Westat

Improving Lives Through Research®

# Maximizing Overlap of NAEP School Samples to Optimize Both Trend and Cross-sectional Estimates

**Lloyd Hicks, Amy Lin, Yiting Long, Keith Rust**

WESTAT @ FCSM 2024

# NAEP Report Card on COVID

The Economist

Menu | Weekly edition | The world in brief | Search

Graphic detail | Daily chart

## Test results in American schools plummeted during the pandemic

Districts with the most virtual schooling had the biggest drop in scores

The Washington Post
Democracy Dies in Darkness

## National test scores plunge, with still no sign of pandemic recovery

5 min | 4315

News | Opinion | The74 | Video | Analysis

Explore | EDUCTION 2024 | FUTURE OF HIGH SCHOOL | ARTIFICIAL INTELLIGENCE | STEM | SCIENCE OF READING

Support The 74 and stories like this one. Donate Today!

NEWS | THE BIG PICTURE

## 'Nation's Report Card': Two Decades of Growth Wiped Out by Two Years of Pandemic

Long-term scores from NAEP show unprecedented declines for 9-year-olds in math and generational literacy loss

**Scores decline during pandemic, remain higher than 1970s**


USA TODAY

HARDEST HIT BY HURRICANES
Hurricane-prone states

GLOBAL GLIMPSES
The day in pictures

TAP INTO THE NEWS
Get the USA TODAY

U.S. | Elections | Sports | Entertainment | Life | Money | Tech | Travel | Opinion

## 'Largest score decline' in reading for nation's 9-year-olds, first-ever drop in math

Kayla Jimenez
USA TODAY

Published 12:25 a.m. ET Sept. 1, 2022 | **Updated 11:36 a.m. ET Sept. 1, 2022**

# Bottom Line Up Front (BLUF)

- Focus: Special sample design and weighting procedures implemented to better evaluate change in student achievement before and after the pandemic
- Application of Keyfitz procedures for sample overlap control
- Procedures are generalizable to other types of establishment-based samples involving at least 2 stages of selection

# Outline

- Describe NAEP, specifically Long-Term Trend
- Describe the problem and solution
- Provide background information
  - NAEP-LTT sample design
  - Keyfitz procedures
  - Special application/modification of Keyfitz
- Describe the effectiveness of solution

# What is NAEP?


- Sponsored by NCES, National Assessment of Educational Progress, produces the Nation's Report Card
- NAEP is congressionally mandated and drives educational policy and programming
- NAEP assessments have been conducted periodically in reading, mathematics, science, writing, U.S. history, civics, geography, and other subjects, beginning in 1969
- NAEP reports academic achievement at the national, state, and district levels
- NAEP has two national assessment programs
  - **Long Term Trend (LTT)** - measure students' educational progress over a long period of time
  - Main NAEP - measures students' knowledge and skills based on the most current curricula and standards

# What is LTT?

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

- Measures student achievement trends via cross-sectional estimates
- Assesses students at 3 age levels: 9, 13, and 17
- Assesses students in 2 subjects: math and reading
- Conducted approximately every 4 years
- Requires precisely replicating past procedures
- Education achievement has relied on LTT results for long time:
  - First reading results were released in 1971
  - First math results were released in 1973

# Objectives of the 2022 NAEP-LTT

NAEP-LTT had two objectives in 2022 that affect the sample design:

**The historical objective of the LTT:**

Generate estimates for 2022 to measure trends

**One-time, time sensitive objective:**

Directly measure the change in student achievement between 2020 (pre-COVID) and 2022 (post-COVID)

# Implementation challenges with the historical objective

Desired outcome

Maximize the consistency with the historic design of LTT studies

Implementation Challenge

Ensure sample design reflects 2022 student population and produces estimates that maintain trend while maximizing analytic power with 2020

# Implementation challenges with the one-time objective

Desired outcome:

Maximize the analytic power to detect differences in student achievement scores between 2020 and 2022

