



Stockholm
University

Data Collection Challenges

Gelaye Worku

Stockholm University, Department of Statistics

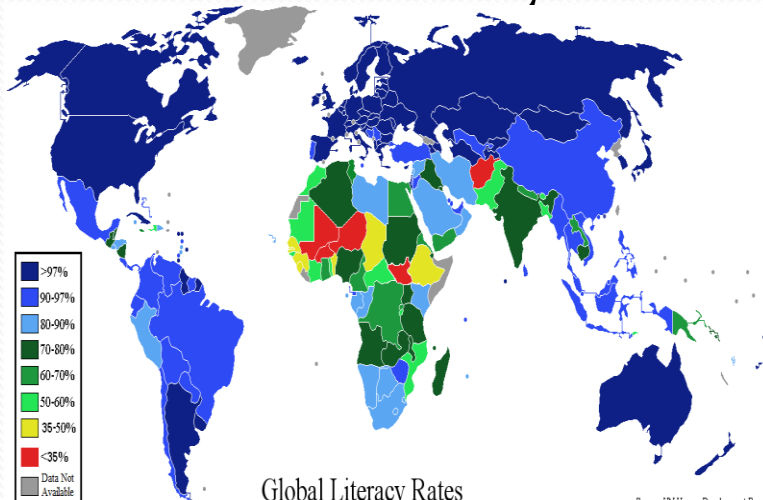
Lars Lyberg

Stockholm University, Department of Statistics

Presented at the International Workshop on Comparative Survey Design and Implementation(CSDI). Bethesda, March 27, 2014

Common data collection challenges

- Lack of central coordination in many 3M
- Lack of QA and QC in many 3M
- Unwillingness to participate
- Hard-to-reach respondents
- Presence of others during interview
- Differences in literacy levels



Most challenging for 3M surveys

- Rate of response between countries.
- Degree of Questions sensitivity
- Choice of data collection mode
- Geography
- Languages
- Cultures
- Religions
- Political
- National financial and methodological resource difference

Most challenging for 3M surveys (Cont'd)

- Translation of all survey materials, not just questions
- Adherence to standards and how to handle deviations
- When to allow flexibility
- Conflict between national and comparative interests
- How to conduct local adaptation
- Lack of skills at local data collection organizations
- Data falsification

Some examples of QA/QC activities and some problems

- **European Social Survey (ESS)**
 - A strong infrastructure with Central Coordinating Team (CCT), Advisory groups, lots of methodology work and a spirit of continuous improvement
 - Setting a minimum response rate of 70% with a maximum noncontact rate of 3% and the difficulties reaching those goals
 - Questionnaire is designed using inputs from the national coordinators.
 - 5-step translation process and Special Quality Packaging (SQP) program is used to check translated questions' reliability and validity.
 - The continuing discussion about single versus mixed mode
 - Differences in frame coverage between countries

Some example(Cont'd)

- **International Social Survey Program (ISSP)**
 - ISSP Methodological Group to handle major survey quality problem
 - Mixed mode
 - Adherence to standards is a problem in
 - Coding as well as weighting to adjust for non-response,
 - Pre-testing translated questionnaire
 - Not including all the core ISSP items in the questionnaire

Some example(Cont'd)

- **Program for International Assessment of Adult Competencies (PIAAC)**
 - Guided by a Board of Participating countries(BPC), supported by a Technical Advisor Group(TAG), and the field tests were monitored by a National Quality Control Monitor (NQCM)
 - Modern translation procedure is used.
 - Minimum response rate requirement resulting in the same kinds of discussions as in the ESS
 - Multiple modes of data collection
 - Data falsification
 - Some countries facing problem in reaching the data collection requirements

Some example(Cont'd)

- **Programme for International Student Assessment (PISA)**
 - No real central team
 - Translation
 - Double translation and back translation
 - Lack of adherence to some of the specifications, especially the sampling standards and the frame standards
 - Data falsification
 - Politically sensitive (national pride)

PISA-The Case of Sweden

- Sweden has dropped considerably in the ranking causing almost a national trauma
- The PISA results have been discussed in Swedish media every day since the beginning of November
- Media are actually running the political agenda when it comes to education and school issues
- These are now the major issues in the upcoming elections, much more important than the economy, taxes, employment etc.
- There is no room for discussing the considerable criticism that exists on how PISA is conducted and analyzed

Some example(Cont'd)

- **World Values Survey (WVS)**
 - Wide range of response rates and sample sizes
 - Response rates vary from 24% to 100% (!) and sample sizes from 300 till 3050.
 - A lot of flexible is allowed vis-à-vis the source questionnaire, where countries can add, delete or rephrase questions
 - No standard translation method is required
 - Single mode is standard but in practice multi-mode is possible
 - Fabricating data

Attention required

- Translation

- In PISA 2006, many multiple choice items were not functioning in equivalent ways in the Asian countries, compared to the test versions in Western languages (Grisa, De Jong, Gebhardt, Berezner, Halleux-Monseur, 2007)
- Variables look similar but have different meanings and variables with the same information measure different things,
 - eg. Social security variables in PIAAC, in WVS 2007 Ethiopia , Mali and Ghana homosexuality is understood as gay men only.
- Some words do not exist in some languages
 - eg. In 2007 WVS the concept of “Euthanasia” has in some cases never been heard of, or it is very hard to understand.

Attention required(Cont'd)

- Mixed mode
 - In PIAAC using multiple modes has caused a problem maintaining the integrity of the concept that should be captured by the variable, but still PIAAC use mixed mode. Typical trade-off issue.
 - ESS uses single mode but has already in 2003 initiated ongoing and considerable research on the effects of mixed mode
 - Martin and Lynn(2011) indicate that there is a good reason to be careful when comparing data from different modes
- Wide range of response rates between countries
 - ESS and PIAAC set a standard unlike for instance WVS
 - Couper and de Leeuw(2003) mentioned that a difference of more than 30% between countries in a 3M survey should give cause for concern that non-response bias may differ across these countries.

Attention required(Cont'd)

- Data falsification
 - In PIAAC in thirteen countries discovered instance of falsification involving one to 22 interviewers (OECD ,2013)
 - In PISA 2009 studies show that out of 70 participating countries, 10 countries have questionable data. In 3 of these the investigation clearly showed that the data had been fabricated (Blasius and Thiesses soon to be published article “Should we trust Survey data? Response simplification and Data Fabrication in the PISA School Principals’ Survey”).
 - Blasius and Thiessen (2012) have found fabricated data in WVS 2005-2008
 - Note that only PIAAC does the QA/QC to find data falsification
- Lack of documentation
 - Even though WVS covers 90% of the world population in their survey, they run short in the documentation department.
 - Eleven participating countries fail to provide documentation for the 2011 ISSP.
 - Documentation would have helped when discussing the lack of trust in the 2009 PISA result.

What do we need to do?

- The Swedish PISA discussions show that it is necessary to strive for a basic quality achieved by a set of enforced standards
- As things are now the basic recommendation from Harvey Goldstein, Sture Holm and others is: Do not reform education based on PISA results only
- We must convince many 3M managers that design, implementation, and analytical deficiencies can have serious effects and that basic QA/QC and a proper process documentation must be in place
- Some error sources such as translation are seriously underrated
- Every serious 3M should have central authorities like ESS and PIAAC
- The communication with users and media must be improved. We must realize that international rankings of different kinds, be they student achievement or happiness, are extremely interesting to media
- We believe that we need to establish a vigorous training program promoting our guidelines among those 3M in need
- How about a CSDI task force?



gelaworku@gmail.com