Sampling in IEA Studies

2011 IEA General Assembly

IRELAND
Introduction

Sampling Challenges

• Target Populations
  • Linking target & survey populations
• Sample Plan Elements
• Sample Plan Implementation
  • Quality standards

Where does sampling fit in the whole study process?
GA↔ International Study Center

Framework

Target Population(s)

Survey Objectives

International Comparable Outcomes
Target Population

- In practice, this is the population for which we want characteristic estimates. It is the population for which information is required.
- IEA target population has to be demonstrably comparable among all participating “countries” over time.
  - Usually students or teachers within participant “countries”
Survey Objectives

- IEA Multi-purpose survey
  - School, student, teacher statistics
- Areas to be covered
  - National, domain statistics
- Kinds of results expected
  - Ratios, totals, proportions
- Users and uses of the data
  - Researchers, government policy analysts
- Levels of accuracy
International Comparable Outcomes

• Sampling should not be an issue in that regard
  • High quality sample (IEA high standards)
  • Coverage & exclusions
  • Response rates
  • Respect of the sample plan (Complete documentation)
Solution

• Ideal scenario: census of all units belonging to the target population
  • Possible for small participating “countries”
• Practical scenario: sample from the above units
  • The only possibility for most participating “countries”
Solution

- Sampling is acceptable in a IEA context if and only if
  - National sampling plans are based on sound and defensible sampling methods
    - Each unit belonging to the target population MUST have a chance to be in the sample (exclusion otherwise)
    - This chance MUST be known (quantifiable) for each unit in the sample
  - Survey results close to census values
Solution

- Sampling is acceptable in a IEA context if and only if
  - National sampling plans are based on sound and defensible sampling methods
  - National sampling plans are thoroughly documented
  - National sampling plans are approved by the ISC prior to implementation
  - Sampling standards are met (implementation)
Elements of a Sample Plan

- Sample design
- Estimation procedures
- Procedures to estimate variances
Sampling Design

Sets of specifications describing:

- Target and survey populations;
- Sampling frame;
- Survey units;
- Sample selection method;
- Size of the sample.
Sampling Design

Survey Population

“The population is the aggregate (or collection) of units to which the survey results apply”

How does this connect to the framework definition known as the “International desired target population”? 
IEA Population structure

Diagram:
- International Desired Target Population
  - National Desired Target Population
    - National Defined Target Population
  - Exclusions from National Coverage
  - Practical Exclusions

Coverage:
- 100% Coverage
- ≤5%
Sampling Frame

• List of units encompassing all individuals and allowing access

• IEA studies: List of schools
  • Lists of students usually not available
  • Takes advantage of the natural population hierarchy (students within classrooms within schools)
  • Reduces collection costs compared to SRS
Sampling Frame

• Sampling frame should have:
  • No incorrect entries, no duplicate entries, or no entries that refer to elements that are not part of the target population

• Information required on the frame:
  • School ID
  • School measure of size (MOS)
  • School status indicator (exclusion categories)
  • Appropriate entries for all stratification variables (public/private, urban/rural, single sex school, etc.)
IEA Population structure

100% Coverage

≤5%

International Desired Target Population

National Desired Target Population

National Defined Target Population

Effective Target Population

Exclusions from National Coverage

School-Level Exclusions

Within-Sample Exclusions
Example

• Difference between the International desired target population and the effective target (survey) population
• TIMSS & PIRLS 2011 Study
Target Population

2010 / 2011 List of sampled schools

2008 / 2009 list of all schools

Prior to school sampling
- International schools
- Private schools
- Remote schools
- Very small schools

New schools since 2009

Within-School Exclusions
- Special needs
- Language

List of sampled schools

Prior to school sampling
- International schools
- Private schools
- Remote schools
- Very small schools

List of all schools
Prior to school sampling
------------------------
International schools
Private schools
Remote schools
Very small schools

New schools since 2009

Within-School Exclusions
------------------------
Special needs
Language
Prior to school sampling
------------------------
International schools
Private schools
Remote schools
Very small schools