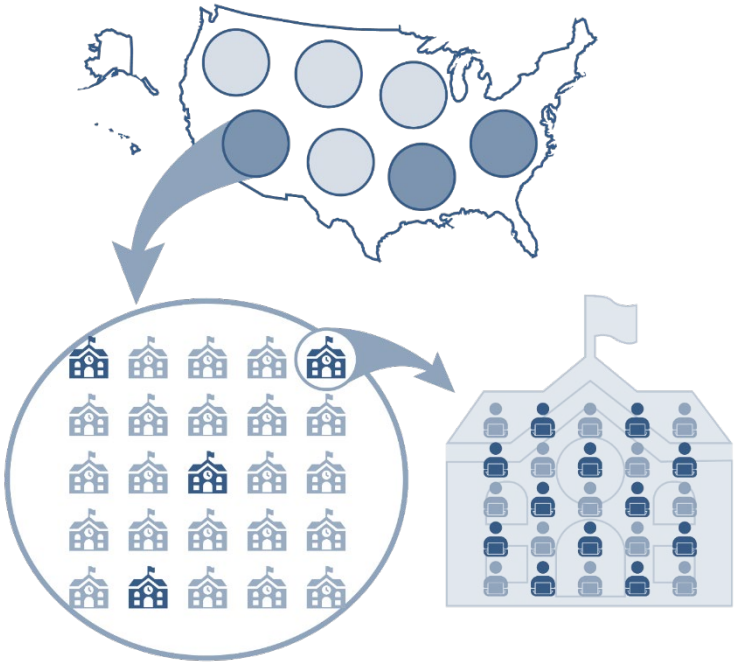
Implementation Challenge

While maximizing power, minimize the potential for increasing overall sample size (e.g., control costs)

# Sample Design Solutions to Compare 2020 to 2022

| Considered Solutions | Implication |
|---|---|
| 1. Use historical LTT sample design which requires sample independence | Will not yield the statistical power needed to detect small differences |
| 2. Use the same school sample as 2020 | Design would not reflect the 2022 student population |
| **3. Select school sample that both represents the 2022 population and achieves maximum overlap with the 2020 school sample (Keyfitz procedure)** | Maximizes power, and controls costs by controlling number of schools included in sample |

# LTT Sample Design | Overview



- Four-step approach:
  1. Selection of PSUs
  2. Selection of schools
  3. Selection of students
  4. Assignment of assessment subject

# LTT Sample Design | Sample Size & School Probabilities

- Target student sample size (assessed, per age)

| School Type/ Subject | 2020 | 2022 |
|---|---|---|
| Public | | |
| Reading | 7,200 | 7,200 |
| Math | 7,200 | 7,200 |
| Private | | |
| Reading | 800 | 800 |
| Math | 800 | 800 |

- School selection probabilities (PPS)

$$P_i = b * MOS_i, where$$

*b* is the "constant of proportionality"

$$b = \sum f(T_O, T_{sch}, MOS_i), where$$

$T_O$: overall target student sample size

$T_{sch}$: within school target student sample size

$MOS_i$: estimated school enrollment for given age

*EPSEM design*

- Number of schools in sample is **not fixed**
- It is determined by target no. of students

2020 and 2022 student sample sizes are same

# Keyfitz requires modification to address dual objectives

- Use Keyfitz Methods to maximize school overlap
  - This is a common procedure used for sample overlap control (Ernst, 1999)
  - This procedure was introduced in 1951 (Keyfitz, 1951)
  - It can be used to minimize overlap or maximize overlap
  - It can be used for overlap control with one sample or several samples (Chowdhury, 2000)

However, Keyfitz methods, by themselves, will not ensure all 2020 schools are retained in the 2022 sample

# Review of Standard Keyfitz Procedure

- Keyfitz procedure is based on Bayes Theorem for conditional probabilities

$$P_i(s_2) = school\ probability\ of\ selection\ for\ sample\ s_2$$

$$P_i(s_2) = P_i(s_2|s_1) \times P_i(s_1) + P_i(s_2|\widehat{s_1}) \times P_i(\widehat{s_1})$$

Legend:
$\mathbf{s_2}$ is the 2022 sample
$\mathbf{s_1}$ is the 2020 sample
$\widehat{s_1}$ is not in the 2020 sample

- In our context, the probability is based on whether a school in the 2022 frame was selected for the 2020 sample or not
- This formula is optimized when $P_i(s_2|s_1) = 1$ and $P_i(s_2|\widehat{s_1}) = 0$

$$P_i(s_2|s_1) = \min[1, \frac{P_i(s_2)}{P_i(s_1)}]$$

$$P_i(s_2|\widehat{s_1}) = \max[0, \frac{P_i(s_2) - P_i(s_1)}{P_i(\widehat{s_1})}]$$

For dual-objectives, we want $P_i(s_2|s_1)$ to resolve to 1 so that schools are retained in sample with certainty

# To meet dual objectives, need to modify the Keyfitz procedure

- **What is the limitation of Keyfitz?**
  - $P_i(s_2|s_1)$ can be less than 1 whenever $P_i(s_2) < P_i(s_1)$
  - Implication: a 2020 sample school is not guaranteed to be in sample in 2022 if its unconditional 2022 school probability is less than its 2020 school probability
- **What can result in $P_i(s_2|s_1) < 1$ ?**
  - This can easily happen when a school's enrollment decreases (recall schools are selected with probability proportional to enrollment)
- **How do we ensure the school is retained in sample with certainty?**
  - Ensure that its 2022 probability of selection is at least as large as its 2020 probability
  - That is, **$P_i(s_2)'$** = max $[P_i(s_2), P_i(s_1)]$

