

Symposium on Options for the Future of the NAEP Long-Term Trend Assessment

Thursday, March 2, 2017

Hilton Alexandria Old Town

Speakers



Joe Willhoft



Ed Haertel



Jack Jennings



Lou Fabrizio



Ina Mullis



Andrew Kolstad



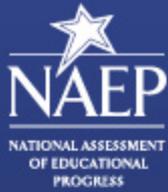
Peggy Carr

Main and Long-Term Trend NAEP

	Main NAEP	LTT NAEP
Origin	Early 1992	Early 1970s
Content	Frameworks	Objectives Booklets
Samples	Grades 4, 8, 12	Ages 9, 13, 17
Reporting	National, State, TUDA	National
Periodicity	2 Years	4 Years+

Today's Program

- ★ Paper presentations
- ★ NCES perspective
- ★ Break
- ★ Panel discussion
- ★ Question-and-answer session
- ★ Recommendations and next steps
- ★ #NAEP



The Future of the NAEP Long-Term Trend Assessments

Edward H. Haertel, Ph.D.

Thursday, March 2, 2017

Hilton Alexandria Old Town

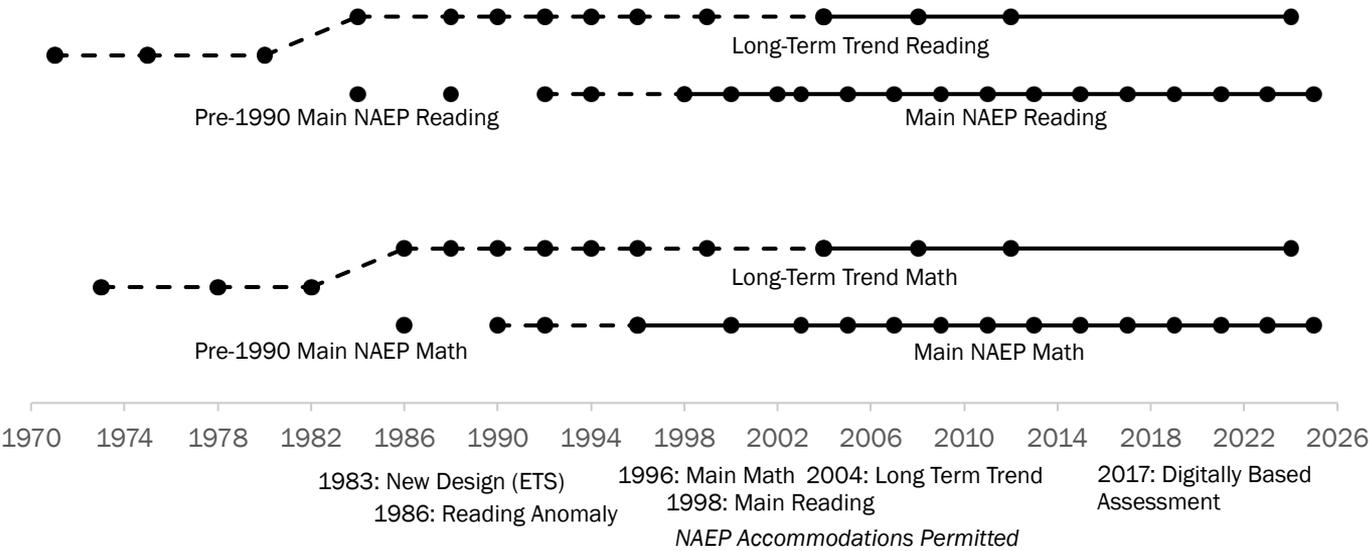
Introduction

- Quick History
 - Objectives-based versus construct-based measurement
 - Differences from main NAEP
- Why Preserve the LTT?
 - Statutory requirement
 - Abiding public/policy interest in “basic skills”
 - Value in tracking age-based cohorts
 - Preserving a legacy
- Shoring Up the LTT
 - Defining LTT constructs
 - Bridging to digital platform and common testing window
 - Ongoing support

