

# Sampling Design of Large-Scale Assessments and Implications for Data Analyses

---

Ting Zhang

American Institutes for Research

NCME | March 2023

# Agenda

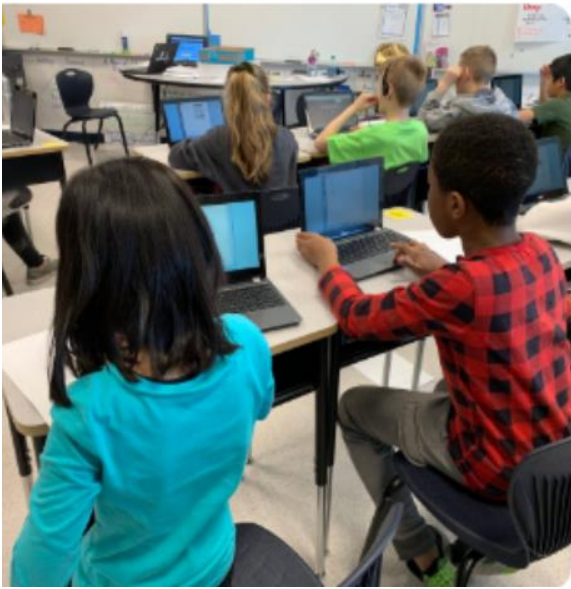
---

- Overview of large-scale assessments (LSA)
- Study designs
  - Complex sample design
  - Implication for variance estimation and sampling weights



# What are educational large-scale assessments (LSA)?

Tests that focus on measuring and monitoring what populations know and can do in academic subject areas



- Populations are usually certain ages/grades in cities, states, countries
- Subpopulation measurement (gender, SES, race/ethnicity) is also prioritized
- Academic subject areas include mathematics, science, reading, social studies, computer literacy, etc.
- Measure contextual factors associated with achievements

# Uses of LSAs

---

Provide information for achievement comparisons between educational systems or jurisdictions

- *How does the performance in one country compare with that of other countries?*

Examine trends in achievement

- *How does one country's achievement increase or decrease over time?*

Improve education by informing policy, research, and practices

- *What factors are associated with educational achievement? What can we learn from others about what works (and what doesn't)? What could be adopted by or adapted?*

# Why Specialized Software Programs are Needed for LSA Data Analysis?

---

# Study Designs

---

- Complex sample design (multi-stage, clustered sample design, sampling weights, sampling variance)
- Complex assessment design (matrix sampling design, use of Item Response Theory (IRT) and Plausible Values (PVs), and measurement variance)



# Sample Design, Sampling Variance and Weights

---

# Why Do We Use Samples?



Impossible to test everyone on everything

- Too many people
- Too many items
- Too expensive

Not necessary to test everyone on everything, e.g.,

- Blood sample
- Soup sample

Source: Sabine Meinck, Ph.D., IEA, Design of International Large-Scale Assessments and Implications for Multilevel Modeling



# You're probably familiar with simple random sampling (SRS)



- Each person has an equal probability of selection
- But, to conduct a study with nationally representative samples, SRS not feasible:
  - Time!
  - Cost!

