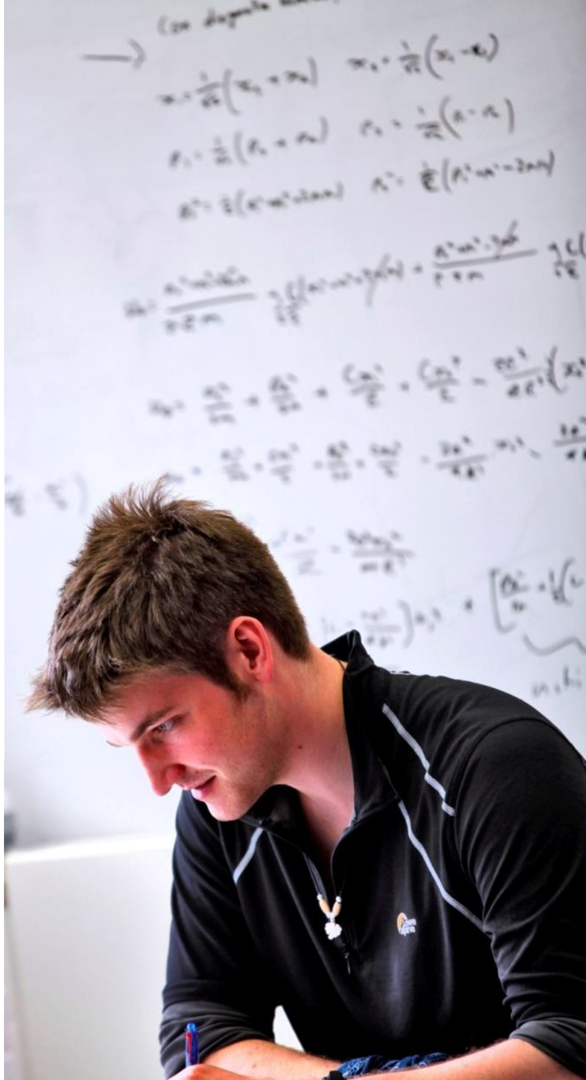
THE UNIVERSITY of EDINBURGH
School of Mathematics

Improving a Mathematics Diagnostic Test

Dr George Kinnear
G.Kinnear@ed.ac.uk
@georgekinnear

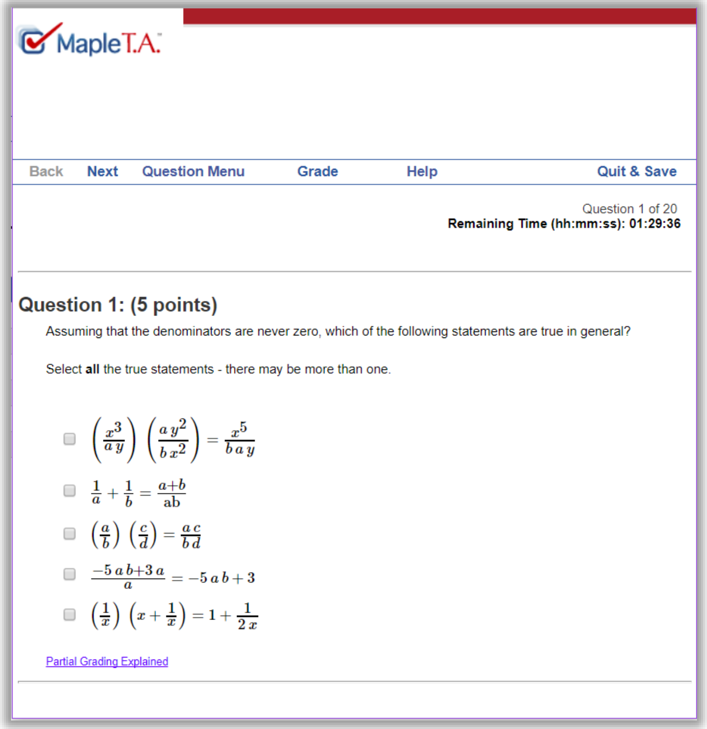


Outline

- Background of the test
- Analysis
- Implementing changes

The Mathematics Diagnostic Test (MDT)

- Administered online to incoming students
  - to help them study
  - to inform decisions
- Multiple choice and numerical answers
- Based on SQA Higher content



THE UNIVERSITY *of* EDINBURGH
School of Mathematics

History of the test

2011

32 questions,
Maple T.A.

2012

Project students
reduce to 20 questions

2013

Project students
check performance

. . .

2017

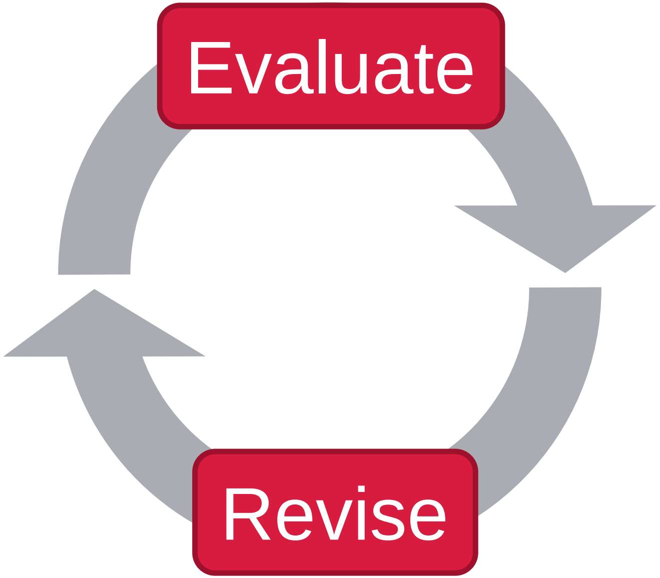
Move to STACK
Review of test content

Summer 2017: Project team

- George Kinnear
- Chris Sangwin
- Toby Bailey
- Tereza Burgetova
- Joanne Ruth Imanuel

