Overview of Census Evaluation Methods

United Nations Statistics Division

United Nations Workshop on Evaluation and Analysis of Census Data,
1-12 December 2014, Nay Pyi Taw , Myanmar

Content

- ❑ Objectives of evaluation of the quality of census data
- ❑ Scope and organization of evaluation programme
- ❑ Sources of census errors
- ❑ Types of census errors
- ❑ Methods for evaluation of errors

Why do we need to evaluate the census?

❑ The census is a huge operation comprised of many stages

❑ It is not perfect and errors can and do occur at all stages of the census operation

❑ Many countries have recognized the need to evaluate the overall quality of their census results and have employed various methods for evaluating census coverage as well as certain types of content error

Aims of evaluation of data

❑ To identify errors and find a solution to correct before releasing the final results

❑ For unavoidable   errors:

  ■ To provide users with a measure of the quality of census data to help them interpret the results

❑ To serve as a basis for constructing the best estimate of census aggregates, such as total population

❑ To provide suggestions and assist the plans for future censuses

Planning a Census Evaluation Program

❑ A census evaluation program should be developed as part of the overall census program and integrated with other census activities

❑ Census errors can happen at all phases of the census operation, including questionnaire design, mapping, enumeration, data capture, coding, editing and imputation

❑ Evaluation of data quality may have two parts:

- Preliminary evaluation will enable the identification of any problem areas that have not been previously detected
- More extensive evaluation should be undertaken on data quality to inform users about unavoidable problems and establishing best estimates

Scope of evaluation

❑ Census evaluation should include at least the followings:
- Analyze consistency in data and between variables
- Analyze evidence of age misreporting
- Analyze the quality of data collected in the census with appropriate methodology such as fertility, mortality, migration, educational and economic chacateristics
- Compare census data with independent data sources (surveys, registers) or previous censuses

Institutional organization

❑ Establishing the census evaluation team

  ■ Team should be trained in the evaluation techniques

  ■ Team should consist of members who have experience in census operations and analysis of census topics-demography , education, housing, labor force, etc.

  ■ Team should have background knowledge of historical events and changes in population structure in the country

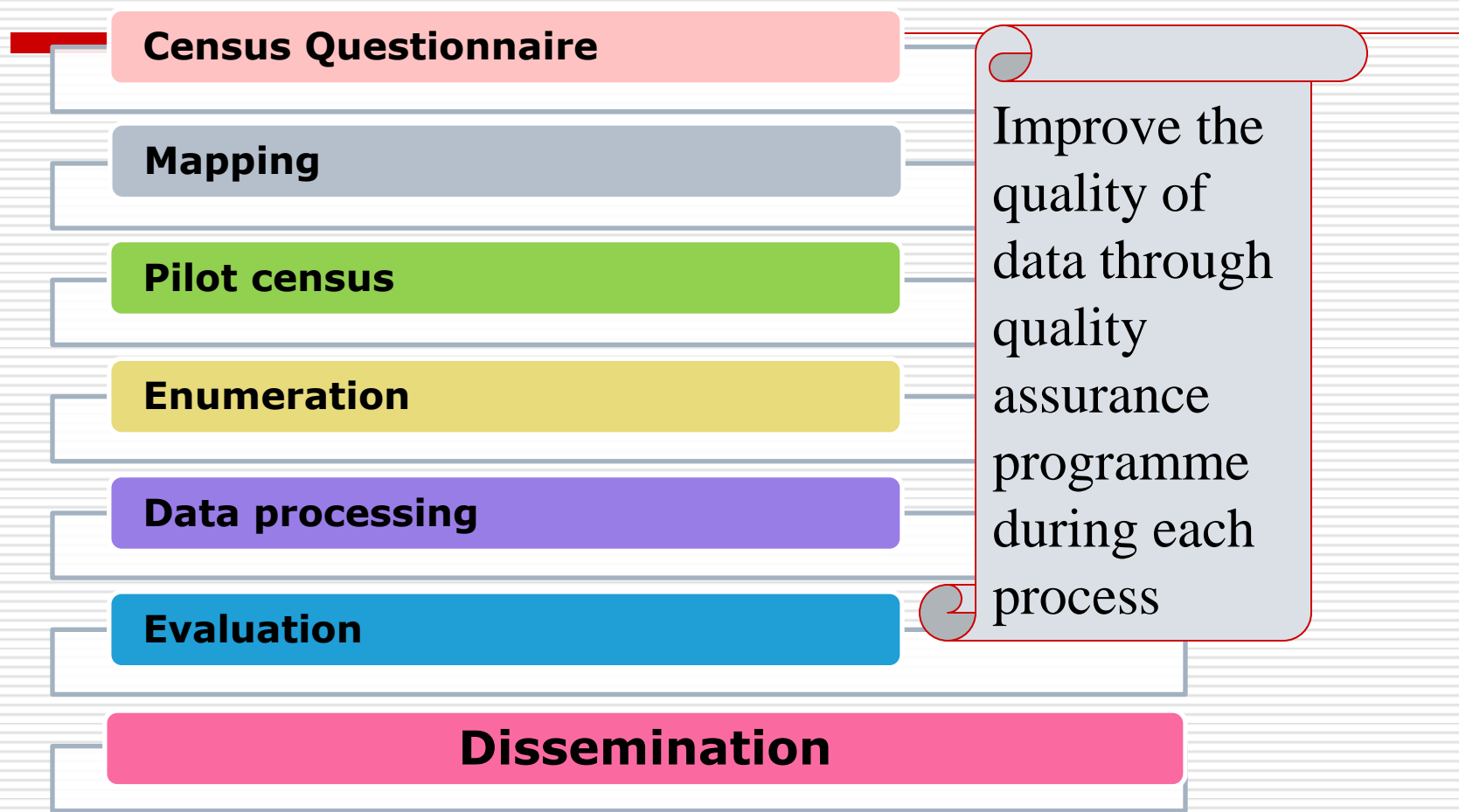 ■ Team should collaborate with related research institutions