Exclusions
------------------------
Special needs
Language

Under Coverage
------------------------
A large group accounting for more than 5% of the target population

Exclusions
------------------------
Small groups each one accounting for less than 5% of the target population

Within-School Exclusions
------------------------
New schools since 2009
• A good list of schools should cover 100% of the target population

• A good sample should be representative of no less than 95% of the international desired target population
### Target Population

<table>
<thead>
<tr>
<th>Countries</th>
<th>Reading Achievement Distribution</th>
<th>Average Scale Score</th>
<th>Years of Formal Schooling*</th>
<th>Average Age</th>
<th>Human Development Index**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy</td>
<td></td>
<td>551 (2.9)</td>
<td>4</td>
<td>9.7</td>
<td>0.940</td>
</tr>
<tr>
<td>Hungary</td>
<td></td>
<td>551 (3.0)</td>
<td>4</td>
<td>10.7</td>
<td>0.869</td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
<td>549 (2.3)</td>
<td>4</td>
<td>10.9</td>
<td>0.951</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td>548 (2.2)</td>
<td>4</td>
<td>10.5</td>
<td>0.932</td>
</tr>
</tbody>
</table>

Users think “all grade 4 students...”
Sampling Frame

Other aspects to consider in defining the international desired target population

• Feasibility
  • Costs, time, resources
• Overlapping with other studies


**Sampling Frame**

- Stratification: grouping of schools
  - To apply disproportionate sample allocation by domains (sub-groups)
    - More reliable domain estimates
  - To ensure proportional representation of specific groups in the sample
  - To improve the reliability of the overall survey estimates by forming homogeneous groups
Stratification: grouping of schools

- To apply disproportionate sample allocation by domains (sub-groups)
- More reliable domain estimates
  - To ensure proportional representation of specific groups in the sample
  - To improve the reliability of the overall survey estimates by forming homogeneous groups

<table>
<thead>
<tr>
<th>Population</th>
<th>Domain A</th>
<th>Domain B</th>
<th>Domain C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15%</td>
<td>15%</td>
<td>70%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>Domain A</th>
<th>Domain B</th>
<th>Domain C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15%</td>
<td>15%</td>
<td>70%</td>
</tr>
</tbody>
</table>
Sampling Frame

- Stratification: grouping of schools
  - To apply disproportionate sample allocation by domains (sub-groups)
    - More reliable domain estimates
  - To ensure proportional representation of specific groups in the sample
  - To improve the reliability of the overall survey estimates by forming homogeneous groups
Stratification: grouping of schools

- To apply disproportionate sample allocation by domains (subgroups)
- More reliable domain estimates
  - To ensure proportional representation of specific groups in the sample
  - To improve the reliability of the overall survey estimates by forming homogeneous groups

<table>
<thead>
<tr>
<th>Population</th>
<th>390</th>
<th>400</th>
<th>505</th>
<th>525</th>
<th>530</th>
<th>650</th>
<th>600</th>
<th>600</th>
<th>600</th>
<th>615</th>
</tr>
</thead>
<tbody>
<tr>
<td>430</td>
<td>450</td>
<td>500</td>
<td>490</td>
<td>550</td>
<td>585</td>
<td>615</td>
<td>630</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>400</th>
<th>430</th>
<th>525</th>
<th>550</th>
<th>600</th>
<th>630</th>
</tr>
</thead>
</table>
Sampling Frame

• Stratification variables
  – Variables defining domains of interest for which:
    • over sampling is required
    • proportional representation is required
  – Variables highly correlated to the school mean scores
  – Variables defining non-response adjustment cells
Sampling Frame

- Stratification guidelines
  - Avoid over-stratification
  - Two to four stratification variables are usually sufficient
  - Each school in the sampling frame must belong to one, and only one, stratum
  - A few divisions of a continuous variable is usually sufficient, i.e., 2 to 5
  - Avoid defining very small strata
  - Sample at least two schools per stratum
Sample Selection Method

• Explicit stratification
  – Consists of building separate school lists according to a set of explicit stratification variables
  – Implements a disproportionate sample allocation to produce more reliable estimates by explicit stratum
**Implicit Stratification**