Summer 2017: Project aims

- Evaluate effectiveness of existing test
- Produce revised test, informed by statistical analysis



Analysis

Classifying the questions

- We applied the “Mathematical Assessment Task Heirarchy” (Smith *et al*., 1996)
- Classification is based on the skills needed to complete the task successfully
- MATH was designed to help construct exams which test a broader range of skills

MATH Taxonomy

Group A Routine procedures	FKFS: Factual Knowledge and Fact Systems
	COMP: Comprehension
	RUOP: Routine Use of Procedures
Group B Using existing mathematical knowledge in new ways	IT: Information transfer
	AINS: Application in New Situations
Group C Application of conceptual knowledge	JI: Justifying and Interpreting
	ICC: Implications, Conjectures and Comparisons
	EVAL: Evaluation

Adapted from Darlington (2014)

Group A example

Question 6: (5 points)

The expression $16\ \cos(x) + 30\ \sin(x)$ can be written in the form $A\ \sin(x+\varphi)$, where $A > 0$ and $-\pi < \varphi < \pi$.

Find the values of A and φ . Give the value of φ , in radians, correct to at least three decimal places.

$A =$ [Num]

$\varphi =$ [Num]

- RUOP: Routine Use of Procedures
- Using a procedure/algorithm in a familiar context

Group B example

Question 14: (5 points)

A curve has equation $y = -\frac{1}{3}x^3 + x^2 - 4x + 9$.

The line $y = mx + c$ is a tangent to the curve at the point (a, b) .

(a) Find the values of m to complete the following statements:

- When $a =$ -1 , $m =$ ______ [Num]
- When $a = 2$, $m =$ ______ [Num]

(b) What is the maximum value of m , over all possible values of a ?.

______

Number

- AINS: Application in New Situations
- Choose and apply appropriate methods/information in new situations

Group A Routine procedures	FKFS: Factual Knowledge and Fact Systems	Recall previously learnt information
	COMP: Comprehension	Decide whether conditions of a simple definition are satisfied
	RUOP: Routine Use of Procedures	Using a procedure/algorithm in a familiar context
Group B Using existing mathematical knowledge in new ways	IT: Information transfer	Transferring information from verbal to numerical or vice versa
		Recognizing applicability of a generic formula in particular contexts
	AINS: Application in New Situations	Choose and apply appropriate methods/information in new situations

THE UNIVERSITY of EDINBURGH
School of Mathematics

Question 3: (5 points)

Functions g and h are defined on suitable domains by $h(x) = \frac{1}{2}\,x^2 + 4$ and $g(x) = 2^{-3\,x}$.

Given that $h(g(x)) = 2^{f(x)} + 4$, find an expression for $f(x)$.

- $6\,x$
- $-3\,x - 1$
- $-\frac{3}{2}\,x^2 + 4$
- $-6\,x - 1$

Question 8: (5 points)

Find the angle between the vectors $(\text{-}3, \text{-}4, 5)$ and $(\text{-}2, \text{-}4, \text{-}5)$.

Give your answer in radians, accurate to at least 3 decimal places.

Number

Question 20: (5 points)

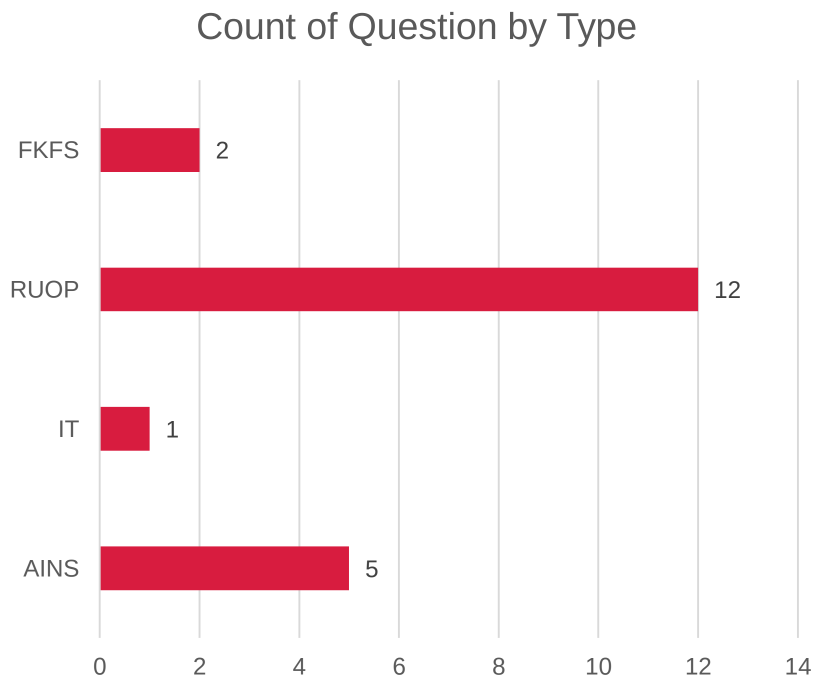
The function $f(x)$ is such that $f(-3) = -7$ and its derivative $f'(-3) = -9$.

