-of-labour-statisticians/19/WCMS\\_230304/lang--en/index.htm](http://ilo.org/global/statistics-and-databases/meetings-and-events/international-conference-of-labour-statisticians/19/WCMS_230304/lang--en/index.htm)

<sup>2</sup> See : <http://www.ilo.org/stat/Areaofwork/Standards/lfs/lang--en/index.htm>



English and translated into French, Russian and Spanish by the ILO. Countries further translated the questionnaires to local languages or adapted the translations to include local expressions, as relevant.

Table 1. Model questionnaires tested by approach and country assignment

Model	Approach	Countries
M1	Work in agriculture start	Cameroon, Ivory Coast, Namibia
M2	Main activity start	Kyrgyzstan, Philippines, Tunisia
M3	Work for pay or profit	Ecuador, Ivory Coast, Moldova, Kyrgyzstan Peru, Philippines, Tunisia, Vietnam
M4	Work by employment type	Namibia, Peru, Vietnam
M5	Has Job or business start	Cameroon, Ecuador, Moldova

### 3. ILO Cognitive interviewing methodology

The 10 NSOs had all an established national LFS, and thus strong subject matter experience. However, none had prior experience with using CI for survey development. The ILO provided capacity building on CI throughout the process. This included a 3-day Training of Trainers delivered by experts in CI from the Office of National Statistics of the United Kingdom (UK ONS) and a hands-on 5 day national training for the team conducting the cognitive interviews. In addition, the ILO participated as observer in the initial cognitive interviews and team debriefings.

With the support of the UK ONS, the ILO developed a CI protocol that specified the questions to be tested; issues to be assessed; cognitive probes and techniques to be used in a semi-structured interview; team composition; target profiles and number of participants; recording method, and process of analysis. Templates were provided to record and summarize each cognitive interview, and to consolidate, analyse and report the results. Table 2 describes key features of the cognitive interviews implemented by the pilot countries.

Table 2. Characteristics of the cognitive interviews implemented by country.

Country	Venue	Language	Team size	Recording	Selection method	Completed interviews		
						Total	Employed	Not emp.
CMR	Household	French, Eton	9	Audio & Notes	Previous survey	41	33	8
ECU	NSO office & household	Spanish	8	Audio & Notes	Previous LFS	31	21	10
CIV	NSO office & household	French, local languages	9	Audio & Notes	Previous survey	40	24	16
KGZ	NSO office & household	Kyrgyz, Russian	4	Audio & Notes	Previous survey	40	28	12
MDA	NSO office	Romanian, Russian	4	Audio	Snowball sampling	38	28	10
NAM	Household	Oshiwambo, English	6	Audio & Notes	Radio Ad & targeted area	40	29	11
PER	Household	Spanish	5	Audio & Notes	Previous LFS	40	31	9
PHL	Hotel meeting room & hhold.	Cebuano, Tagalog	12	Audio & Notes	Snowball sampling	30	24	6
TUN	Recording room	Tunisian Arabic	5	Audio & Notes	Snowball sampling	40	22	18
VNM	NSO office	Vietnamese	-	Audio & Notes	Snowball sampling	40	32	8

Global findings were identified by the ILO through a review of the completed analysis templates that listed the results from each interview including the verbatim answer provided by the respondent and the



findings from the cognitive probes for each stage of the question-answer process (comprehension, recall, judgement and response) as well as the analysis reports submitted by countries. The global findings were further discussed and validated with the 10 pilot countries in a workshop held once country implementation of the pilot studies had been completed in November 2016.

#### 4. Main findings

Overall, the results of the cognitive interviews were encouraging in that no major differences were observed in how respondents 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. CI provided evidence of construct overlap across countries regarding underlying concepts such as “work for pay or profit”, “job search”, “availability” and “desire to work”. Likewise, it showed that the 5 questionnaires, while using different structures, question order and wording, were functionally equivalent in capturing diverse employment activities as well as participation in own-use production work and labour underutilization.

