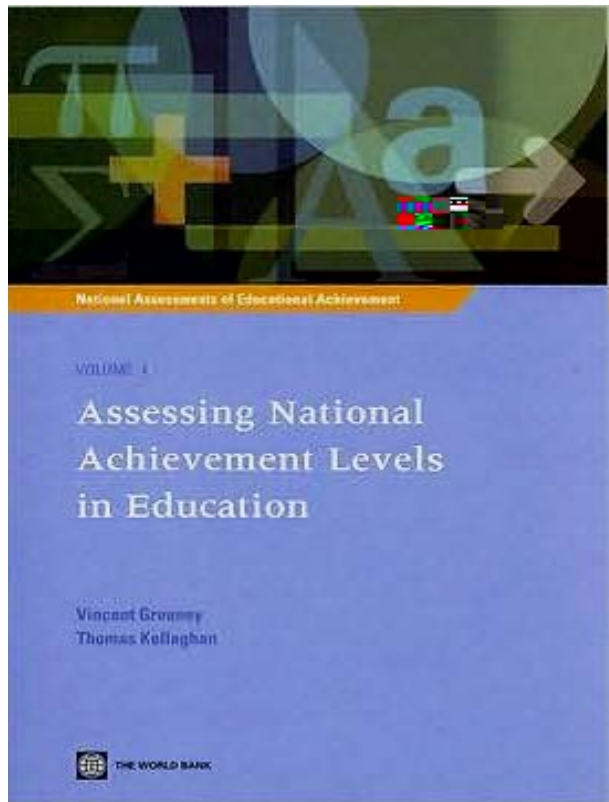


Developing and Measuring National Learning Standards

Vincent Greaney and Fernando Cartwright
4th READ Global Conference

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

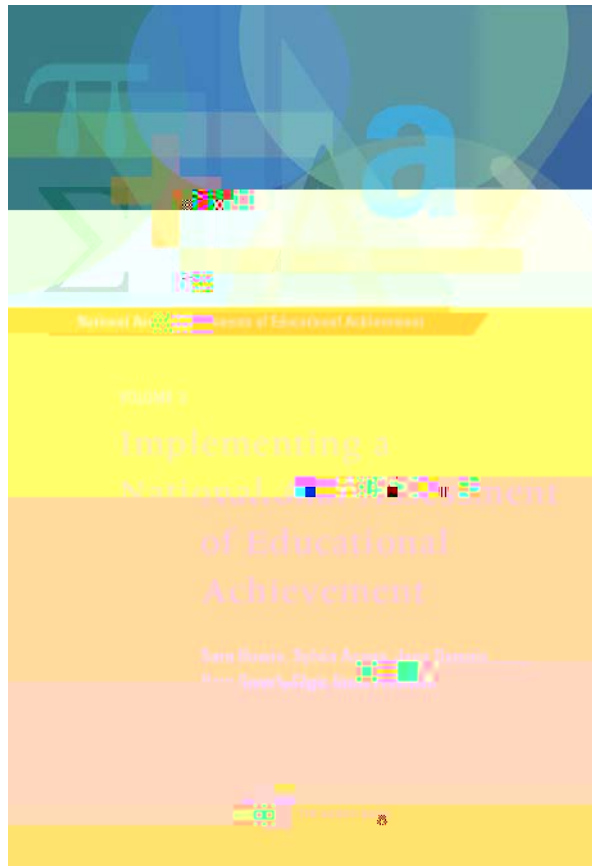
Volume 2

Chapters include



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

Analyzing Data from a National Assessment of Educational Achievement

Volume 4

Drafted and reviewed .

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



Standard Setting : What is it?

-

Rainfall standards

Can be uniform over time

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

Rain guage



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
-

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
- Modified Angoff's method
- **Bookmark method**



Item Centered

- Borderline Group method
- Contrasting Groups method



Person Centered

(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
- "

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 -

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