Conceptual Understanding and How to Assess It

Ian Jones
Mathematics Education Centre
Loughborough University

University of Auckland, June 2018

Loughborough University

Mathematics Education Centre

Plan

- How to make a maths exam.
- Comparative judgement.
- Case study 1: fractions.
- Case study 2: calculus.
- Case study 3: general achievement.
- Common objections.
- Wider research basis.

How to make a maths exam.

number

algebra

geometry

statistics

number

problem solving

algebra

geometry

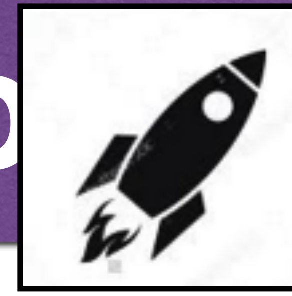
statistics

functional maths

number

$2 + 3$

problem solving



algebra

$a > c + 1$

geometry

$\sin \theta$

statistics

$P(A) = \frac{1}{2}$

functional maths

30% discount

nu 2 + 3 er

pro  m solving

algebra

Procedural knowledge

"execute actions to solve problems, not generalisable"

(Rittle-Johnson, Siegler & Alibali, 2001)

ge

stati P(A) = ½

func 30% discount aths

number

problem solving

ge

Conceptual understanding

“fundamental principles,
network of relationships”

(Hiebert & Lefevre, 1986; Rittle-Johnson et al., 2001)

statistics

functional maths

Why is conceptual understanding hard to assess?

What is an equation? Give examples of how equations can be useful.

What is an equation? Give examples of how equations can be useful.

£10 + £5.25 = £15.25p

Small ~~equasions~~ equasions like the one above arn't hard enugh to put into a collum So thats how we set it out

Equasions ~~are~~ are sums based on +, −, X, ÷ and ( ) we set it out like that ~~to~~ ~~je~~ its to easy

to put into a collum or you simply can't put it in a collum.

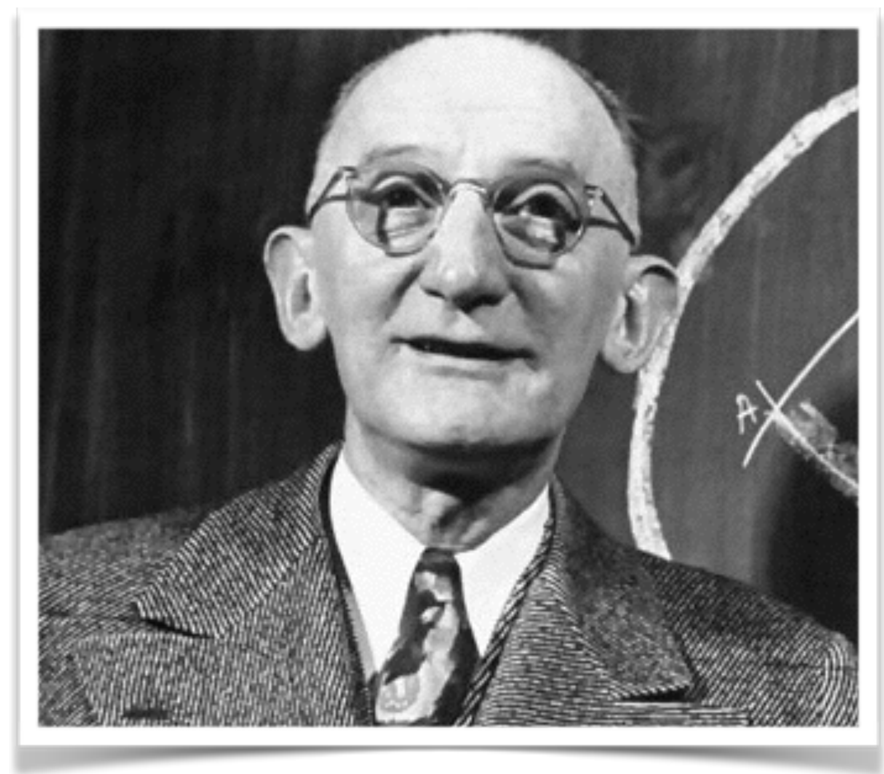
Comparative judgement.