- Consists of sorting the school lists by a set of implicit stratification variables
- Usually used only with the PPS systematic sampling method
- Ensures a strictly proportional allocation of the school sample across all implicit strata

<table>
<thead>
<tr>
<th>Population</th>
<th>Domain A</th>
<th>15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain B</td>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>Domain C</td>
<td>70%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>Domain A</th>
<th>15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain B</td>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>Domain C</td>
<td>70%</td>
<td></td>
</tr>
</tbody>
</table>

- Ensures strictly proportional allocation of the school sample across all implicit strata
Sample Selection Method

Stratified multi-stage PPS systematic sampling (cluster sampling)

- School sampling
  - Description of the process
  - Multi-stage, PPS, systematic
  - Replacement schools
- Within-School sampling
School Sampling

• Multi-stage sampling (clustered sampling)
  • Sample of schools is first drawn (PSU)
  • Within selected schools, a sample of classes (teachers) could then drawn,
  • Within selected classes, a sample of (usually all) students is chosen

• Within explicit strata, PSUs are selected in a systematic way (random start) by opposition to have each PSU chosen randomly
School Sampling

Multi-stage stratified cluster sample selection
School Sampling

Multi-stage
School Sampling

Multi-stage
School Sampling

Multi-stage stratified

Stratification consists of grouping schools into strata
Examples: Regions, zones, urbanization, socio-economic status, school types
School Sampling

- Probability proportional to size (PPS) sampling
  - IEA’s usual approach
    PSUs (schools) are selected with probability proportional to their size

- A measure of size (MOS) is required for each first stage sampling unit (PSU)
School Sampling

PPS sampling:

- The larger the school, the bigger the chance to sample it
- Very large schools are automatically (always!) in
  - Perception of a good sample
  - Important response burden for the schools but not for the students within these schools
- Gain in reliability (when the size of the school is correlated to the characteristic of interest)
School Sampling

PPS sampling:

- Tends to produce equal sampling probabilities at the ultimate level (students, teachers) for units belonging to the same stratum (when the same number of classes (teachers) is sampled within each school)

- Small schools are selected with equal probability however
  - The measure of size is too unstable
If PPS is used to select one school out of 1000 schools (for a total of 60,000 students) then that school has MOS chances out of 60,000 of being selected. In this example, the 50 sampled students would represent the 60,000 found in the population.
School Sampling

PPS school sampling

If this school is selected however, then the 160 sampled students would represent the 60 000 found in the population.
School Sampling

PPS sampling

• Selection with PPS produces a sample in which each selected units (school) represent the same number of elements (students) in the population. Thus if you are selecting 2 schools from a population of 60 000 students, each selected school represents 30 000 students.
School Sampling

PPS school sampling

R  R  R  R  U  R  U  U  U  U  U  R

50 95 60 75 160 95 130 85 100 130 100 ...

30 000

SRS : 2 schools out of 1000

500 X 50 500 X 160

25 000 80 000

30 000

MOS
School Sampling

Replacement schools
- Two replacement schools are assigned to each original selected school (when possible)
  - One replacement for the Field Test
- Attempt to reduce the risk of bias without reducing the sample size too much
  - Has to be documented
  - Can’t choose your own replacements
School Sampling

Replacement schools

• Only refusals can be replaced
• Nothing is better than the original schools (possible bias)
• A priori identification of replacement schools is the standard approach in IEA studies (sizes & characteristics)
• Usually no substitution is allowed past the first stage sampling (schools)
School Sampling

Replacement schools

• Participation Rates (Standards)

  Minimum school participation rate of 85%
  Minimum classroom participation rate of 95%
  Minimum student participation rate of 85%

OR

Minimum overall participation rate of 75%
Bias effects (sampler nightmare)

Pop: M=500, S=100

80% response rate

M = 480, M = 505

80% resp. rate

M = 490, M = 509
Bias effects (sampler nightmare)
Pop: M=500    S=100
Sample: Standard error of the mean = 5
Confidence levels by bias scenarios

