

The fairness of tests: Item-level DIF analysis of high-stakes exams in England

Ben Smith & Ruth Johnson



theguardian
website of the year

GCSE results: biggest gap in 11 years
between boys and girls A*-C pass rate

CMPO Working Paper Series No. 03/084

Girls Rock, Boys Roll: An Analysis of the Age 14-16 Gender Gap in English Schools

Simon Burgess*
Brendon McConnell**
Carol Propper*
Deborah Wilson**

*CASE, LSE, CEPR and CMPO, The University of Bristol
**CMPO, The University of Bristol

BBC

Sign in

NEWS

Gender gap

The gender gap has widened further - with 71.3% of entries by girls getting a C grade or above compared with 62.4% of boys.

The Telegraph

HOME

Lifestyle | Men

[Thinking Man](#) | [Active](#) | [Fashion and Style](#) | [Relationships](#) | [The Filter](#)

Home > Lifestyle > Men > Thinking Man

How do we close the education gender gap?

**Gender and
education:
the evidence on
pupils in England**



department for
education and skills

Creating Opportunity
Releasing Potential
Achieving Excellence

Policy interventions

- “The pupil premium is additional funding for publicly funded schools in England to raise the attainment of disadvantaged pupils of all abilities and to close the gaps between them and their peers.” gov.co.uk

Raising Boys' Achievement

Mike Younger and Molly Warrington
with
John Gray, Jean Rudduck, Ros McLellan,
Eva Bearne, Ruth Kershner and Pat Bricheno