An alternative approach

- Abandon attempts to specify conceptual understanding in rubrics.
- Abandon attempts to anticipate every possible student response in advance.
- Trust ourselves to know conceptual understanding when we see it.
- Get some help from psychophysics.

Law of Comparative Judgement

"Judgement is inherently comparative."



L.L. Thurstone
(1887 - 1955)

Psychophysics

How much does this shoe cost?

Psychophysics

Which shoe costs more?

Psychophysics

Giuseppe Zanotti Suede
Peep Toe Bootie
$399.95

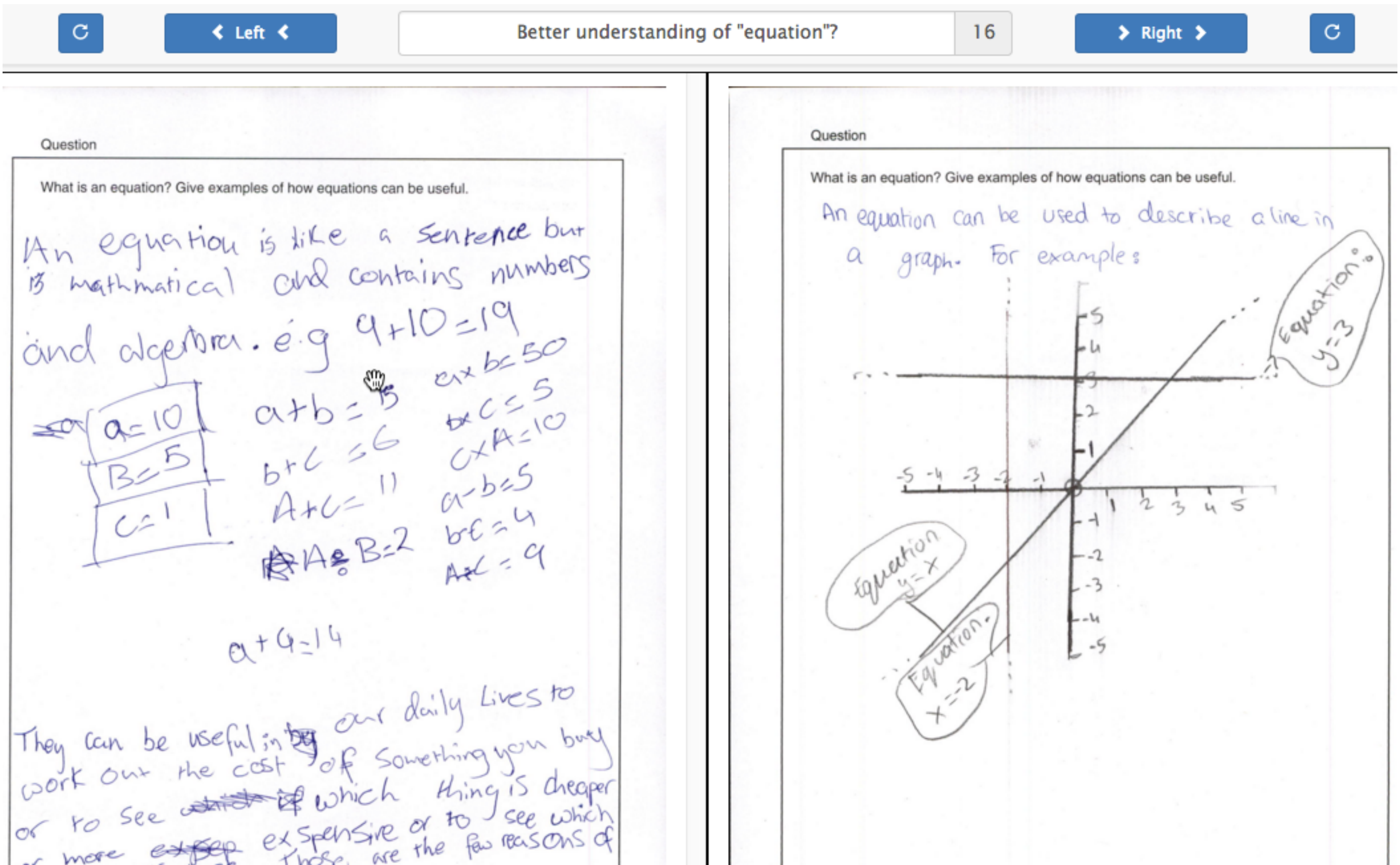
Qupid Carol-13X
Striped Flat
$9.94

www.dsw.com

Comparative Judgement

- Humans good at comparing two objects, poor at judging one object in isolation.
- Comparison eliminates bias.
- Promise for assessing difficult-to-specify but important learning outcomes such as "conceptual understanding".

