Process Data — the New Frontier for Assessment Development:

Rich New Soil or a Quixotic Quest?

Stephen Provasnik
National Center for Education Statistics, U.S. Department of Education
6 May 2019

Overview

- What we discussed at the ETS symposium on Process Data in Washington, DC last December
- What I think we might be able to agree on in regards to uses for Process Data
- Where we could go by venturing into this new land

Logfiles vs. process data

- Logfiles - everything captured in a digital-based assessment (DBA)
  - from the order and speed of inputs (e.g., clicks and keystrokes) to the VPN of the device used to take the assessment.
- Process Data - the empirical data that reflect the process of working on a test question
  - reflecting cognitive and noncogitive, particularly psychological, constructs.

TIMSS Video Studies (late 1990s)

Recordings of selected classroom lessons were coded to indicate (among other many things):

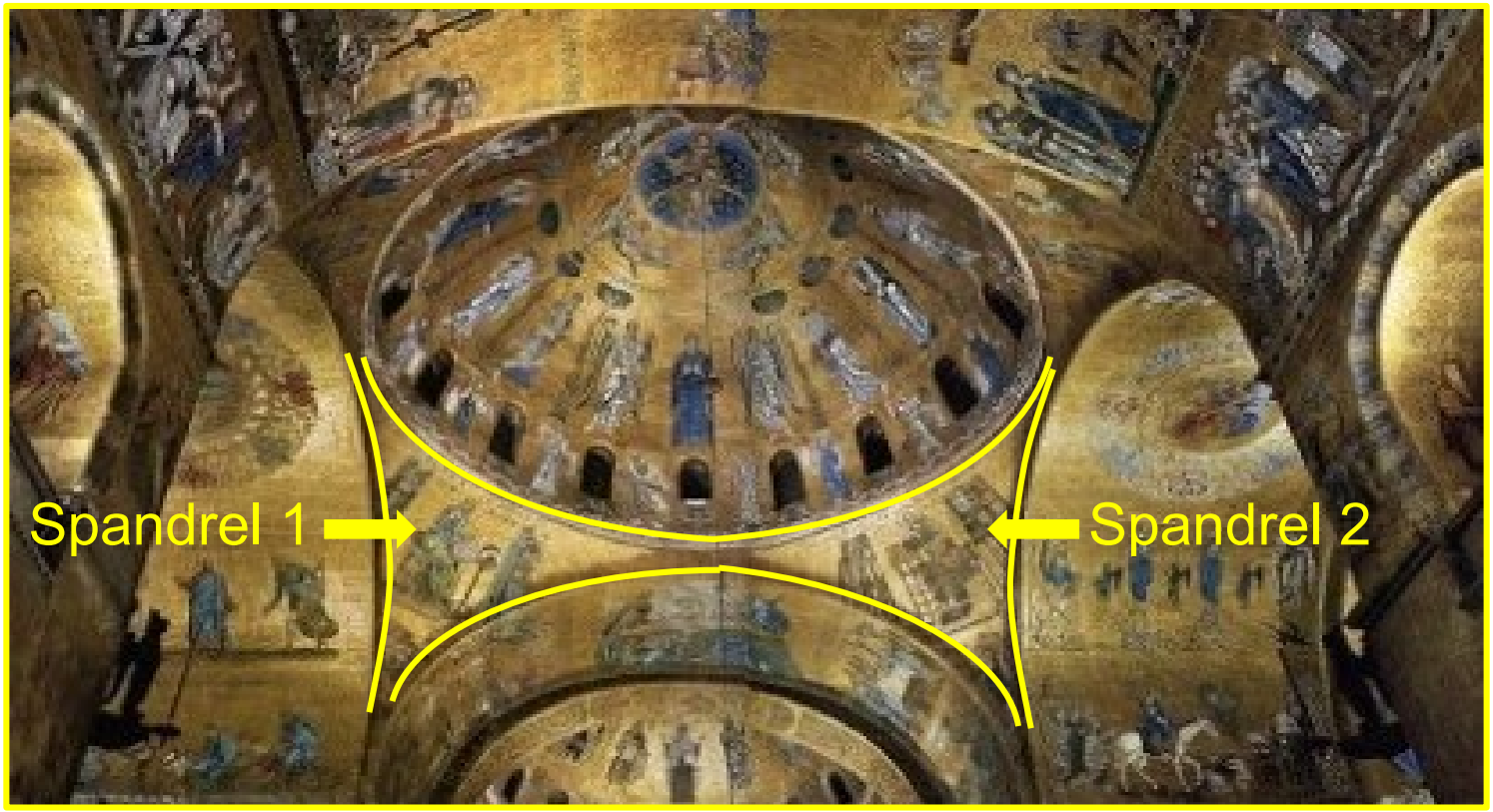
- the assigned type of work, with categories of “whole-class work,” “individual work,” “pair/partner work,” and “small-group work”
- the “number of words” the classroom teacher used in a ratio to the number of words students used, when talking to the whole class
- the proportion of the lesson spent on review of previous content vs. spent explaining new content

Conclusions from the last December's symposium

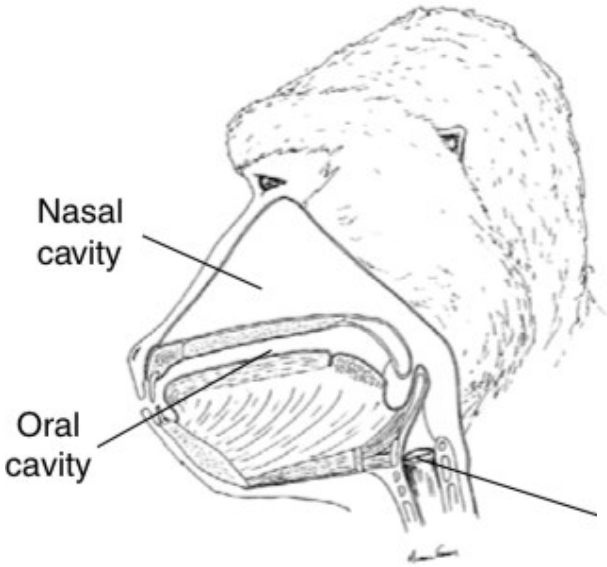
1. Develop a systemic approach to logfiles—to answer the question of what exactly logfiles should capture
2. Develop a theory for Process Data—to answer the question of how to use process data
3. Develop guidelines and standards for how to convert logfiles into process data