Information on census processes

❑ It is necessary that the evaluation team have a good understanding of the census process

- Which population groups were included/excluded

- Whether and how the data should be weighted

- Any known problems with the enumeration and/or data entry and editing processes

- If and how missing values have been edited

  ☐ If there are no missing values on age and sex, the data has almost certainly been imputed

  ☐ Imputed values should ideally be flagged

  ☐ Editing rules for logical imputation, hot-decking or any other method that was used should be well understood and their effects carefully considered
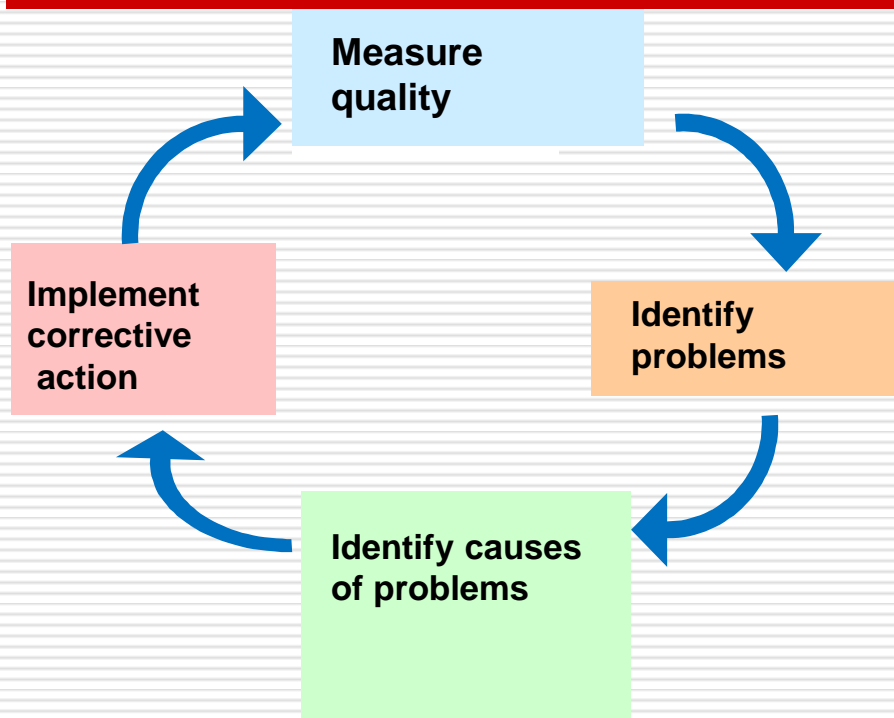
Main census phases

- **Census Questionnaire**
- **Mapping**
- **Pilot census**
- **Enumeration**
- **Data processing**
- **Evaluation**
- **Dissemination**