Application to education



Outcomes

- Pairwise decisions statistically modelled.
- Unique score for each student.
- Reliability, misfit, validity measures.

Case study 1: fractions.

Jones, I., Inglis, M., Gilmore, C., & Hodgen, J. (2013). Measuring conceptual understanding: The case of fractions. In A. M. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education.* (Vol. 3, pp. 113-120). Kiel, Germany.

Test question

Write down these fractions in order of size from smallest to largest.
Underneath, describe and explain your method for doing this.

$$\frac{3}{4} \qquad \frac{3}{8} \qquad \frac{2}{5} \qquad \frac{8}{10} \qquad \frac{1}{4} \qquad \frac{1}{25} \qquad \frac{1}{8}$$

Which student displays the better conceptual understanding of fractions?

$\frac{1}{25}$, $\frac{1}{8}$, $\frac{1}{4}$, $\frac{3}{8}$, $\frac{2}{5}$, $\frac{3}{4}$, $\frac{8}{10}$

To know which

find a common denomonator and times the bottom + top by the same. Then put them in order. Another way is its the top ÷ bottom and the smallest decimal comes first.

$\frac{1}{25}$, $\frac{1}{8}$, $\frac{3}{8}$, $\frac{1}{4}$, $\frac{2}{5}$, $\frac{3}{4}$, $\frac{8}{10}$

I did this by imagining a circle cut up into as many pieces as the bottom number then I imagined the ~~amoun~~ Same amount as the top number ~~colores~~ coloured in.

E.g $\frac{7}{8}$

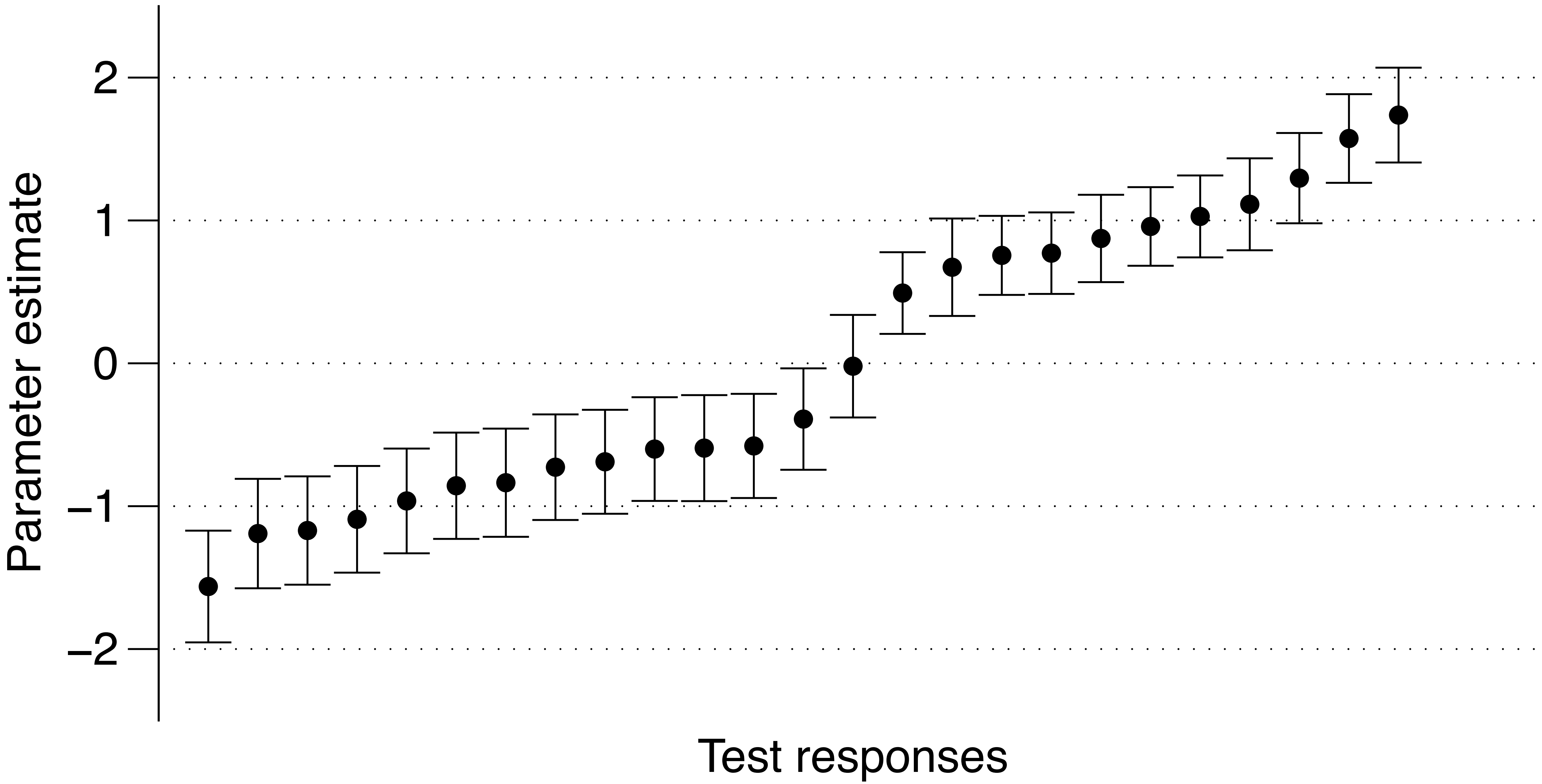
The 8th that is left white is the bottom fraction and the shaded ones are the top fraction.