University of Cambridge Faculty of Education

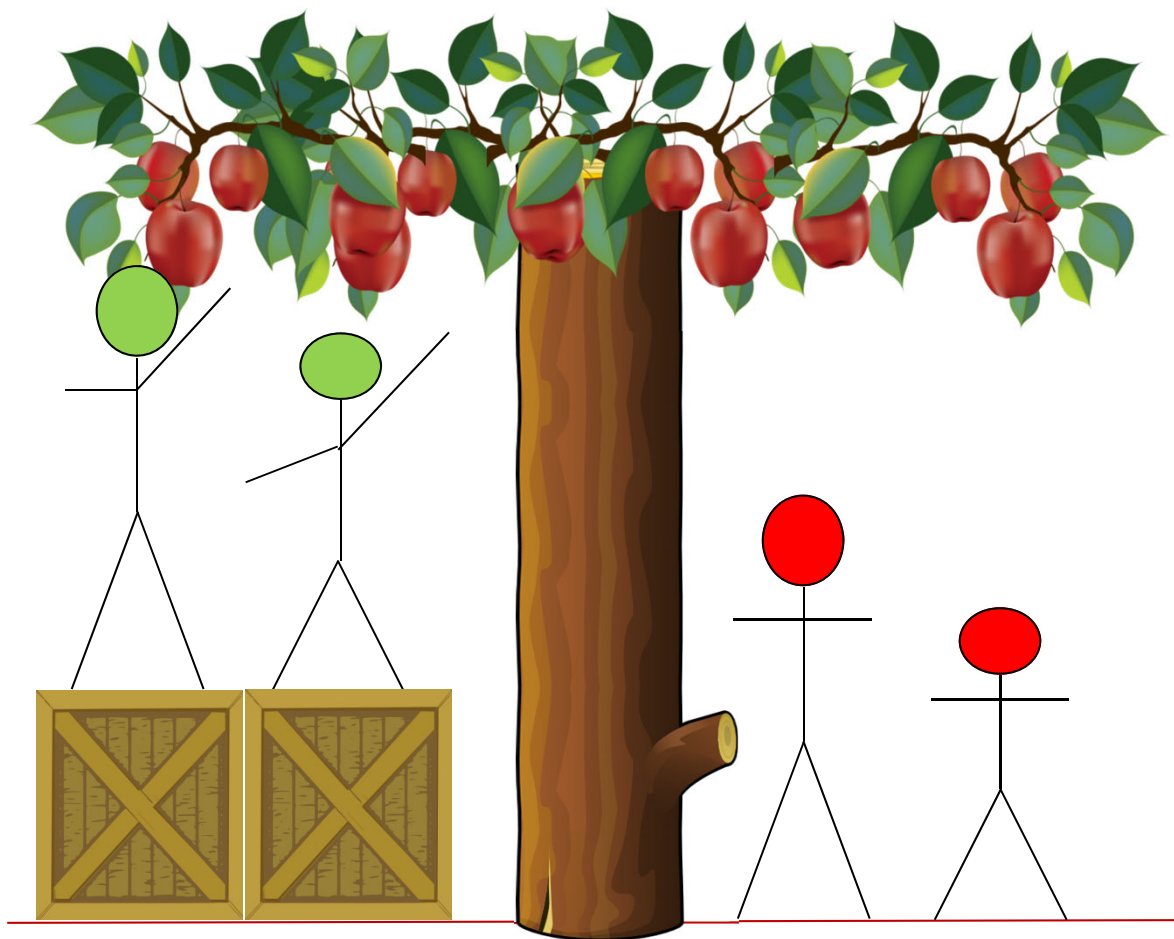
cerp.org.uk



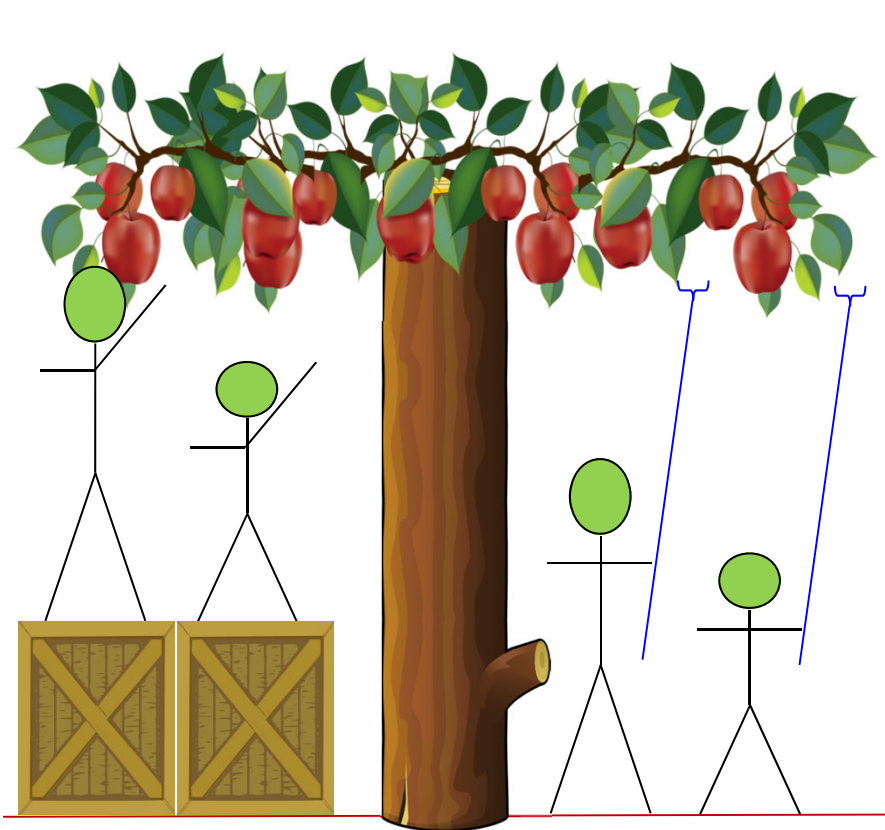
Promoting the education of looked after children

Statutory guidance for local authorities

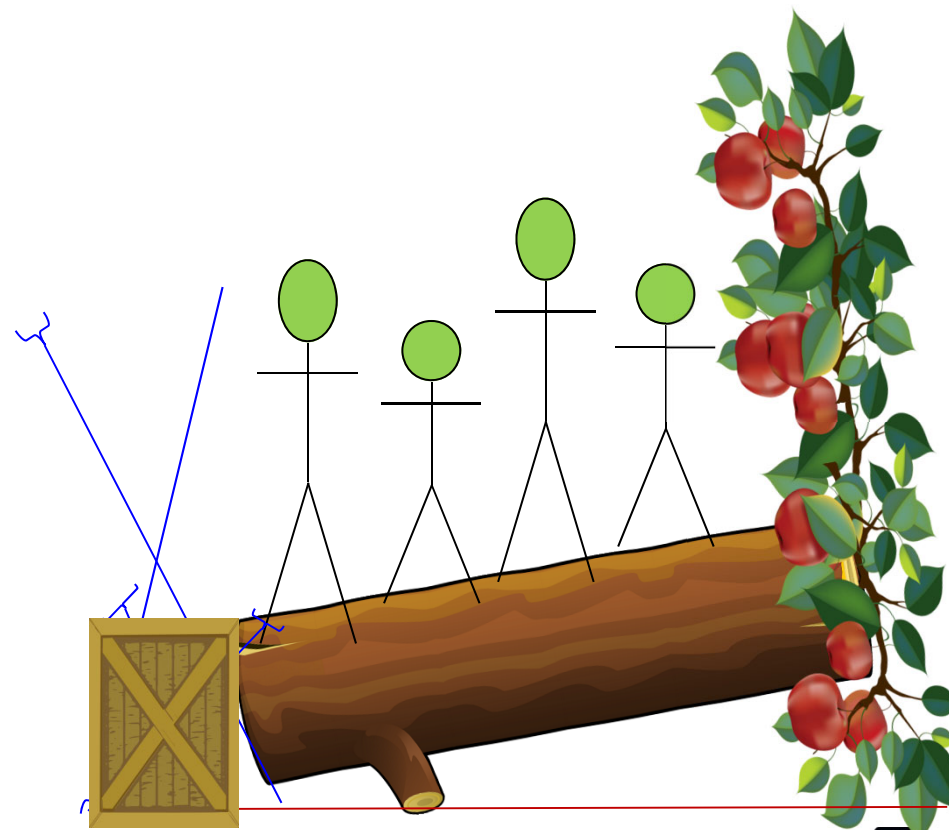
What do we mean by 'fairness' in assessment?



How can we make things more equitable?



cerp.org.uk



AQA

DIF: differential item functioning

- DIF has been found over numerous variables:
 - Gender (Elwood, 1995; Carlton, 2000; Twist & Sainsbury, 2009)
 - Social class/deprivation (Cooper, 1998; Cooper and Dunne, 1998)
 - Ethnicity (Demack, Drew and Grimsley, 2000; Rudduck, 1999)
 - Language (Pollitt, Marriott and Ahmed, 2000; Theodosiou-Zipiti & Lamprianou, 2015)
- Often in specific types of item:
(Chilisa, 2000; Elwood, 1994; Gipps and Murphy, 1994)
 - MCQs
 - Essays
 - Content

Our research

- Gaps in research:
 - Most 'attainment gap' research at qualification level – not item level
 - Little work is intersectional
 - Little recent work focuses on GCSEs
- Research questions
 - What DIF exists in GCSEs?
 - Can we use this knowledge to inform design of items and question papers?

Data and variables of interest

- Item-level marks for items from two tiered papers:
 - AQA GCSE English Language
 - AQA GCSE Mathematics B (Linear)
 - Non-calculator paper
- Variables
 - Gender
 - Ethnicity (minor ethnic group code)
 - 1st language (minor 1st language code)
 - SEN: special educational needs status
 - IDACI: income deprivation affecting children index
 - CLA: In care status (for at least 1 day this year)

Data analysis

- Latent class analysis
 - Identifies unmeasured (latent) class membership using the observed variables
 - Tested how many classes best fit the data based upon:
 - All demographic DIF variables (gender, ethnicity, 1st language, SEN, IDACI, CLA)
 - **And** mark obtained on the item
 - Controlled for age, school type, absence