Given that $g(x) = x\,f(x)$, what is the value of $g'(-3)$?

Number

MATH Taxonomy

- Overall in the MDT:
  - 70% were Group A (FKFS/RUOP)
  - 30% were Group B (IT/AINS)

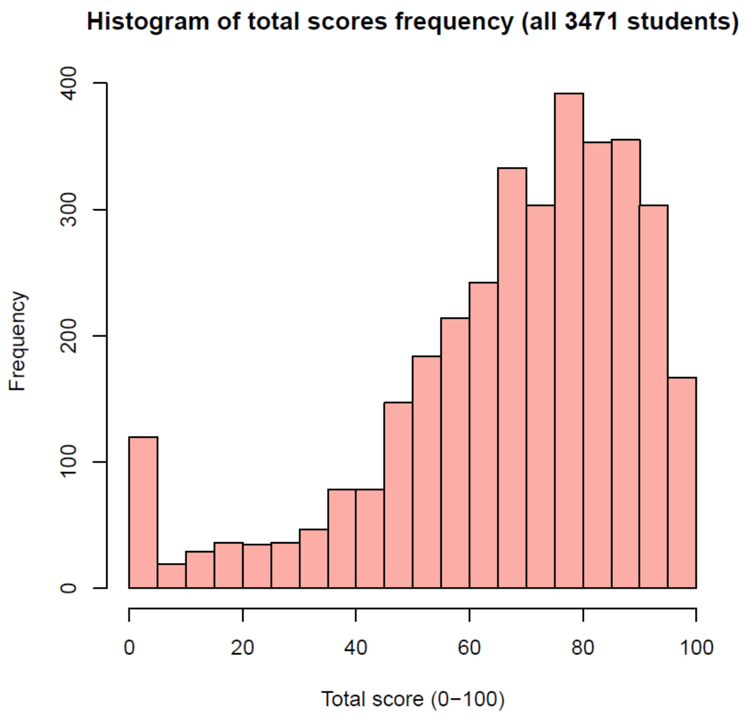


What we learned

- The test might benefit from more emphasis on Group B tasks
- Group C was completely missing

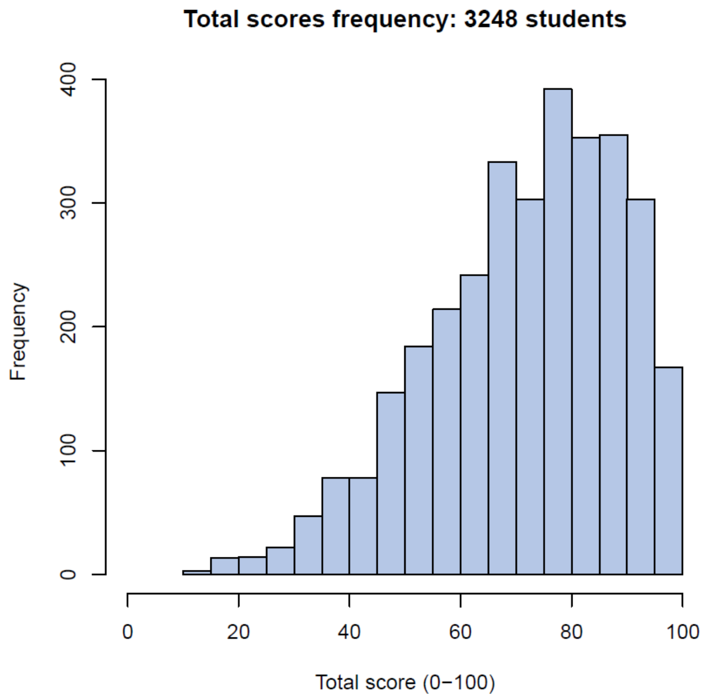
The data

- Raw scores for tests taken in 2013-2016
- Linked to student records (gender, entry qualifications, course results, …)



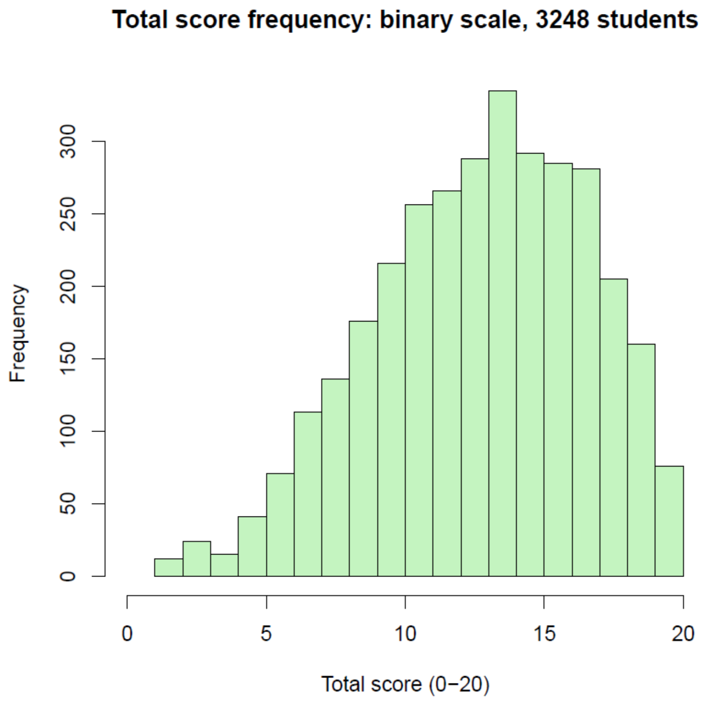
The data

- “Non-serious” attempts were identified and removed



The data

- Raw scores (5 marks per question) were turned into “binary” scores
- 1 mark for each question
- Must be completely correct to get the mark