Spandrels of San Marco

Spandrels of San Marco

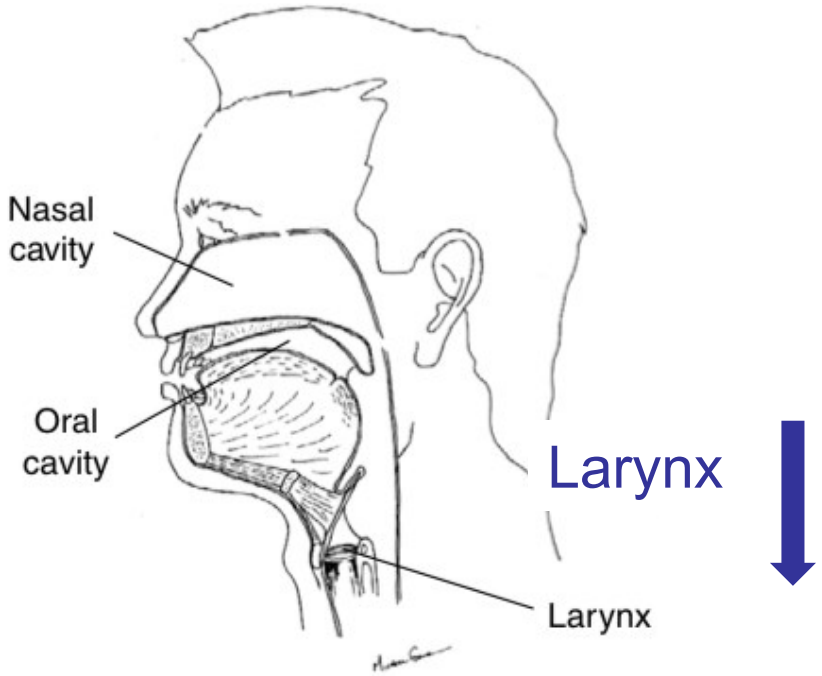


Spandrels of San Marco



Source: https://www.cell.com/current-biology/pdf/S0960-9822(08)00371-0.pdf

Current Biology Vol 18 No 11
R458



Source: https://www.cell.com/current-biology/pdf/S0960-9822(08)00371-0.pdf

Ongoing Evolution in Assessment

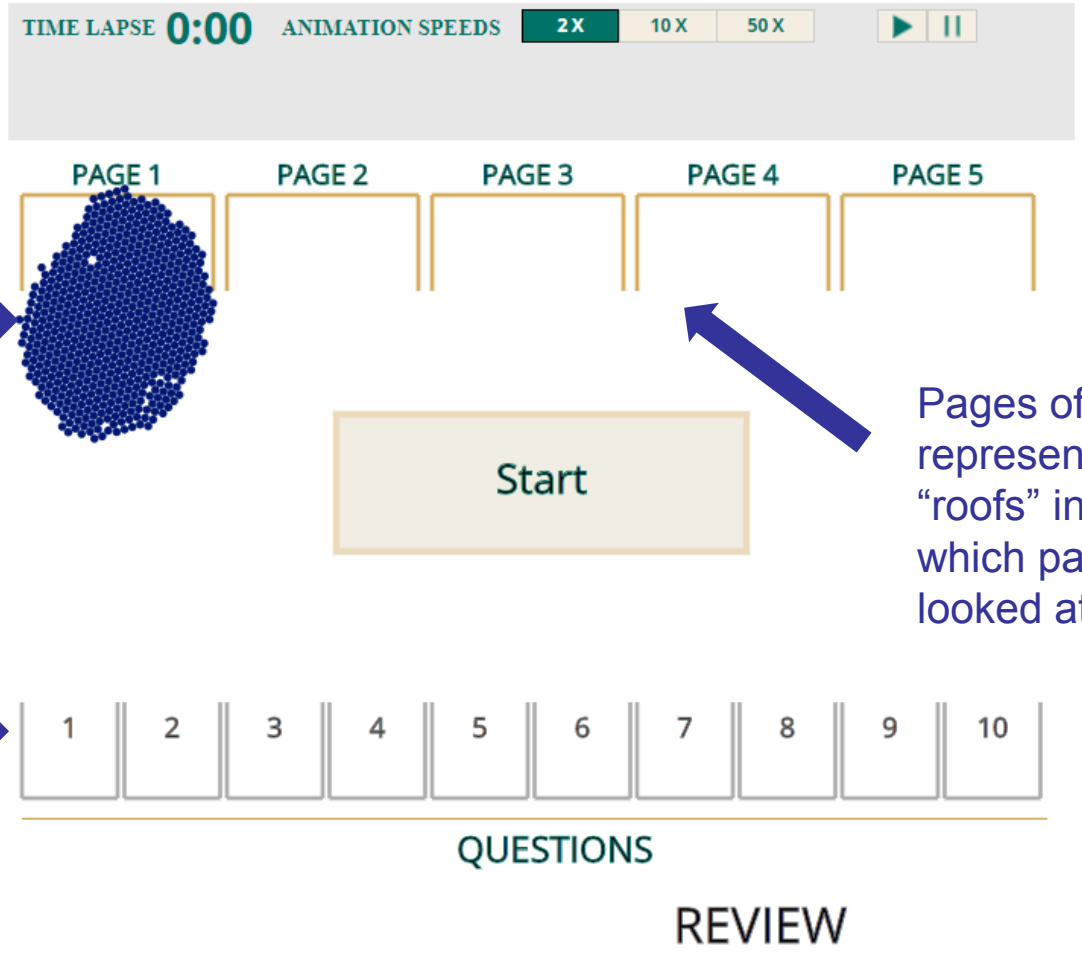
	Past	Present	*Future*
Item Development	Labor Intensive	Labor Intensive	*Automatized*
Item types	Generic	Enhanced	*Real-life*
Test design	Static	Semi-static	*Data-driven*
Test assembly	Labor Intensive	Semi-automatized	*Automatized*
Accessibility	Limited	Universal design	*Adaptive*
Timing	Not measurable	Measured	*Used*
Pathways	Not observable	Observable	*Modeled*
Validity	Content/core-based	Construct based	*Process based*
Feedback	Summative	Summative	*Diagnostic*

Diagnostic or forensic applications

These include using logfiles and process data to improve data quality

- by helping understand how items function and what variables make items more difficult or more reliable items
- by distinguishing among “missing” answers which are
  - “not reached” (never seen)
  - “omitted” (seen, taken time over, but ultimately skipped)
  - “not attempted” (seen, but not time taken before being skipped)
- by identifying student guessing or cases that are outliers, which may indicate possible cases of cheating, or cases of programming error