While problems were observed with the use of specific terms or phrases, interpretation of reference periods, and identification of particular groups, such as contributing family workers, these issues were not specific to a single country. Rather, the issues observed pointed to problems with translation or with the formulation of questions and the structure of sections in the source questionnaires. Without CI, many of these issues would not have come to light as they related to approaches generally considered as common practice in labour force surveys. Below selected examples are presented to illustrate the usefulness of CI in identifying potential sources of response error and assessing the equivalence of the different questionnaires and across contexts.

##### *Questions to identify employed persons*

All five model questionnaire used a different question sequence to identify the employed. Nevertheless, all included key terms such as “pay”, “profit”, “business”, “in kind”, “at least 1 hour” which are crucial elements of the employment definition, as well as recovery questions for small or casual jobs and for contributing family workers. CI showed that across countries, recovery questions were necessary to identify all employed, and in particular to capture contributing family workers. At the same time, consistent problems were identified with the interpretation of key terms. Payment “for profit” was often not understood or misunderstood, for example, as “god given” in Tunisia or as “government benefits” in Peru; “business” was generally taken to mean registered businesses with a fixed location; “in kind” was interpreted often as having negative connotations, including as referring to “illegal activities” or to payment with “personal services.”

In the case of the phrase “for at least 1 hour” (which is meant to include all types of jobs, even those done for 1 hour), several countries reported respondents that misunderstood this as referring to “additional work”, “overtime”, or “secondary activities”. For example, in Vietnam one respondent said: “*made me think that the question refers to a second job, not the main job*”. Similar cases were observed in Ecuador and Peru. In addition, Cameroon and Moldova indicated that the inclusion of this phrase led some respondents to consider only activities with low numbers of hours.

##### *Questions to capture job search*

Across the 10 pilot countries, the findings showed a clear and consistent understanding of the initial questions on “job search in the last [4 weeks/30 days/month]” as referring to doing something to find work to generate income including paid employment or self-employment. In Ivory Coast, for example, when asked to paraphrase the question a participant replied “*Whether I tried to find a paid job or trading business.*” In Peru, when asked what type of work the question referred to a participant indicated “*any type of business, it can be for half a day, for a few hours, two or three times a week.*”



Further probing confirmed that respondents who replied positively indeed referred to having taken various steps to find a job. As illustration, in the Kyrgyz Republic one participant noted *“I have been reading newspapers, looking for job advertisements, googled through internet, and asked my parents for assistance in my job search”*; in Tunisia another participant reported *“I applied directly to shops selling household appliances, visited the governorate and municipality offices, and met with a social worker to discuss my case;”* in Peru a participant seeking to start her own business said *“I thought about a business I wanted to start, looked for ideas, planned how it could happen, looked for a place.”* The findings confirmed that respondents not only consider formal methods or job search, but rather a variety that include traditional, modern and informal channels. They also indicated that methods not recognized in the international standards as “active” steps such as “reading newspapers” were taken into account when replying to these questions.

In the case of the reference period “the last [4 weeks/month/30 days],” all pilot countries reported observing inconsistent interpretations. Evidence from Tunisia, showed that the “last (calendar) month” tended to be interpreted as from the “1<sup>st</sup> to 30<sup>th</sup> of the current month [September]”, and in one case, as the period starting on the “1<sup>st</sup> of the previous month [August] until the end of the current month [September]”. Likewise problematic were the interpretations of the “last 4 weeks” and “last 30 days.” Cameroon reported interpretations ranging from “from today to the same day four weeks ago” to “in the last (calendar) month.” Yet in Namibia, some respondents did not find the period to be specific enough.

#### *Question to capture desire to work*

Being a new criterion introduced by the 19th ICLS, a main focus was to assess how respondents understood a question on “desire to work”, whether they had a formed opinion, and how they decided on their answer. Across countries, replies indicated that respondents generally understood the question as referring to their “wish to work.” The question was found to be easy and respondents had a formed opinion. The answers given indicated that respondents consider their financial needs or interest in generating an income as the primary factor for wanting to work. A respondent from the Philippines for example stated *“earning income is my primary reason”*. Nevertheless, other factors were also taken into consideration. Among younger respondents, gaining experience or using their education was also reported. Gaining financial independence from relatives was likewise mentioned, a respondent in Vietnam indicated that she wanted to work because *“right now I depend on my husband.”* Yet others mentioned avoiding boredom or idleness as a reason for wanting to work. A woman in Ivory Coast for example noted *“I want to work because I get bored. Retirement is not easy. I would like to raise livestock.”* Furthermore, no evidence was found of respondents saying “No” to this question because of a perceived lack of work opportunities or specific working conditions in their context. Rather, factors considered for not wanting to work referred to personal circumstances such as “needing rest”, “being retired”, “being a student”, “being aged”, “lack of strength.” These findings confirmed that a question on desire to work was relevant across contexts and population groups, and that respondents consistently interpreted the question as intended.