Introduction

- Vision for the Future
 - Documenting and supporting past and current uses
 - Envisioning new uses
 - Expanding LTT potential

Overview of LTT History



From Objectives to Constructs

- Early NAEP (Pre-ETS days)
 - Lists of objectives, each operationalized by exercises (items)
 - Individual exercises viewed as being of intrinsic interest
 - Reporting focus is performance on single exercises
- Modern NAEP (since 1983; especially since 1992)
 - Frameworks and specifications define assessment constructs
 - Items viewed as interchangeable indicators of constructs
 - Reporting focus is distribution of scale scores

LTT Today versus Main NAEP

- Age-based versus Grade-based Sampling
 - 1/1 to 12/31 for 9-year-olds
 - 1/1 to 12/31 for 13-year-olds
 - 10/1 to 9/30 for 17-year-olds (9 months younger)
- Different Testing Windows
 - Fall for 13-year-olds
 - Winter for 9-year-olds (closely matches main NAEP)
 - Spring for 17-year-olds
- Different Contents/Processes Assessed
 - Lower-level, simpler
- Different Item Formats
 - More multiple-choice on LTT
 - Differences increasing as main NAEP moves to digital platform

Why Preserve the LTT?

- Statutory requirement
- Abiding public/policy interest in “basic skills”
 - Less “aspirational” than main NAEP
 - More stable than main NAEP
 - May be better targeted to younger (age-based) cohorts
 - Quantifies potential trade-offs as curricula aim higher
- Value in tracking age-based cohorts
 - Policy questions re differential grade retention, trends in school enrollment age
- Preserving a legacy
 - Longer trend lines are increasingly valuable
 - LTT is a neglected resource which, if developed, can enhance meaningfulness and utility of NAEP program

Shoring Up the LTT

- Defining LTT constructs
 - Task for NAGB
 - Begins with review of LTT exercises and objectives
 - Guided by curriculum expertise and empirical analyses
 - Process of creation as well as discovery
- Bridging to digital platform and common testing window
 - One two-group bridge study at each of three age levels
 - Needs to happen soon
- Ongoing support
 - Modest but sustained investment in item development
 - Regular and predictable schedule for LTT administrations

Vision for the Future

- Documenting and supporting past and current uses
 - NCES and other LTT reports
 - Secondary analyses of LTT data
 - Policy interpretations and uses
- Envisioning new uses
 - More fully exploiting contrasts between LTT age cohort trends and main NAEP grade cohort trends
 - Linkages to create super-long trend lines
 - ???
- Expanding LTT potential
 - Enhancements to support longitudinal research
 - Relating (to be created) LTT Frameworks to contemporary workplace needs, CCR, policy concerns

Thank You



Jack Jennings

Former President and CEO, Center on Education Policy

Is it Time to Retire Long-Term Trend?

Louis M. (Lou) Fabrizio, Ph.D.
Director of Data, Research & Federal Policy
NC Department of Public Instruction
NAGB Symposium, March 2, 2017

Preface

- Over 20 years working for the state of North Carolina
- My opinions
- Not speaking on behalf of the State Board of Education, State Superintendent of Public Instruction or NAGB
- Offering state perspective in support of main NAEP (national, state-by-state and TUDA)

Main Conclusions

- LTT should be discontinued
- Convince U.S. Congress to remove requirement for LTT
- Pursue research into enhancing reporting of main NAEP (and correcting problems of private school participation)

Main Reasons

- LTT is out of date
- LTT doesn't provide useful information at the state level
- Anti-testing environment
- LTT not worth the resources required to update

Main Reasons (continued)

- Main NAEP generates benchmark data for states and large school districts (TUDA)
- Main NAEP information more useful to state-level policymakers and general public within states
- Congress supports main NAEP (required reporting in NCLB and ESSA)

Content of the Long Term Trend Assessments Compared to Main NAEP

Ina V.S. Mullis, Ph.D.

