



USAID
FROM THE AMERICAN PEOPLE

Research Design and Sampling Frameworks

Elena Vinogradova
Education Development Center

April 27, 2016

About the Presentation

- Prepared for the USAID workshop/webinar “Release of the *EGRA Toolkit, Second Edition: Updated Guidance and Tools for Conducting Early Grade Reading Assessments*,” Bethesda, Maryland, April 27, 2016.
- This presentation and the workshop/webinar were funded by USAID under the Education Data for Decision Making (EdData II) project, led by RTI International, Task Order No. AID-OAA-BC-12-00003 (RTI Task 20).
- The initial version of this presentation was prepared for the USAID “EGRA Basics and Beyond” workshop/webinar, November 2015, Bethesda, Maryland.

Session Objectives

- Highlight differences among snapshot, performance, and impact evaluations
- Review requisites for survey samples
- Understand importance of these pieces for data analysis, post data collection

Toolkit Section 5

Toolkit Annexes B-D

Toolkit Section 11

1. **Snapshot assessment** – to obtain a diagnostic of student performance at a single time point; data for benchmarking, research on interplay between mother tongue and language of instruction, etc.
2. **Performance evaluation** – to evaluate whether changes occurred in learners' performance over a period of time, based on initial and follow-up assessments
3. **Impact evaluation** – to evaluate the impact of a program or intervention on learners' performance over a period of time, based on a comparison of treatment and control/comparison groups

- Sample consists of one or more samples drawn from the population(s) at one point in time
- Allows researchers to describe characteristics of the population(s) and to identify correlations between characteristics at that point in time
- Uses complex/cluster sampling

Example:

EGRA administered once to a random sample of students in a nationally representative sample of schools in order to get a snapshot of student performance at that time.

- Design is based on the evaluation questions
- Measures contribution
- Cannot be used for establishing true impact of a project
- Uses simple random or complex/cluster sampling
- Consists of two or more measurements
- Examples include:
 - Cohort design
 - Longitudinal design

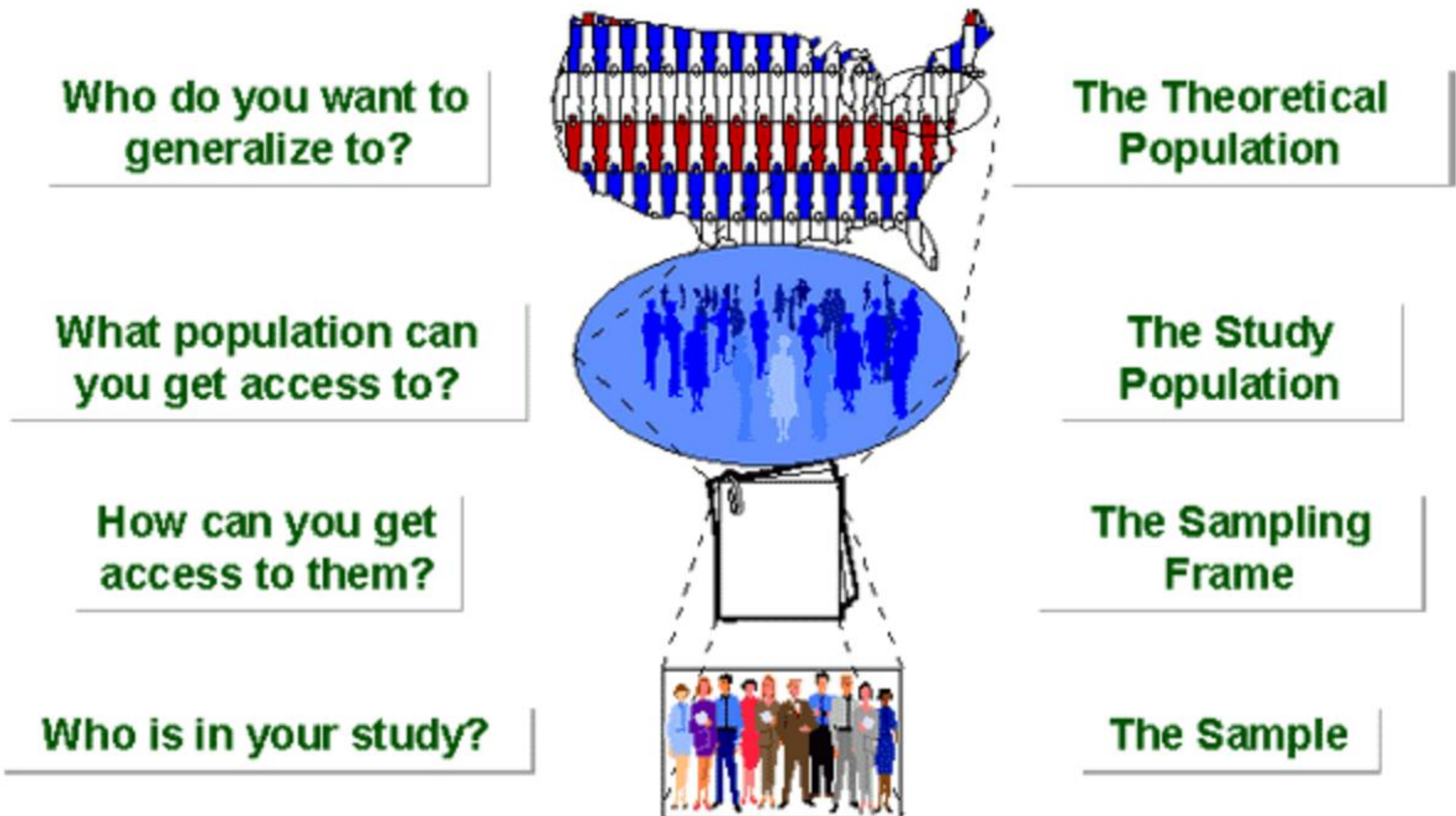
Example:

EGRA administered at the end of the school year at the beginning, middle, and end of the project, to a random sample of students in a representative sample of project schools, in order to measure changes in student performance.

- What changes do we see **as a result** of the intervention?
 - Does the intervention affect groups of students differently?
 - What moderating/mediating factors affect results?
- Common designs include:
 - Experimental: randomized controlled trial
 - Quasi-experimental:
 - Regression discontinuity design
 - Matched comparison group design

Example: EGRA administered to a random sample of students in a representative sample of schools at baseline, and to another random sample of students at endline, in order to monitor change in overall performance over time as the result of an intervention.

What is a sample?



Source: William M. Trochim. 2006. "Sampling Terminology." *Research Methods Knowledge Base* (2nd ed.). <http://www.socialresearchmethods.net/kb/sampterm.php>

**If life is like eating a box of chocolates....
sampling is like ladling from a pot of soup.**



You don't have to drink the whole pot of soup to know what it tastes like, but you do have to stir it well to make sure you get a good taste. In other words, you need to sample properly to ensure you get an accurate picture.

For example, if you want to know children’s reading abilities in grade 2 and grade 3, you need to “ladle from two pots of soup”—i.e., sample children from each class.



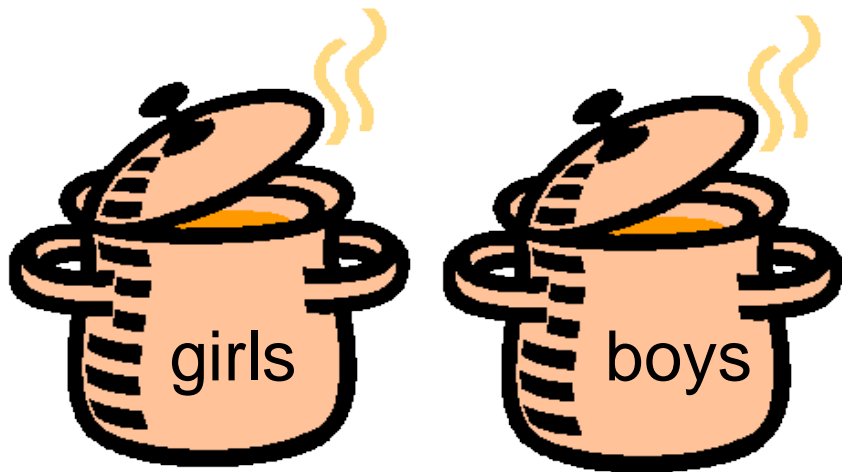
Grade 2



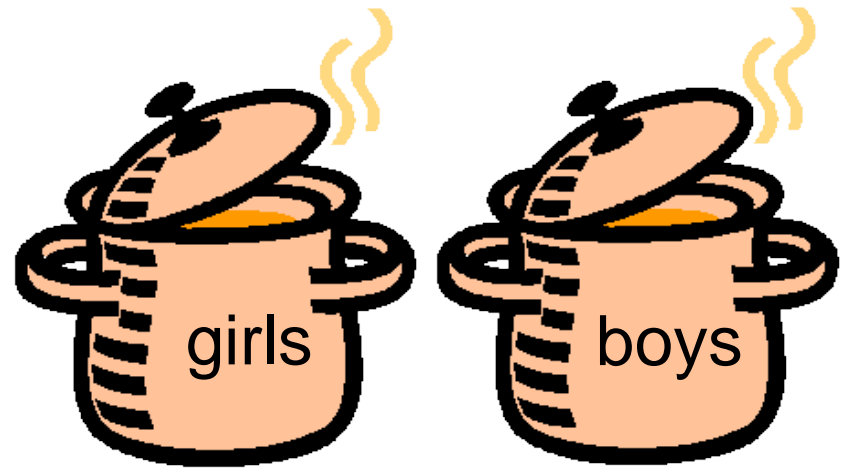
Grade 3

What if you want to know how girls in grade 2 compare with boys in grade 2? And what about boys versus girls in grade 3?

