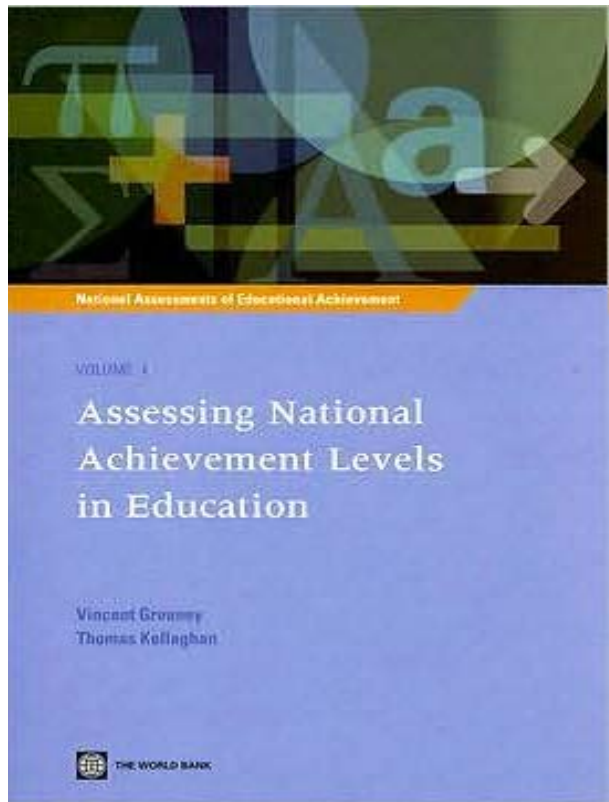# Developing and Measuring National Learning Standards

Vincent Greaney and Fernando Cartwright
4th READ Global Conference
St. Petersburg, Russia
May 12-16, 2014

Assessing National Achievement Levels in Education
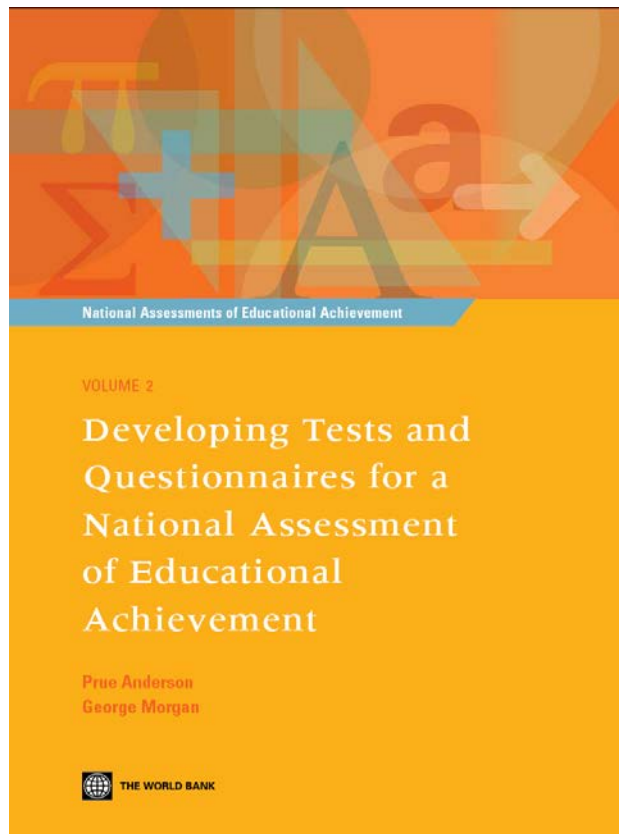
Volume One

Chapters include

- Reasons for doing a national assessment
- Key decisions to be made
- Descriptions of
  - national assessments
  - international assessments

Developing Tests and Questionnaires

Volume 2

National Assessments of Educational Achievement

VOLUME 2

Developing Tests and Questionnaires for a National Assessment of Educational Achievement
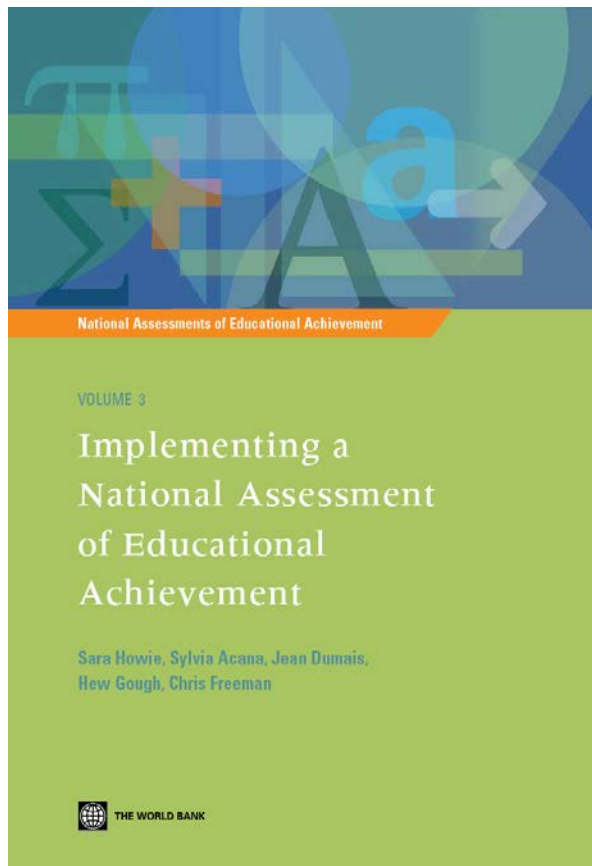
Prue Anderson
George Morgan