# Considerations in Applying our Approach

- By increasing the $P_i(s_2)$ and consequently the Keyfitzed probability, the expected number of schools in the 2022 sample will be larger than 2020
  - Increases the cost of fielding the student assessments
  - The extent of the increase depends on the number of schools where enrollment declined from 2020 to 2022
- To avoid biasing the sample, must apply this maximum function to **all** schools on the 2022 frame:
  - schools sampled in 2020,
  - schools **not** sampled in 2020
  - schools new to the frame in 2022

# Benefit of Our Solution

- Meets dual-objectives
- Minimizes the number of additional schools sampled
- Increases the analytic power to detect differences (reduces standard errors)
- Simple to apply

# Implications on Weighting & Variance Estimation

- School weight ($W_i$) computed as

$$W_i = \frac{1}{P_i(s_2)'}, where$$

$$P_i(s_2)' = \max[P_i(S_2), P_i(S_1)]$$

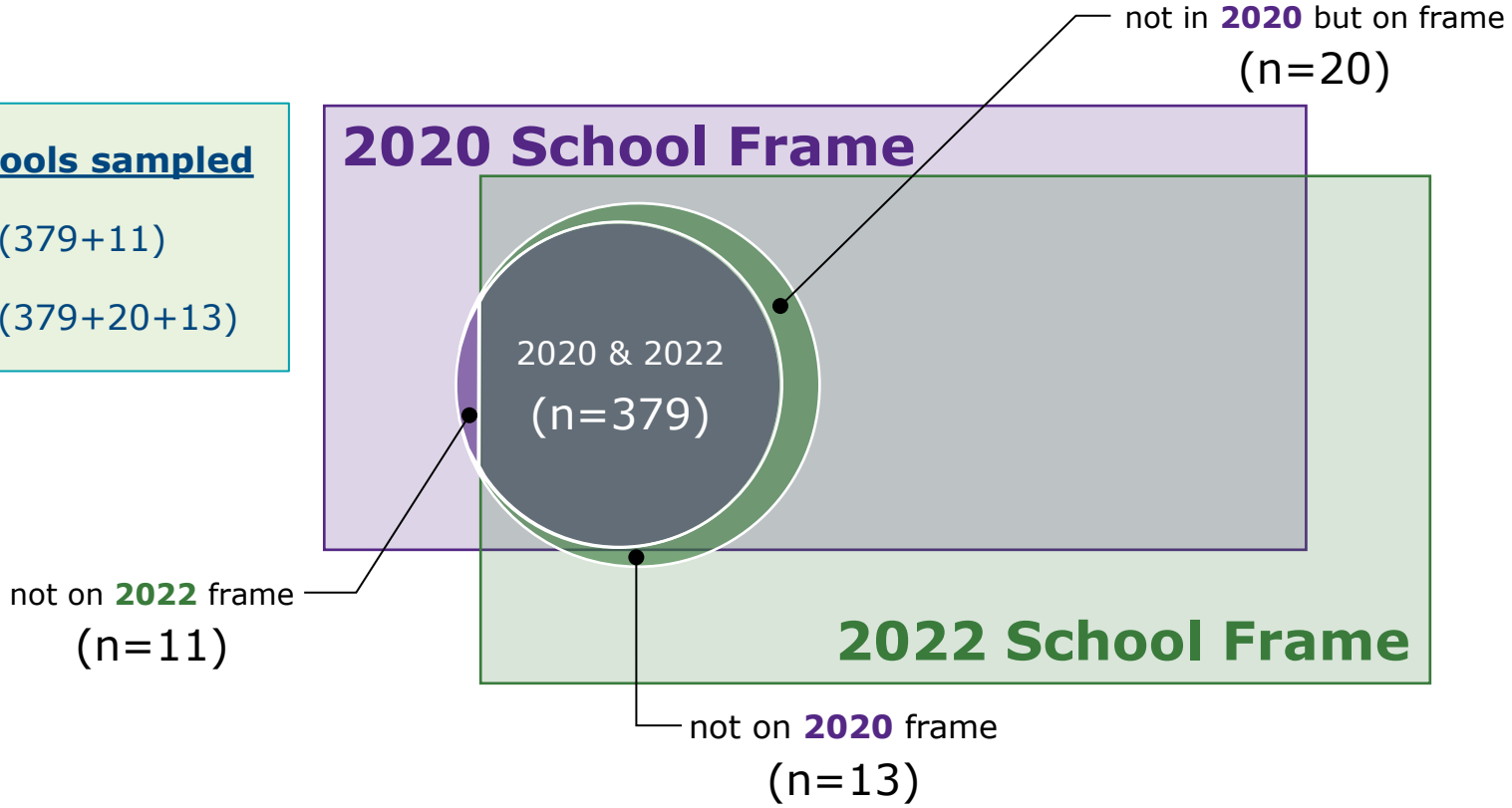
- Used same variance structure as 2020 to capture covariance component in the variance calculation of differences