Boston College

National Assessment Governing Board Symposium on Options for
the Future of the NAEP Long-Term Trend Assessment

March 2, 2017

Main NAEP World Class

- Nation and states
 - Innovative digital formats
 - Challenging content
 - Forward-looking frameworks
-

LTTs Out of Date

- Both LTTs almost wholly multiple choice – no innovative items
 - Reading passages short, not authentic texts, no current topics or online material
 - Mathematics little or no problem solving
-

LTTs Too Easy

- Large proportion overlap items: ages 9-13, 13-17, and 9-13-17
- Puts a ceiling on item difficulty
- Especially noticeable at age 17 – TIMSS 8th grade level

LTTs Misunderstood

Urban legend

- LTTs tested same items since 1971
- Sadly, not true
- Worse yet, difficult to tell what the LTTs actually do measure

Origins of the LTTs

- In 1990, NAEP state-by-state assessments required completely new, modern assessments – main NAEP
- New frameworks led by CCSSO and NAEP Governing Board
- Subsets of existing 1980 assessments kept to measure trend back to 1971 – LTTs

LTTs 1970s to present

- NAEP had a rocky start – lots of specific objectives, each item would tell its own story
- Some items released and others retained – keeping items only way to measure trends
- By 1980s at ETS, assessments improved but still some original items (small budget)
- Subsets of 1980s content “frozen” until today

LTTs Irrelevant and Possibly Invalid

Today's global society greatly changed
from 45 years ago

- LTTs outdated and misunderstood
- Can “saving” the LTT's be the best use of scarce resources?

Conclusion

- High quality main NAEP nearly 20 years of trends
- No longer high quality LTTs have done their job of linking back to 1971

Let them retire gracefully – use available resources for future NAEP

Thank You!

Ina V.S. Mullis, Ph.D.

Boston College

National Assessment Governing Board Symposium on Options for
the Future of the NAEP Long-Term Trend Assessment

March 2, 2017



A Rescue Plan for NAEP's Long-Term Trend Assessments

Andrew Kolstad, Ph.D.

*Principal Statistician, P20 Strategies LLC, and former Senior
Technical Advisor to the Assessment Division, NCES*

Questions about the NAEP-LTT

1. Should they be continued or dropped?
2. If continued, can the LTT and main NAEP be integrated into a single, dual-purpose assessment?
3. If continued, should the LTT switch to a digital assessment format and a common testing window?
4. How can continuity of the trends be assured across such administrative changes?
5. If continued, how should the defects of the limited materials that serve as frameworks be remedied?

Integrated into main NAEP?

- The technical challenges in merging the main NAEP and long-term trend assessments are insurmountable
 - The cognitive measurements of the two NAEP assessment programs are too different to integrate
 - The target populations of the two NAEP assessment programs are too different to be merged

Switch to digitally based forms?

- The long-term trend assessments should switch to a digital format, while maintaining their current content
 - Resource savings would derive from not having to maintain separate facilities for paper test forms
 - NAEP's capacity to handle paper-based test forms may be lost before 2024 — the next scheduled administration

A common testing period?

- The long-term trend assessments for 13- and 17-year-olds should be administered during the normal testing period of January to March
 - This change would conserve future resources, because many operations could be shared with main NAEP
 - The average ages of the student populations at the time of being tested would change

Need for a bridge study

- In a bridge study, the long-term trend assessments would be administered to two different samples
 - Each would be drawn from a common age-based population and would use common test items, but
 - One would use paper forms during the old testing periods and the other would use digital forms during the main NAEP testing period

Repairing the LTT frameworks

- The goal of such a Board-managed project would be to retrofit explicit frameworks and item specifications to the existing objectives
- Rewritten frameworks would better inform the public about the content of these assessments and their differences from main NAEP