THE WORLD BANK

Chapters include

- Item writing
- Pretesting
- Constructing questionnaires
- Developing the administration manual
- Examples: national and international assessments

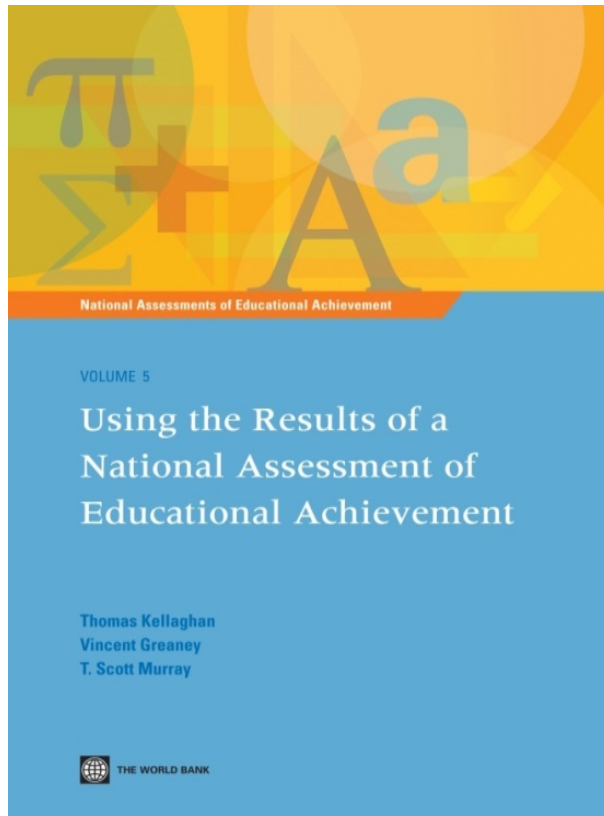# Implementing a National Assessment of Educational Achievement

Volume 3

Chapters include

- Administration and staffing
- Sampling and weighting
- Tasks following administration
- Managing and cleaning data

Using the Results of a National Assessment of Educational Achievement

Volume 5

Chapters include

- Drafting reports
- Using the results to influence
  - policy/management
  - curriculum
  - teaching
  - public awareness

Analyzing Data from a National Assessment of Educational Achievement

Volume 4

Drafted and reviewed .

To be published later this year

Chapters include

- Basic statistical analysis
  - Describing achievement group comparisons, correlation, charts and graphs
- Item and test analysis
  - Pilot and final data,IRT, performance levels, rotated booklets monitoring change

Item and Test Analysis (IATA)

- Developed by Fernando Cartwright
- Program has been pilot tested in a range of countries and modified
- Designed to be
  - Technically adequate
  - User friendly

IATA and Pilot Testing

- Introduce IATA
  - review the results of a pilot test
  - set performance levels and cut scores
  - compare assessments results
- Aim of pilot testing
  - to identify an adequate set of good quality items to be included in an assessment
- Note you must have an adequate number of pilot items to test.

Pilot testing

- Critical to ensure the quality of the national assessment
- Inadequate attention given to
  - Selecting good item writers
  - Reviewing curriculum content
  - Reviewing cognitive levels
- Often far too few items pilot tested

Standard Setting : What is it?

- It is a process for defining levels of proficiency or achievement.
- Examples:
  - **NCLB**: Below basic, Basic, Proficient and Advanced (NCLB)
  - **PISA** reading: Below Level 1, Level 1, Level 2, Level 3, Level 5, Level 5
  - **TIMSS**: Low (400), Intermediate (475) High (550) and Advanced (625)

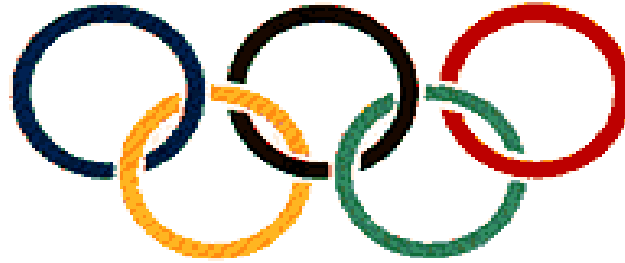- It is not purely a statistical or methodological process

Rainfall standards

Can be uniform over time

Rain guage

- < 10mm
- 11 – 20mm
- 21 – 30mm
- > 31mm

Standards Can Vary

Beijing Olympics qualification standards.
Athletes had to meet the specified standard (cut scores)
- 1 min 46 seconds
- 13 min 21.5 seconds
- 2 hours 15 minutes
- 5.70 metres

Standard Setting : Who does it?