# Results | Public Schools for Age 9

**Public schools sampled**

2020: 390 (379+11)

2022: 412 (379+20+13)



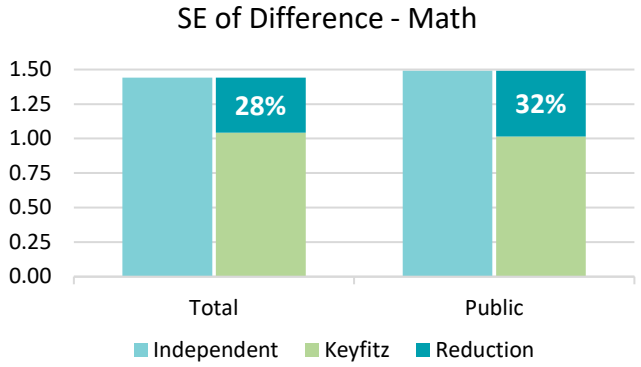
# Results | Counts of Sampled Schools for Age 9

## Total count of schools sampled, by Public and Private

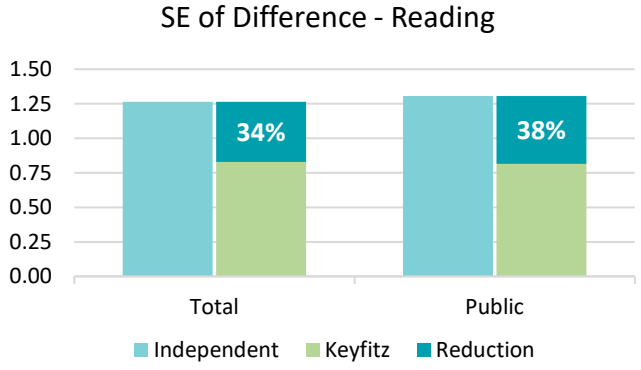
| | Public | Private | Total |
|---|---|---|---|
| **2020 (LTT design)** | **390** | **153** | **543** |
| No longer exist in 2022 | 11 | 21 | 32 |
| 2022 (usual LTT design) | 392 | 153 | 545 |
| **2022 (modified Keyfitz)** | **412** | **163** | **575** |
| In 2020 sample | 379 | 132 | 511 |
| Not in 2020 but on frame | 20 | 8 | 28 |
| Not on 2020 frame | 13 | 23 | 36 |

# Results | Mean Achievement Scores and SEs for Age 9

| Math | Mean | | SE | | Diff | SE (Diff) | |
|---|---|---|---|---|---|---|---|
| | 2020 | 2022 | 2020 | 2022 | | Ind. | Keyfitz |
| Total | 241 | 234 | 0.8 | 1.2 | -7.4 | 1.4 | 1.0 |
| Public | 241 | 233 | 0.9 | 1.2 | -7.7 | 1.5 | 1.0 |




| Reading | Mean | | SE | | Diff | SE (Diff) | |
|---|---|---|---|---|---|---|---|
| | 2020 | 2022 | 2020 | 2022 | | Ind. | Keyfitz |
| Total | 220 | 215 | 0.8 | 1.0 | -5.1 | 1.3 | 0.8 |
| Public | 219 | 213 | 0.9 | 1.0 | -5.5 | 1.3 | 0.8 |



# Thank you

LloydHicks@westat.com
AmyLin@westat.com
YitingLong@westat.com
KeithRust@westat.com

westat.com

Photos are for illustrative purposes only. All persons depicted, unless otherwise stated, are models.

# NAEP Report Card on COVID

- Test results in American schools plummeted during the pandemic (source: The Economist)
- National test scores plunge, with still no sign of pandemic recovery (source: Washington Post)
- 'Largest score decline' in reading for nation's 9-year-olds, first-ever drop in math (source: USA Today)
- 'Nation's Report Card': Two Decades of Growth Wiped Out by Two Years of Pandemic (source: The 74)

# Citations


- L. R. Ernst, The maximization and minimization of sample overlap problems: A half century of results, Proceedings of the International Statistical Institute (1999) 168–182.
- N. Keyfitz, Sampling with Probabilities Proportional to Size: Adjustment for Changes in Probabilities. Journal of American Statistical Association (1951), 46, 105-109.
- S. Chowdhury, A. Chu, S. Kaufman, Minimizing overlap in NCES surveys, Proceedings of the Survey Research Methods Section, American Statistical Association (2000) 174–179.