Assessment in Education

Barry McGaw
Director for Education
Organisation for Economic Co-operation and Development

A Question of Confidence: Maintaining Trust in National Examination Systems

Cambridge Assessment Network, Cambridge, UK
17 October 2005

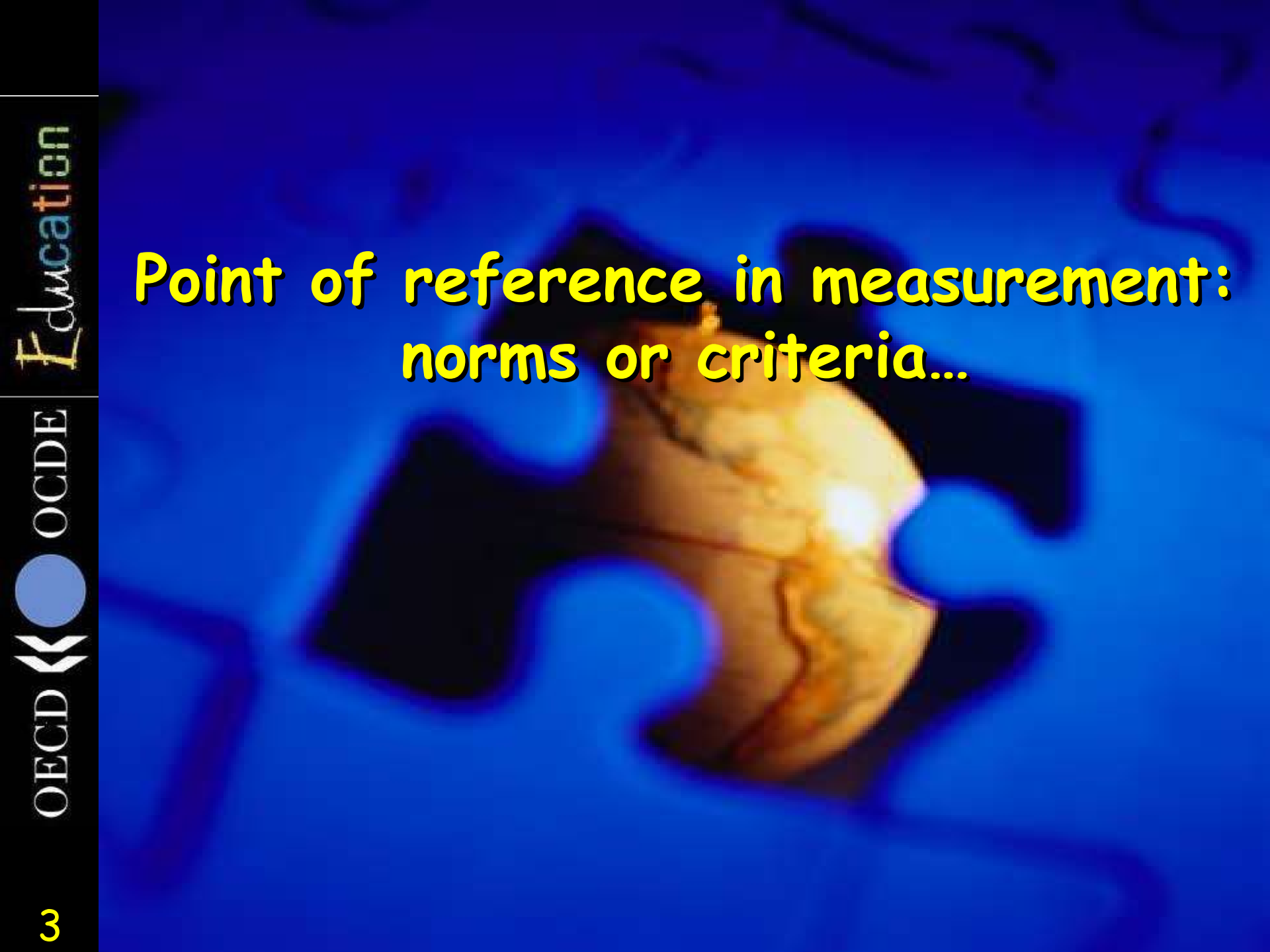
OECD Education

Purpose of educational assessment

E. F. Lindquist (Ed). (1951). Educational Measurement

"the functions of educational measurement are concerned…with the facilitation of learning" (Cook, 1951).

"educational measurement is conceived, not as a process quite apart from instruction, but rather as an integral part of it" (Tyler, 1951).

