Assessing high-order learning in mathematics: a comparative judgement approach

Ian Jones
Loughborough University, UK

University of Warwick
March 2022

Loughborough University
Mathematics Education Centre

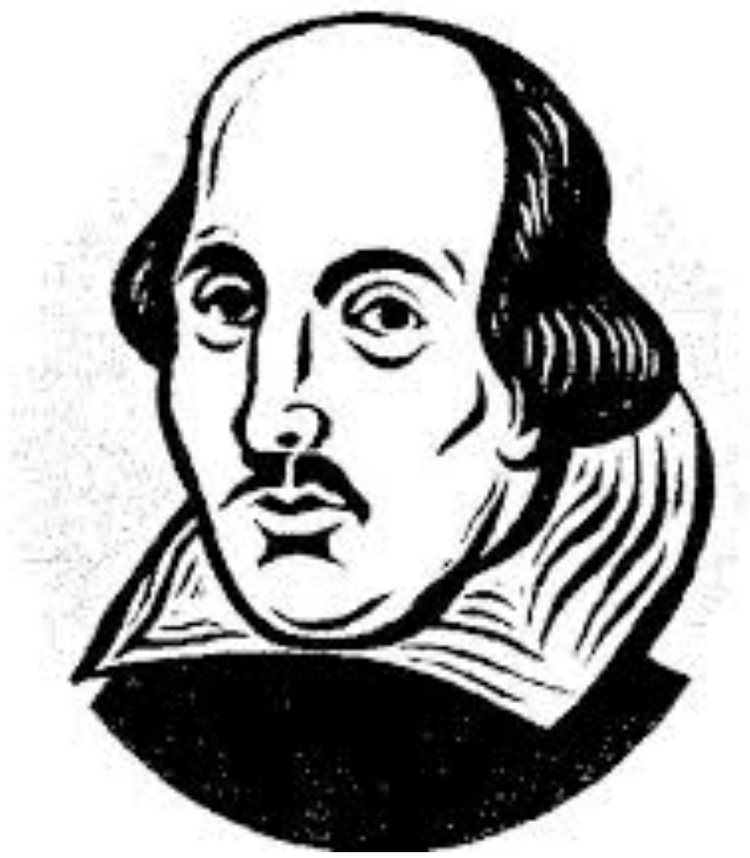
High order, but hard to assess, learning constructs

- Conceptual understanding
- Proof comprehension
- Problem solving skills
- Mathematical beauty

The essence of comparative judgement

- Place complex and varied objects on a unidimensional scale

Shall I compare thee to a summer's day?



Shall I compare thee to a summer's day?

Shall I compare thee to a summer's day?

Shall I compare thee to a summer's day?

Shall I compare thee to a summer's day?

Thou art more lovely and more temperate.

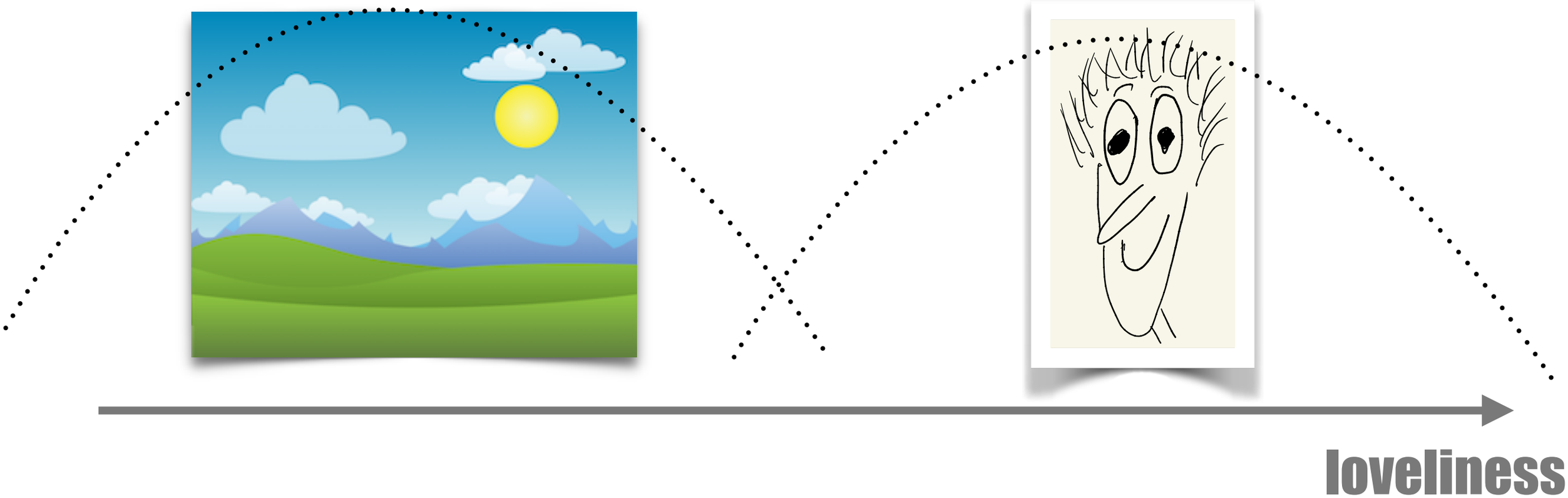
Shall I compare thee to a summer's day?