Visualization of NAEP reading patterns from sampled logfiles



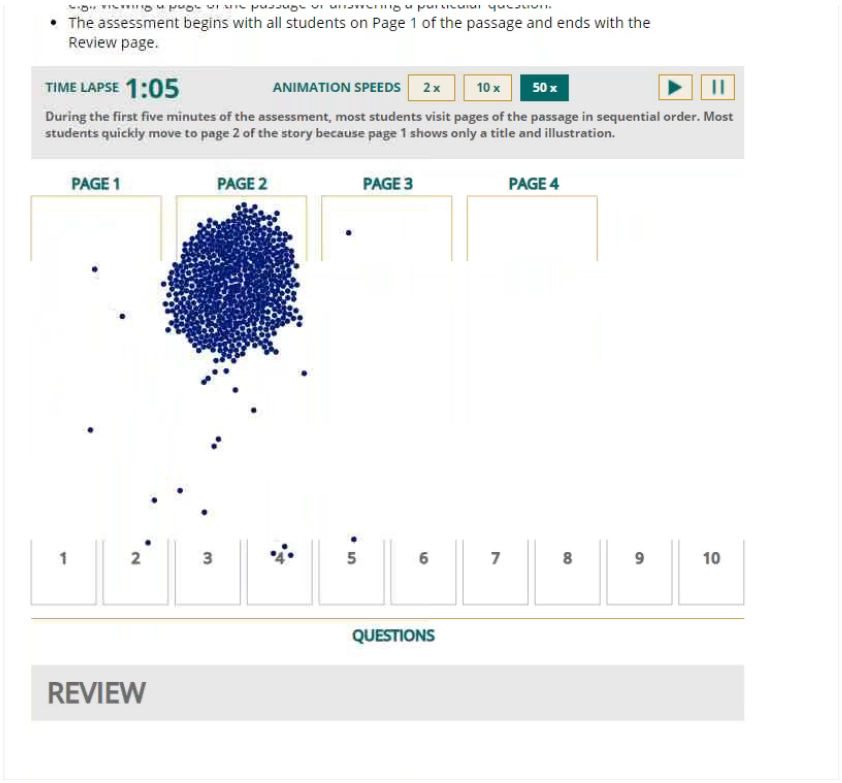
Each sampled student represented by a blue dot.

Pages of text represented by "roofs" indicating which page being looked at.

Ten test questions represented by "buckets"

NATIONAL CENTER

Visualization of NAEP reading patterns from sampled logfiles



Diagnostic or forensic applications

These include using logfiles and process data to improve data quality

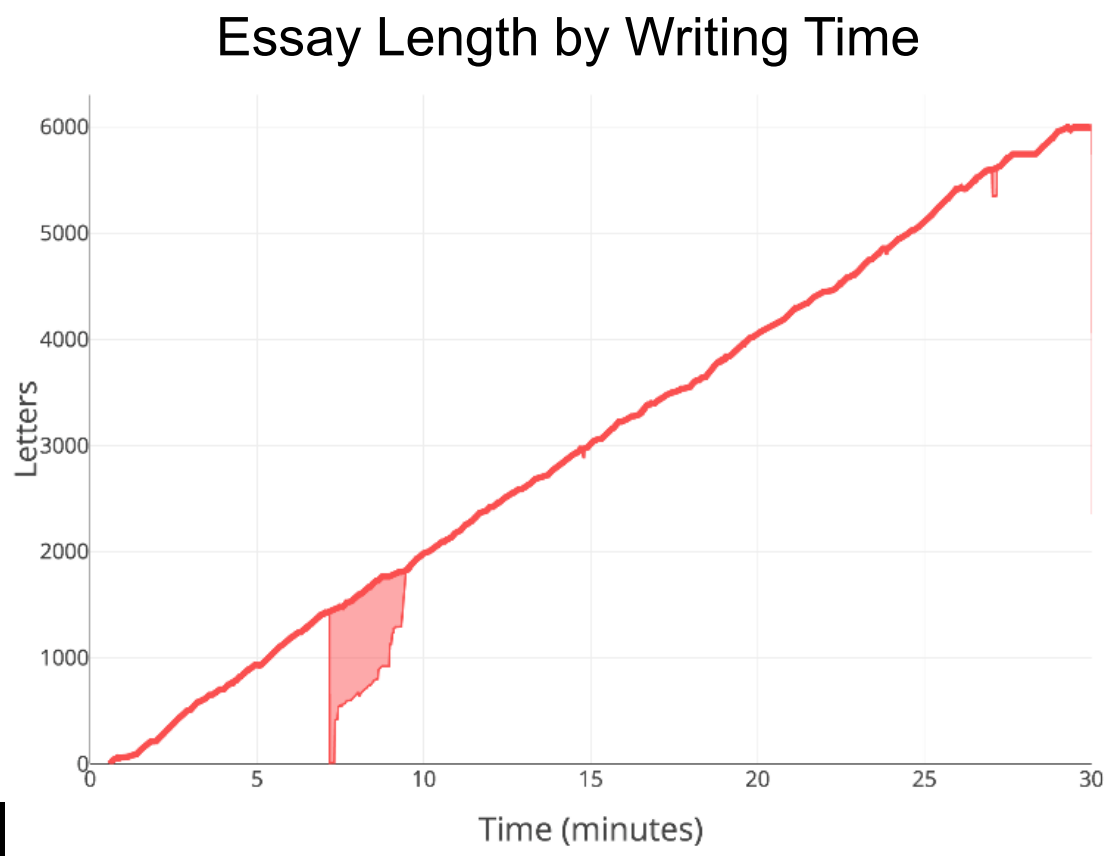
- by helping understand how items function and what variables make items more difficult or more reliable items
- by distinguishing among “missing” answers which are truly “not reached” (never seen), which should be “omitted” (seen, taken time over, but ultimately skipped), and which are “not attempted” (seen, but not time taken before being skipped)
- by identifying student guessing or cases that are outliers, which may indicate possible cases of cheating, or cases of programming error