This is what I imagine in my head

Method

- *Materials.* 25 responses to test question, pupils aged 12/13 ($N = 10$) or 14/15.
- *Judges.* 8 maths education experts with previous experience using CJ to assess mathematics.
- *Procedure.* Each judge completed 50 live pairwise judgements online.

Outcome



Outcome was internally consistent

- Scale Separation Reliability = .88 (internal consistency, c.f. Cronbach's α).
- Judge 'misfit' figures all within two standard deviations of the mean.
- Script 'misfit' figures within two standard deviations of the mean, bar one marginal misfit figure.

Outcome was reliable

- Split-halves technique to estimate reliability.
- Judges randomly split into two groups, two new rank orders produced and correlated (Pearson's r on scripts' parameter estimates).
- Repeated 36 times for every possible combination of judges in two groups.
- range $r = .79$ to $.95$, mean $= .87$.

Predictors

- **General mathematical achievement**
  based on teachers':
  - predicted grades (A* - F) for older children;
  - dichotomous (high/low) assessment for younger children.
- **Procedural score**
  based on children's fractions ordering accuracy, from 0 (fractions in order)
  to 7 (fractions completely out of order).

Outcome was valid

- Mathematical achievement was a significant predictor;
- Procedural score was not a significant predictor.

	Total variance explained	Mathematical achievement	Procedural score
Older children	**53%** $F(2,12) = 6.69$, $p = .011$*	**β = .40** $t(12) = 2.64$, $p = .022$*	**β = -.07** $t(12) = -.52$, $p = .613$
Younger children	**68%** $F(2, 7) = 7.33$, $p = .019$*	**β = 1.38** $t(7) = -3.83$, $p = .006$*	**β = -.23** $t(7) = -1.84$, $p = .108$