Gender

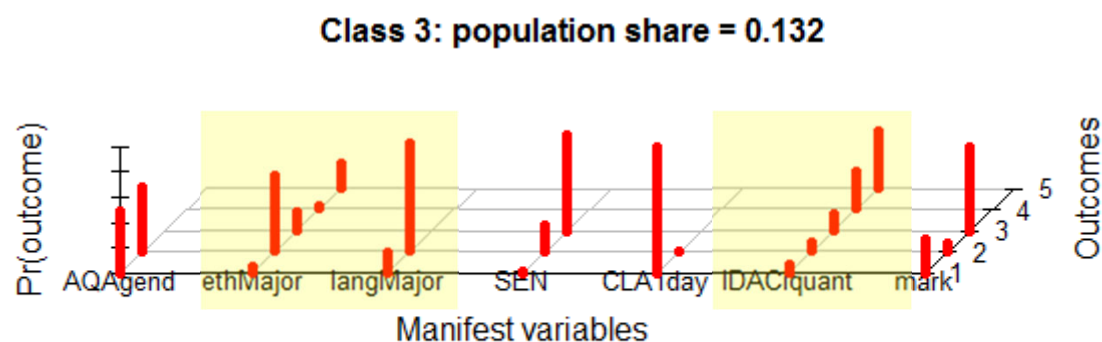
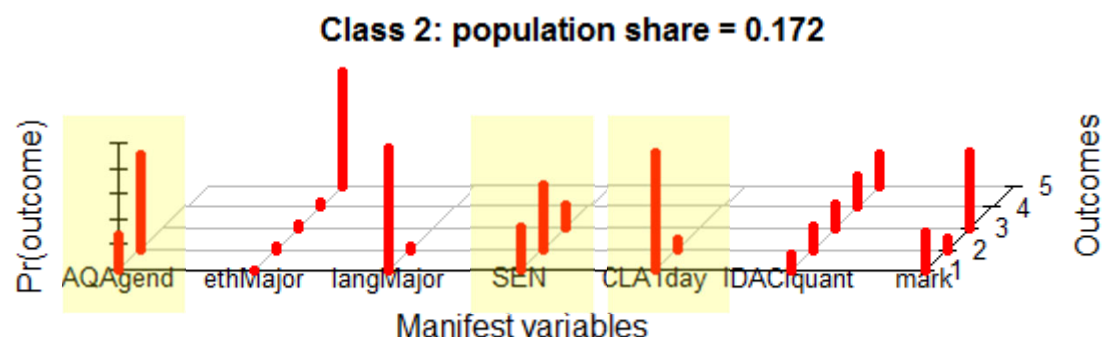
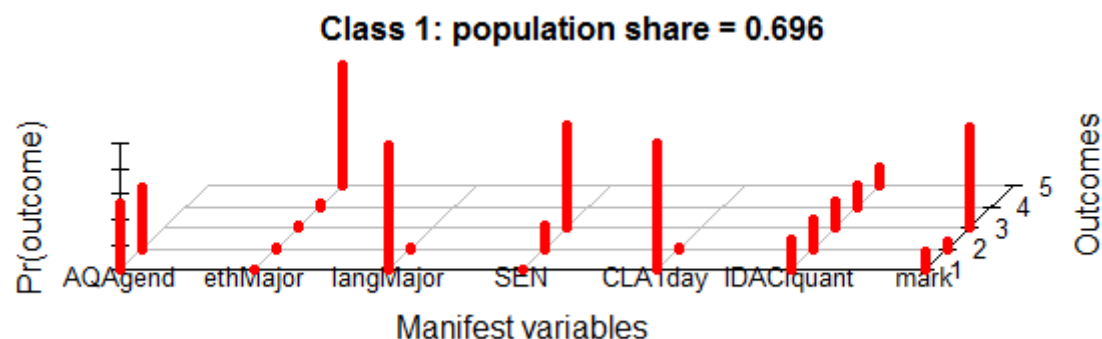
1 = Female
2 = Male

Ethnicity

1 = Other
2 = Asian
3 = Black
4 = Mixed
5 = White

1st language

1 = English
2 = Other



SEN

1 = EHCP
2 = SEN support
3 = None

CLA

1 = Not in care
2 = In care

IDACI

1 = Least deprived
...
5 = Most deprived

Mark

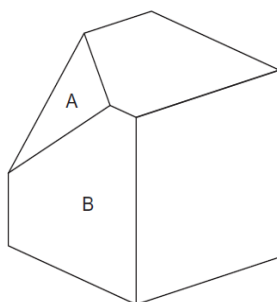
1 = 0 marks
2 = 1 mark
Etc...

‘Typical’ findings

- 3 classes:
 - Second 15-20%, worst performing. Characterised by more likely to be SEN/CLA, and more deprived.
 - Third 10-15%, middling performance. Characterised by primarily (~90%) non-white and non-English speakers, and even higher levels of deprivation.
- 4th class (if present) splits baseline group by gender
- If only two classes: baseline and non-Eng class
- Maths: 2nd class more likely to be male in Foundation, female in Higher
- English: harder to establish baseline – fewer items

Maths - Foundation

- 4 A cube has one corner cut off as shown.



- 4 (a) What shape is face A?
Circle your answer.

[1 mark]

hexagon octagon pentagon triangle

- 4 (b) What shape is face B?
Circle your answer.