Research into understanding respondent behaviors and cognitive strategies

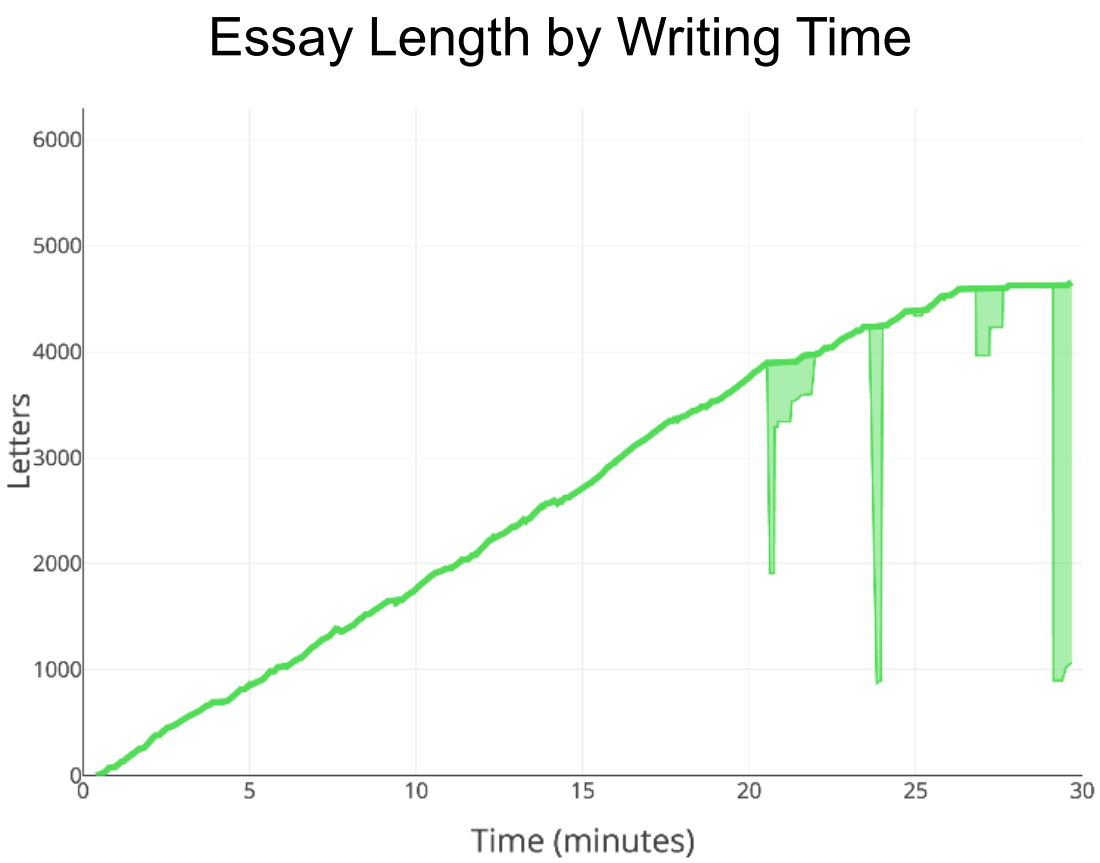
For example

- to improve teaching and learning with specific information on how different students think/perform
- to better understand factors that distinguish high- and low-performers, or expert from novice strategies
- to better understand the relationship of motivation and performance.