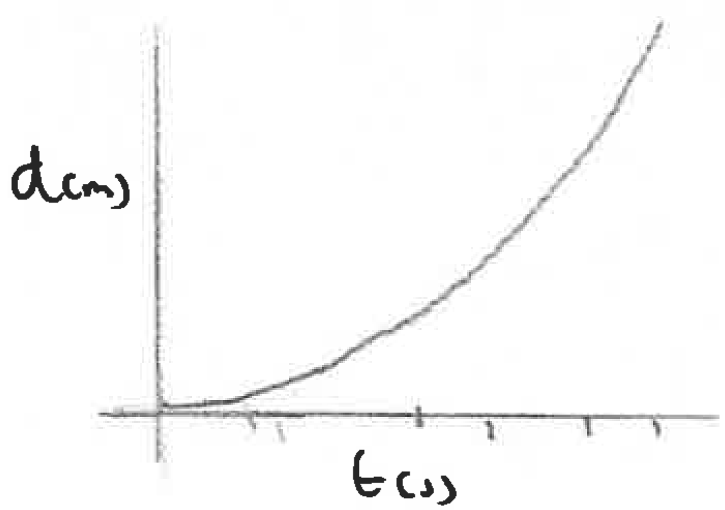
Case study 2: calculus

Bisson, M., Jones, I., Gilmore, C. & Inglis, M. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education, 2*, 141-164.

Test question

Explain what a **derivative** is to someone who hasn't encountered it before. Use diagrams, examples and writing to include everything you know about derivatives. **Write only in the box below.**

A derivative in algebraic ~~physical~~ terms is the change in variable x, over change in variable y. For example if we think of a car accelerating and plot a graph for ~~at~~ the distance ~~each second~~ covered per unit ~~in each~~ time.



we can see that the car is accelerating as the gradient is increasing per unit time this therefore means there is a larger increase in y per change in x, which is a derivative. So by finding the derivative of the equation we can find another variable for this case the cars velocity at time t.

$x = s$
$y = m$

$\frac{\Delta y}{\Delta x} = \frac{dy}{dx} = \frac{m}{s} = ms^{-1} = v$

we can differentiate again to find another variable which is relevant to the problem for this case the case acceleration.

Method

- *Materials.* 40 undergraduate responses to test question, from a Mathematical Methods in Chemical Engineering module.
- *Judges.* 30 mathematics PhD students.
- *Procedure.* Each judge completed 42 live pairwise judgements online.

Reliability and validity

- Internal consistency, *SSR* = .94.
- Inter-rater (split-halves, 20 iterations), *r* = .87.

Reliability and validity

- Internal consistency, *SSR* = .94.
- Inter-rater (split-halves, 20 iterations), *r* = .87.
- Validity: Calculus Concept Inventory, α = .40.

	CCI	A-level (*N*=33)	Module
CJ	**0.09**	**0.44***	**.37***
CCI		**0.09**	**0.28**

Case study 3:
general achievement

25 test questions

Write one or more questions for a friend that involves using a chart or diagram to represent data.

Give examples of large and small angles. Explain what you know about different sizes of angles.

What do the symbols =, < and > mean? Write down and explain as many examples as you can.

How are fractions and decimals similar? Give as many examples as you can to show how they are similar.

Below is an example of a sequence. Explain what is meant by a sequence and give examples of other sequences.

7,11,15,19,…

Method

- *Materials.* 668 responses to test questions, pupils (N = 197) aged 11 to 14 from one school. 1 to 5 tests per pupil (mode = 4).
- *Judges.* 11 mathematics PhD students.
- *Procedure.* Each judge completed 600 judgements.

Reliability and validity

- Internal consistency, *SSR* = .87.
- Inter-rater (split-halves, 100 iterations), *r* = .75.

Reliability and validity

- Internal consistency, *SSR* = .87.
- Inter-rater (split-halves, 100 iterations), *r* = .75.
- Validity: KS2 Maths & English (*N* = 148).

Total variance explained	Mathematics achievement	English achievement
22% $F(2,145) = 20.04$, $p < .001$**	**β = 0.40** $t(145) = 4.94$, $p < .001$**	**β = 0.10** $t(145) = 1.02$, $p = .310$