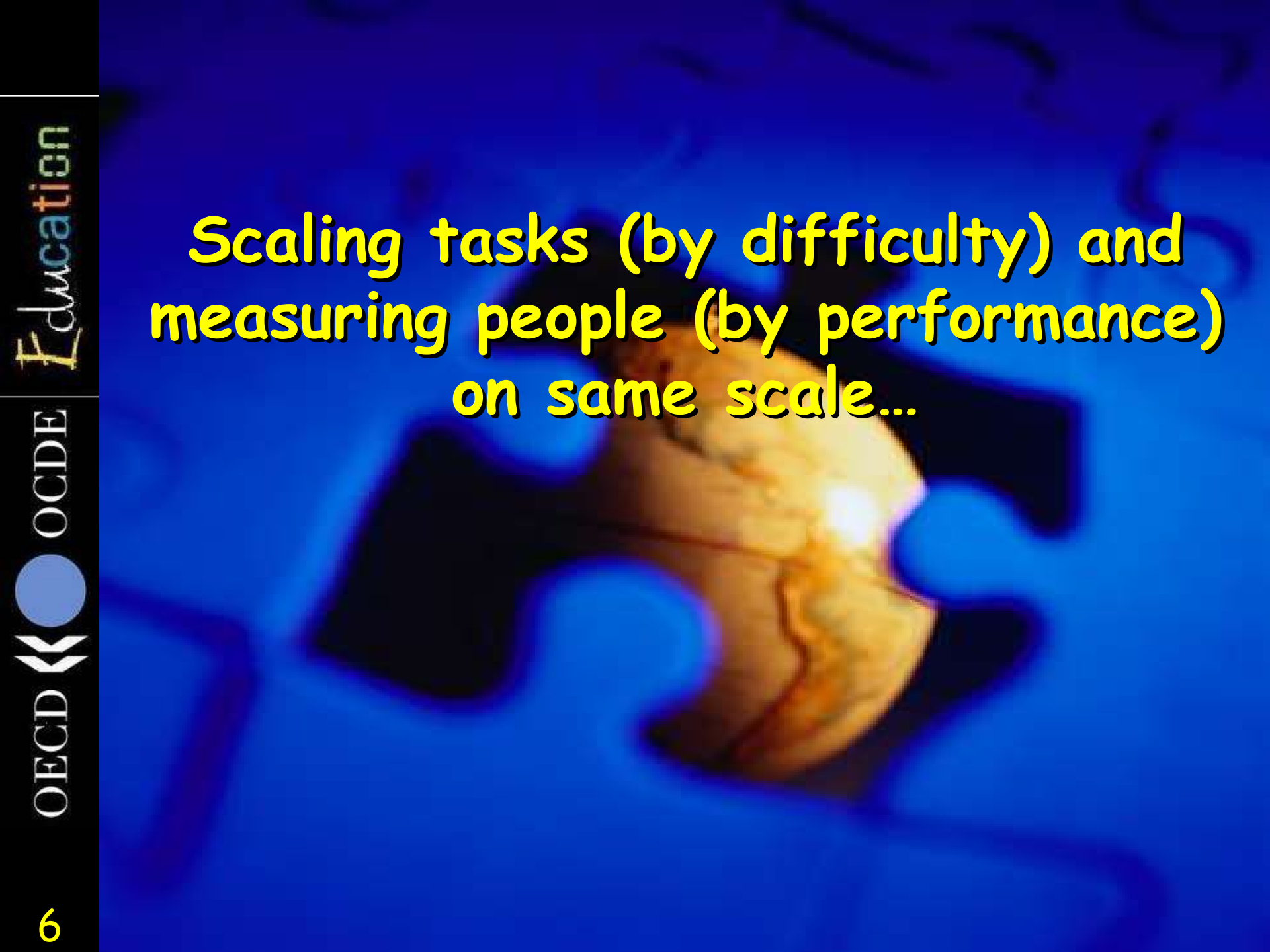

Point of reference in measurement:
norms or criteria…

Point of reference for judging individuals

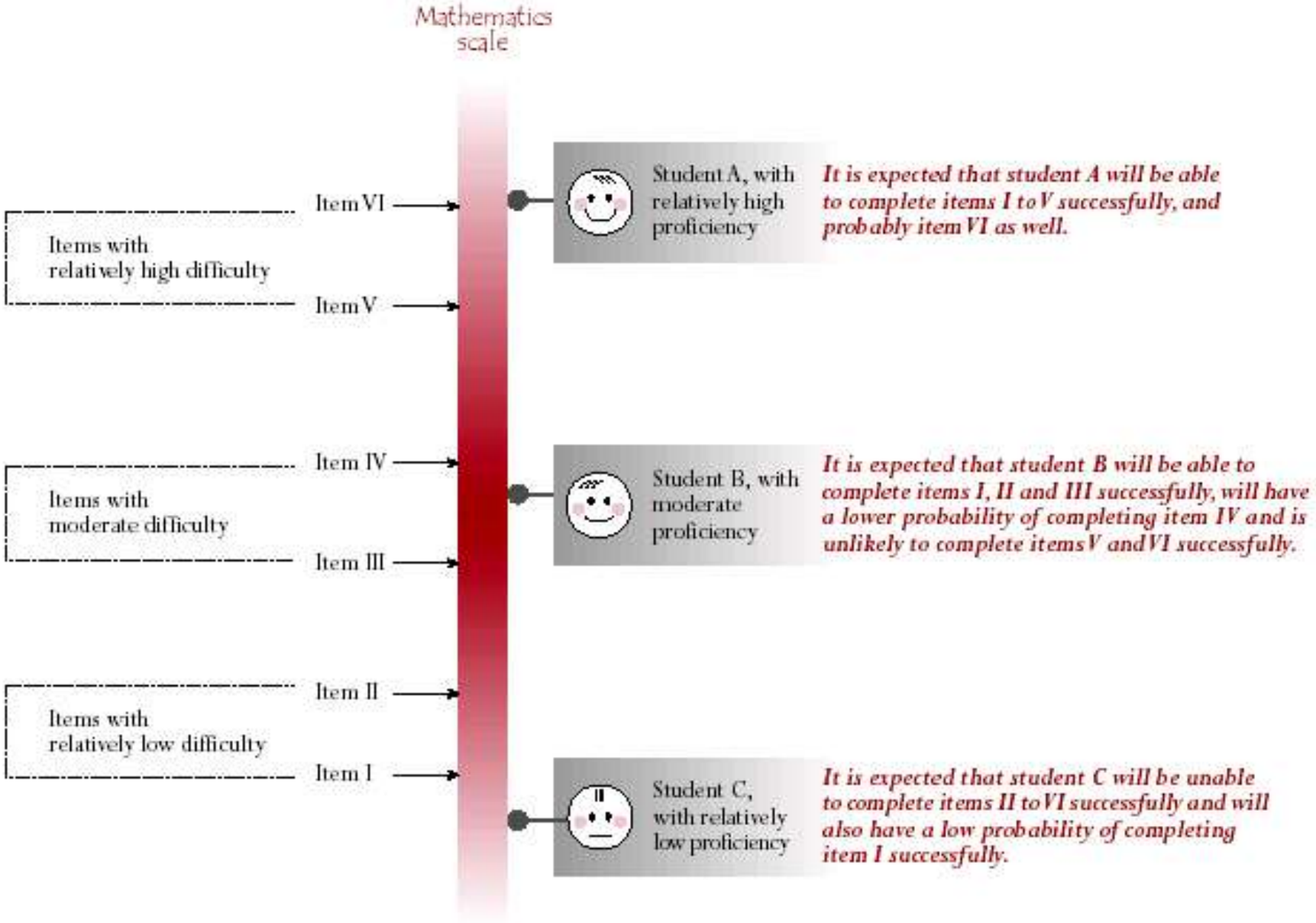
- Using an independent external measure (Psychophysics)
  - Judgements of phenomenon (e.g. brightness of light)
    - requiring judgements of differences, not absolute values
    - comparing judgements with direct measures of phenomenon
    - developing a scale of human judgement of phenomenon
  - Interest in nature of human judgement not phenomenon
- Using performance of others to judge individuals
  - Psychological phenomena without external measure
    - developed in the context of studies of individual differences
  - Individual performances judged in relation to:
    - the performance of others
    - in particular, the average performance of others (or norm)
  - Want to look better? Choose different company.

In search of an external criterion

- Criterion-referenced measurement
  - specify learning required
  - judge students against requirements (not each other)
- New psychometric models
  - Simultaneous scale construction and measurement
    - locate tasks on scale by difficulty
    - locate individuals on same scale by performance
    - interpret performance with reference to tasks
  - Model wars
    - do the data fit the model (lack of fit indicates poor data)
    - does the model fit the data (more complex model to improve fit)


Scaling tasks (by difficulty) and measuring people (by performance) on same scale…

Locating tasks and persons on same scale



OECD (2004), Learning for tomorrow's world: First results from PISA 2003, p.45.

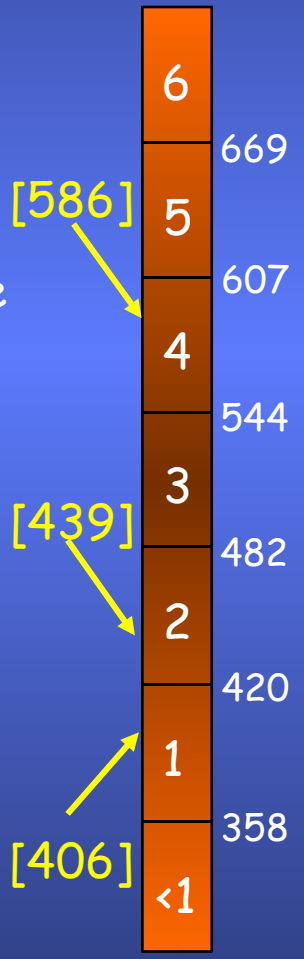
PISA 2003 mathematics item

Mei-Ling from Singapore was preparing to go to South Africa for 3 months as an exchange student. She needed to change some Singapore dollars (SGD) into South African rand (ZAR).

