Improving an online diagnostic test via item analysis

George Kinnear, University of Edinburgh (G.Kinnear@ed.ac.uk)

THE UNIVERSITY of EDINBURGH

Diagnostic tests are widely used in university mathematics departments. Here, we demonstrate a set of methods for evaluating the performance of such a test, based on the example of improving the test used by the University of Edinburgh.

Aims of the test

- Inform staff about the preparedness of incoming undergraduates.
- Show students the sort of skills we expect them to have, and give them a chance to self-diagnose topics requiring further study.
- Inform decisions about course choice, e.g. direct entry to Year 2.

Project timeline

2011 Original 32 question test created using Maple T.A.

2012 Test is reduced to current form of 20 questions

2013 Project evaluates performance of the new test

2016 There are now ≈1000 test attempts each year

2017 With the test moving to the STACK online assessment system, the current project re-evaluates its performance and implements changes

Test format

- 20 questions on core high-school mathematics
- 90 minute time limit, one attempt allowed (a practice quiz is available)
- Questions are a mixture of numerical response and multiple choice
- All questions make use of randomisation
- Feedback is limited to marks and whether correct (no worked solutions)

Example questions

Q3

Functions g and h are defined on suitable domains by $g(x) = 3^{-3x}$ and $h(x) = \frac{x^2}{9} + 2$.

Given that $h(g(x)) = 3^{f(x)} + 2$, find an expression for $f(x)$.

Which one of the following is the correct expression for $f(x)$?

- $-6x - 2$
- $2 - \frac{x^2}{3}$
- $12x$
- $-3x - 2$

Q6

The expression $15\sin(x) + 8\cos(x)$ 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 =

φ =

Q8

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

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

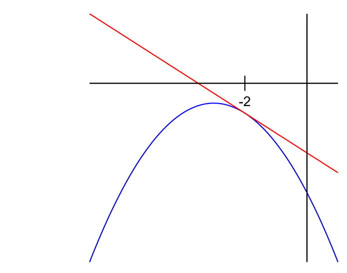
Q17

Which one of the following is the derivative of $(2x^2+6)^3$ with respect to x?

- $6(2x^2+6)^2$
- $\frac{(2x^2+6)^4}{16x}$
- $48x^5$
- $12x(2x^2+6)^2$

Q13

The curve with equation $y = -t^2 - 6t - 11$ has a tangent at $t = -2$, as shown in the diagram.



The tangent has equation $y = mt + c$. What are the values of m and c?

$m =$

$c =$

Q14

A curve has equation $y = -\frac{x^3}{3} + 3x^2 - 33x - 8$.

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 $a = -2$, $m =$
- When $a = 1$, $m =$

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

Q20

The function $f(x)$ is such that $f(4) = 11$ and its derivative $f'(4) = -7$.

Given that $g(x) = xf(x)$, what is the value of $g'(4)$?

$g'(4) =$

Evaluation methods

1. Content

The MATH taxonomy (Smith et al., 1996) was applied to judge the type of mathematical skill assessed by each question. The results show the test is predominantly based on "routine procedures", which is perhaps not surprising for a test of this type.

To make the test more balanced, three Group A questions were removed, and replaced by two Group B and one Group C question.

	Old	New
Group A Routine Procedures	14	11
Group B Using knowledge in new ways	6	8
Group C Mathematical arguments	0	1

New Group C question

Given two integers a and b,

- $\max(a,b)$ denotes the maximum of a and b, e.g. $\max(10,20) = 20$,
- $\min(a,b)$ denotes the minimum of a and b, e.g. $\min(10,20) = 10$.

(a) Evaluate the following expressions:

$\max(\min(10,8),12) =$

$\min(\max(10,8),12) =$

(b) Give values of a, b, c for which the following inequality is **false**:

$$\max(\min(a,b),c) > \min(\max(a,b),c).$$

$a =$

$b =$

$c =$

NQ11

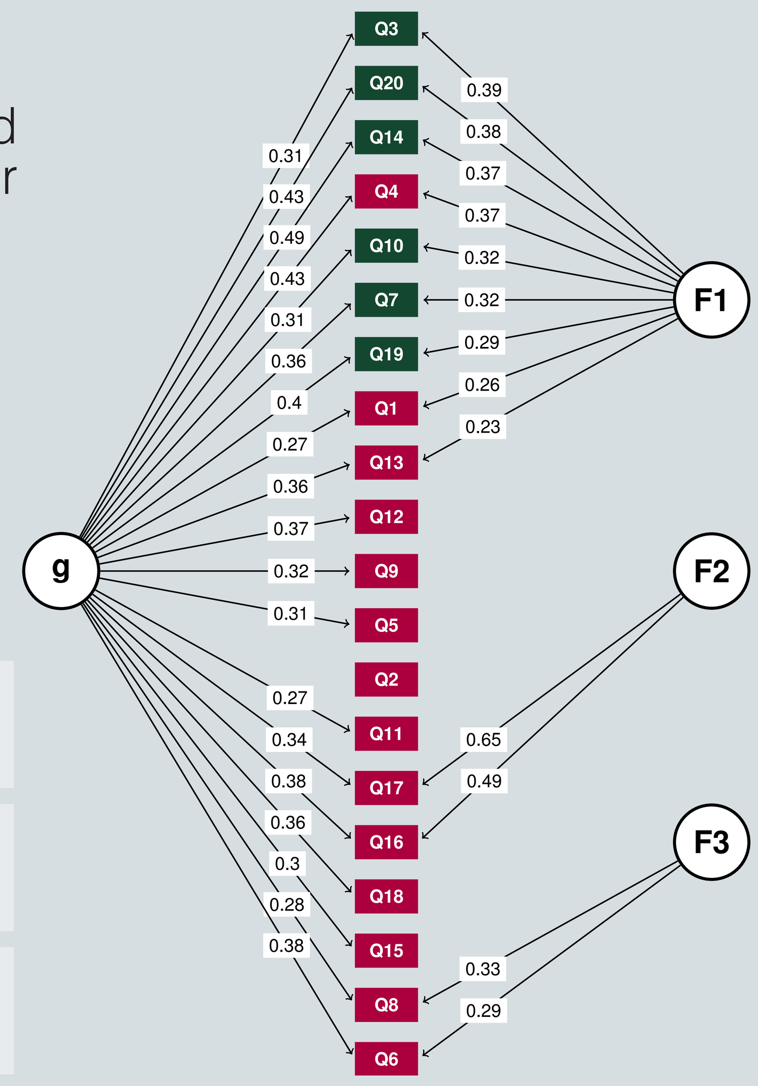
2. Factor analysis

Factor analysis aims to explain observed scores in terms of underlying factors or latent traits, e.g. mathematical ability.

Results from the R package *psych* (Revelle, 2018) showed:

- good loading on a general factor, g, indicating that all of the questions are assessing a common construct
- three subsidiary group factors:

F1	Predominantly questions from MATH Group B
F2	Chain rule, perhaps calculus more generally
F3	Questions requiring the use of a calculator



3. Psychometric analysis

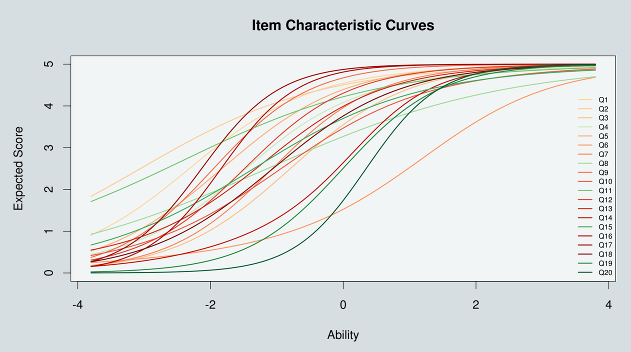
Assuming the test is measuring a single trait, e.g. mathematical ability, item response theory uses the response data to estimate the trait level for each student.