Uses of retrofitted frameworks

1. Guidance for SD/ELL students on appropriate accommodations
2. Guidance for item writers ...
 - to develop cognitive exercises that replace those that are released
 - to fill intended areas with exercises that are currently missing from the item pool
3. The ability to set achievement levels for the long-term trend assessments

Symposium on Options for the Future of the NAEP Long-Term Trend Assessment

PEGGY G. CARR, PhD
Acting Commissioner
March 2, 2017

INSTITUTE OF EDUCATION SCIENCES
NATIONAL CENTER FOR
EDUCATION STATISTICS

FACTORS TO CONSIDER

1. Mode

- Paper and Pencil
- Digital
 - Transadapted
 - 'Paper-to-screen' items

2. Testing window and test length

- Maintain current testing window and test length
- Move to main NAEP window
 - Maintain test length* and spiral within school
 - Adjust test length and spiral within session

3. Test content and framework

* LTT currently features three 15 minute blocks

1. MODE: P&P vs. DIGITAL

- Disappearing infrastructure supporting paper-and-pencil assessments
 - Greater cost
 - More difficult risk management
- Digital options and related validity issues
 - Transadaptation versus 'paper-to-screen' items

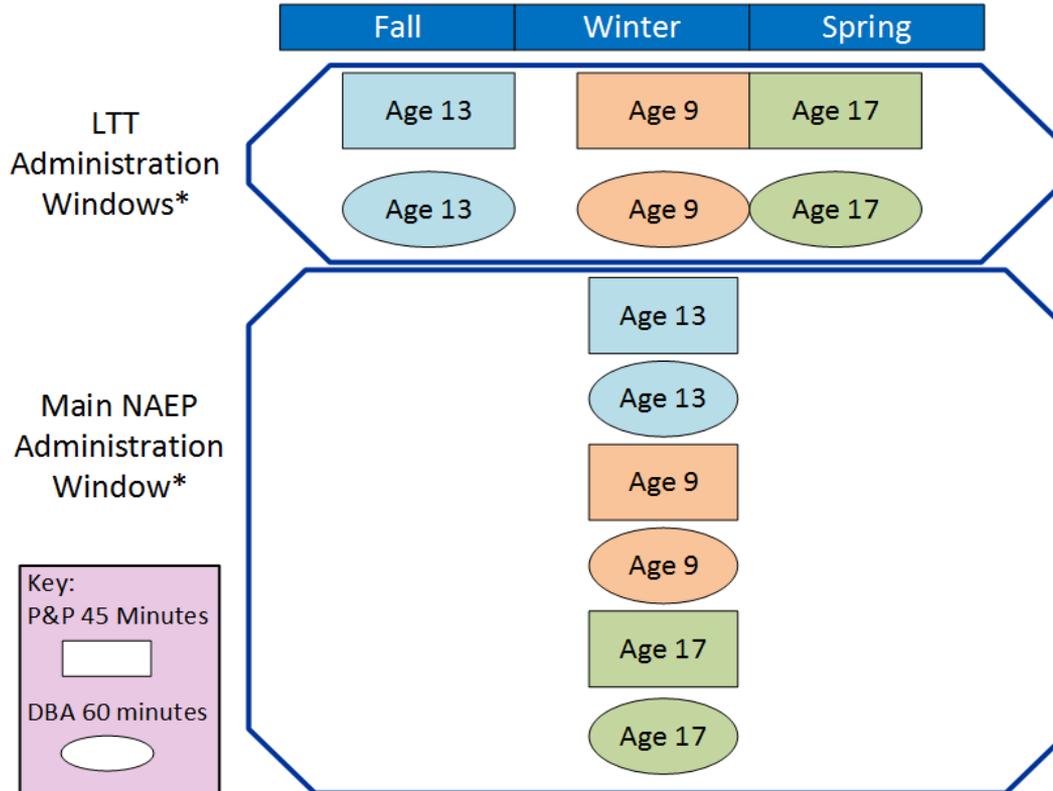
2. TESTING WINDOW & TEST LENGTH

- Challenges with moving testing window to main NAEP window:
 - Cost inefficiencies with:
 - Age-based samples in different schools from main NAEP
 - Training administrators
 - Different test length
 - LTT: Three 15-minute blocks
 - Main NAEP: Two 30-minute blocks
 - Need for multiple bridge studies to account for changes in modes, windows, test length