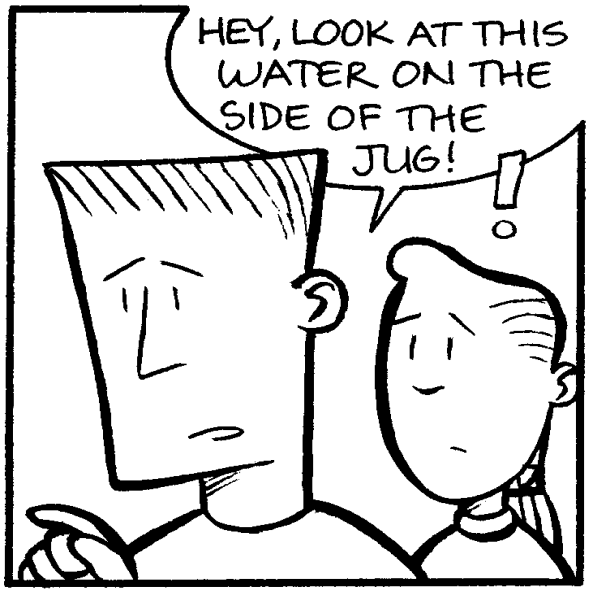
During these 3 months the exchange rate had changed from 4.2 to 4.0 ZAR per SGD. Was it in Mei-Ling's favour that the exchange rate now was 4.0 ZAR instead of 4.2 ZAR, when she changed her South African rand back to Singapore dollars? Give an explanation to support your answer.

On returning to Singapore after 3 months, Mei-Ling had 3 900 ZAR left. She changed this back to Singapore dollars, noting that the exchange rate had changed to: 1 SGD = 4.0 ZAR. How much money in Singapore dollars did Mei-Ling get?

Mei-Ling found out that the exchange rate between Singapore dollars and South African rand was: 1 SGD = 4.0 ZAR. Mei-Ling changed 3000 Singapore dollars into South African rand at this exchange rate. How much money in South African rand did Mei-Ling get?



Tapping science beliefs



WHERE HAS THE WATER ON THE OUTSIDE OF THE JUG COME FROM?

Classification of responses

Content of response	Score
condensation when air cools	4
from atmosphere - no mechanism	3
condensation - no atmosphere	2
liquid on outside comes from inside	1
liquid passed through sides of jug	0
uninterpretable responses	0

Percent of responses by type

Grade	5	9
condensation when air cools	2	16
from atmosphere - no mechanism	7	30
condensation - no atmosphere	36	23
liquid on outside comes from inside	23	15
liquid passed through sides of jug	6	3
uninterpretable responses	25	13

