

MAINTAINING STANDARDS IN NATIONAL CURRICULUM TESTS

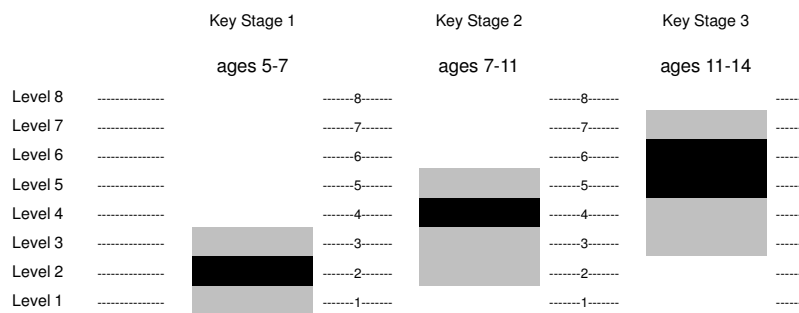
Malcolm Hayes


February 2012


ALWAYS LEARNING

PEARSON

National Curriculum Measures



 During the key stage, most children will work within this range

 By the end of the key stage, most children will have reached this level

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING

PEARSON

Standards:

Suppose that last year 25% of children at ks2 gained a level 5
This year the figure is 30%

- Standards are lower because getting a 5 means less.
- Standards are higher because more children jumped this particular hurdle

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING

PEARSON

Standards:

Suppose that last year the level 5 cut-score was 78
This year we put it at 76 and more pupils gain level 5

- Standards are lower
 - because the bar has dropped
- Standards are higher
 - because more pupils 'passed'
- The standard is the same
 - because we've equated the tests

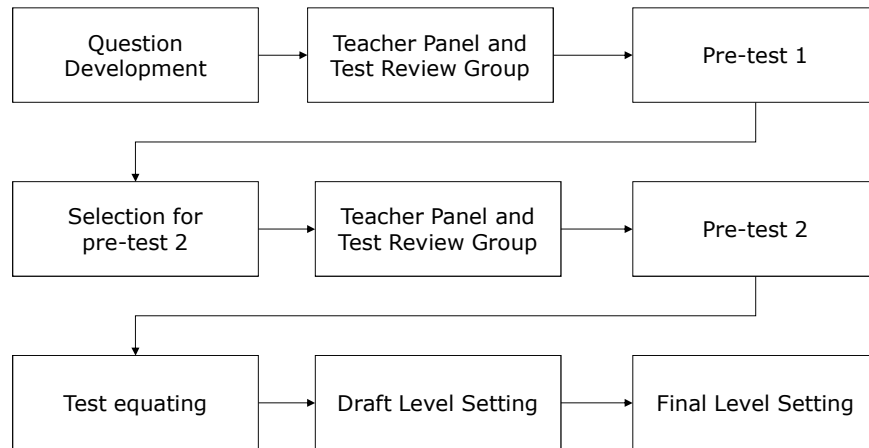
The expected outcome for any child taking a test
should be independent of the version of the test
taken

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING

PEARSON

Key processes in Test Development



Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING

PEARSON

Key Stage 2 Maths Test Structure

- One test for the whole ability range
- Levels 3, 4 and 5
- Approx one third of questions at each level
- One calculator not allowed written paper (40 marks)
- One calculator allowed written paper (40 marks)
- One mental mathematics paper (20 marks)

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING

PEARSON

The test development cycle

Draft level setting

- Statistical evidence from pre-testing
 - Recommendations from Teacher Judgement
 - Additional information from test developers
- ↓
- Draft levels set
 - Script scrutiny ranges set

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING

PEARSON

The test development cycle

Final level setting

- Draft Levels
- Script scrutiny recommendations
- Live test data (approx 30,000 cases)



- Final levels set

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING

PEARSON

Key Stage 2 Equating

- Equating of pre-test to pre-test
- IRT model – 2 parameter
- 33 item anchor test
- Pre-test effect accounted for by model
- Live test data used for checking the model

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING

PEARSON

Assumptions for test equating

- Tests are measuring the same construct
- Tests are *unidimensional*
- Tests are valid and reliable
- Tests are built to the same specification each year
- Experimental conditions are constant

ks2 mathematics Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

10

ALWAYS LEARNING

PEARSON

Outline of equating design

- Item Response Theory (IRT) model
- Common-item, non-equivalent groups design
- Common items in the form of an *external anchor*
- *Separate* estimation of parameters
- *Mean/Sigma* transformations

ks2 mathematics Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

11

ALWAYS LEARNING

PEARSON

Pre-test design

- Two versions of the test
- Pupils take all three components from one version
- All pupils take the 33-item anchor test
- Sampled at class level – 1 class per school
- Samples balanced by pupil ability
- Sample size >800 pupils for each version

ks2 mathematics Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

12

ALWAYS LEARNING

PEARSON

Pre-test administration

- Anchor test administered by class teacher
- Pre-test administered by a consultant
- Pre-test takes place within a 2-week window
- Test window is immediately before the live test
- Schools encouraged to complete all three components on the same day
- Mental test administered via CD