Cronbach’s alpha

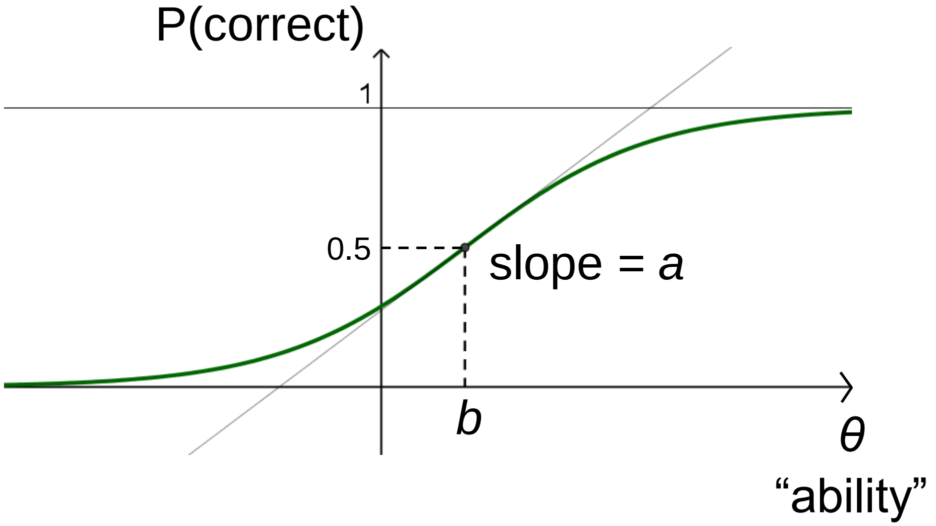
- A measure of the reliability of the test
  - Split the test into two halves
  - What is the correlation between the two halves?
  - Take the average of this over all possible splits
- For the MDT, α=**0.7848**

Item response theory

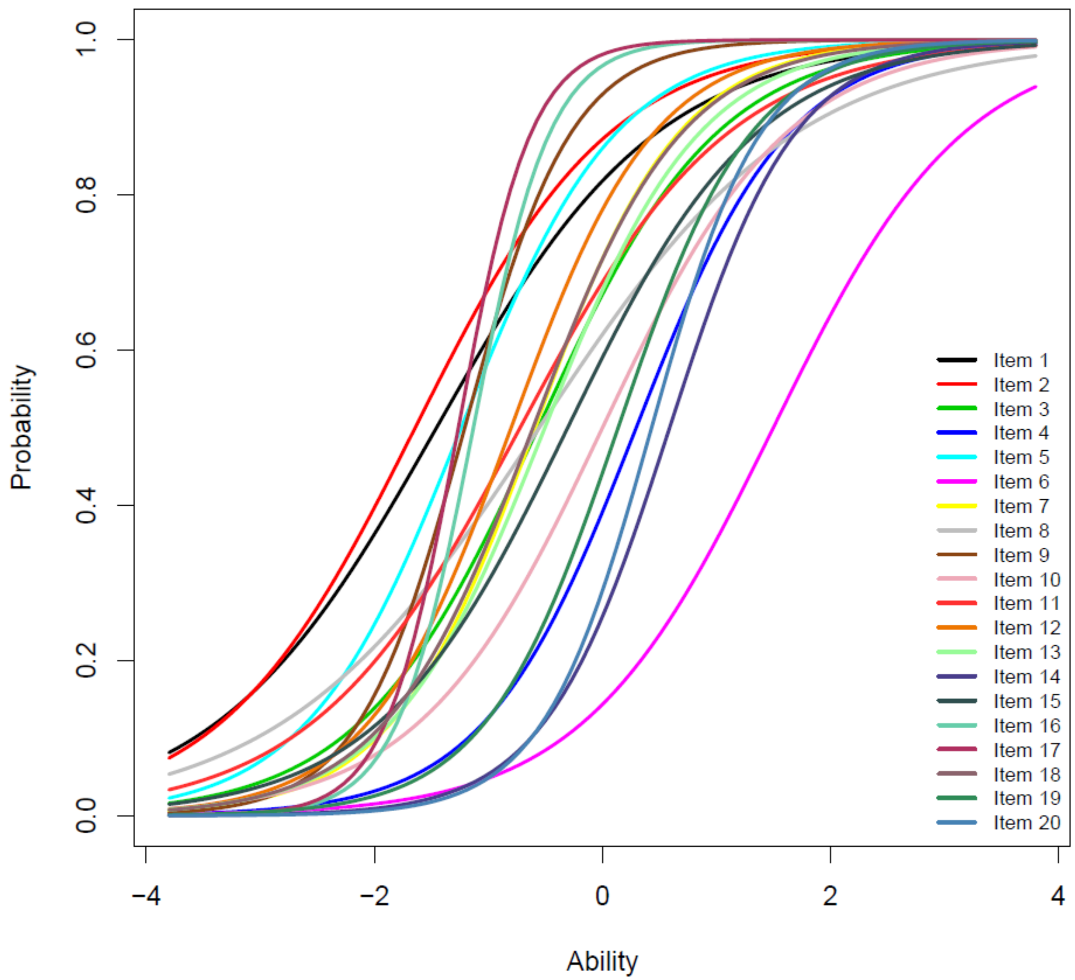
- A sophisticated model, assuming students' scores depend on their ability as well as properties of the question
- The probability of a student with ability θ answering correctly is modelled as:

$$P(\theta, b, a) = \frac{\exp\left[a(\theta - b)\right]}{1 + \exp\left[a(\theta - b)\right]}$$

where b is the difficulty and a is the discrimination

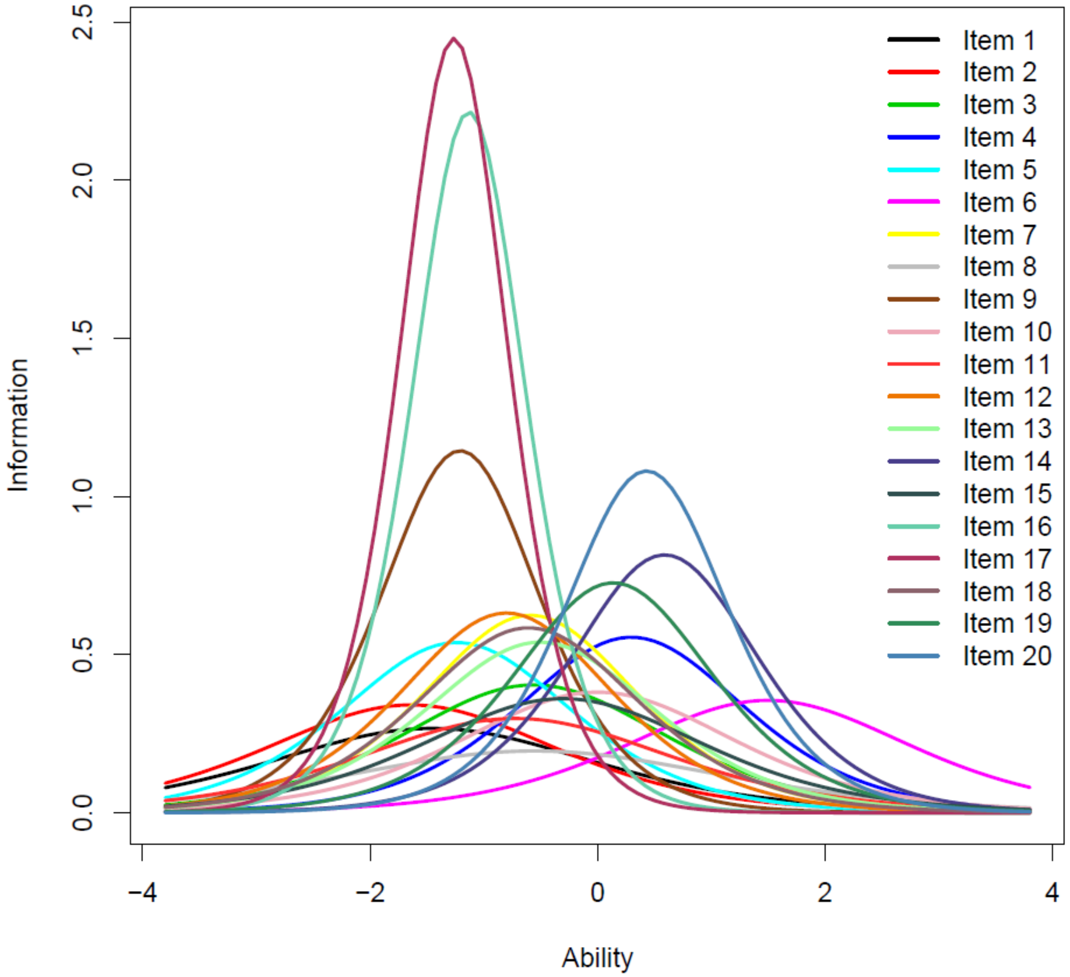


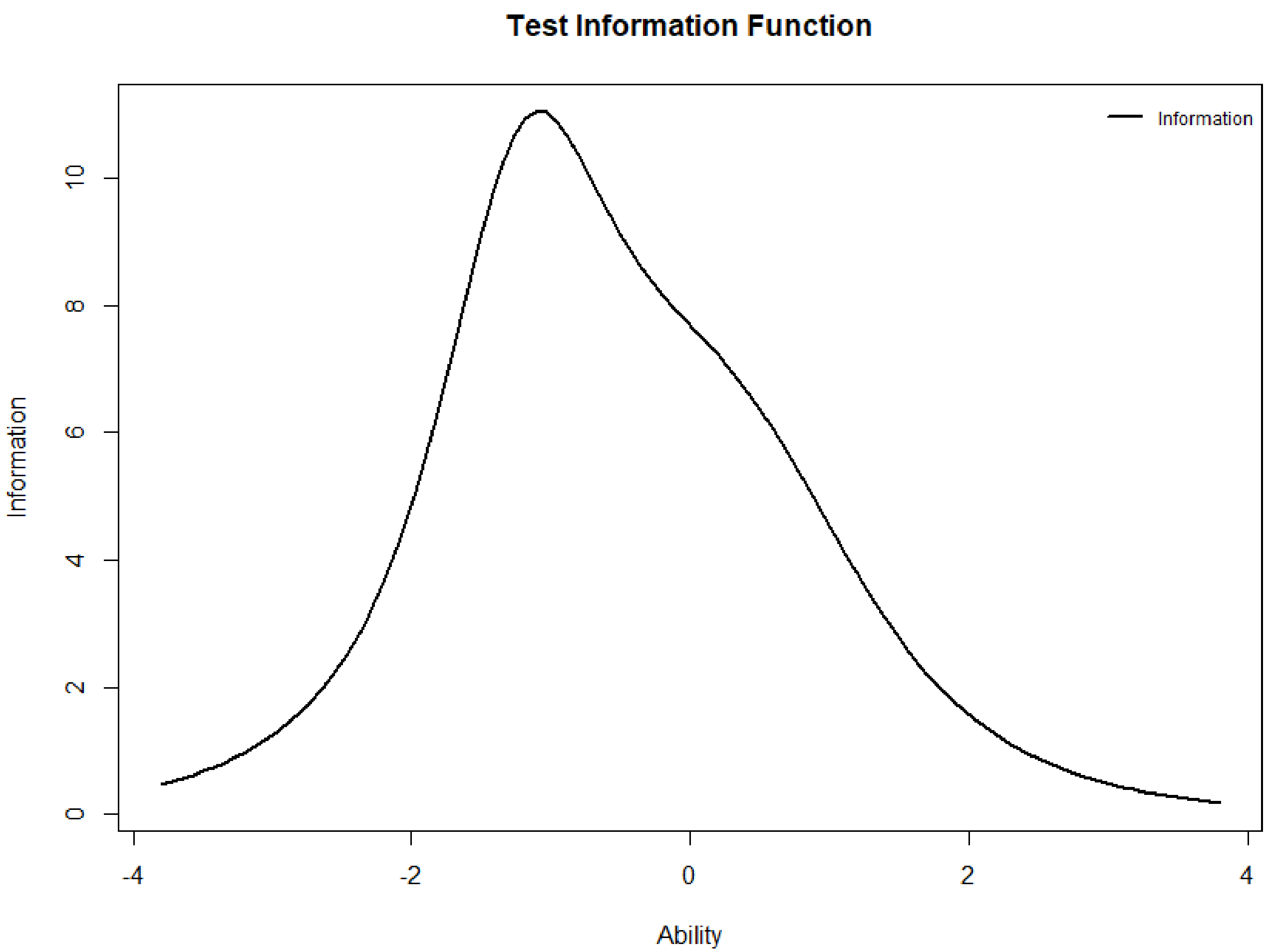
Item Characteristic Curves



THE UNIVERSITY of E
School of Mathem

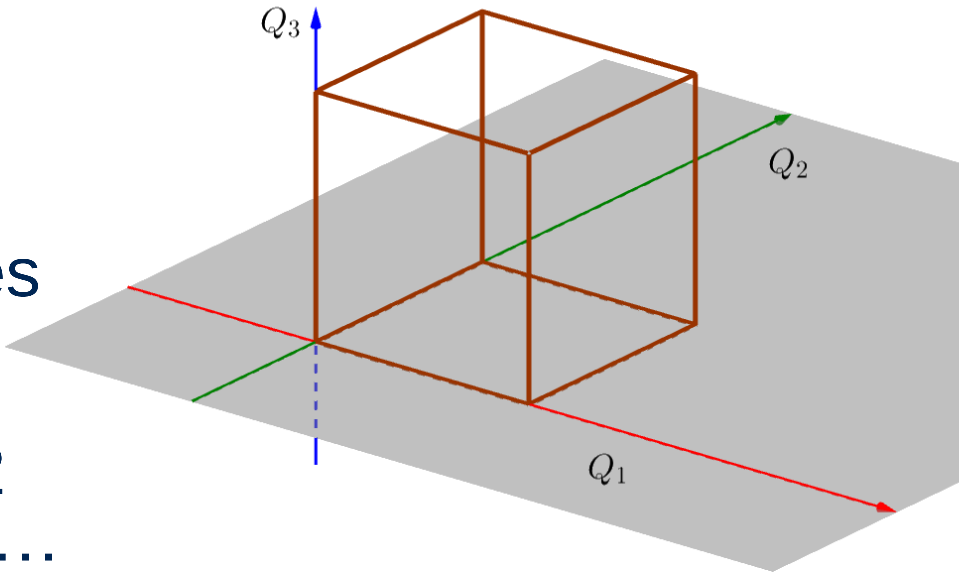