Category boundaries

condensation when air cools

63

from atmosphere - no mechanism

61

condensation - no atmosphere

60

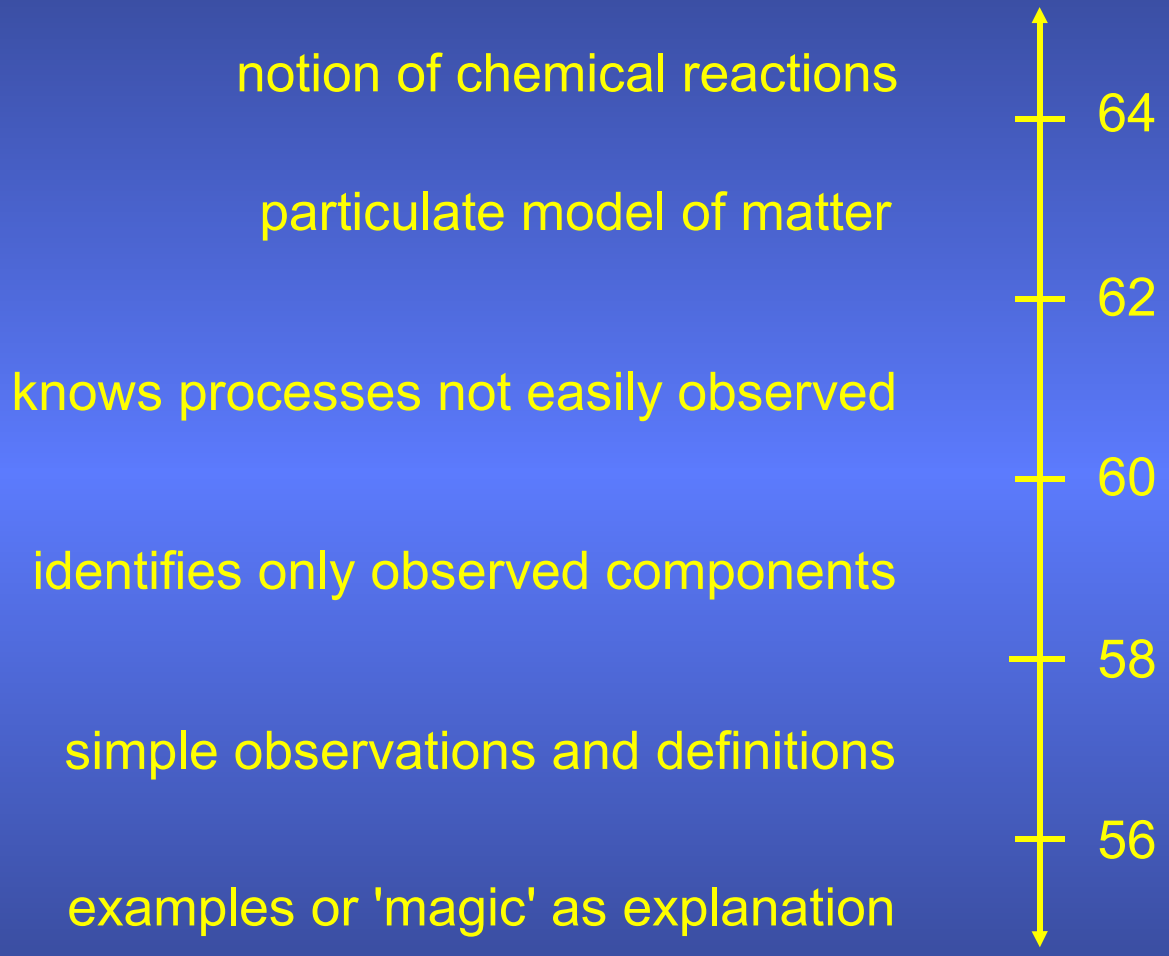
liquid on outside comes from inside

58

liquid passed through sides of jug

uninterpretable responses

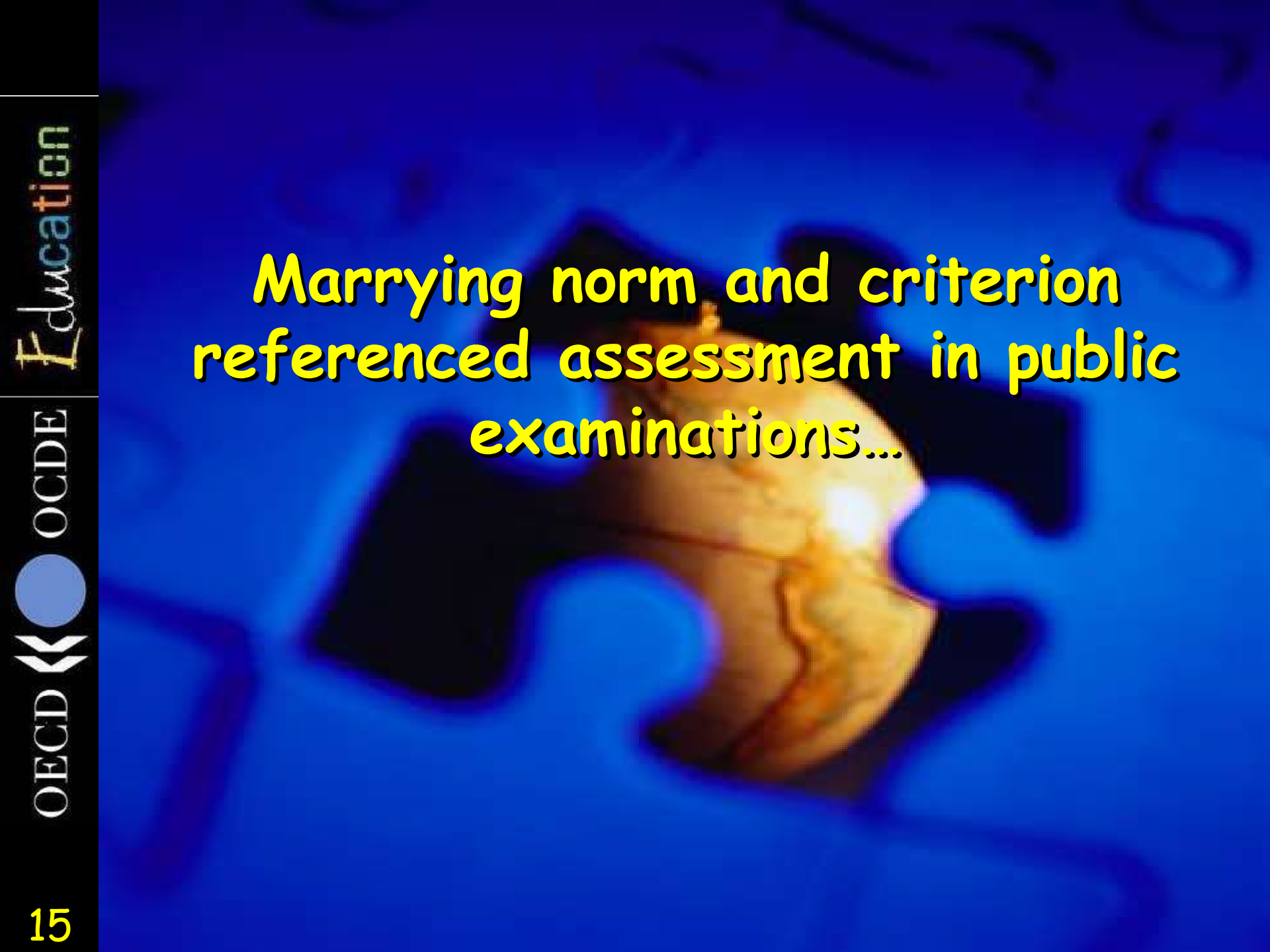
Structure of matter scale (using further items)



Map of selected PISA 2003 mathematics tasks

Level	Space and shape Figures 2.4a–c	Change and relationships Figures 2.7a–b	Quantity Figures 2.10a–b	Uncertainty Figures 2.13a–c
6	CARPENTER *Question 1 (687)*	WALKING *Question 5 – Score 3 (723)*		ROBBERIES *Question 15 – Score 2 (694)*
668.7				
5		WALKING *Question 5 – Score 2 (666)*		
606.6		WALKING *Question 4 (611)*		TEST SCORES *Question 6 (620)*
4		WALKING *Question 5 – Score 1 (605)* GROWING UP *Question 8 (574)*	EXCHANGE RATE *Question 11 (586)* SKATEBOARD *Question 13 (570)* SKATEBOARD *Question 14 (554)*	ROBBERIES *Question 15 – Score 1 (577)* EXPORTS *Question 18 (565)*
544.4				
3	NUMBER CUBES *Question 3 (503)*	GROWING UP *Question 7 – Score 2 (525)*		OECD average = 500
482.4			SKATEBOARD *Question 12 – Score 2 (496)*	
2			SKATEBOARD *Question 12 – Score 1 (464)* EXCHANGE RATE *Question 10 (439)*	EXPORTS *Question 17 (427)*
420.4	STAIRCASE *Question 2 (421)*	GROWING UP *Question 7 – Score 1 (420)*		
1			EXCHANGE RATE *Question 9 (406)*	
358.3				
Below Level 1				

OECD (2004), Learning for tomorrow's world: First results from PISA 2003, p.48.