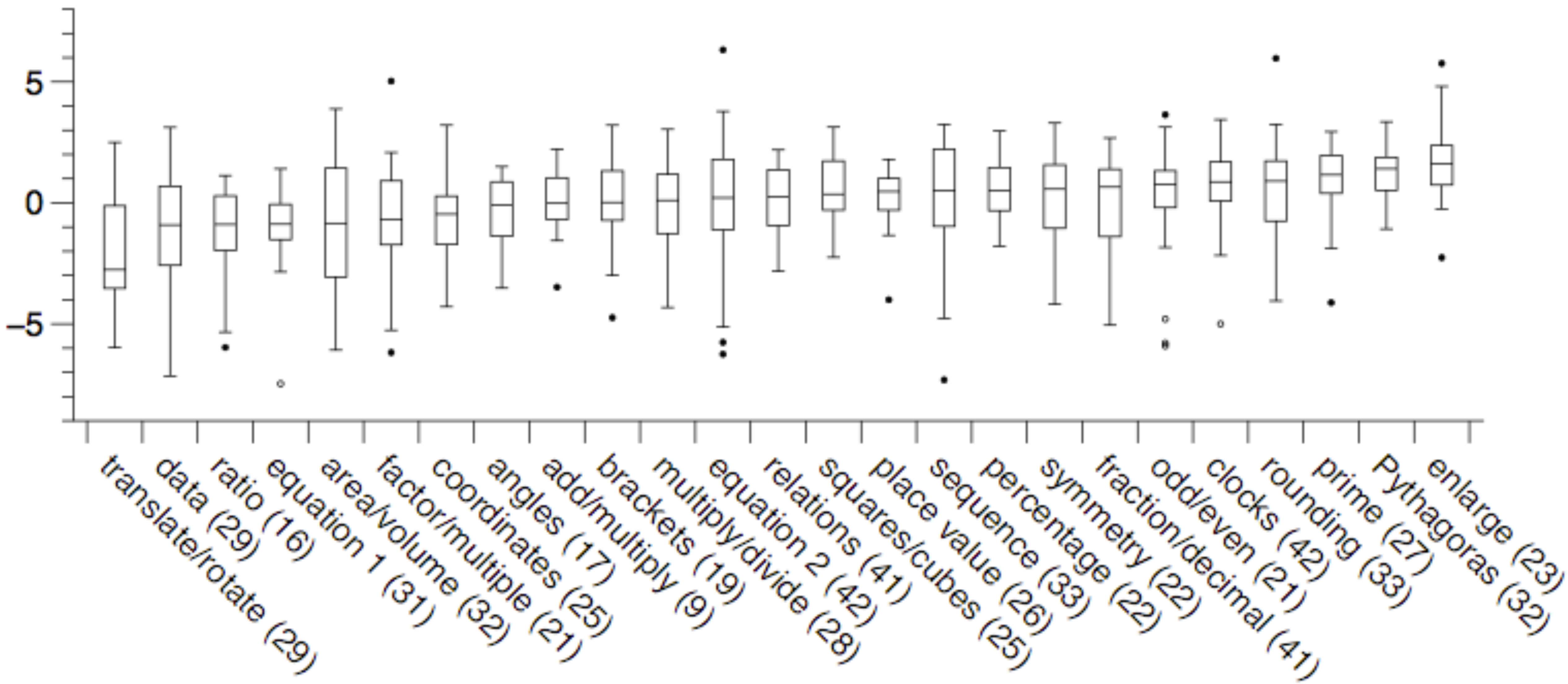
Reliability and validity

- Internal consistency, $SSR = .87$.
- Inter-rater (split-halves, 100 iterations), $r = .75$.
- Validity: KS2 Maths & English ($N = 148$).

Total variance explained	Mathematics achievement	English achievement
22% $F(2,145) = 20.04$, $p < .001$**	**$\beta = 0.40$** $t(145) = 4.94$, $p < .001$**	**$\beta = 0.10$** $t(145) = 1.02$, $p = .310$

- Younger pupils ($N = 61$), 34% variance explained.
- Older pupils ($N = 87$), 13% variance explained.

Question performance



$F(24, 661) = 5.07$, $p < .001$, $\eta^2 = 0.155$

Research basis.

Funding.

- The Royal Society, £208,003.
- Nuffield Foundation, £155,098.
- Education Endowment Fund, £20,000.
- No More Marking Ltd., £12,777.
- Assessment and Qualifications Alliance, £11,409.
- Higher Education Academy, £1430.

Peer-reviewed research.