Item Information Curves




Test Information Function
Information
Information
0
2
4
6
8
10
-4
-2
0
2
4
Ability

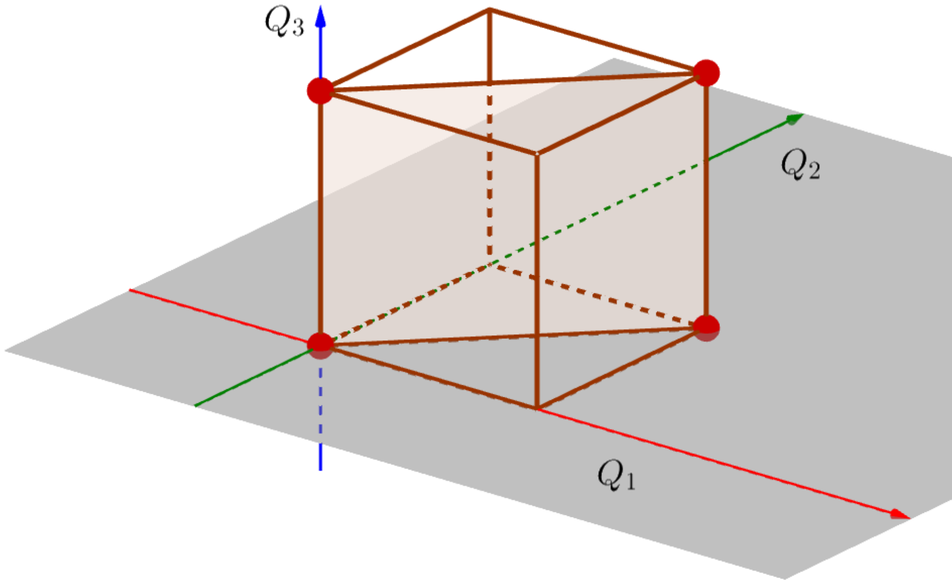
Factor analysis

- Suppose we had 3 questions, scored 0 or 1
- The possible student responses are the vertices of the unit cube
- Now suppose Q1 and Q2 are related, but Q3 is not…