Marrying norm and criterion referenced assessment in public examinations…

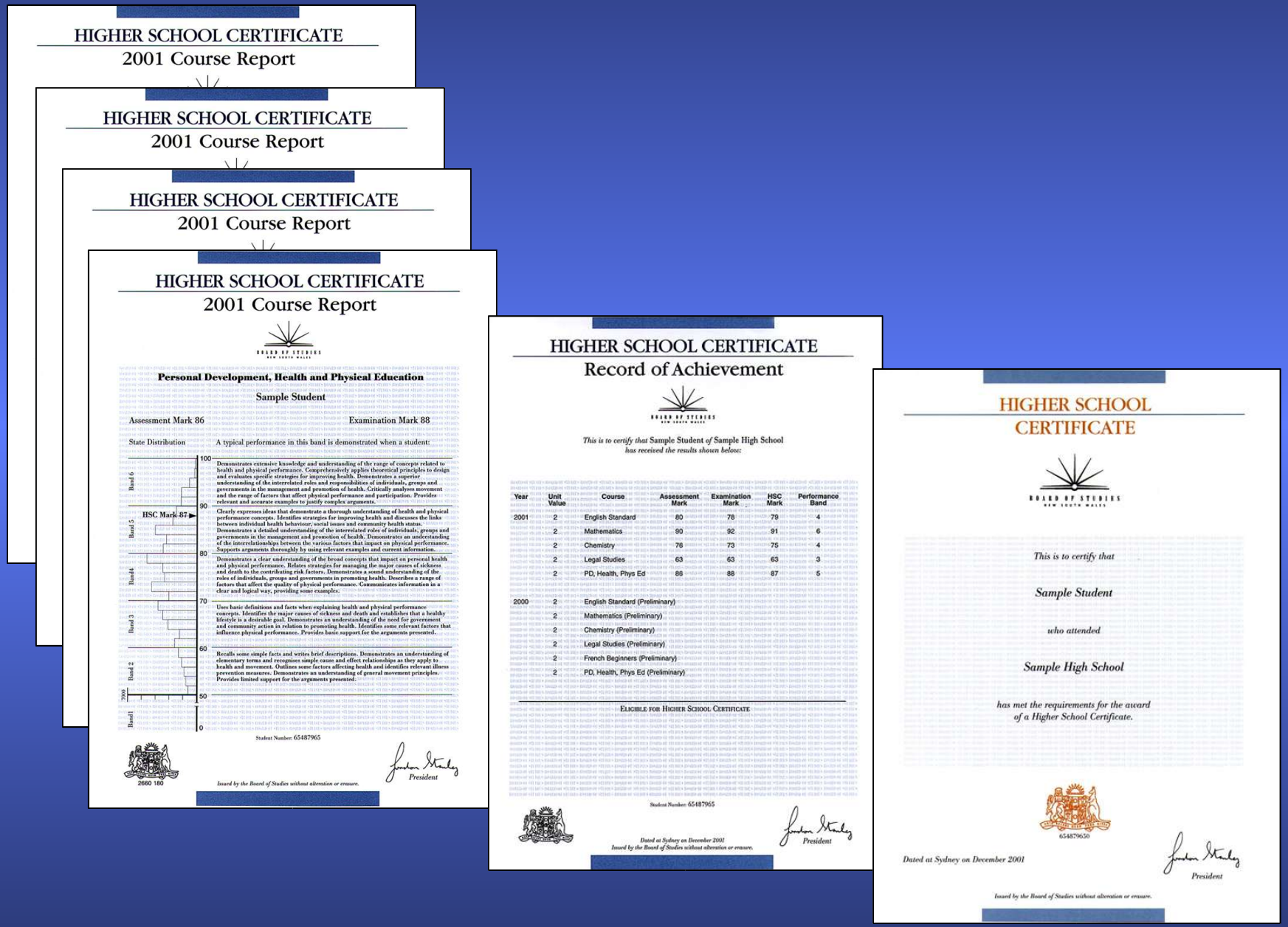
Public examinations

- High-stakes assessments based on curriculum
  - secondary certification and university entrance
  - selection for highly competitive courses (top 1½%)
- The comparability-over-time problem…
  - Grade distributions used to monitor standards
    - failure rate used as a measure of 'standards'
    - claim that if participation rates grow, grades should decline to ensure that an 'A' still and 'A', etc
    - do enough students fail?
  - Criterion (standard) and norm (cohort)-referencing
    - 'standards' are never absent (in curriculum, examination)
    - 'standards' are ignored in the norm-based award of results
    - cannot use link items over time, whole test must become public
    - marrying criterion and norm-referencing with judgments

Marrying criterion and norm-referencing

- England
  - use of criteria defined for some grade boundaries
  - review of previous years' scripts at grade boundaries
  - reference to prior grade distributions
  - reference to evidence of change in student cohort to justify shifts in grade distributions between years
- Australia (New South Wales)
  - development of band descriptors
  - 'consistent' definition of bands over years.
  - reporting with norm and criterion-referencing

NSW Higher School Certificate Documents



HIGHER SCHOOL CERTIFICATE
2001 Course Report

HIGHER SCHOOL CERTIFICATE
2001 Course Report

HIGHER SCHOOL CERTIFICATE
2001 Course Report

HIGHER SCHOOL CERTIFICATE
2001 Course Report

BOARD OF STUDIES
NEW SOUTH WALES

Personal Development, Health and Physical Education

Sample Student

Assessment Mark 86 — Examination Mark 88

State Distribution — A typical performance in this band is demonstrated when a student:

HSC Mark 87

100 — Demonstrates extensive knowledge and understanding of the range of concepts related to health and physical performance. Comprehensively applies theoretical principles to design and evaluates specific strategies for improving health. Demonstrates a superior understanding of the interrelated roles and responsibilities of individuals, groups and governments in the management and promotion of health. Critically analyses movement and the range of factors that affect physical performance and participation. Provides relevant and accurate examples to justify complex arguments.

90 — Clearly expresses ideas that demonstrate a thorough understanding of health and physical performance concepts. Identifies strategies for improving health and discusses the links between individual health behaviour, social issues and community health status. Demonstrates a detailed understanding of the interrelated roles of individuals, groups and governments in the management and promotion of health. Demonstrates an understanding of the interrelationships between the various factors that impact on physical performance. Supports arguments thoroughly by using relevant examples and current information.

80 — Demonstrates a clear understanding of the broad concepts that impact on personal health and physical performance. Relates strategies for managing the major causes of sickness and death to the contributing risk factors. Demonstrates a sound understanding of the roles of individuals, groups and governments in promoting health. Describes a range of factors that affect the quality of physical performance. Communicates information in a clear and logical way, providing some examples.

70 — Uses basic definitions and facts when explaining health and physical performance concepts. Identifies the major causes of sickness and death and establishes that a healthy lifestyle is a desirable goal. Demonstrates an understanding of the need for government and community action in relation to promoting health. Identifies some relevant factors that influence physical performance. Provides basic support for the arguments presented.

60 — Recalls some simple facts and writes brief descriptions. Demonstrates an understanding of elementary terms and recognises simple cause and effect relationships as they apply to health and movement. Outlines some factors affecting health and identifies relevant illness prevention measures. Demonstrates an understanding of general movement principles. Provides limited support for the arguments presented.

50

0

Band 6 · Band 5 · Band 4 · Band 3 · Band 2 · Band 1

2000

Student Number: 65487965

2660 180

Issued by the Board of Studies without alteration or erasure.

President

HIGHER SCHOOL CERTIFICATE
Record of Achievement

BOARD OF STUDIES
NEW SOUTH WALES

This is to certify that Sample Student *of* Sample High School
has received the results shown below:

Year	Unit Value	Course	Assessment Mark	Examination Mark	HSC Mark	Performance Band
2001	2	English Standard	80	78	79	4
	2	Mathematics	90	92	91	6
	2	Chemistry	76	73	75	4
	2	Legal Studies	63	63	63	3
	2	PD, Health, Phys Ed	86	88	87	5
2000	2	English Standard (Preliminary)				
	2	Mathematics (Preliminary)				
	2	Chemistry (Preliminary)				
	2	Legal Studies (Preliminary)				
	2	French Beginners (Preliminary)				
	2	PD, Health, Phys Ed (Preliminary)				

ELIGIBLE FOR HIGHER SCHOOL CERTIFICATE

Student Number: 65487965

Dated at Sydney on December 2001
Issued by the Board of Studies without alteration or erasure.

President

HIGHER SCHOOL
CERTIFICATE

BOARD OF STUDIES
NEW SOUTH WALES

This is to certify that

Sample Student

who attended

Sample High School

has met the requirements for the award of a Higher School Certificate.

65487965

Dated at Sydney on December 2001

President

Issued by the Board of Studies without alteration or erasure.