Thou art more lovely ~~and more temperate.~~

loveliness

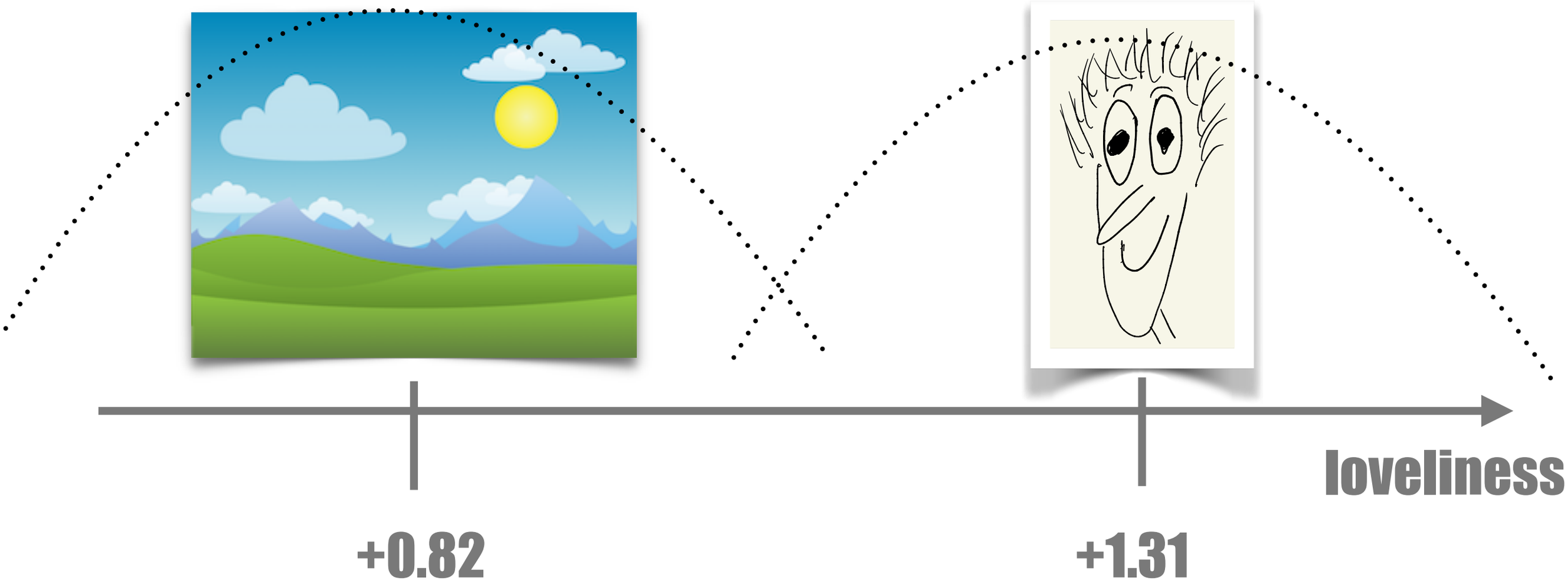
Shall I compare thee to a summer's day?

Thou art more lovely ~~and more temperate.~~



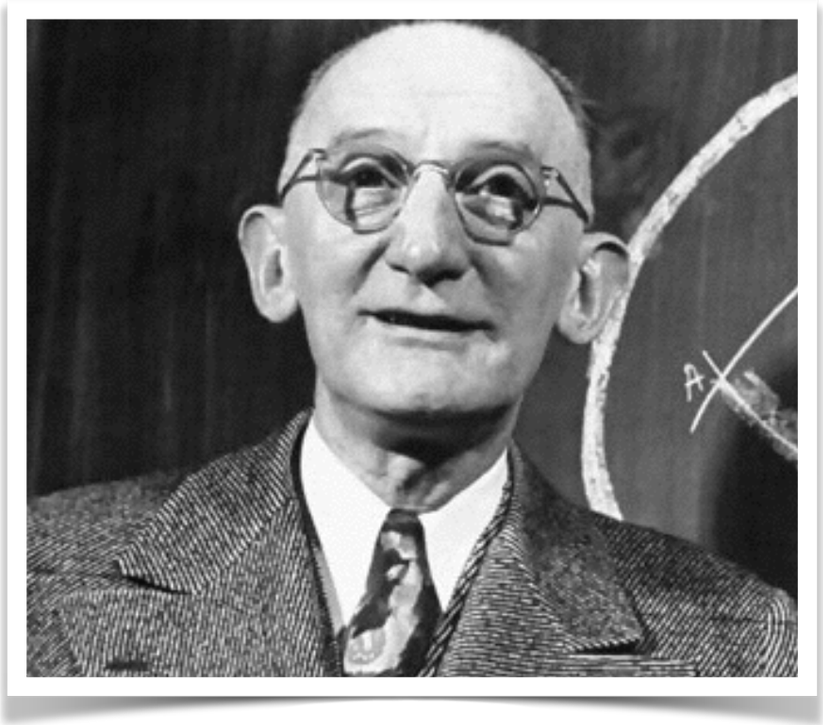
Shall I compare thee to a summer's day?