Improve the quality of data through quality assurance programme during each process
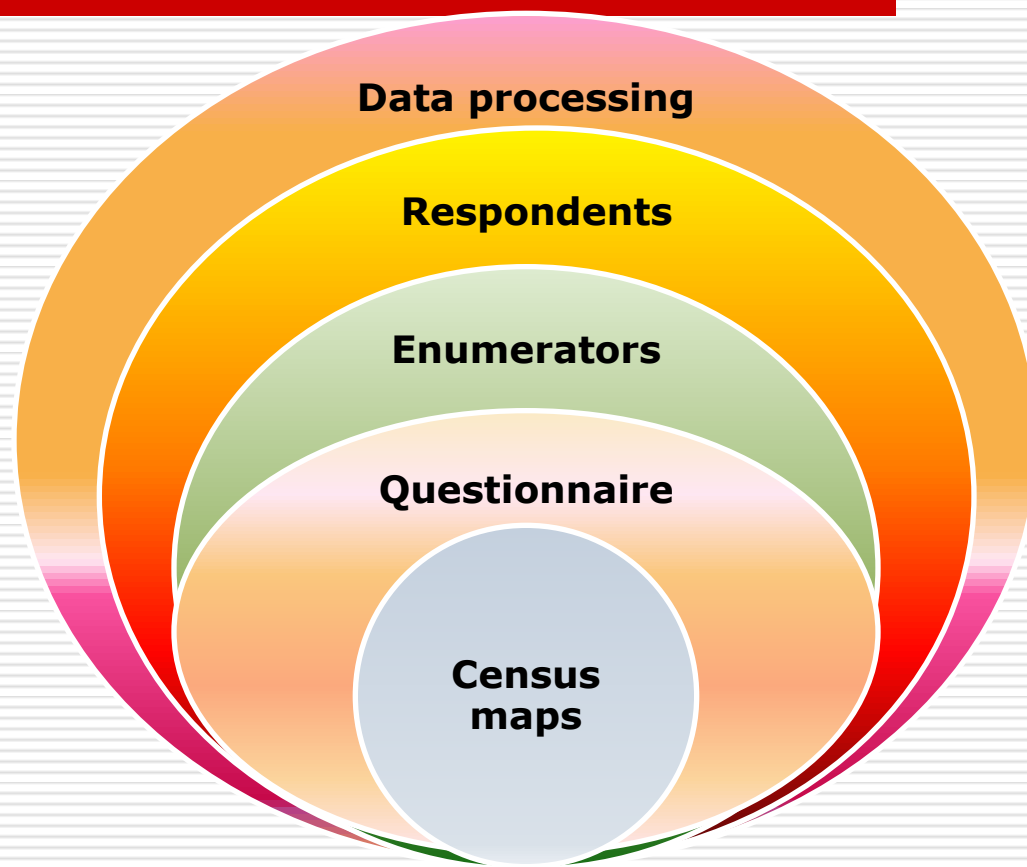
Quality assurance programme for controlling errors

Without such a programme, the census data may contain many errors which can severely diminish the usefulness of the results

Overview of sources of errors

Data processing

Respondents

Enumerators

Questionnaire

Census maps

Errors can be many kinds from different sources

Sources of errors

❑ <u>Errors in mapping and listing living quarters</u>

- ▪ Incomplete or inaccurate maps and/or listing
- ▪ Inaccurate demarcation of enumeration areas
  - o Overlapping or missing some areas
  - o Unclear boundaries of enumeration areas

Sources of errors

❑ <u>Errors in questionnaire design</u>

- Poorly designed questions or instructions
- Poor sequencing of the questions
- Poor communication between respondent and enumerator
- Skip pattern- not clear or not placed appropriately

P2b   National Identity Card Number

P3   Relationship to head of household

Head	01
Wife/husband	02
Son/daughter	03
Son/daughter-in-law	04
Grandchild/great grand child	05
Mother/Father of head or spouse	06
Other relative	07
Domestic employee	08
Boarder	09
Non relative	10
Clergy	11
Visitor	12

P4   Sex

Male	1
Female	2

P5   Date of birth

If the exact year of birth is not known, write an estimated year of birth and write '88' for the

---

Sri Lanka Chetty — 7
Bharatha — 8
Other (specify) — 9

P9   Citizenship

If Sri Lankan mark '77'. If not known mark '88'. If a citizen of another country write the name of the country and its code.

Sri Lankan   77   Not known   88

Other country (specify)

P10   District of birth

If the mother was usually residing in this district at child birth, mark '77'. If mother resided in another district, write the name of the district and its code. If born in another country, write the name of the country and its code.

This district   77

District / Country

P11   Duration (in years) of residence in this district

If since birth, mark '98' and go to P14. If less than 1 year write '00'. If a person migrated more than once to this district/country, state the duration since last move.