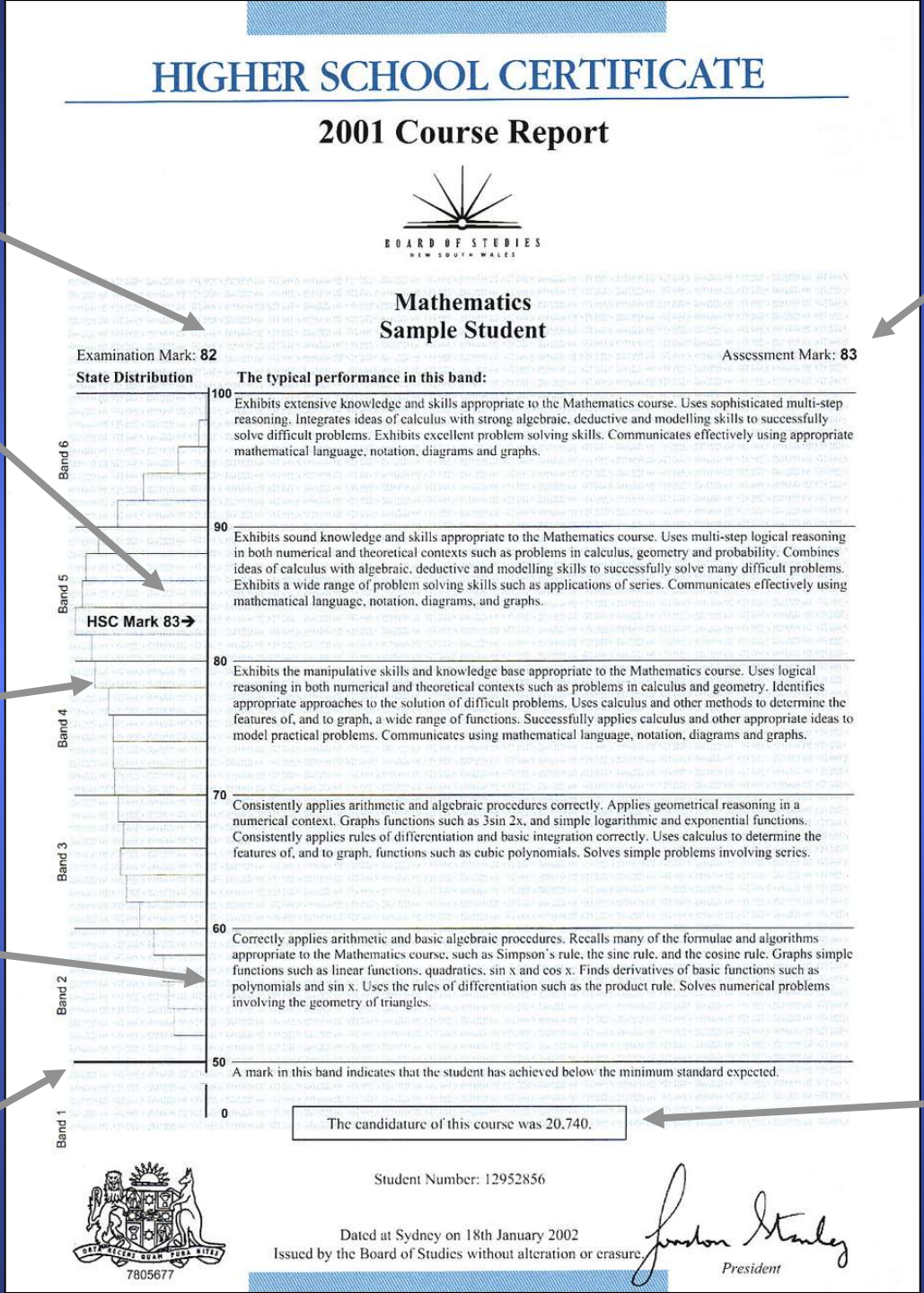
Student's examination mark

Student's overall mark

Norm referenced: distribution of results for all students

Mark Range: 0–100

Minimum standard expected (50)



HIGHER SCHOOL CERTIFICATE

2001 Course Report

BOARD OF STUDIES
NEW SOUTH WALES

Mathematics
Sample Student

Examination Mark: 82

Assessment Mark: 83

State Distribution

The typical performance in this band:

Band 6 — 100
Exhibits extensive knowledge and skills appropriate to the Mathematics course. Uses sophisticated multi-step reasoning. Integrates ideas of calculus with strong algebraic, deductive and modelling skills to successfully solve difficult problems. Exhibits excellent problem solving skills. Communicates effectively using appropriate mathematical language, notation, diagrams and graphs.

Band 5 — 90
Exhibits sound knowledge and skills appropriate to the Mathematics course. Uses multi-step logical reasoning in both numerical and theoretical contexts such as problems in calculus, geometry and probability. Combines ideas of calculus with algebraic, deductive and modelling skills to successfully solve many difficult problems. Exhibits a wide range of problem solving skills such as applications of series. Communicates effectively using mathematical language, notation, diagrams, and graphs.

HSC Mark 83→

Band 4 — 80
Exhibits the manipulative skills and knowledge base appropriate to the Mathematics course. Uses logical reasoning in both numerical and theoretical contexts such as problems in calculus and geometry. Identifies appropriate approaches to the solution of difficult problems. Uses calculus and other methods to determine the features of, and to graph, a wide range of functions. Successfully applies calculus and other appropriate ideas to model practical problems. Communicates using mathematical language, notation, diagrams and graphs.

Band 3 — 70
Consistently applies arithmetic and algebraic procedures correctly. Applies geometrical reasoning in a numerical context. Graphs functions such as 3sin 2x, and simple logarithmic and exponential functions. Consistently applies rules of differentiation and basic integration correctly. Uses calculus to determine the features of, and to graph, functions such as cubic polynomials. Solves simple problems involving series.

Band 2 — 60
Correctly applies arithmetic and basic algebraic procedures. Recalls many of the formulae and algorithms appropriate to the Mathematics course, such as Simpson's rule, the sine rule, and the cosine rule. Graphs simple functions such as linear functions, quadratics, sin x and cos x. Finds derivatives of basic functions such as polynomials and sin x. Uses the rules of differentiation such as the product rule. Solves numerical problems involving the geometry of triangles.

Band 1 — 50
A mark in this band indicates that the student has achieved below the minimum standard expected.

0

The candidature of this course was 20,740.

Student Number: 12952856

Dated at Sydney on 18th January 2002
Issued by the Board of Studies without alteration or erasure