#### *Question to capture availability to work*

The new standards retained the criterion of “availability to work” as part of the identification of the unemployed, however, the reference period was expanded to cover “the last week and the two weeks after”. A single question that included the alternative reference periods as separate response options was developed for testing. The findings from CI revealed that the question as formulated was rather complex to apply and, in turn, to respond. In Ecuador, for example a participant indicated *“The question is clear, however, you are making me think backward and forward in time.”* Another participant in the Philippines similarly noted *“The question is easy, but a calendar is needed for easy counting.”* The complexity observed appeared not to be related to the intended meaning of the question, but rather to the recall process required by the two reference periods, that were often read without a pause in between. In some cases, this caused confusion and the question had to be repeated by the interviewer.



CI further showed that respondents generally interpreted the question as intended; that is as having time or being ready to start working. In Vietnam, for example a participant replied “No, I am not available now. I want to find part-time work. The start time will depend on when I can combine it with my studies”. Another respondent in Tunisia indicated that she was not available last week because she had family responsibilities, but she would be available in the next two weeks. Similarly, a respondent in Moldova who answered “No” when probed indicated that the question was asking about “being available to go to work” and that “[she] could start working only in September as [her] children were [at the moment] on vacation and [she] had to take care of them.” The findings further indicated that while respondents thought about specific types of jobs when answering, their replies were not influenced by the likelihood of finding such work in their local area, but rather focused on their immediate circumstances affecting their time availability.

Findings like those illustrated above shed light on potential elements of the questionnaires that could lead to potential sources of error, whether due to problems of comprehension, recall, judgement or sensitivity. These issues were addressed and revised versions of the questionnaires were prepared in advance of the field tests. Owing to the complexity of the cross-country pilot study assessment, it was not possible to conduct a second iteration of the cognitive interviews.

## 5. Conclusions

A common concern in cross-country CI is the extent to which the existing methods –which have predominantly been developed in the USA and UK, apply to other settings. For survey developers it is instrumental that any differences documented during the testing phase reflect actual variations in the way questions are interpreted or answered across countries rather than differences in how the testing protocol worked in each setting (Willis 2015). This is a crucial issue not only in terms of identifying potential sources of error, but also to establish questionnaire equivalence and ultimately, assess the comparability of the results across countries.

The ILO pilot studies did not identify issues with the CI methods that would question its applicability in the different settings where the tests took place. The findings served to identify potential sources of response error that are particularly relevant for labour force survey design. In addition, some of the challenges encountered during both the planning and conduct of the cognitive tests highlighted a number of lessons learned to take into account when planning CI for cross-country surveys. These include the importance of translation and country adaptation, hands-on training, development of a simplified but comprehensive structured or semi-structured cognitive guide, support in identifying practical participant recruitment and selection methods, and practical tools to support each step in the analysis of the results.

## References

- Collins D. (2003). *Pretesting survey instruments: an overview of cognitive methods*. Quality of Life Research.
- Fitzgerald R., Widdop S., Gray M. and Collins D. (2009). *Testing for equivalence using cross-national CI*, Working Paper Series, no. 01. Center for Comparative Social Surveys.
- International Labour Organization (2013). *19th International Conference of Labour Statisticians (Geneva, 2-11 October 2013), Report of the Conference*; International Labour Organization
- Willis, G. (2015). *The Practice of Cross-Cultural CI*, Public Opinion Quarterly, Vol 79.
- Miller K., Fitzgerald R., Padilla J., Willson S., Widdop S., Caspar R., Dimov M., Gray M., Nunes C., Prüfer P., Schöbi N., and Schoua-Glusberg (2011). *Design and Analysis of Cognitive Interviews for Comparative Multinational Testing*, Field Methods. Sage.