<table>
<thead>
<tr>
<th>Sch–Std-Total</th>
<th>10 pts drop</th>
<th>20 pts drop</th>
<th>30 pts drop</th>
</tr>
</thead>
<tbody>
<tr>
<td>resp. rates %</td>
<td>m</td>
<td>level</td>
<td>m</td>
</tr>
<tr>
<td>100-100-100</td>
<td>500</td>
<td>95%</td>
<td>500</td>
</tr>
<tr>
<td>95-95-90</td>
<td>501</td>
<td>94%</td>
<td>502</td>
</tr>
<tr>
<td>90-90-81</td>
<td>502</td>
<td>93%</td>
<td>504</td>
</tr>
<tr>
<td>85-85-72</td>
<td>504</td>
<td>89%</td>
<td>507</td>
</tr>
<tr>
<td>80-85-68</td>
<td>504</td>
<td>86%</td>
<td>509</td>
</tr>
<tr>
<td>75-85-64</td>
<td>505</td>
<td>83%</td>
<td>510</td>
</tr>
<tr>
<td>70-85-60</td>
<td>506</td>
<td>77%</td>
<td>512</td>
</tr>
</tbody>
</table>
Bias effects (sampler nightmare)

Pop:
- M = 500
- S = 100

80% response rate
- M = 480
- M = 505
- M = 509

M = 490

80% resp. rate
School Sampling

• Within school sampling
  • Whenever possible, class selection is recommended (students)
  • When possible, ‘take all’ approach should at least be considered (teachers)
  • Substitution should not be allowed
  • Participation rates should be set high
  • Exclusions should be clearly identified prior to sampling (and kept to a minimum)
IEA International Standards:

• An effective sample size resulting in the following 95% confidence limits for sample estimates of population
  • MEANS: $m \pm 0.1s$ (s being the population standard deviation)
  • PERCENTAGES: $p \pm 5\%$
  • CORRELATIONS: $r \pm 0.1$

• This is equivalent to a simple random sample of 400 units
Sample Size

A 95% confidence interval can be described as follows: If sampling is repeated indefinitely, each sample leading to a new confidence interval, then in 95% of the samples the interval will cover the true population value (G.W. Snedecor, W G. Cochran)
95%
Sample Size

Factors to consider:

1. Data requirements (levels & precision)
2. Sample plan
Sample Size

Factors to consider:

1. Data requirements (levels & precision)
2. Sample plan

Design effect = \frac{\text{Variance under current design}}{\text{Variance under a SRS}}

- Estimates of design effects (intraclass correlation coefficients) are usually derived from previous studies
Sample Size

Factors to consider:

1. Data requirements (levels & precision)
2. Sample plan

\[
\text{Margin of error} = 2 \times S \sqrt{\frac{\text{Deff}}{n}}
\]

- Where \( n \) gives the student sample size
- Deff of 5 and standard deviation of 100 are usually assumed for new countries
Sample Size

Factors to consider:

1. Data requirements (levels & precision)
2. Sample plan

• Margin of error is inversely proportional to the square root of the sample size

• For example, if 150 schools give a 7 points margin of error than under a similar design effect, 35 to 40 schools will buy you a 14 points margin.

• For reasonably stable domain estimates, a sample size of at least 30 schools is preferable.
### List of schools

<table>
<thead>
<tr>
<th>Domain A</th>
<th>Domain B</th>
<th>Domain C</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>10%</td>
<td>80%</td>
</tr>
</tbody>
</table>

### Sample of schools

<table>
<thead>
<tr>
<th>Domain A</th>
<th>Domain B</th>
<th>Domain C</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 schools</td>
<td>15 schools</td>
<td>120 schools</td>
</tr>
</tbody>
</table>

+ 20

+20
Large variation between schools combined with little variation within school = High intraclass correlation coefficient
Little variation between schools combined with large variation within school = High intraclass correlation coefficient
Sample Size

Factors affecting sample sizes:

1. Data requirements (levels & precision)
2. Sample plan
3. Variability of characteristics in the population (Main variable(s) of interest)
4. Non-response
5. Cost and time (operational constraints)
6. Study constraints.
7. Size of the population
Sample allocation between strata

- When variability among units does not differ much from stratum to stratum and the same can be said about the collection costs, proportional allocation is usually recommended unless strata themselves are of principal interest

Examples:
- Regions are to be equally represented
- Under allocation of very remote areas
## Sample Size

### Sample allocation between strata

<table>
<thead>
<tr>
<th>POP:</th>
<th>90%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>4000 sampled units</td>
<td>Str 1 (M.E.)</td>
<td>Str 2 (M.E.)</td>
</tr>
<tr>
<td>Proportional allocation</td>
<td>3600 (10.5)</td>
<td>400 (31.6)</td>
</tr>
<tr>
<td>Equal allocation</td>
<td>2000 (14)</td>
<td>2000 (14)</td>
</tr>
<tr>
<td>IEA approach (min. fixed sizes)</td>
<td>3300 (11)</td>
<td>2000 (14)</td>
</tr>
<tr>
<td>Disproportional allocation</td>
<td>5400 (8.6)</td>
<td>100 (63)</td>
</tr>
</tbody>
</table>
Sample Selection Method
(Overview within explicit strata)
Warning: Schools with no targeted students, with only excluded students, or closed schools are not replaced

**MOS**
Cumulative MOS

SI = 350

Sampled Schools

Respondent schools
Sampling stages

Sampled Schools

Sampled Classes

Sampled students

Participating

Respondants

MOS

SI = 350

Cumulative MOS
Sample Design

Why do we need one?

- To make sure each unit of the “effective” target population has a known probability of being in the sample
- To make sure we know how to compute this probability
- To allow for UNBIASED mean and standard error estimates about the “effective target population”
- To get an estimate of the level of coverage.
Element of a Sample Plan

Estimation procedures

- Sampling weights
- What are they?
- Why using them?
  - An example
Estimation Procedures

Sampling weights (What are they?)

- Values assigned to all sampling units
- Based on the selection probabilities applied at each sampling stage
- Adjusted to take into account non-response
- The number of times students are selected differs. This is usually the case!
  - Sample design (stratification, MOS over time, etc)
  - Non response
• Sample estimates $\neq$ Census estimates for all variables of interest
However, on average we have the following relation for any variable of interest:

\[
\sum_{\text{all samples}} \frac{\text{Weighted sample estimates}}{\text{Total number of samples}} = \text{Census value}
\]
- Using sampling weights compensates for the fact that some units are selected more often.
- On average over all possible samples, all units have the same importance (as for a census)
Not using sampling weights gives more importance (weight!) to those students selected more often over all samples.
Not using sampling weights shifts survey estimates away from the corresponding census value by an amount that is unknown.
• Using sampling weights provides...
  – Unbiased estimates of population characteristics
Census Value

2 s.e. 2 s.e.
Census Value

2 s.e. 2 s.e.

95%
Sampling Weights

- Why?
  Using sampling weights:
  - Produce unbiased statistic estimates from the sample (equal to census statistics on average).
  - Produce reliable estimates of standard error for these statistics (tells us how much confidence one can have in sample statistics to reproduce census statistics from a particular sample).
Sampling Weights

- If you don’t use weights...

Weighted

Unweighted

Census Value

Your sample

All samples (unknown)
In a perfect world (sampler universe)

60 (3 classes of 20 kids)

160 (8 classes of 20 kids)

\[ \frac{350}{60} \times \frac{3}{1} \times \frac{20}{20} = \frac{350}{160} \times \frac{8}{1} \times \frac{20}{20} = 350 \]

SI = 350

Third stage sampling
Quality standards

Documentation (A must!)
- Describe sampling methodology
- Describe sampling implementation
  - Coverage
  - Sample sizes
  - Exclusion rates
  - Participation rates

Field Test
- All efforts should be made to “field test” all sampling procedures
- Allows for a first glance at potential problems with the frame.
I have good reasons to believe that the census value for this country is between 545 and 557 (551 ± 2 X 3.0)