7805677

President

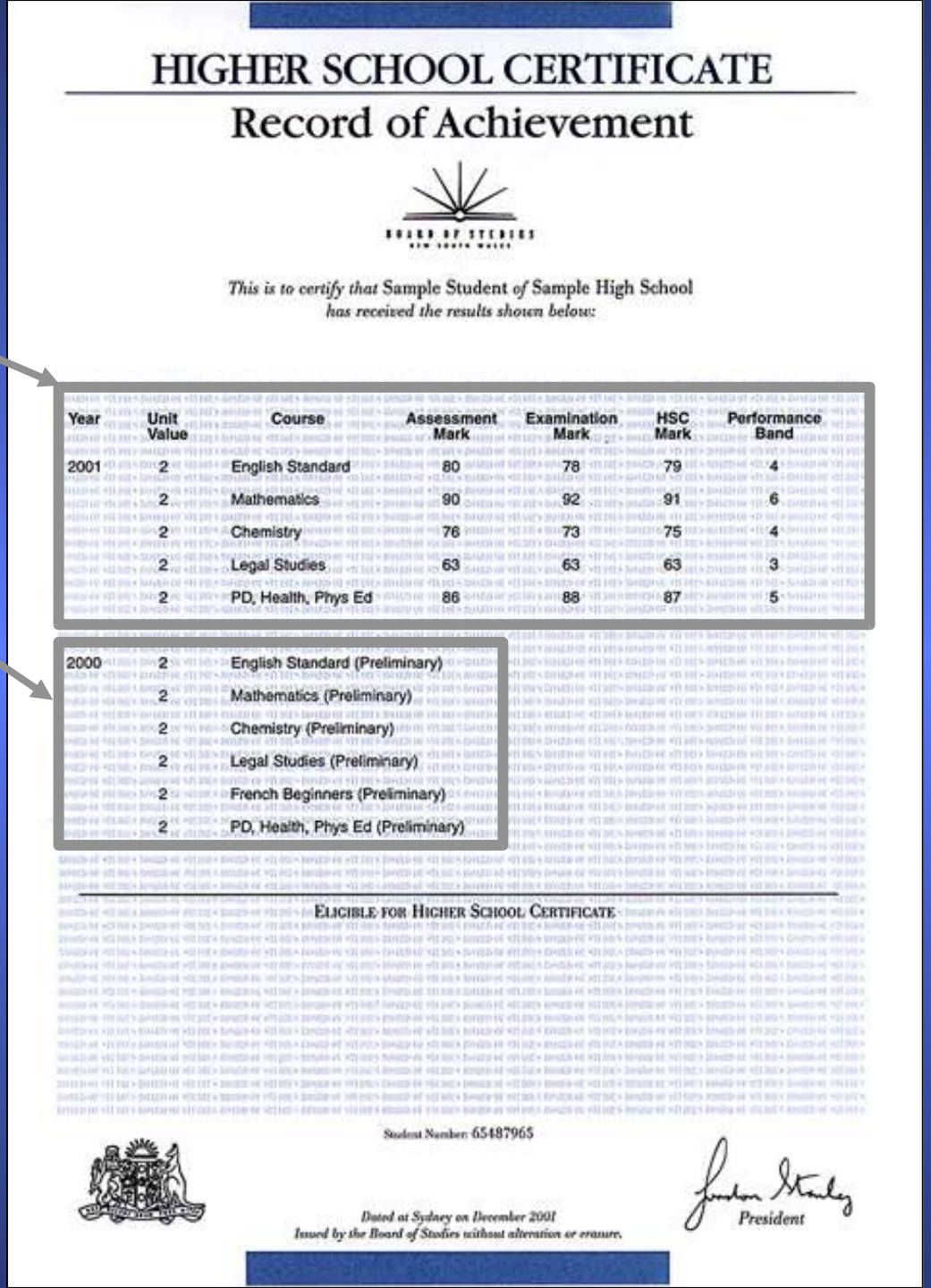
Student's school assessment mark

Standards referenced: bands describing what students know and can do

Number of candidates

All HSC courses listed with: (School) Assessment Mark, Examination Mark, (Overall) HSC Mark, Performance Band

All Preliminary courses listed



HIGHER SCHOOL CERTIFICATE

Record of Achievement

BOARD OF STUDIES
NEW SOUTH WALES

This is to certify that Sample Student *of* Sample High School
has received the results shown below:

Year	Unit Value	Course	Assessment Mark	Examination Mark	HSC Mark	Performance Band
2001	2	English Standard	80	78	79	4
	2	Mathematics	90	92	91	6
	2	Chemistry	76	73	75	4
	2	Legal Studies	63	63	63	3
	2	PD, Health, Phys Ed	86	88	87	5
2000	2	English Standard (Preliminary)				
	2	Mathematics (Preliminary)				
	2	Chemistry (Preliminary)				
	2	Legal Studies (Preliminary)				
	2	French Beginners (Preliminary)				
	2	PD, Health, Phys Ed (Preliminary)				

ELIGIBLE FOR HIGHER SCHOOL CERTIFICATE

Student Number 65487965

Dated at Sydney on December 2001
Issued by the Board of Studies without alteration or erasure.

President

How New South Wales got there…

- Review and recommendations for change
  - New NSW Higher School Certificate
    - McGaw, (1997). Shaping their future: Recommendations for reform of the Higher School Certificate. Sydney: Department of Training and Education Co-ordination
  - Scaling process
    - standards-referencing to curriculum and over-time
    - Bennett, J. (2001), Standards-setting and the NSW Higher School Certificate www.boardofstudies.nsw.edu.au/manuals/pdf_doc/bennett.pdf
- Developing grade descriptors
  - Used past examinations
    - experienced examiners for each subject
    - reviewed examination papers and students' marked papers
  - Developed band descriptors
    - described performance for Band 6 to 2, low Band 1 not described

Using grade descriptors

- Stage 1
  - Examiners independently form 'image of band'
  - Set cut mark for each band boundary on each question
- Stage 2
  - Examiners work together to reach agreement on boundary locations for bands on each question
  - Boundary locations for total scores also established
- Stage 3
  - Student work at boundaries on total scores reviewed
  - Cut points reviewed and determined
  - Boundaries located on mark scale
    - 5/6 boundary set to 90
    - 4/5 boundary set to 80
    - …
    - 1/2 boundary set to 50


Purpose of assessment:
formative or summative…

Distinguishing purposes of assessment

- Summative
  - To provide summary assessments of a student's at a particular stage.
  - Status of stages can vary:
    - Key stages: low stakes for students but high stakes for schools
    - Annual reports to parents from schools: low stakes
    - Public examinations: high stakes
- Formative
  - Frequent, interactive assessments of a student to identify learning needs and shape teaching
  - Barriers to widespread use:
    - tension between classroom-based formative assessment and high-visibility summative assessments
    - lack of connection between systemic, school and classroom approaches to assessment and evaluation

Benefits of formative assessment

Paul Black & Dylan Wiliam (1998). *Assessment and Classroom Learning. Assessment in education: Principles, policy and practice*, **5**, 7-74.

"Assessment becomes 'formative assessment' when evidence is actually used to adapt the teaching work to meet student needs."

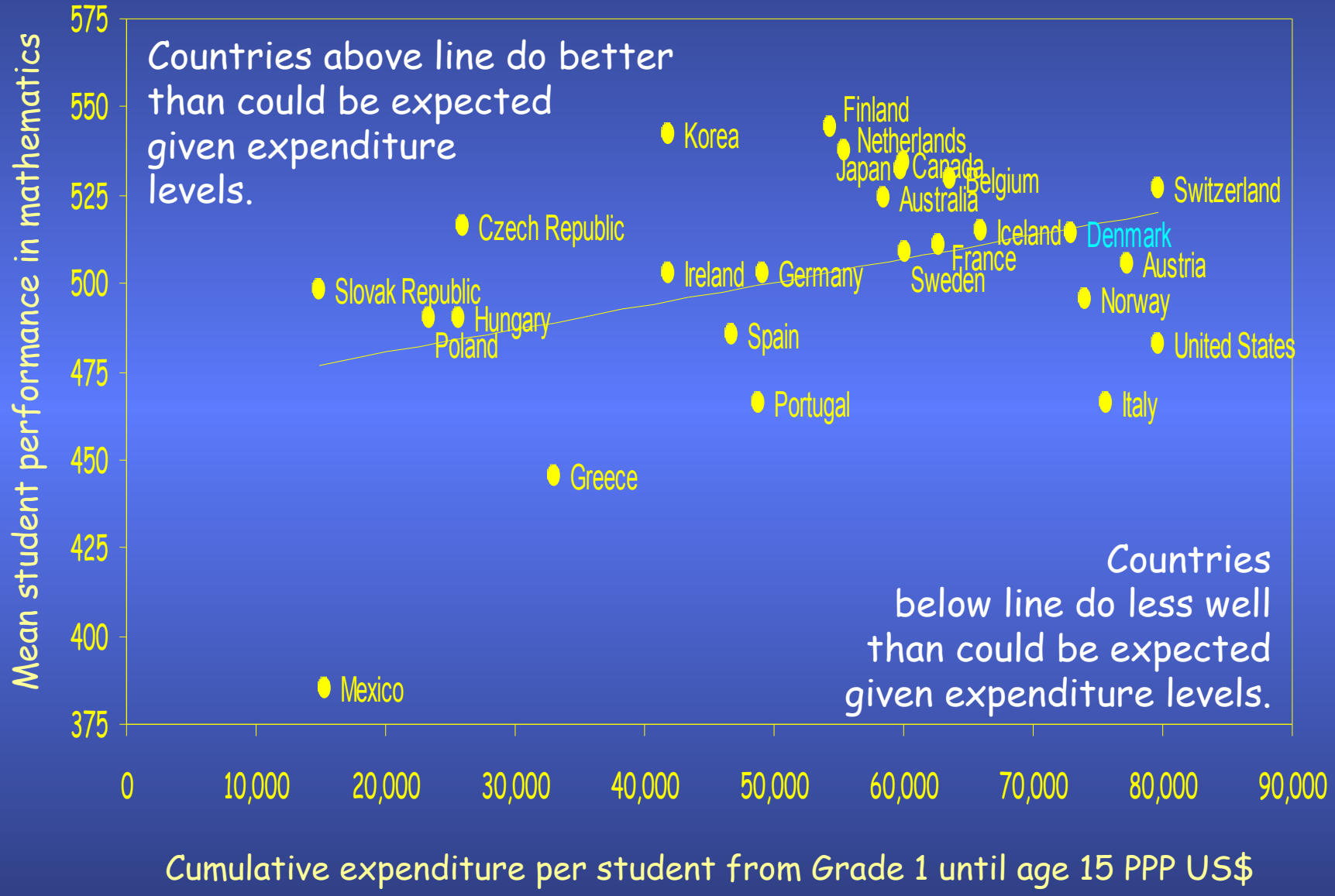
Formative assessment experiments produce effect sizes of .40 - .70, larger than found for most educational interventions.