Use of Process Data from NAEP Writing

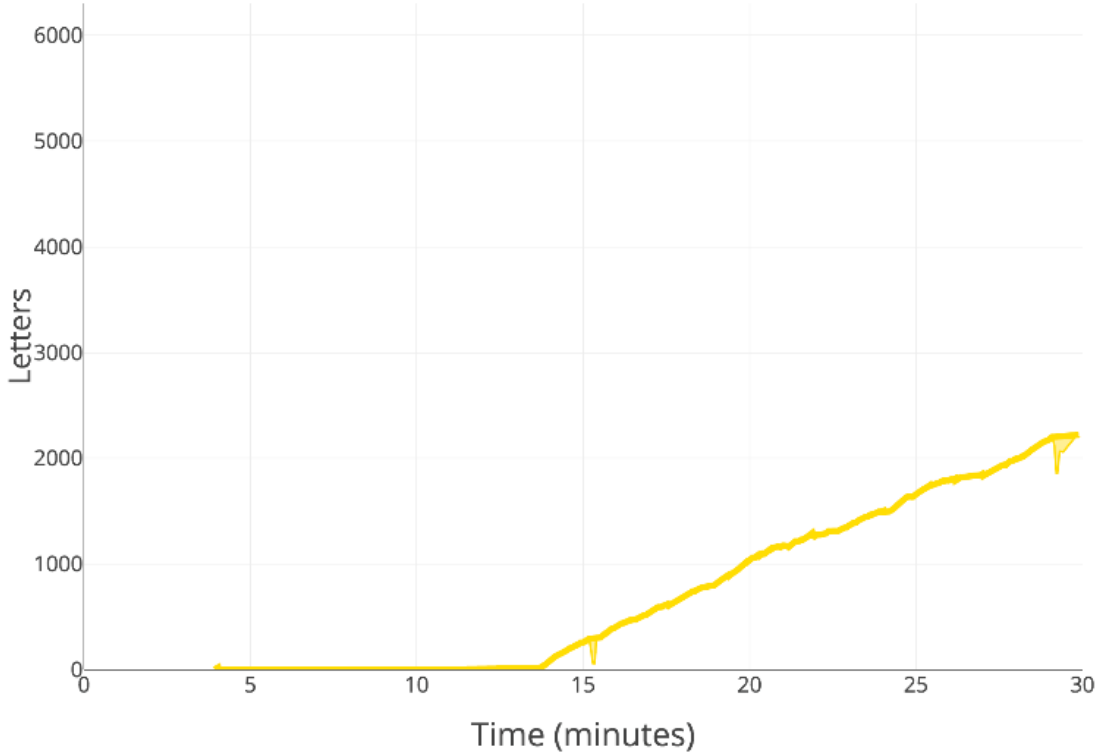


Expanded Use of Process Data



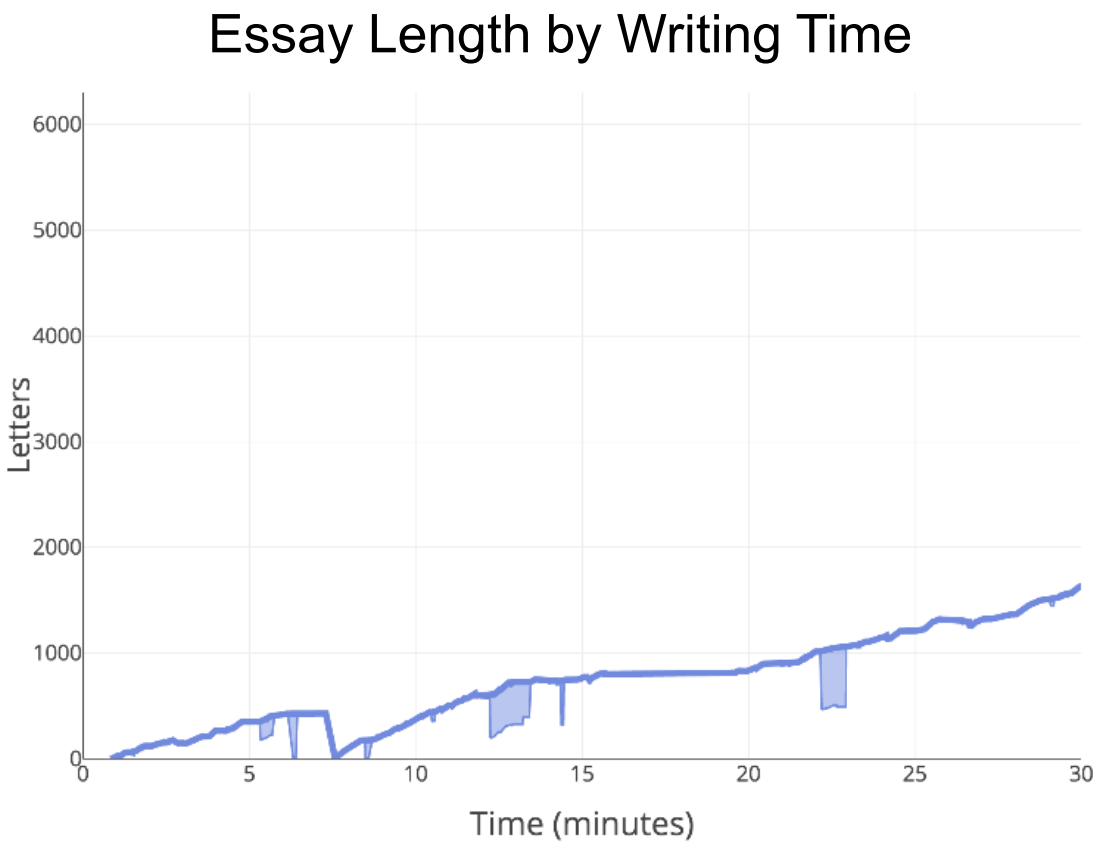
Expanded Use of Process Data

Essay Length by Writing Time



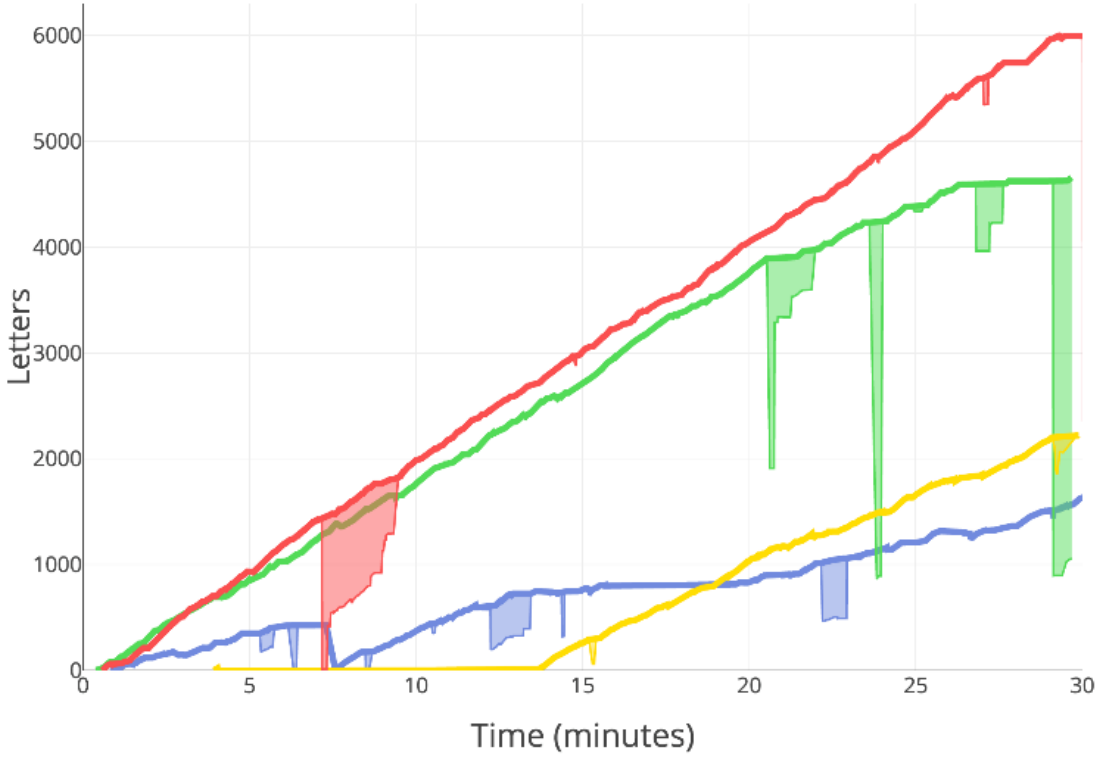
NATIONAL CENTER

Expanded Use of Process Data



Expanded Use of Process Data