- Aim : Ensure the test results will be useful

- Standard setting should involve
  - policy makers,
  - test developers
  - and measurement specialists

Cut scores

- Use cut scores to define levels or standards

- At each level a cut score is the cut off point that determines if a student is performing at a particular level

- Challenge : Determine the score that classifies students who get a lower score into one level and other students into a higher level

Related to Test Development

- Tests should have
  - appropriate content and
  - appropriate level of difficulty
- Part of test development
  - Need good item writers who understand the process of setting cut scores
- The cut scores should be both consistent with the intended educational policy and psychometrically sound (Bejar, 2008)

Approaches to Standard Setting

- Nedelsky's method
- Angoff's method
- Ebel's method   Item Centered
- Modified Angoff's method
- **Bookmark method**

- Borderline Group method   Person Centered
- Contrasting Groups method

(Zieky, Perie, & Livingston, 2008)

Bookmark Method

- Rank order items ordered by difficulty called Response Probability (RP) in IATA.

- Subject matter experts place a "bookmark" in the list of items to indicate the cutscore. This is the point where the student who is on the boundary would be unlikely to answer the remaining items correctly.

- IATA allows you to make adjustments.

Describe each Proficiency Level

- Subject matter specialists should exam the items at each level and write clear descriptions of students levels of knowledge and understanding at each proficiency level

- "Students know some basic facts about the solar system and show an initial understanding of earth's physical characteristics and resources. .." (*TIMSS Intermediate International Benchmark, Science*)

# Document the rationale and the process

Existing professional standards:

- "The rationale and procedures used for establishing cut scores should be clearly documented"

Comparing Results to the Results of Other Assessments

- Of great interest to national policy makers
  - Getting value for money spent on education?
- In particular governments need to know if student achievement levels are
  - Generally increasing
  - Staying more or less the same, or
  - Generally decreasing

Russia : Changes in PISA Mean Reading Achievement Score 2000 -2012
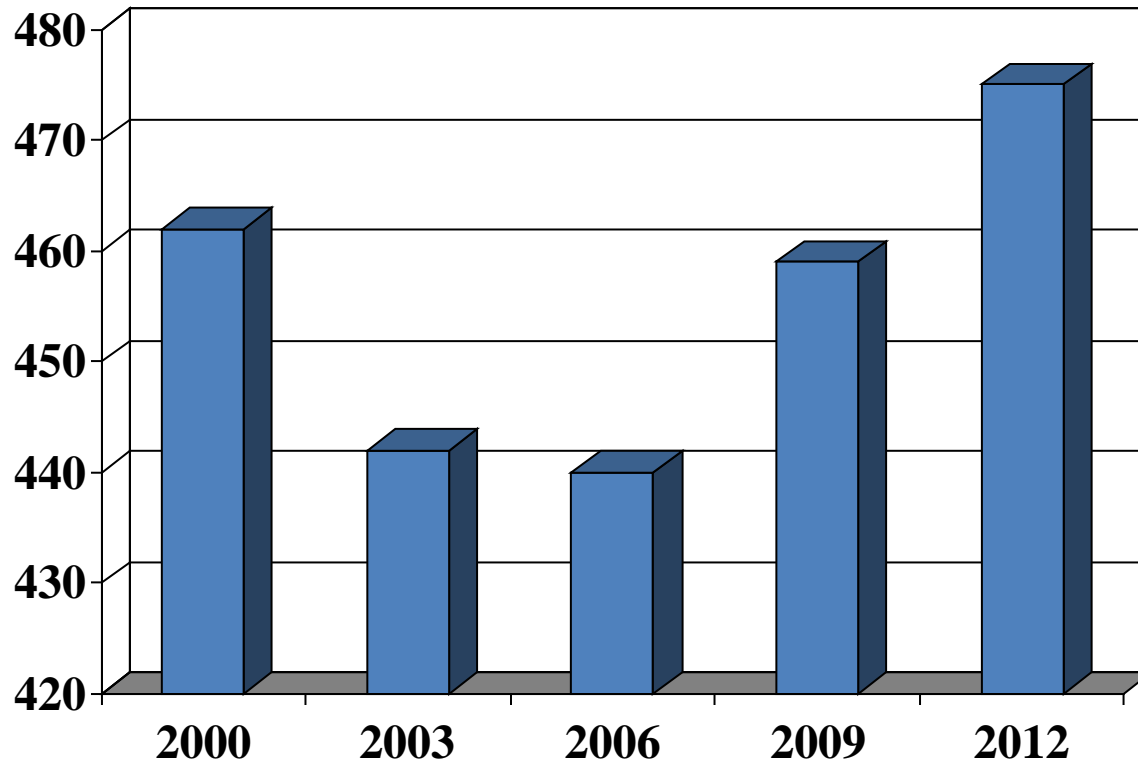
Linking to Comparing Results to the Results of Other Assessments

- Examples
  - Compare tests used in different states (oblasts) that share some common items
  - Use some items from international assessments in national assessment
- Essential:
  - Must be able to use items common to both assessments
  - Must have item statistics for both assessments