Many studies show that improved formative assessments help low achievers most.

Formative assessment not a guarantee of quality

- Denmark's experience
  - Denmark's legislative commitment
    - requires schools to make comprehensive, versatile assessments
    - integrated into teaching to guide students, help teachers adapt

Denmark's PISA 2000 results



OECD (2004), Learning for tomorrow's world: First results from PISA 2003, Table 2.6, p.358.

Formative assessment not a guarantee of quality

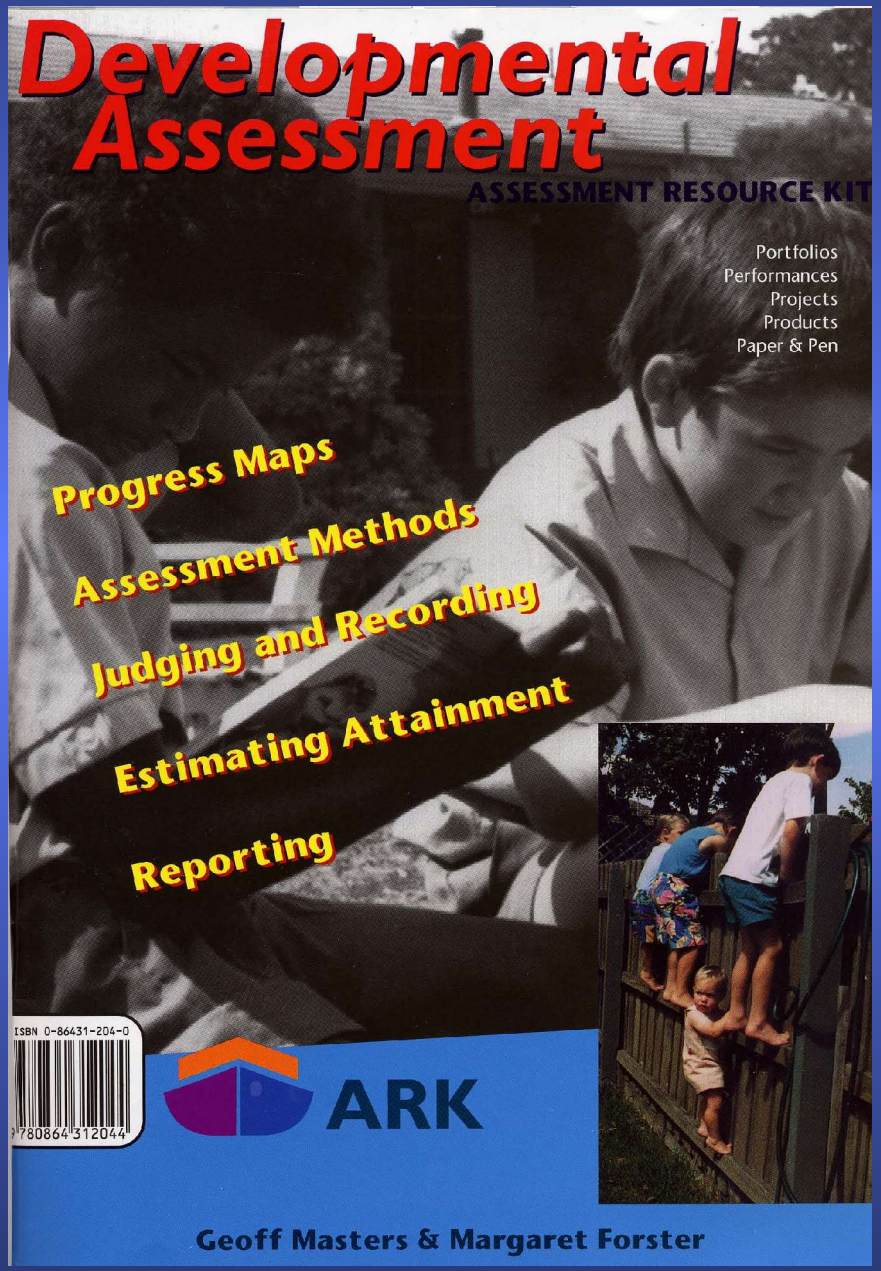
- Denmark's experience
  - Denmark's legislative commitment
    - requires schools to make comprehensive, versatile assessments
    - integrated into teaching to guide students, help teachers adapt
  - Denmark's PISA 2000 results
    - average performance, high expenditure per student
    - inefficient (cf Finland: top performance, average expenditure)
  - OECD review of Denmark
    - system lacks a culture of accountability
    - major recommendations for change, including system monitoring

Advancing formative assessment

- Legislation giving it priority
- Encouraging formative use of summative data
- Guidelines embedded in curriculum materials
- Provision of tools and exemplars
- Investment in initiatives incorporating it
- Investment in teacher professional development

OECD (2005), Formative assessment: Improving learning in secondary classrooms.

Exemplary materials



- Progress maps
  - Describe skills, understandings, and knowledge in the sequence in which they typically develop.
  - Give a picture of what it means to 'improve' in an area of learning.
- Assessment methods
- Judging and recording
- Estimating attainment
- Reporting


Aligning formative and summative assessment…

Aligning formative and summative assessment

- Scope of alignment
  - Basic: ensure policies do not conflict
  - Sophisticated: formative and summative reinforce other
- Strategies
  - Ensure summative assessments measure key skills on which development is expected to occur
  - Convince teachers that use of formative assessment will lead to better summative assessment results
  - Encourage risk-taking in teachers as they explore better ways of assessing and teaching
  - Broaden basis for judging teachers to include, for example, student motivation, students' capacity to judge own progress and (possibly) progress of others.

Benefits of good assessment

- Makes goals clear to learners
- Makes improvement clear
  - Normative assessment offers only improvement in rank (thus improvement at the expense of others)
  - Criterion (or standards) referenced assessment shows improvement in terms of knowledge and skills
- Teachers learners how to monitory own learning
  - Key meta-cognitive capacity
  - Builds the base for lifelong learning

Thank-you

Contact

Barry.McGaw@oecd.org

OECD education website

www.oecd.org/edu