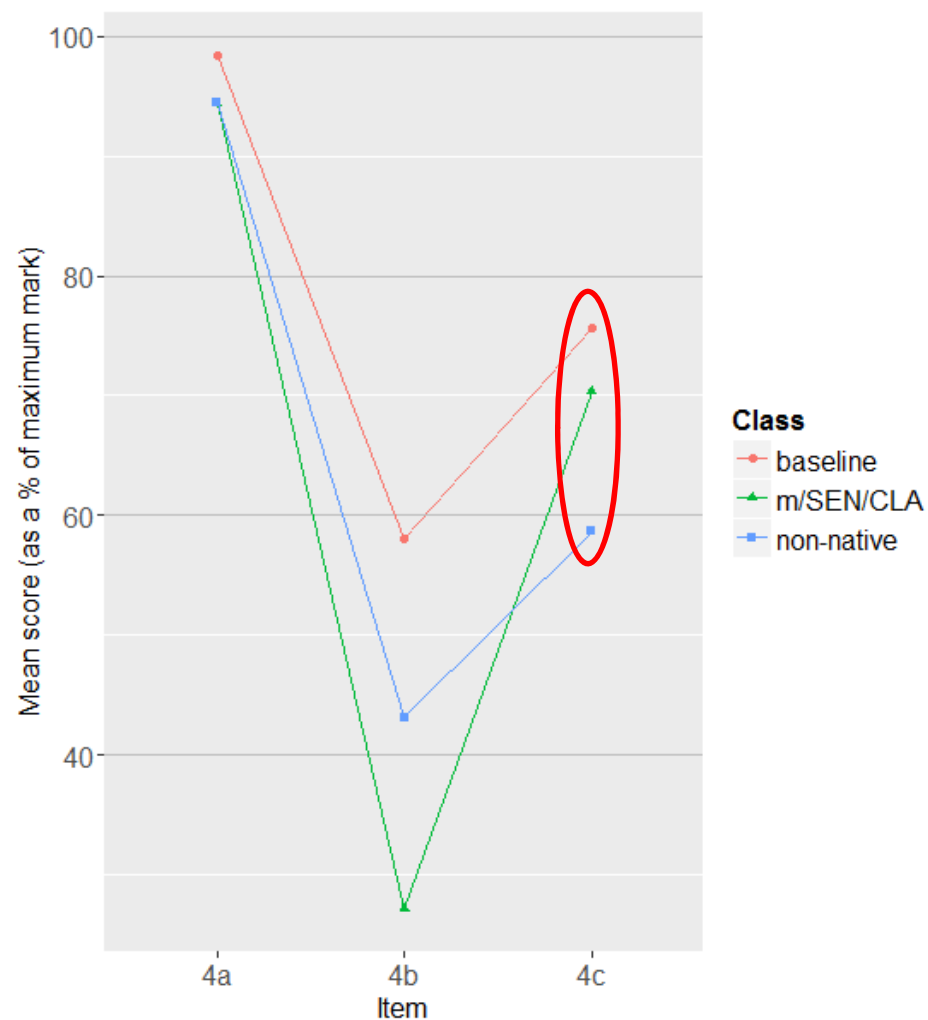
[1 mark]

hexagon pentagon rhombus trapezium

- 4 (c) Face C is opposite face B.

What shape is face C?

[1 mark]



Maths - Foundation

- 15 (a) Which **one** of these is **not** a factor of $24xy$.
Circle your answer.

[1 mark]

3

$4y$

$10xy$

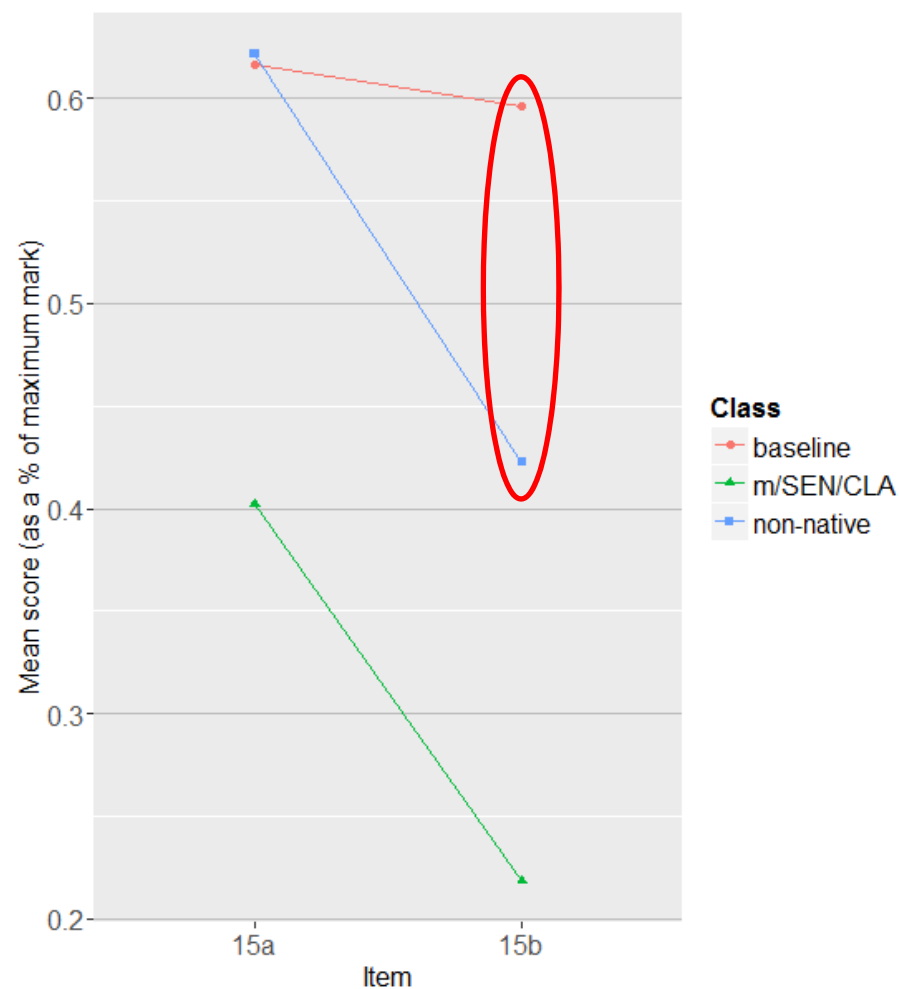
$12x$

- 15 (b) Work out the highest common factor (HCF) of 18 and 30

[2 marks]

.....
.....

Answer



Maths - Higher

1 (a) Multiply out $5(3x + 7)$ [1 mark]

.....
.....