- Conceptual understanding/problem solving.

Bisson, M., Jones, I., Gilmore, C. & Inglis, M. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education, 2,* 141-164.

Jones, I. & Inglis, M. (2015). The problem of assessing problem solving: Can comparative judgement help? *Educational Studies in Mathematics, 89*, 337-355.

Jones, I., Swan, M. & Pollitt, A. (2014). Assessing mathematical problem solving using comparative judgement. *International Journal of Science and Mathematics Education, 13*, 151-177.

- Peer assessment.

Jones, I. & Sirl, D. (2017). Peer assessment of mathematical understanding using comparative judgement. *NOMAD: Nordic Studies in Mathematics Education, 22*, 147-164.

Jones, I. & Wheadon, C. (2015). Peer assessment using comparative and absolute judgement. *Studies in Educational Evaluation, 47*, 93-101.

Jones, I. & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education, 39*, 1774-1787.

McMahon, S. & Jones, I. (2014). A comparative judgement approach to teacher assessment. *Assessment in Education: Principles, Policy & Practice, 22*, 368-389.

Peer-reviewed research.

- Standards over time.

  Jones, I., Humphries, S., Wheadon, C. & Inglis, M. (2016). Fifty years of A-level Mathematics: Have standards changed? *British Educational Research Journal, 42*, 543–560.

- RCT measures.

  Jones, I., Bisson, M., Gilmore, C. & Inglis, M. (submitted). Measuring conceptual understanding in randomised control trial studies: can comparative judgement help? *British Educational Research Journal.*

  Bisson, M., Gilmore, C., Inglis, M. & Jones, I. (submitted). Measuring conceptual understanding: The case of teaching differential calculus using contextualised and decontextualised representations. *Research in Mathematics Education.*

Peer-reviewed research.

- Other research teams.

Tarricone, P., & Newhouse, C. P. (2016). Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability. *International Journal of Educational Technology in Higher Education,* 1-11.

van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V., & De Maeyer, S. (2016). Validity of comparative judgement to assess academic writing: examining implications of its holistic character and building on a shared consensus. *Assessment in Education: Principles, Policy & Practice*, 1-16.

Steedle, J. T., & Ferrara, S. (2016). Evaluating Comparative Judgment as an Approach to Essay Scoring. *Applied Measurement in Education, 29*, 211-223.

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice, 19*, 281-300.

Special issue (2012). *International Journal of Technology and Design Education, 2*.

Heldsinger, S., & Humphry, S. (2010). Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher, 37*, 1-19.

Bramley, T., & Gill, T. (2010). Evaluating the rank-ordering method for standard maintaining. *Research Papers in Education, 25*, 293-317.

Impact.

- National examination standards monitoring. (Ofqual, exam boards).
- International standards comparison: New Zealand, England, Northern Ireland. (CCEA).
- Writing moderation: *Sharing Standards*. (No More Marking Ltd.).
- Tertiary mathematics assessment. (Me!)

Common objections.

It’s norm-referenced.

- Objection: It only measures students relative to one another.
- No - multiple methods for grading against criteria using comparative judgement.
- These methods are superior to traditional marking.

It’s opaque.

- Objection: Marking is transparent and readily communicated.
- Marking: Validity resides in rubrics. ‘Transparency’ at cost of straight-jacketed assessments.
- Comparative judgement: Validity resides in the collective understanding of the relevant community of practice. ‘Opacity’ enables richer yet reliable assessments.

It’s not mathematics.

- Complaint: Open-ended responses require generic skills such as written communication, creativity and organising information.
- True! Comparative judgement assumes a broader view of mathematics.
- Validity analyses evidence that it is maths.
- Moreover these skills are inherent to the mathematics of professional mathematicians.

Alcock, L. (2012). *How to Study for a Mathematics Degree*. Oxford: Oxford University Press.
Houston, K. (2009). *How to Write Mathematics*. Available online at www.kevinhouston.net/pdf/htwm.pdf (accessed 1 April 2018).
Yoon, C. (2017). The writing mathematician. *For the Learning of Mathematics, 37*, 30–34.

Let’s have a go.

tinyurl.com/NZCERmaths

tinyurl.com/NZCERenglish