Since birth   98 ➔ go to P14

	No difficulty 1	Difficult 2	Not possible 3
(a) Seeing (even with the use of glasses if they are used)			
(b) Hearing (even with the use of hearing aids if they are used)			
(c) Walking a short distance or up/down about 12 steps in a stairway			
(d) Cognition (remembering and concentrating)			
(e) Day-to-day selfcare such as getting dressed, washing etc.			
(f) Communication in their own language due to physical or mental reasons			

For persons born in 2008 November or before (3 years & over)

P15   Educational activity involved in the last 30 days

Pre-school education	1
School education	2
Undergraduate /Postgraduate edu.	3
Vocational/ Technical education	4

12. Name of the household members present in the census night (Start from head of household)	13. Member(s) Serial No.	14. Age (Completed years)	15. Relationship with Head of Household					16. Sex		17. Marital Status				18. Religion					19. Type of disability 0=None 1=Speech 2=Vision 3=Hearing 4=Physical 5=Mental 6=Autistic	20. Student (Currently)		21. Highest class passed (Write class passed code)	22. Field of Education				23. Can write a letter?	
			Head	Wife/ Husband	Child	Other Relative	Non-Relative	Male	Female	Unmarried	Married	Widowed	Divorced/ Separated	Muslim	Hindu	Christian	Buddhist	Other		Yes	No		General	Vocational/ Technical	Religious	Not Applicable	Yes	No
	1		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐
	2		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐
	3		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐
	4		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐
	5		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐
	6		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐
	7		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐
	8		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐
	9		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐
	0		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		☐	☐		☐	☐	☐	☐	☐	☐

For members age 7 ye...

	FOR ALL MEMBERS OF THE HOUSEHOLD										AGE 5 AND ABOVE		
IDENTITY CARD	**MIGRATION**										**EDUCATION**		
10. Type of identity card	Place of Birth		Place of Usual Residence		15. Duration in place of usual residence (in years) *If less than 1 year write "00"*	16. Reason for movement to this township (usual residence)	Place of Previous Usual Residence		19. Can (Name) read and write in any language?	20. Is (Name) currently attending, previously attended or never attended school/ college?	21. What is the highest education grade/level (Name) completed?		
	11. Township *If born here write "000", if not write Township code*	12. Urban or Rural	13. Township *If here write "000", if not write Township code*	14. Urban or Rural			17. Township *If here write "000", if not write Township code*	18. Urban or Rural			None - 00 Grade - 01-11 College - 12 Vocational training - 13 Undergraduate diploma - 14 Graduate - 15 Postgraduate diploma - 16 Masters Degree - 17 PhD - 18 Other - 19		

Serial Number column with rows 1–7.

Row identity card options (per row): Citizenship Scrutiny Card (pink), Associate Scrutiny Citizenship Card (blue), Naturalised Scrutiny Citizenship Card (green), National Registration Card (three fold card, green for men, pink for women), Religious Card, Temporary Registration Card (white), Foreign Registration Card (FRC), Foreign Passport, None of the documents above, Child below 10 years — coded 1 2 3 4 5 6 7 8 9 10.

Column subtitles: Enter code from manual (11), Urban/Rural (12), Enter code from manual (13), Urban/Rural (14), Employment/in search for employment / Education / Marriage / Followed family / Conflict / Did not move / Other (16, coded 1 2 3 4 5 6 7), Enter code from manual (17), Urban/Rural (18), Yes/No (19, coded 1 2), Currently attending / Previously attended / Never attended (20, coded 1 2 3).

Sources of errors

- **Enumerator errors**
  - Not fully explaining the meaning of the questions to the respondents or changing the wording of the questions
  - Making errors in recording the responses
  - Not asking some questions and creating unknown data
- **Respondent errors**
  - Misunderstanding or deliberate misreporting
  - Proxy responses –when someone other than the person to whom the information pertains- provides the responses to the questions

Sources of errors

- **Data entry errors** : Invalid entries or mistakes in scanning and capturing data
  - ☐ Data capture system can ensure that the value of each field is within the permissible range of values for that item
- **Coding errors:** giving wrong code to the information

- **Errors in editing/imputation** :
  - ☐ The editing process changes or corrects invalid and inconsistent data by imputing non-responses or inconsistent information with plausible data
  - ☐ Any of these editing operations can introduce new errors

Types of census errors

❑ Coverage errors:
  ■ Errors in the count of persons or housing units resulting from cases having been "missed" or "counted erroneously" or "double counting"

❑ Content errors:
  ■ Errors in the recorded characteristics of persons, households or housing units

Coverage error

❑ **Omissions :** Missing housing units, households, and/or persons during census enumeration