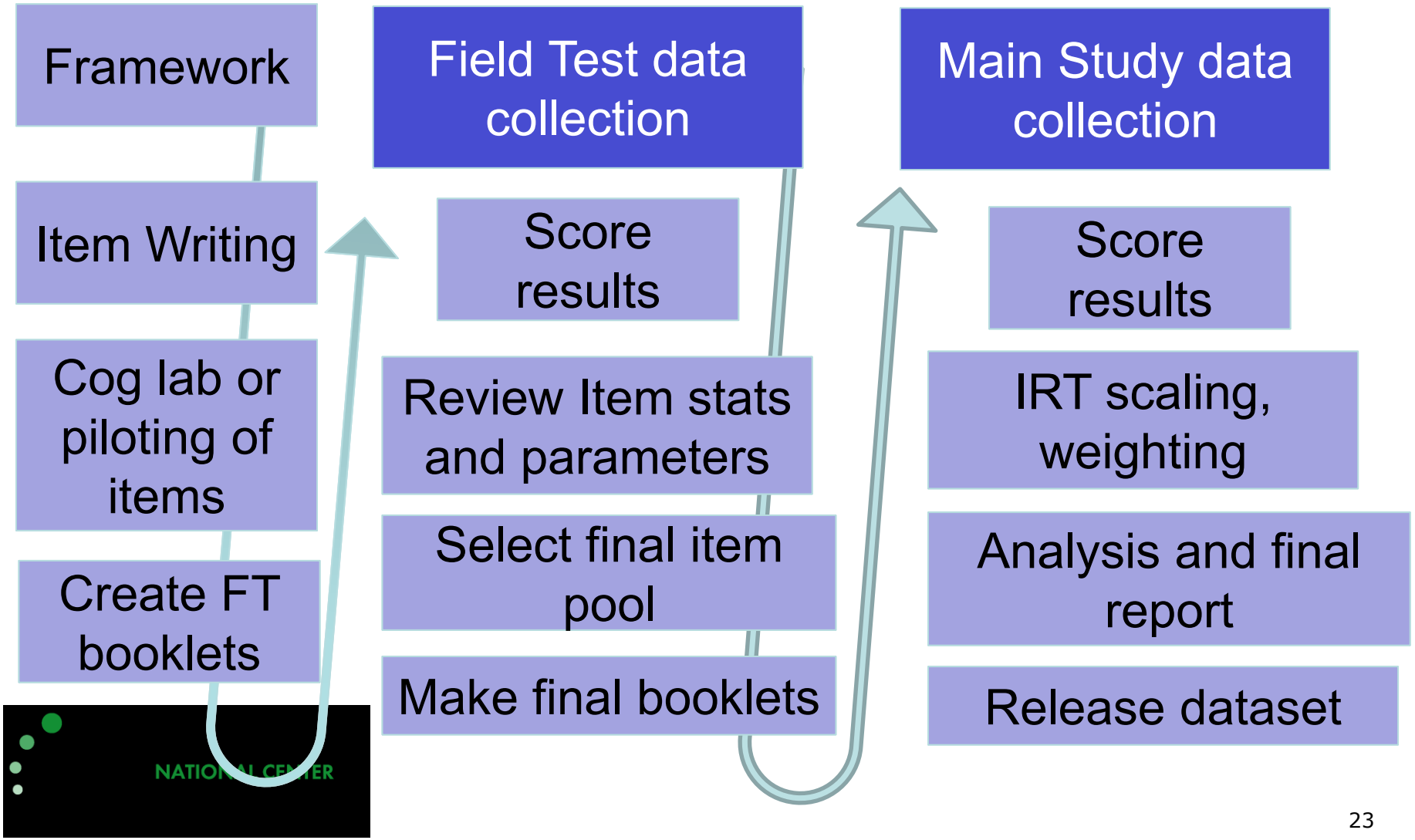
Essay Length by Writing Time



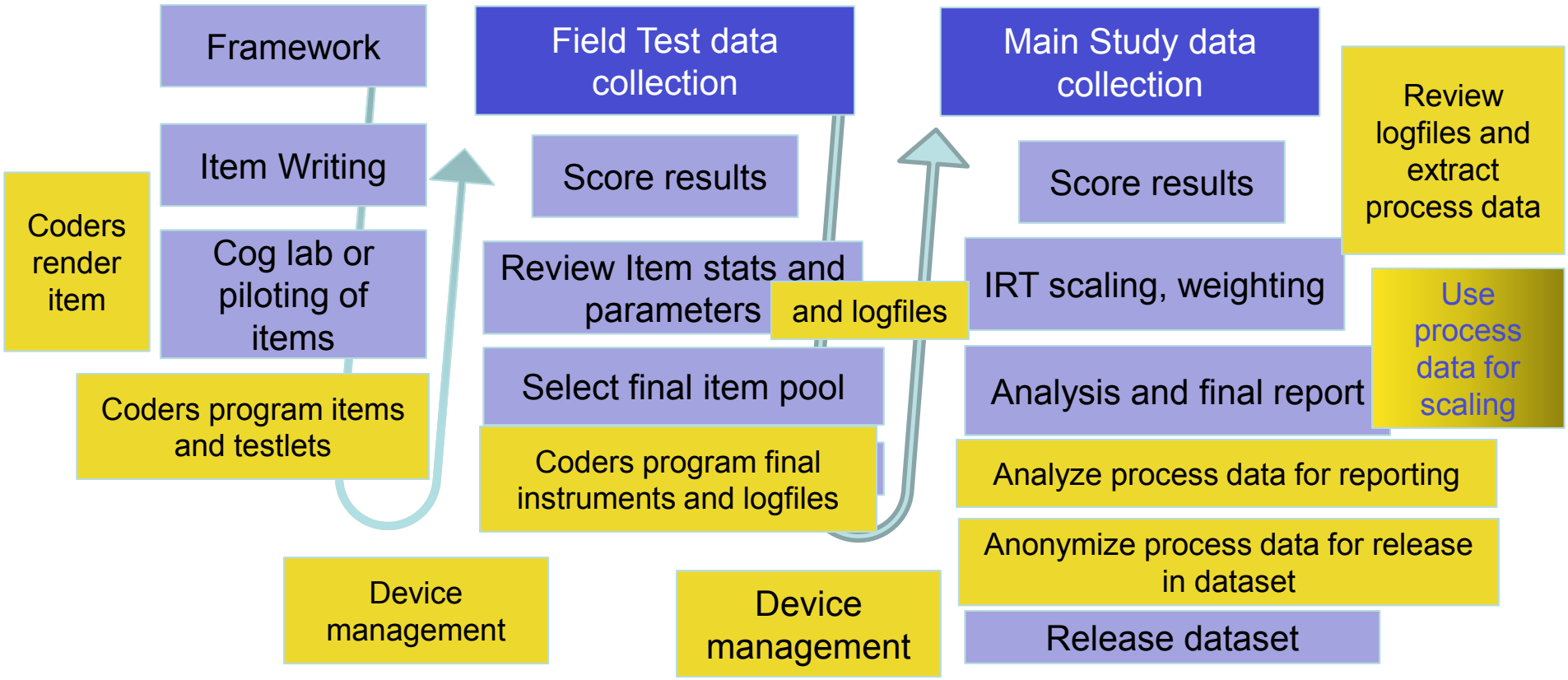
NATIONAL CENTER

NCES Example of Process Data Analysis

Test Development Before DBA

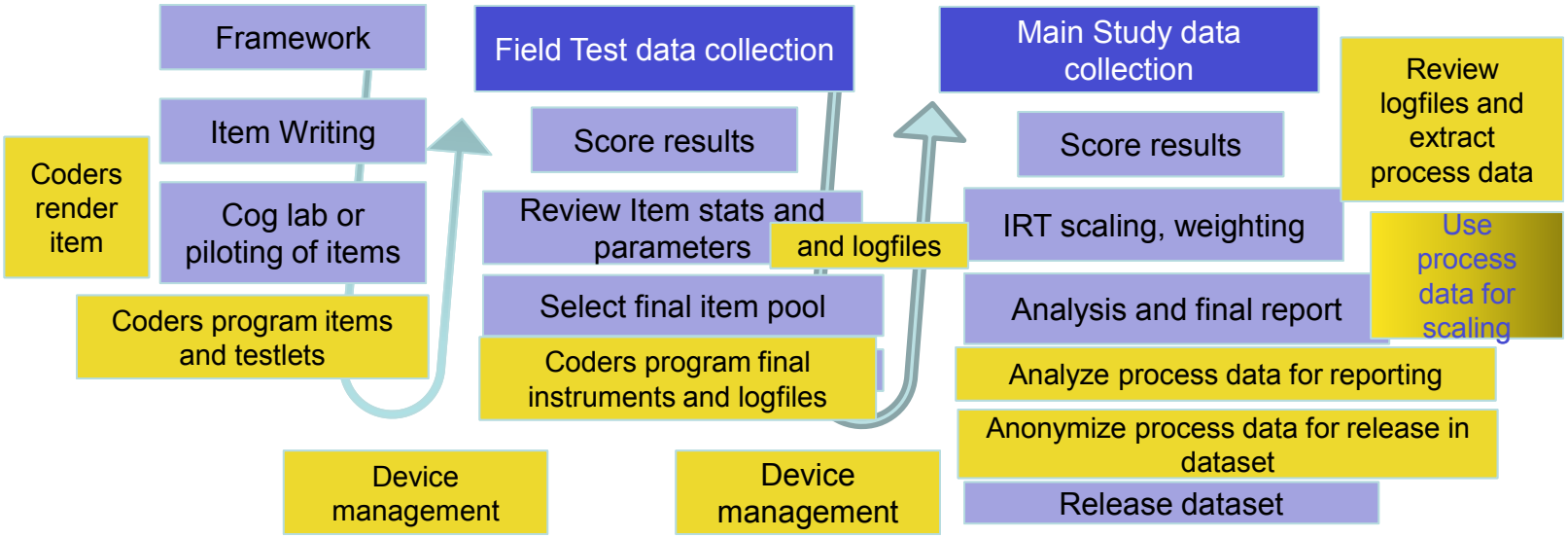


Test Development for DBA

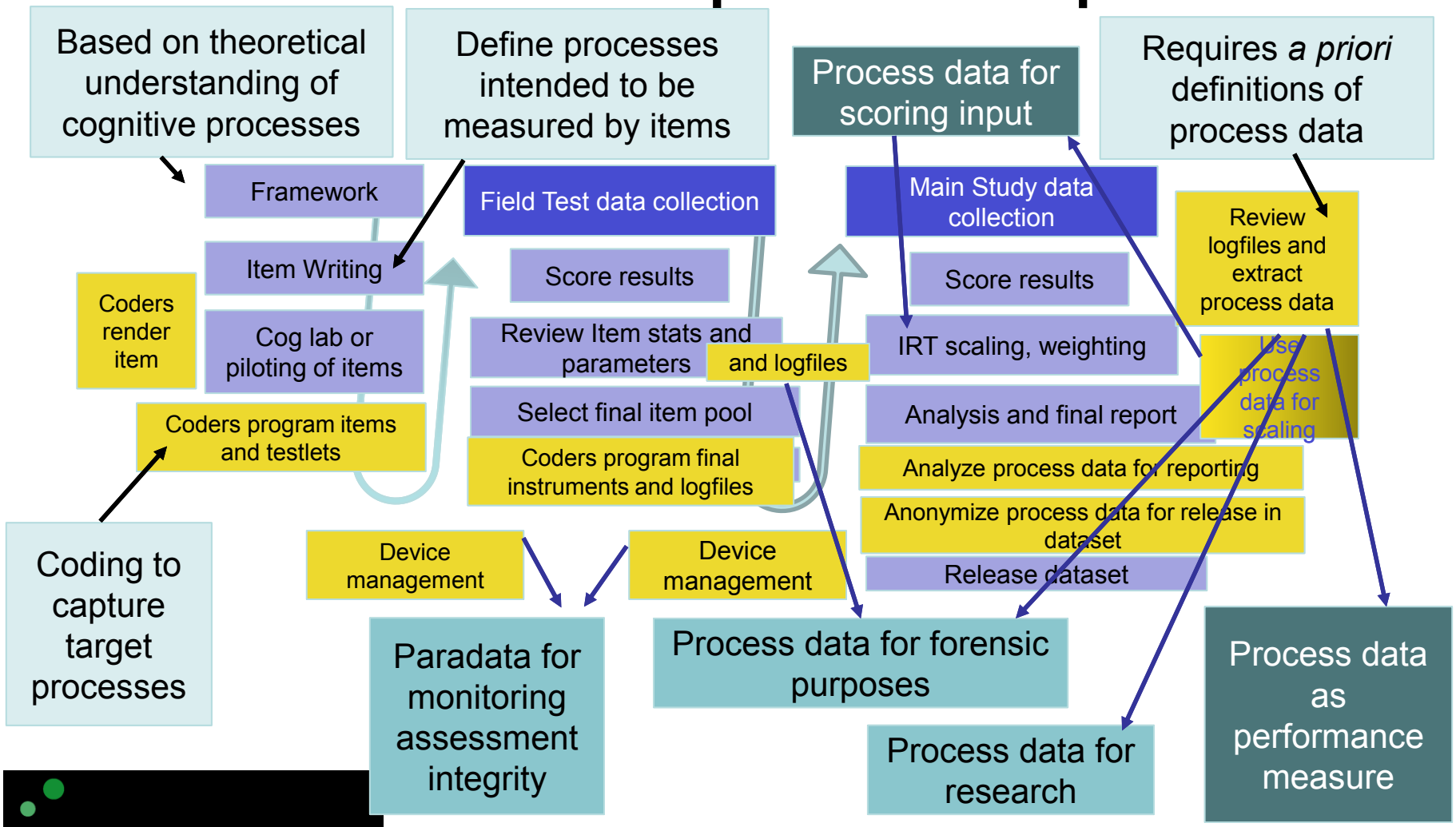


NATIONAL CENTER

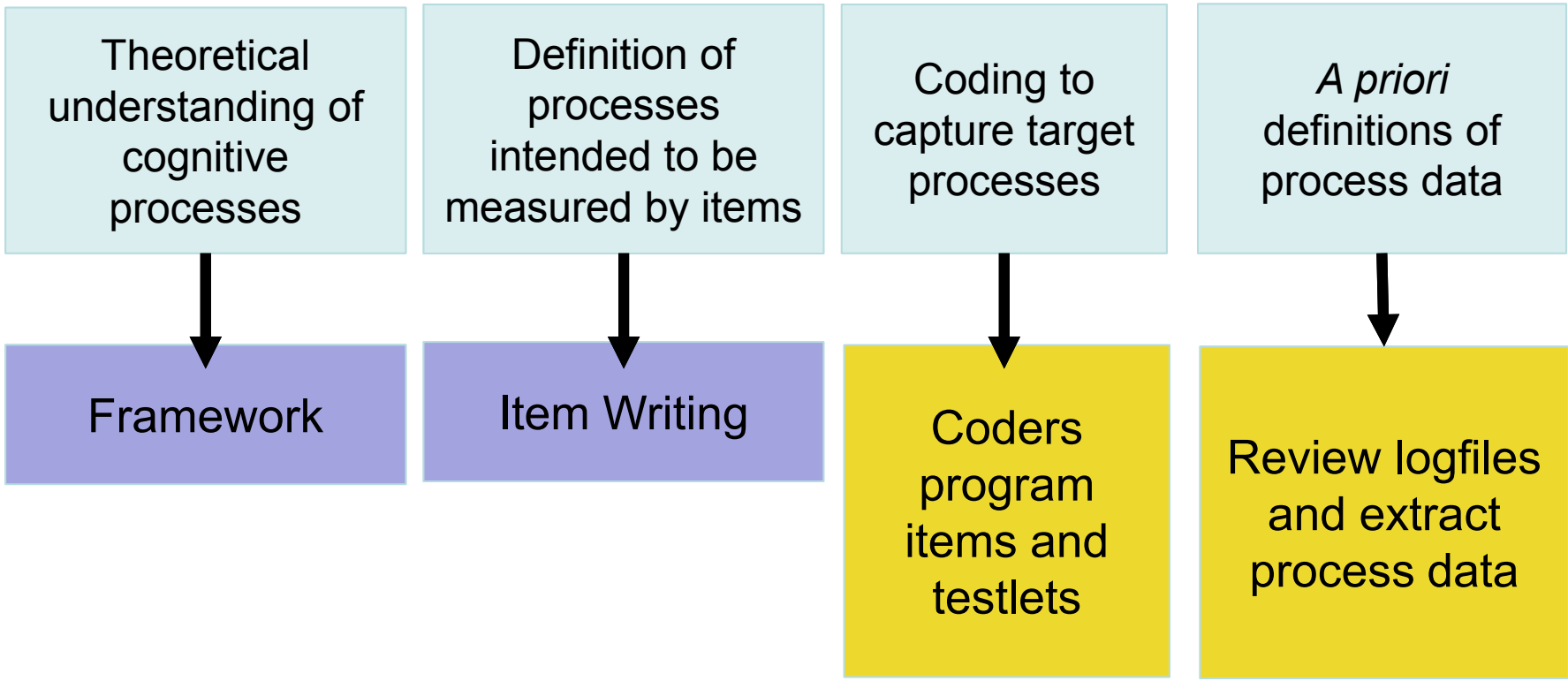