Answer

1 (b) Make w the subject of the formula $z = w + 3$ [1 mark]

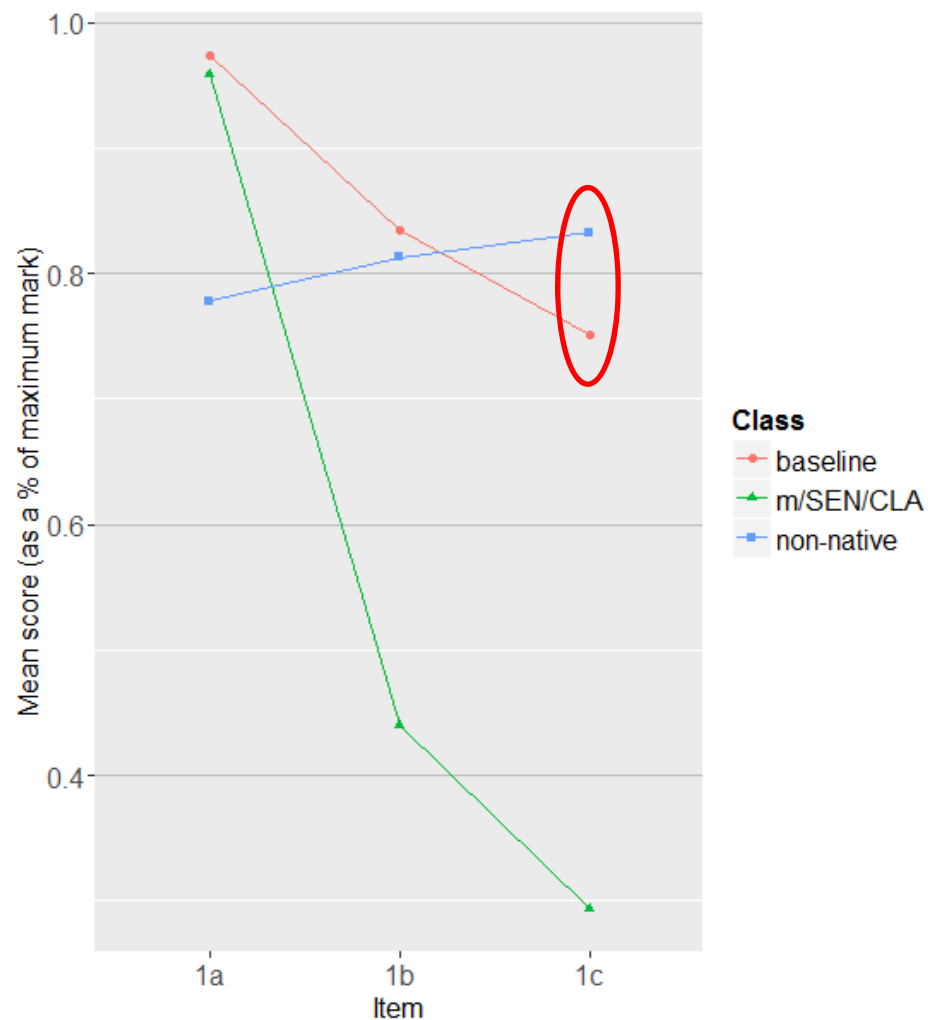
.....
.....

Answer

1 (c) Factorise fully $4y^2 + 6y$ [2 marks]

.....
.....

Answer



Maths - Higher

20 (a) Simplify fully $\frac{m^3 \times m^5 \times m}{m^2 \times m^4}$ [1 mark]

.....

.....

Answer

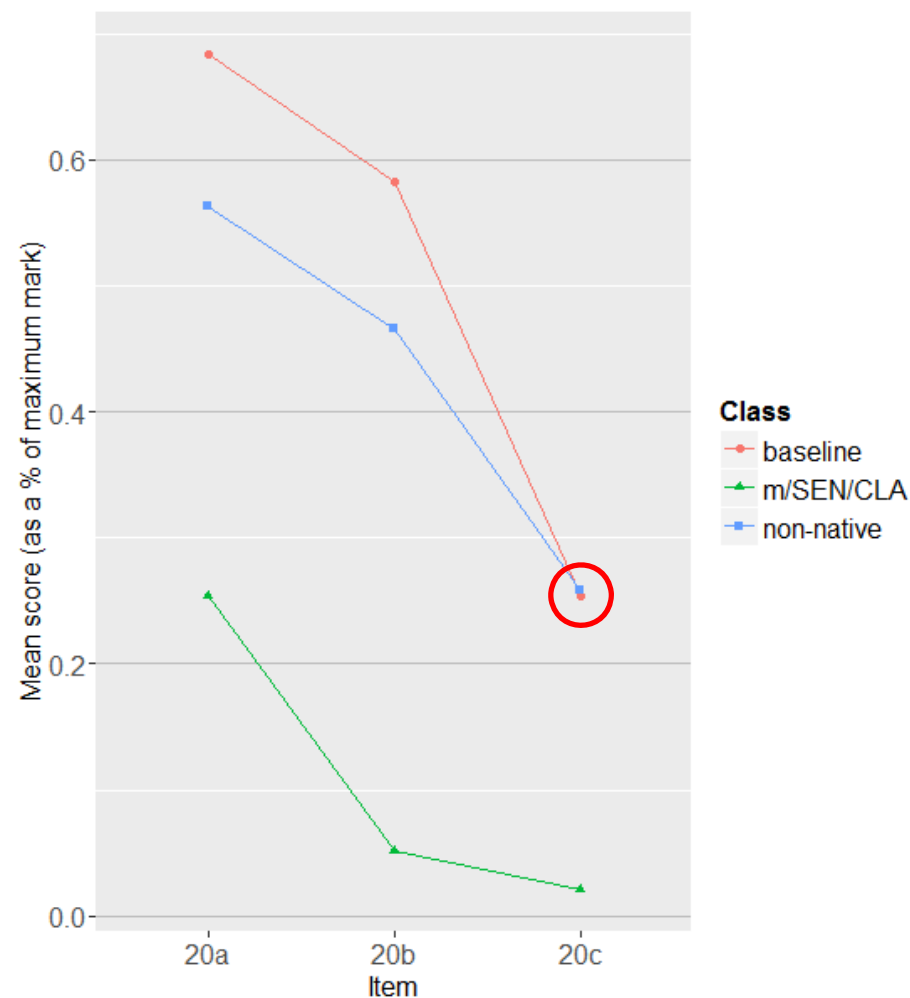
20 (b) Expand and simplify $(3 + \sqrt{2})(5 - \sqrt{2})$ [2 marks]

.....

.....