❑ **Erroneous inclusions :** Housing units, households and persons enumerated when they should have not been enumerated in specific EA

❑ **Duplications:** Occur when persons, households or housing units are counted more than once/ or captured more than ones

Coverage error

❑ Sources of coverage error:

- ■ Incomplete or inaccurate maps or address lists of enumeration areas,
- ■ Failure by enumerators to canvas all the units in their assignment areas or all the individuals in the units
- ■ Duplicate counting of some units or individuals,
- ■ Erroneous enumeration of certain categories of persons such as visitors or non-residents

Coverage errors

Gross error

❑ Sum of duplications, erroneous inclusions and omissions

Net error

❑ Difference between over-counts and under-counts

- Under-count if the number of omissions ("missing" people) exceeds the number of duplicates and erroneous enumerations

- Over-count if total of the number of duplicates and erroneous enumerations exceeds the number of omissions

Content errors

- ❑ Content errors arise from the <u>incorrect reporting or recording</u> of the characteristics of persons, households and housing units

- ❑ Every phase of census data collection and processing has the potential for introducing content errors into the census results
  - ■ Enumerators, respondents, scanning, data capture, coding, editing/imputation

Methods for the evaluation of census errors

Single Source of Data (rely only on the census being evaluated)

❑ Demographic analysis

- ■ *Consistency checks*
- ■ *Analysis of distribution or ratios of particular census topics*

Multiple Sources of Data

❑ Non-matching studies

- ■ *Demographic analysis using multiple census rounds*
- ■ *Comparison with administrative sources or existing surveys*

❑ Matching studies – not covered in this workshop

- ■ *Post Enumeration Surveys*
- ■ *Record checks*

Source: U.S. Census Bureau, 1985. *Evaluating Censuses of Population and Housing*

Multiple Sources of Data – Matching studies – Record checks

❑ Census records are matched with a sample of records from official registration systems such as the vital registration system

❑ The relevant respondents to the census questionnaire are traced to the time of the census

❑ Sources include:
  ■ Previous censuses
  ■ Birth registration
  ■ School enrollment
  ■ National identification cards/registers
  ■ Immigration registers
  ■ Voter registration lists
  ■ Health or social security records

Multiple Sources of Data – Matching studies – Record checks

❑ Both coverage and content errors can be measured through the above comparisons

To evaluate coverage efficiently the following preconditions are essential:

❑ A large and clearly-defined segment of census population (if not the entire population) should be covered by the registration system

❑ The census and registration systems should be independent of one another

❑ There should be sufficient information in the records to be able to match them with census respondents accurately

Multiple Sources of Data – Matching studies – Record checks

To evaluate content efficiently the following preconditions are essential:

❑ The register system should contain relevant items covered in the census such as age, sex, education, relationship, marital status etc.

❑ Definitions of variables should be identical between the census and the register

Record checks – strengths and weaknesses

- ❑ Can provide separate estimates of coverage and content error, net and gross error
- ❑ With the right data, more characteristics can be evaluated compared to what can be done with non-matching studies
- ❑ Calls for a high level of technical skill and registration system
- ❑ Matching is expensive
- ❑ In many countries, registration systems are not sufficiently complete for this method to be feasible

Multiple Sources of Data – Matching studies – Post-Enumeration Surveys (PES)

❑ A PES entails the complete re-enumeration of a representative sample of the population, which is then matched to the corresponding records from the census enumeration

❑ PES can fulfill multiple objectives:

- Assess the degree of coverage of the main enumeration
- Assess implications of coverage error for usefulness of the data
- Examine characteristics of those who have been missed by the main enumeration
- Develop recommendations for design of future censuses and surveys

Multiple Sources of Data – Matching studies – Post-Enumeration Surveys (PES)

❑ The PES should be independent of the census

- ■ A survey is conducted using a sampling frame independent of the census. Persons from this survey are then matched to the census to estimate the number of persons missed or erroneously enumerated in the census

Advantages:

- ■ The results of a PES can be used to separately evaluate coverage vs. content error and net vs. gross error

- ■ Incorporates matching of individuals or units between the census and PES – this allows for a direct comparison of results

- ■ Its results are generally more reliable than those of the census

Multiple Sources of Data – Matching studies – Post-Enumeration Surveys (PES)

Challenges:

❑ Requires highly skilled field and professional staff

❑ Matching is complex and costly

❑ To be valid, the PES has to be conducted in a short time after the census to limit the complicating effects of population change, recall bias etc.

Conclusion

➢   A number of methods exist for carrying out census evaluation

➢   In practice, many countries use a combination of such methods in order to improve the quality of evaluation programme