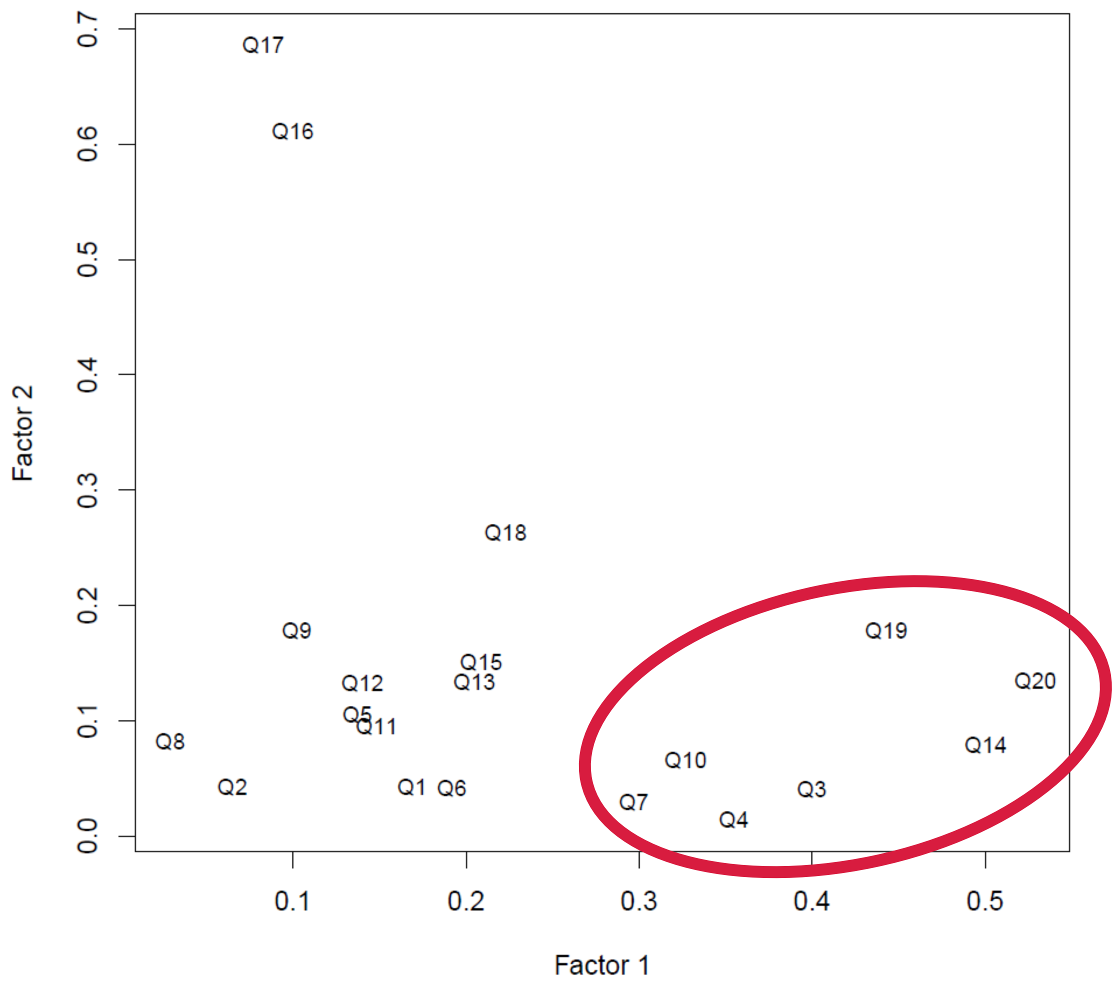


Factor analysis

- Most of our data points will lie on the vertices with Q1=Q2
- So rather than 3D data, it's essentially 2D