RISKS AND COSTS ASSOCIATED WITH BRIDGE STUDIES

- Errors are cumulative and difficult to disentangle
- Rapidly changing levels of students' digital literacy require conducting bridge studies sooner than 2024
- Cost implications:
 - Many samples needed for data collection
 - Scheduling and training of test administrators

POSSIBLE BRIDGE DESIGN

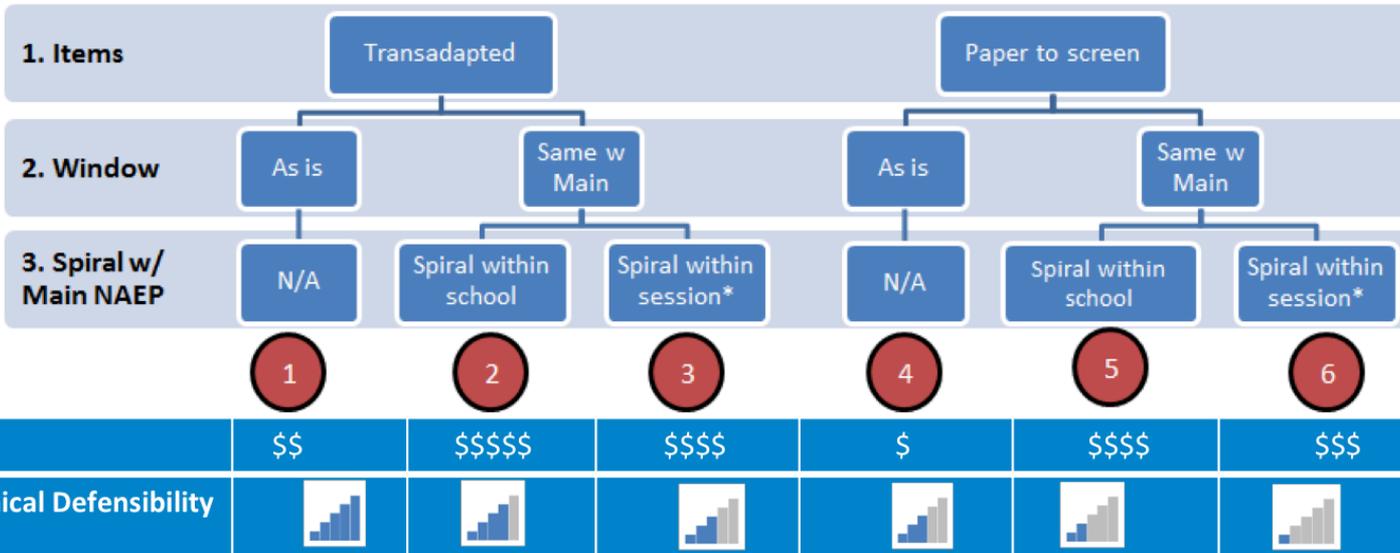


*Sample size: 90,000 students; 6,000 schools

3. TEST CONTENT AND FRAMEWORK

- Long-Term Trend more skills-oriented compared to main NAEP
- More challenges in reading
 - Major shifts in instruction leading to misalignment
 - An updated framework will risk altering construct being measured

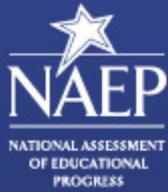
DESIGN OPTIONS



* Requires adjusting test length (currently three 15 min blocks)

INSTITUTE OF EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS



Symposium on Options for the Future of the NAEP Long-Term Trend Assessment

Panel Discussion