Answer

20 (c) Work out the value of $25^{-\frac{1}{2}} \times 81^{\frac{3}{4}}$ [3 marks]



Maths - Higher

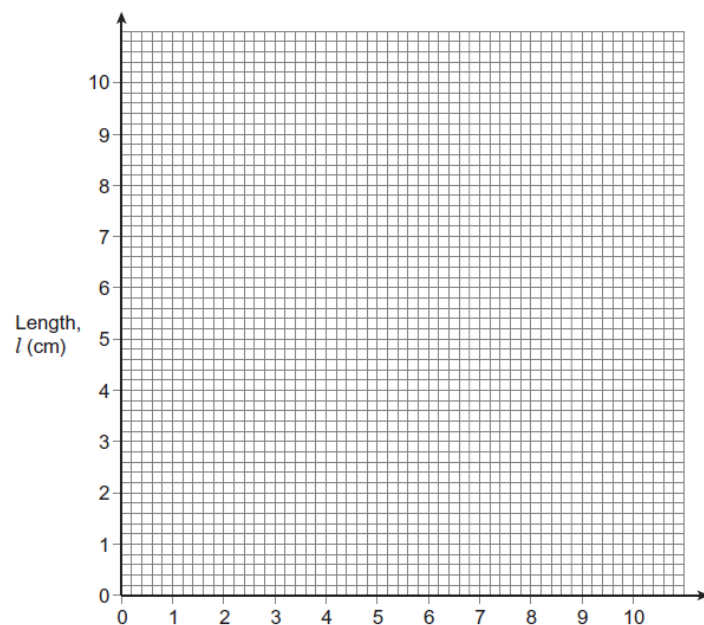
12 (a) w and l are such that

$$1 \leq w \leq 9$$

$$w + l = 10$$

Show this information on the graph.

[2 marks]



12 (b) Use the graph, or otherwise, to work out the value of w when $l = 3w$
You must show your working.

[2 marks]

.....

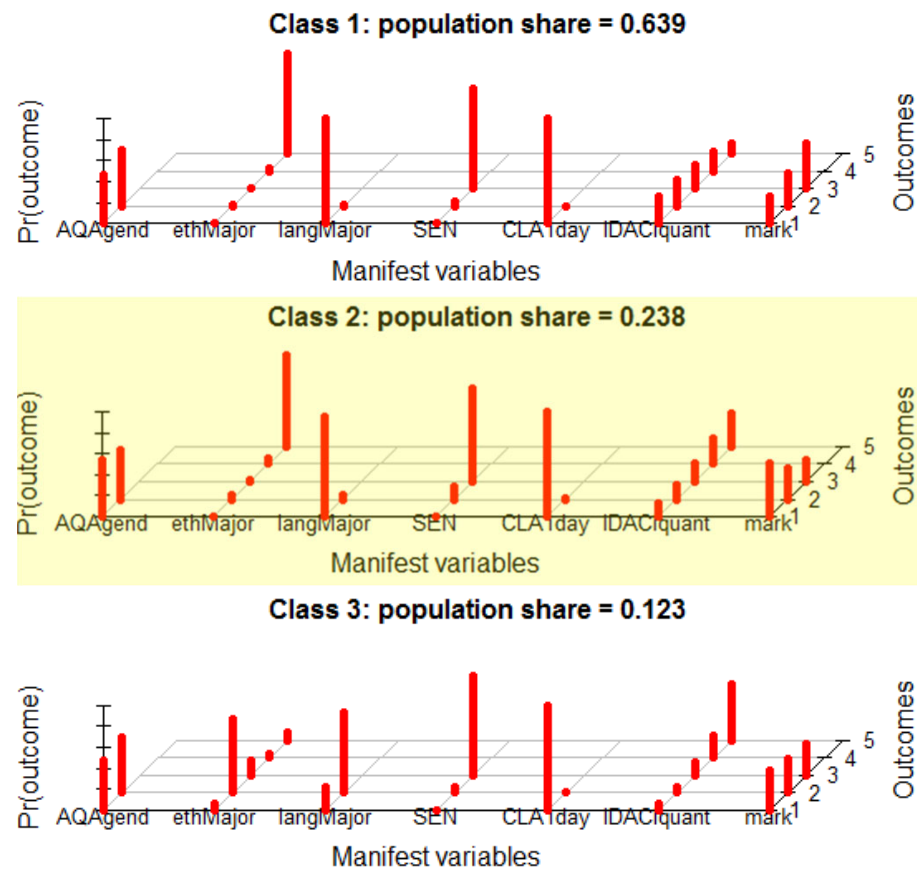
.....

.....

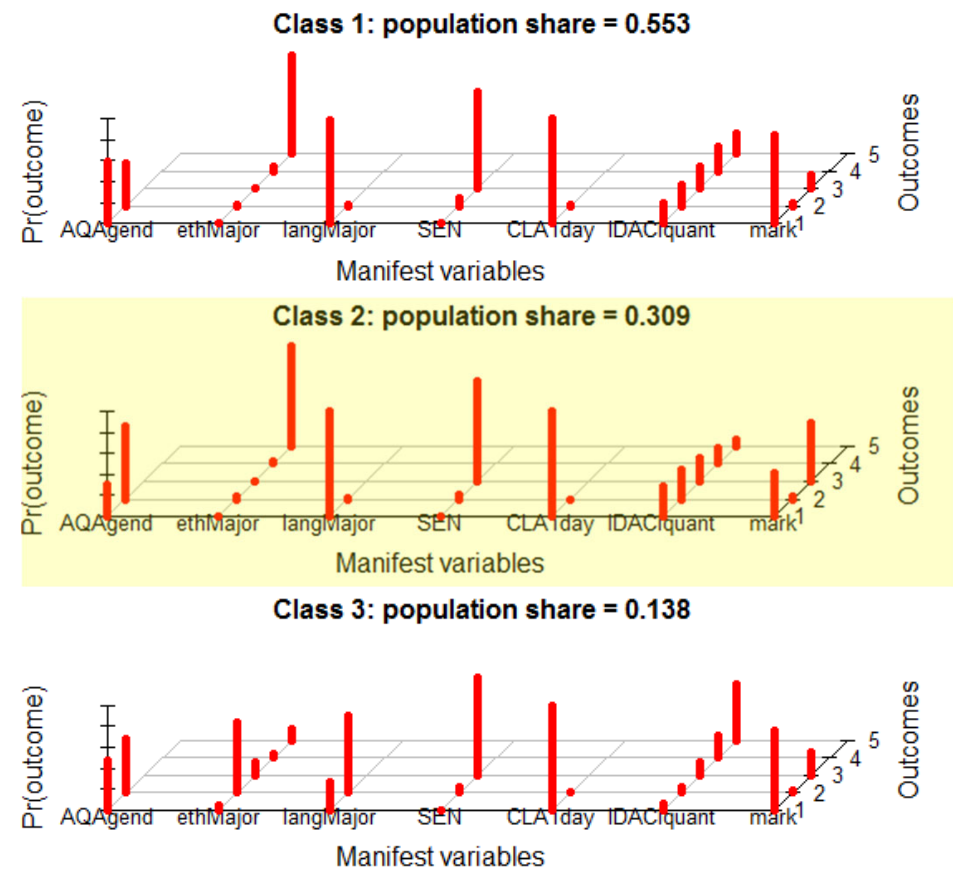
.....