Questions factor loadings on Factor 1 vs on Factor 2



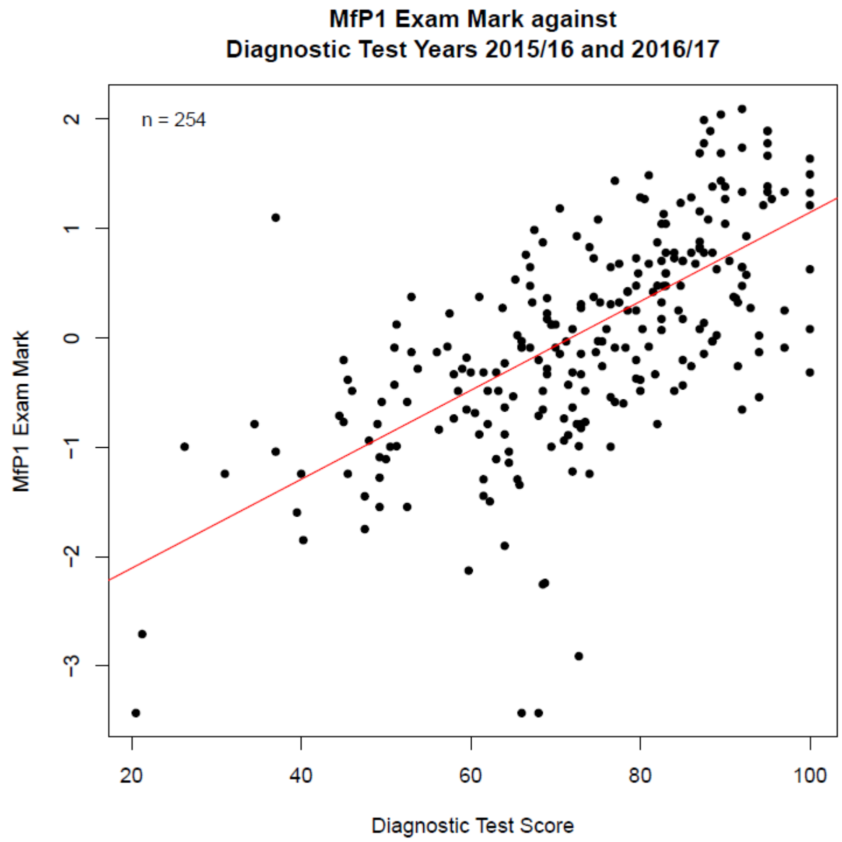
MATH
Group B
(mostly)

What we learned

- The reliability of the test is acceptable
- Most items are performing very well, but some are poor discriminators
- The test could be better at distinguishing students of medium-to-high ability
- We can see a distinction between Group A and B questions in the student response data

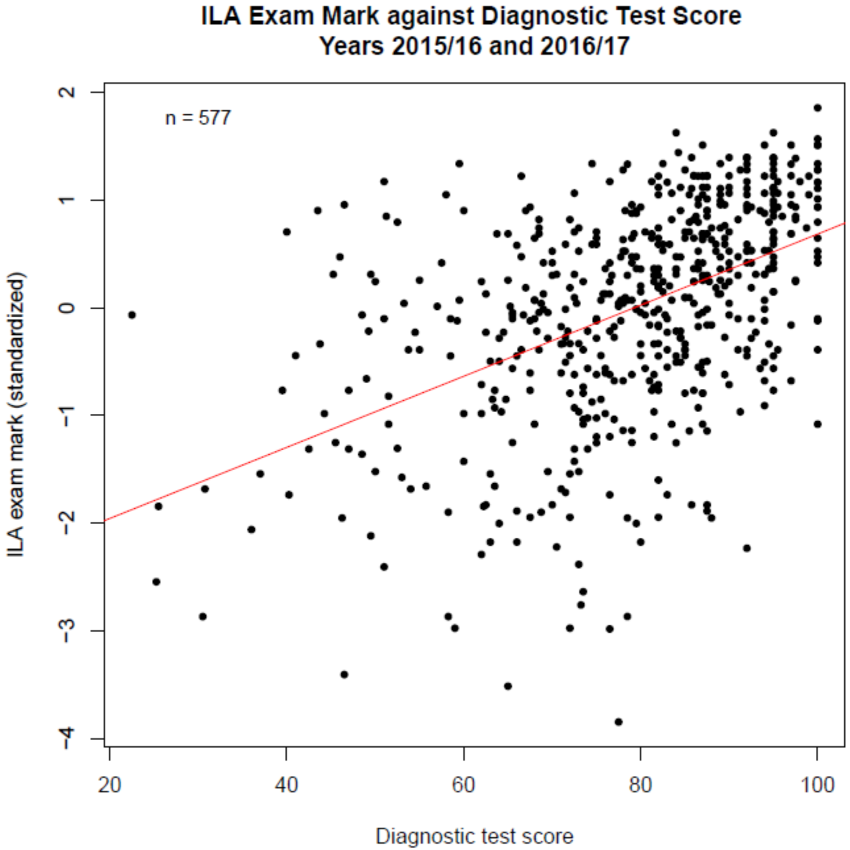
Relationship to later performance

- The test is a reasonably good predictor of Year 1 performance
- The strongest correlation was with Mathematics for Physics 1 (0.643)



Relationship to later performance

- Correlation with Introduction to Linear Algebra is 0.477
- Analysis of variance suggests that Group B questions are the best predictors



Implementing changes

Goals

- Remove poorly performing items
- Introduce:
  - a greater proportion of Group B questions
  - at least one Group C question
- Try to add items with good discrimination at higher ability level

Results

- 941 attempts so far
- From data generated by Moodle:
  - Cronbach's alpha: 0.8595 (up from 0.7848)
  - The two new Group B questions seem to be among the more difficult questions
- More detailed analysis to follow in 2018…

Conclusion

- The MATH taxonomy can be a useful tool when thinking about test design

- Statistical tools can also help to produce a more focused test
  - Cronbach's alpha
  - Facility/discrimination/IRT
  - Factor analysis

Thank you!

References

- Darlington, E. (2014) ‘Contrasts in mathematical challenges in A-level Mathematics and Further Mathematics, and undergraduate mathematics examinations’, *Teaching Mathematics and its Applications*. Oxford University Press, 33(4), pp. 213–229. doi: 10.1093/teamat/hru021.
- Smith, G. *et al.* (1996) ‘Constructing mathematical examinations to assess a range of knowledge and skills’, *International Journal of Mathematical Education in Science and Technology*. Taylor & Francis Group , 27(1), pp. 65–77. doi: 10.1080/0020739960270109.