Grade 2



Grade 3



For a snapshot study

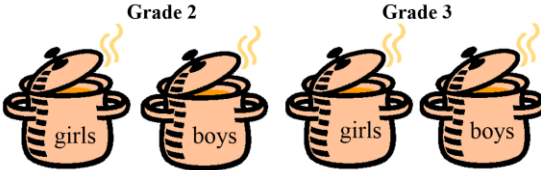
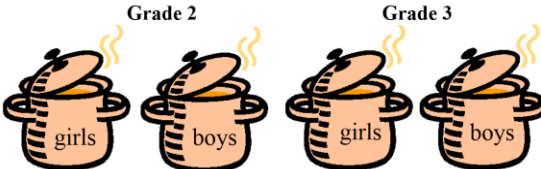
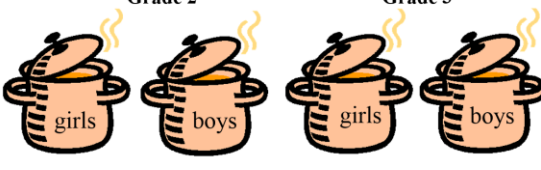

Sample cell inputs:

- Variability estimate
- Confidence interval
- Confidence level (95%)
- Design effect (for clustered samples)

Sample for an evaluation

Sample cell inputs:

- Type of data analysis
- Intraclass correlation coefficient
- Alpha (type I error)
- Power (1 – type II error)
- **Effect size**

- Province A 
- Province B 
- Province C 
- Province D 

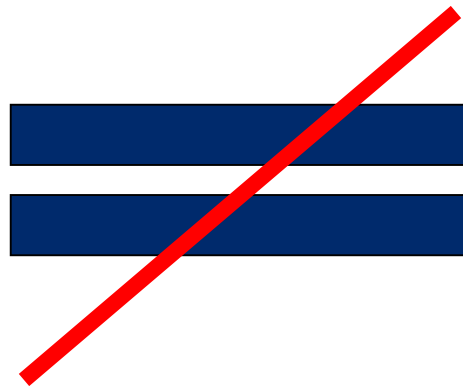
FINAL SAMPLE
= Sample cell
size multiplied by
the number of
disaggregation
categories

	Grade 2	Grade 3	Total
Public schools	400	400	800
Private schools	400	400	800
Nonformal schools	400	400	800
Total	1,200	1,200	2,400 pupils

Weights are used in data analysis to compensate for disproportionate representation of different ***types*** of clusters.

Calculating Size of a Sample Cell: Clustered

Assessing 10 students in 100 schools...



...is not the same as assessing 20 students in 50 schools.

Students tend to vary more between schools than within, so sampling more schools is usually preferable to sampling more students within the same school to obtain a target sample size.

Additional Considerations

- Amount of time available to collect data
- Distance to cover to reach schools (may need to cluster sample)
- Number of assessors you can hire
- Amount of equipment you can purchase
- **BUDGET available**

Best Practice: Engage with the Ministry of Education to obtain necessary information about schools (e.g., location, number of teachers per grade, number of students per grade, and schedule).

General EGRA Sampling Guidance

- There are **no rules of thumb**. The ***purpose*** of assessment will drive both research design and sampling.
- Sampling design is a ***compromise*** between the desire to increase precision and budgetary considerations.
- Increase in a sample size improves ***accuracy*** of population estimates.
- Expected ***effect size*** has the largest impact on sample size
- When possible, engage a statistician with good knowledge of research objectives for computing a sample.

Refer to EGRA Toolkit Annexes B–D for more detailed sampling guidance.

Research Design and Sampling: Practical Takeaways

- Research design structures
 - EGRA administration timing
 - EGRA administration intervals
 - Sample design
 - Type of data analysis
 - What we we learn
- Every design has limitations
- There are no “good” or “bad” designs but there are more appropriate or less appropriate designs for the ***purpose of research/evaluation***
- Budgets will impact designs

Data analysis

- Descriptive statistics
- Inferential statistics

Reporting principles

- Transparency
- Accessibility
- Usefulness
- Completeness

For more information...

<https://www.eddataglobal.org/>