Answer

12a



12b



Maths - Higher

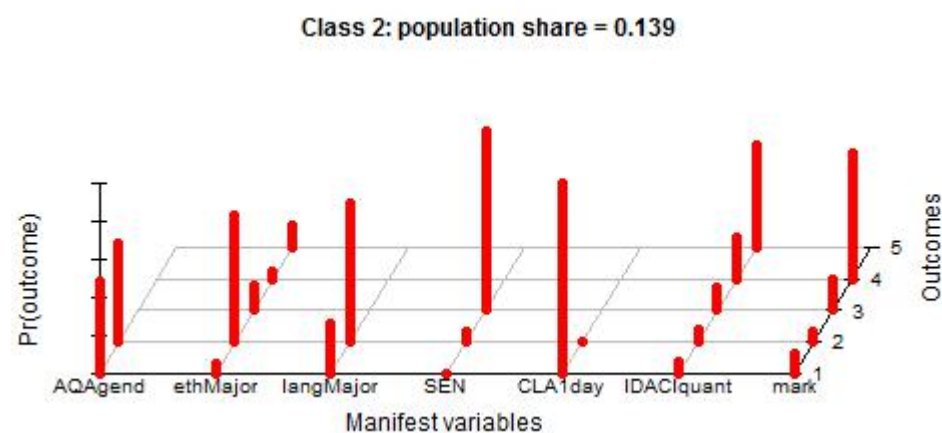
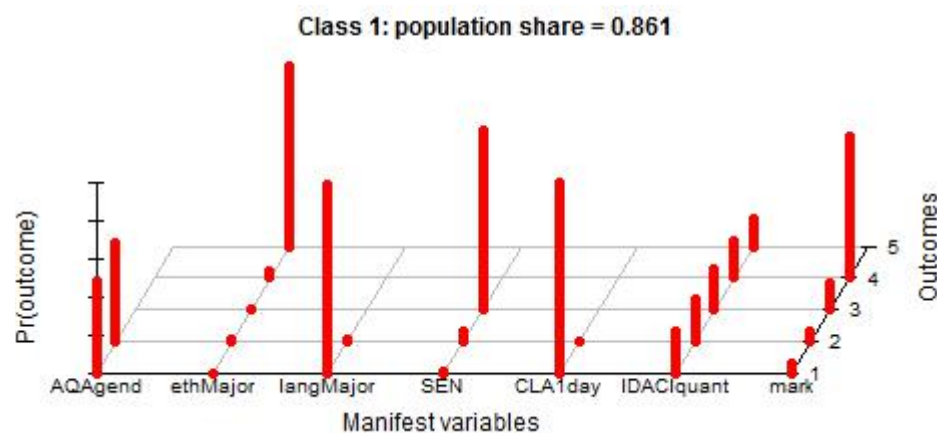
7 The first buses to X and Y leave a bus station at 7 am

Buses to X leave every 25 minutes.

Buses to Y leave every 20 minutes.

When will the buses to X and Y next leave at the same time?