Test Development for DBA



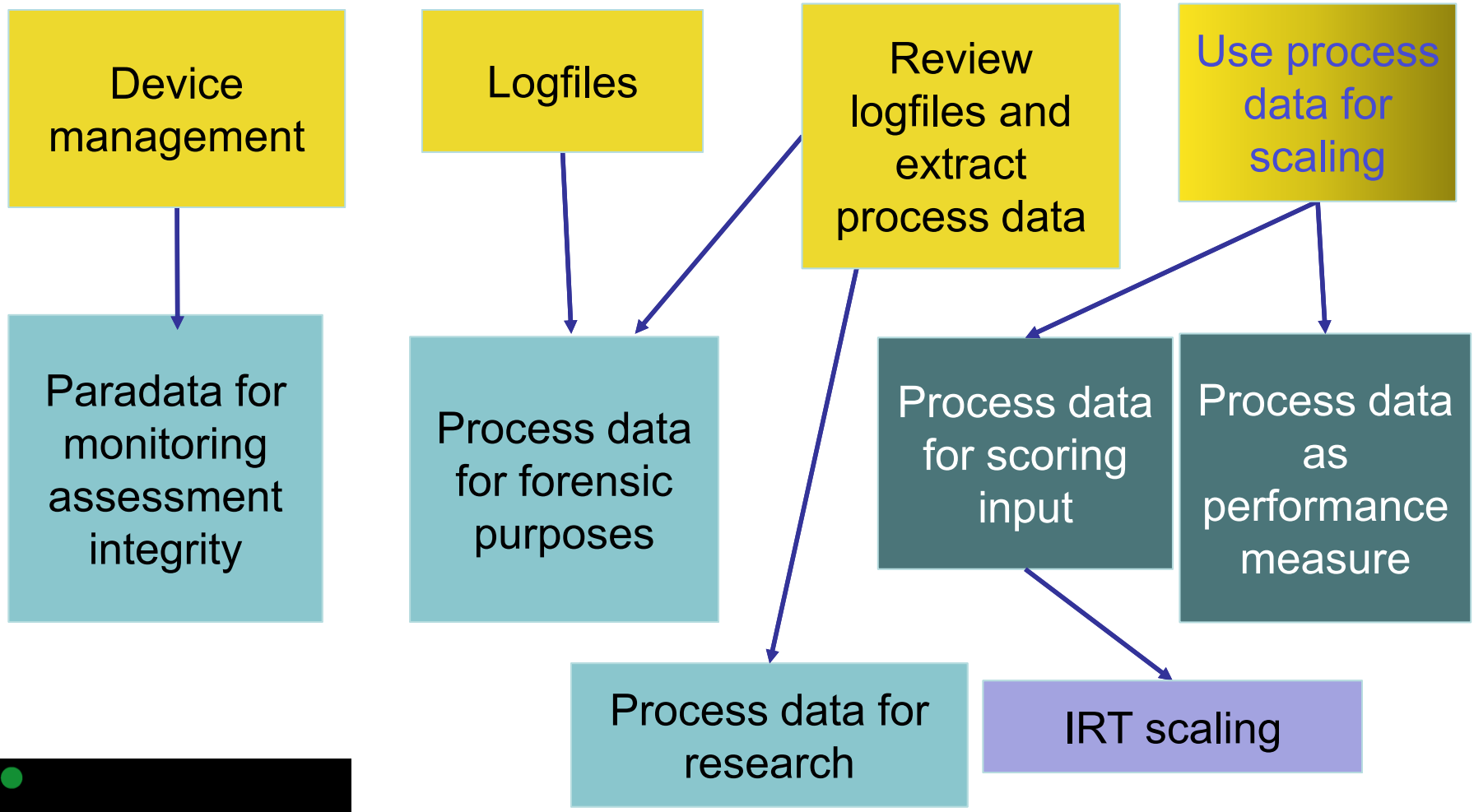
Process Data Inputs and Outputs



Shaping Process Data



Process Data Inputs and Outputs



Inappropriate uses of process data

For example,

- Overgeneralizing from one item to all items, or one process to many processes
- Concluding that strategies associated with higher performance are the strategies that all students should be taught
- Making classroom and formative assessments turn on process data in such a way that students lose unstructured opportunities to try out new ways of thinking and doing

Thank you

Stephen Provasnik
National Center for Education Statistics, U.S. Department of Education
Stephen.Provasnik@ed.gov

NATIONAL CENTER