ks2 mathematics Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

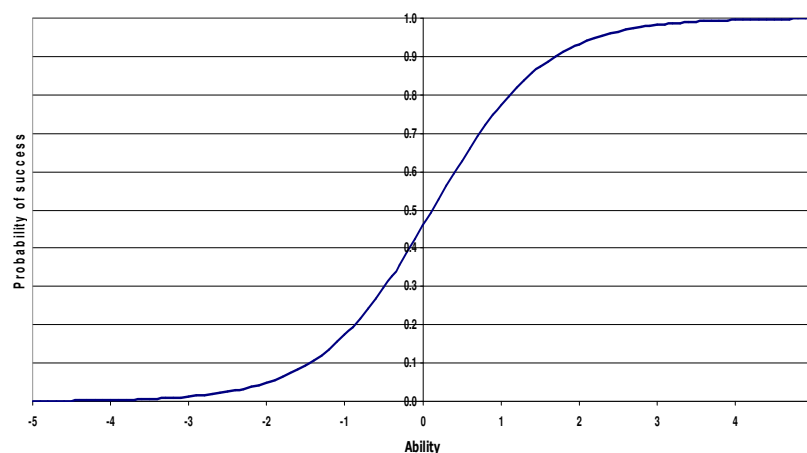
13

ALWAYS LEARNING

PEARSON

Modelling item performance

Example of an Item Characteristic Curve (ICC)



ks2 mathematics Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

14

ALWAYS LEARNING

PEARSON

Item Response Model

$$P(X_i = 1 | \theta) = \frac{1}{1 + e^{-1.702a_i(\theta - b_i)}}$$

Where:

X_i = outcome for item i

a_i = discrimination parameter for item i

b_i = difficulty parameter for item i

θ = ability of the pupil

ks2 mathematics Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

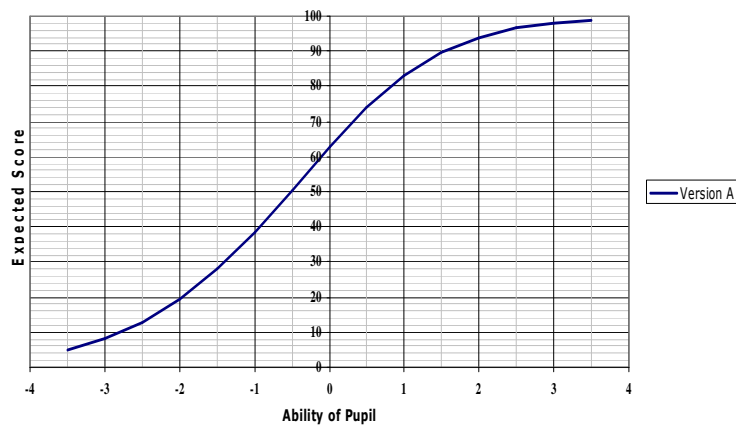
15

ALWAYS LEARNING

PEARSON

Modelling test performance

Illustration of Test Characteristic Curve



ks2 mathematics Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

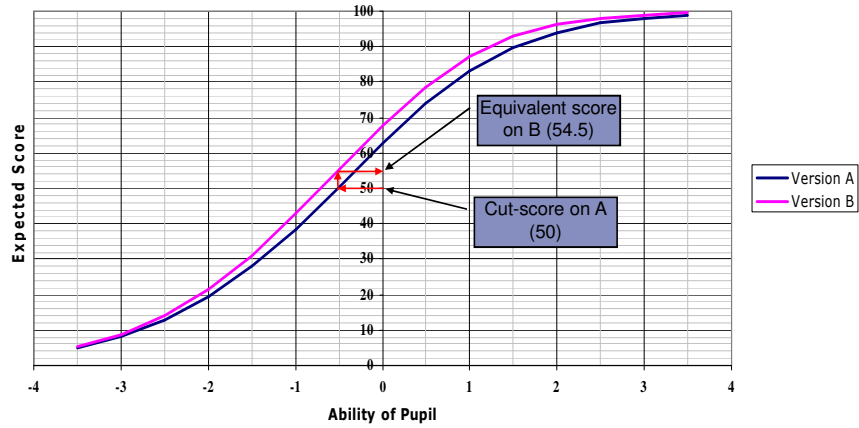
16

ALWAYS LEARNING

PEARSON

Modelling test performance

Illustration of Comparison of Test Characteristic Curves



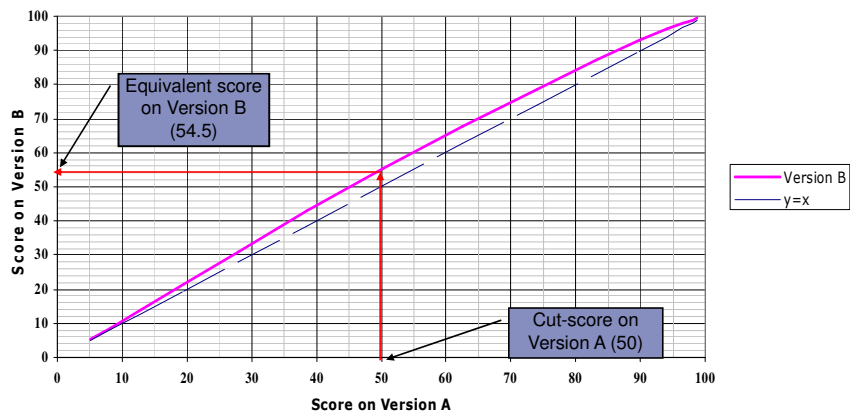
ks2 mathematics Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

17

ALWAYS LEARNING

PEARSON

Illustration of Comparison of Expected Scores



ks2 mathematics Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

18

ALWAYS LEARNING

PEARSON

Issues

- Choice of model
- Method of anchoring
- Weighting of strands of evidence
- Practical considerations
- Confidence bands on equating

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING

PEARSON