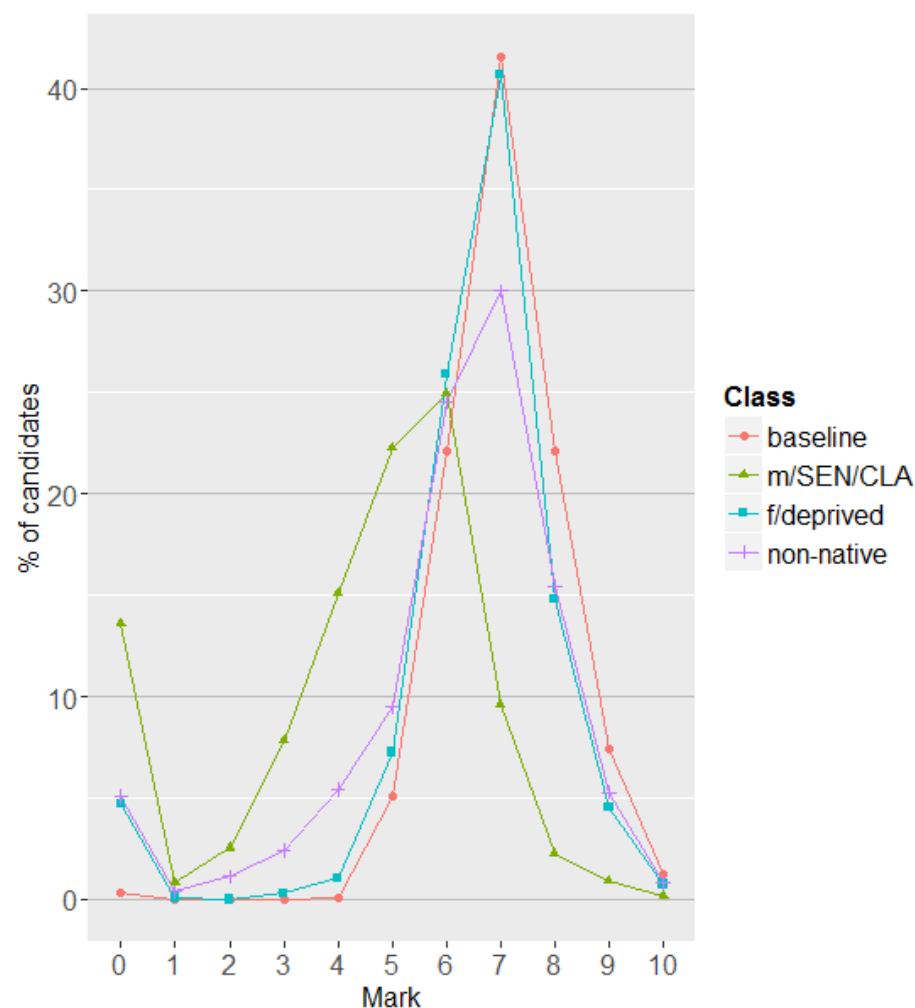
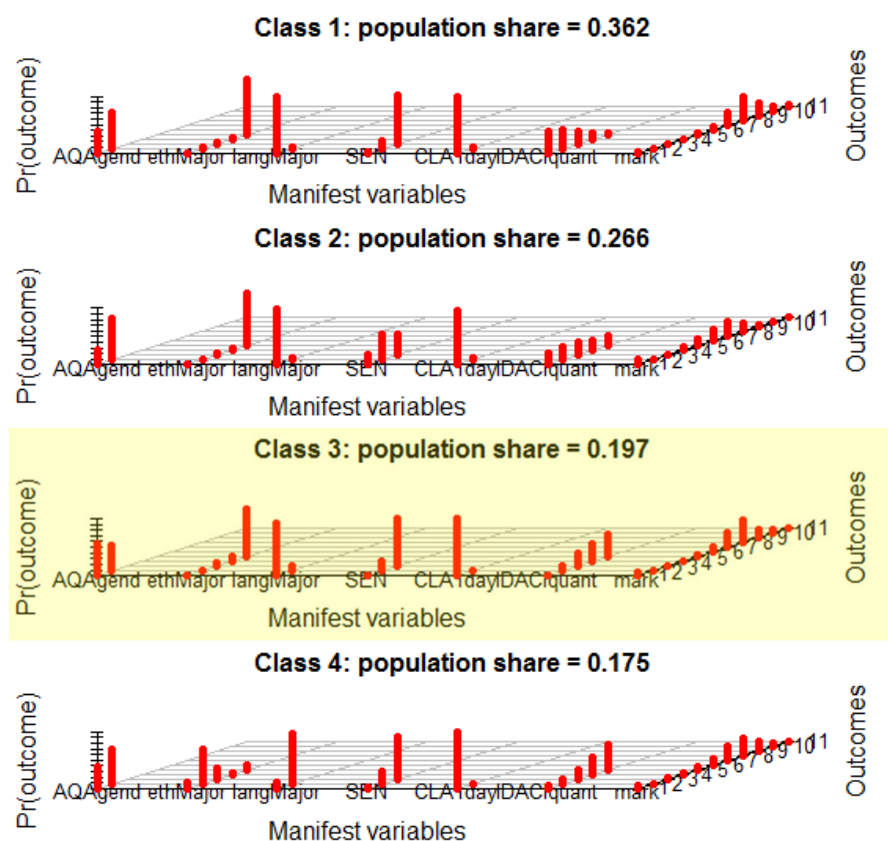
[3 marks]



DIF in Maths items

- Primarily non-white/Eng 1st lang class makes notable gains on baseline in:
 - MCQ items
 - Algebra Qs
- Some items show classes with more boys/girls performing well/poorly, but hard to discern a pattern based on item features. Potentially girls do less well in questions with:
 - A spatial element
 - A complex context

English - Foundation



- 5 Choose a good thing that you have done in your life and write an entry for your blog, explaining why it made a difference to others.

Remember to:

- write an entry for your blog
- use language to explain.

Try to write approximately one page.

[16 marks]

Foundation

- 5 Choose a good thing that you have done in your life and write an entry for your **blog**, **explaining** why it made a difference to others.

Remember to:

- write an entry for your **blog**
- use language to **explain**.

Try to write approximately one page.

[16 marks]

Higher

- 5 Write an entry for your blog which **describes** a place that you have visited which has **affected you** or **stayed in your memory**, and explain why this is so.

[16 marks]

Foundation

Now read **Source 3**, 'Manchester to Blackpool Night Ride', a leaflet from the British Heart Foundation.

3 How does the writer use **language features** for effect in the leaflet?

Remember to:

- give some examples of language features
- explain the effects.

[12 marks]

Now look again at **Source 1** and **Source 3**.

4 Compare the ways they use **presentational features** for effect.

Remember to:

- write about the ways the sources are **presented**
- explain the **effect** of the presentational features
- compare the way they **look**.

[12 marks]

Higher

Read **Source 1**, '**UN unveils plans to eliminate child labour by 2020**' by Randeep Ramesh.

1 What do you understand from the article about the issues of child labour?

[8 marks]

Now read **Source 2**, '**Admit it. You love cheap clothes. And you don't care about child slave labour**' by Gethin Chamberlain.

2 Explain how the headline and picture are effective and how they link to the text.

[8 marks]

DIF in English items

- DIF very complex in English, can emerge based on:
 - Text type (and perhaps pictures)
 - Emotive content
 - Question type/command word
 - QoWC marks (quality of writing)
- Appears to be interactions between tiers:
 - Foundation shows “typical” pattern of 3 classes
 - Gender DIF much more striking in Higher tier
 - Candidates without English as their first language are potentially better at different item types on each tier
 - May be an artefact of tiering

“Is DIF a bad thing?”

- It depends! For example...
 - Non-native speaker DIF inevitable
 - As identified in Maths papers, MCQ and algebra questions reduce this DIF
 - If an assessment was 100% MCQ, non-native speaker DIF could be eliminated (or minimised)
 - *But* this compromises the validity of the assessment!
 - (At least in terms of how we define validity in the UK...)

The first step to thinking about whether (and how) we address DIF is knowing whether (and where) it exists.

Next steps?

- Drill down into classes of interest
- Examine more papers (other subjects, years)
- Investigate data further:
 - Which distractor was chosen?
 - Which text type/command was chosen?
 - Which mark was missed?
- Use item typology to delineate trends in DIF
- Publication?

cerp.org.uk

Copyright © AQA and its licensors. All rights reserved.

AQA Education (AQA) is a registered charity (registered charity number 1073334) and a company limited by guarantee registered in England and Wales (company number 3644723). Registered address: AQA, Devas Street, Manchester M15 6EX.