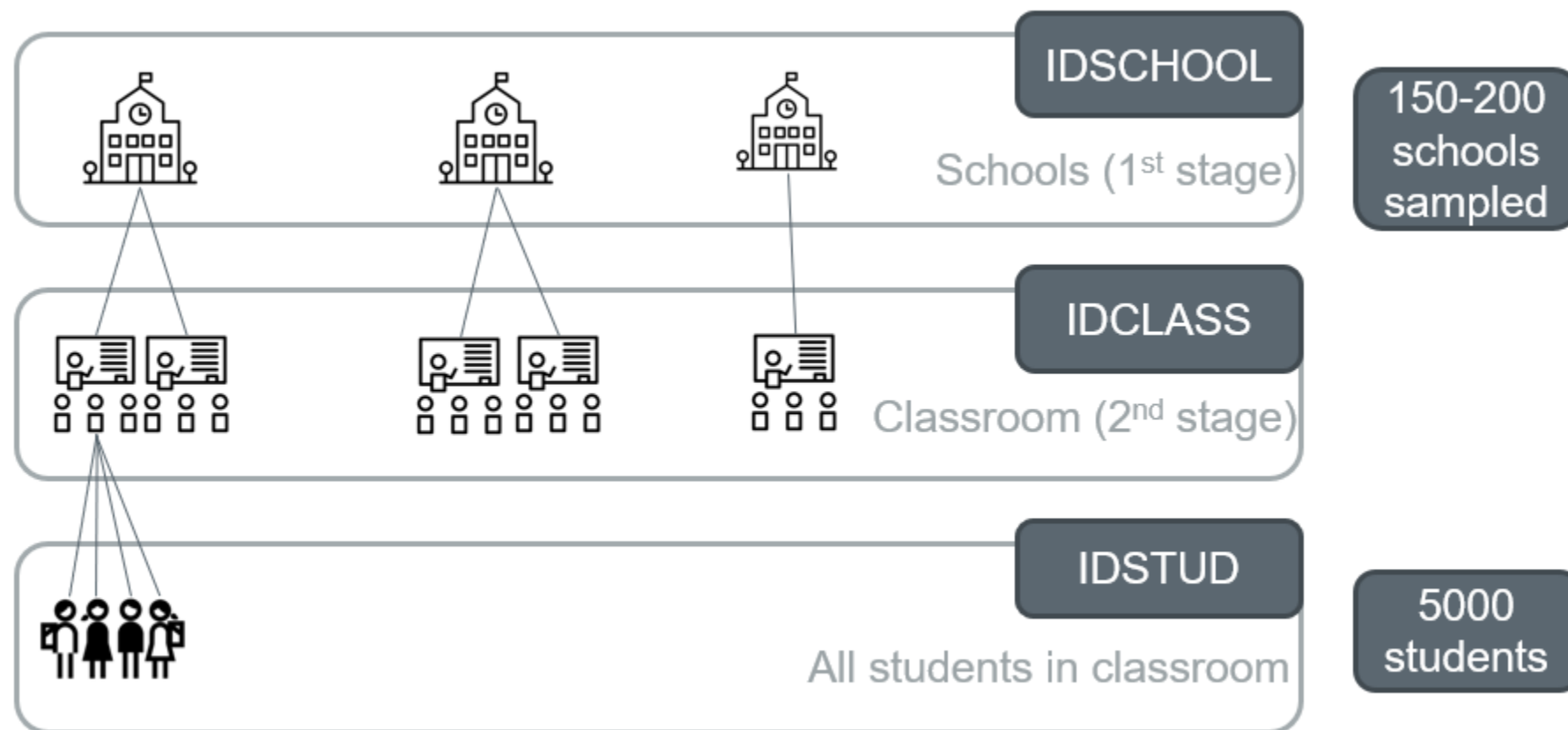
# Complex sample designs

---

Why “complex”?

- Multiple stages of stratified sampling
- Homogenous clusters are sampled
- Selection probabilities differ for different sampling units

# Multiple sampling stages (TIMSS)



# What are the implications of multistage cluster sampling?



**Cluster effect:** Individuals within classrooms/schools tend to be more similar than individuals between schools

# What are the implications of multistage cluster sampling?

- In general, the sampling variance of a clustered sample tends to be larger than the sampling variance of a simple random sample of the same size.
- In studies using a complex sampling design, standard errors tend to get larger, partly due to the sampling variance.
- Variances are essential in statistical tests of significance
  - Biased variances could make differences between scores appear significant when in fact they are not

# How to handle clustering:

---

## Three recommended methods

- Replication methods. TIMSS and NAEP use jackknife repeated replication
- Taylor series approximations
- Hierarchical linear models

# Sampling schools and students: Unequal probabilities of selection

---

- Schools are sampled with **probability proportional to their size** from the list of all schools in the population
  - Larger schools are more likely to be selected than small ones
  - Students in large schools have a lower probability of selection than classes in smaller schools
  - Result: the overall probability of selection of students is more similar across different size schools

When we sample with unequal probabilities of selection, aren't we introduce bias into our analytical results?

# Sample Weights

---

- To account for unequal probabilities of selection, sampling weights should be used in all statistical calculations where inferences are made to populations
  - Otherwise, population estimates (e.g., means, percentages) will be biased
  - Most standard software have ways to specify weights



# TIMSS Student Weights

**Final student weights = school-level sampling probability \* student-level sampling probability\*non-participation adjustments**

- Students are assigned sampling weights to adjust for over- or under-representation of particular groups in the final sample
- **Student weight is the inverse of the probability of selection**
- Students with higher weight values are representing more people in the target population
- Use of sampling weights is necessary for computation of sound, representative estimates
- Weights adjust for nonparticipation
- **Sum of the overall student weights equals the number of students in the target population**

# Probabilities and Weights

- **Student weight is the inverse of the probability of selection**
- Suppose a school has a probability of selection of 0.1 and each student within a school has a probability of selection of 0.2. What is the students' probability of selection?

School prob

Student within school prob

Joint Prob

1/10

2/10

2/100 or 1/50

- Student weight = inverse probability

$$50/1 = 50$$

**Q: If a school has a 0.2 probability of selection and once school is selected the student has a 0.05 probability of selection, what is the students weight?**

**A. 100**

**B. 200**

**C. 20**

**D. 500**

# Implications for Large-Scale Data Analysts

---

- Many statistical software packages assume the data being analyzed come from a simple random sample with independent observations and equal selection probability
- LSAs clustering of observations in schools and classes prevents them from being independent
  - Need to use the JK, Taylor series or HLM method for sampling variance
- Complex sampling leads to unequal selection probability
  - Weights need to be applied for unbiased estimates
- Statistical software exists for “complex samples”
  - NCES’s EdSurvey and Dire!

# Resources

---

[IEA TIMSS International Database](#)

[IEA International Database Analyzer \(IDB Analyzer\) and its tutorials](#)

[NCES International Data Explorer \(IDE\)](#)

[NCES \*EdSurvey\* R statistical package](#)

[NCES \*Dire\* R statistical package](#)

[NCES Distance Learning Dataset Training Modules \(DLDT\)](#)

[NCES International Activities Program website](#)

[TIMSS and PIRLS International Study Center website](#)

[IEA website](#)

[ILSA Gateway](#)

[OECD PISA website](#)

## Questions?

---

Ting Zhang, PhD  
American Institutes for Research  
tzhang@air.org