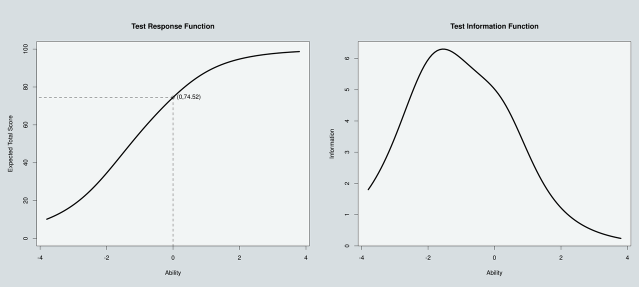
For each question, the expected score for any given ability is modelled by a logistic (S-shaped) curve, with two parameters:

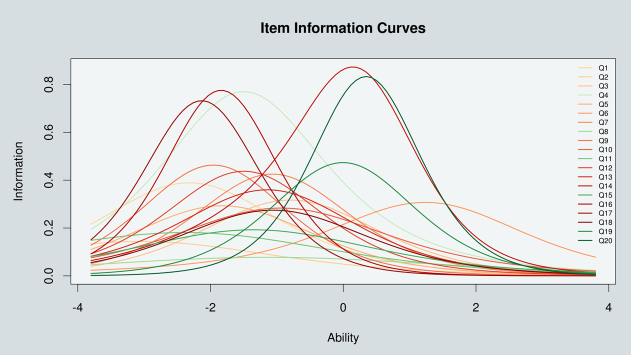
- **difficulty** – horizontal shift (left=easier)
- **discrimination** – the maximum slope

The three items selected for replacement were among the least discriminating items.

The information curve for each question shows where it is most discriminating on the ability scale. Their sum, the test information function, shows the test is most informative at the lower end of the ability scale.



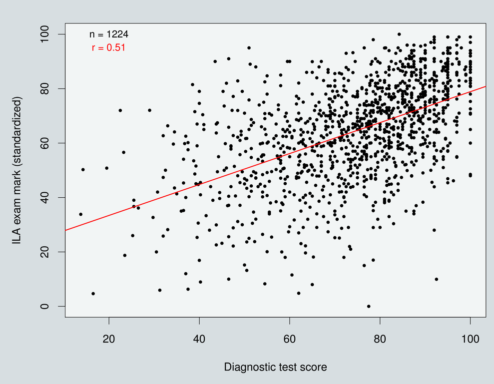




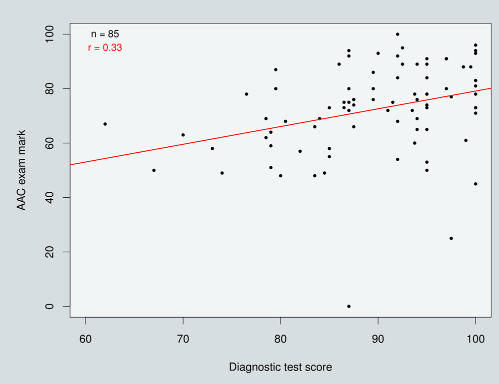
4. Predictive validity

Diagnostic Test scores are used in practice to help inform course choice, so it is important to understand their relationship with subsequent performance. The scores correlated moderately with the first course results in three different tracks:

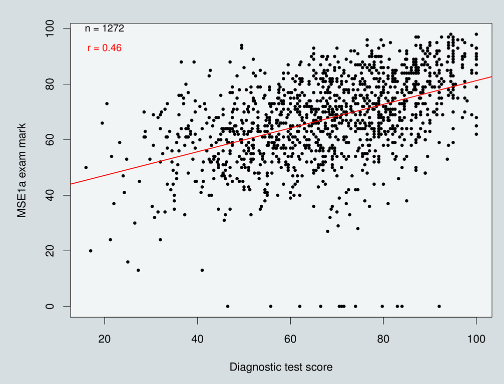
Year 1 Mathematics r=0.51



Year 2 Direct Entry r=0.33



Year 1 Engineering Mathematics r=0.46



Interestingly, slightly higher correlations were found using the sub-score on the MATH Group B questions, e.g. r=0.53 for Year 1 Mathematics. This contributed to the decision to rebalance the test in favour of Group B.

References

Revelle, W. (2018). How To: Use the psych package for Factor Analysis and data reduction. [online] Available at: http://personality-project.org/r/psych/HowTo/factor.pdf [Accessed 31 Jul. 2018].

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.

Acknowledgements

This project was supported with funding from the Principal's Teaching Award Scheme at the University of Edinburgh. Thanks are due to the other project team members: Tereza Burgetova, Joanne Ruth Imanuel, and Chito Wang (who conducted most of the statistical analysis), Chris Sangwin and Toby Bailey.