Thou art more lovely ~~and more temperate.~~



Law of Comparative Judgement

People are inconsistent when making *absolute* judgements, and consistent when making *relative* judgements



L.L. Thurstone
(1887 - 1955)

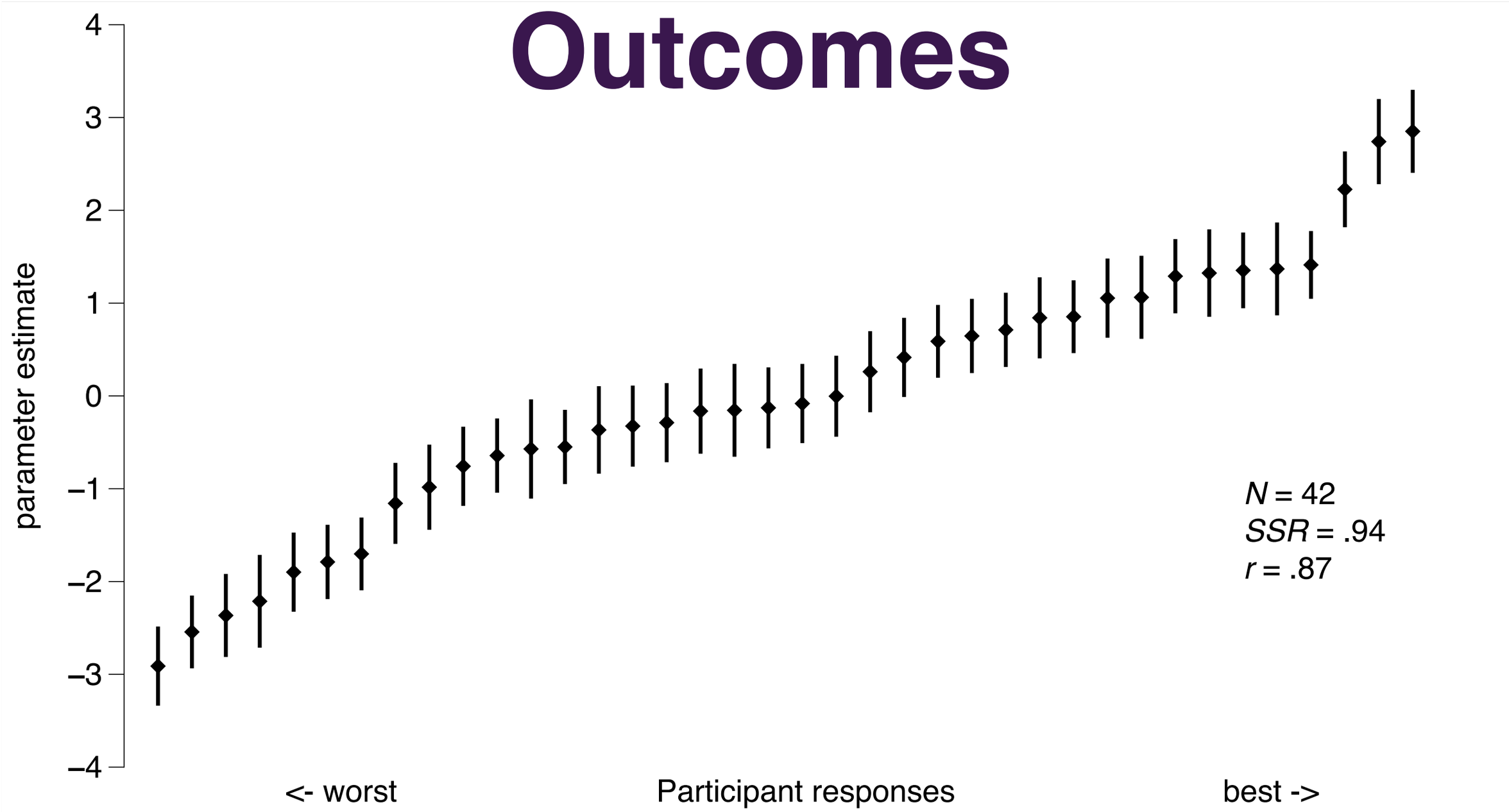
To assess high-order learning, we need

- Extended responses to open-ended prompt
- No rubrics, no coding scheme
- One high-order construct
- Judges (lecturers, peers)
- Direct pairwise comparisons of responses, via a comparative judgement engine
  - No More Marking Ltd.
  - Moodle Plug-in

Typical outcomes

- Pairwise decisions
  (> 10 × number of submissions)
- Fitted to the Bradley-Terry model
- Unique score for each participant
- Reliability measures

Outcomes



- *SSR* is Scale Separation Reliability (analogous to Cronbach's alpha)
- *r* is inter-rater reliability (split-halves technique, 20 iterations)

Bisson et al. (2016)

Use case

Foundation Mathematics
at
Loughborough University

Use case

Formative peer assessment 1

Formative peer assessment 2

Formative peer assessment 3

Summative peer assessment

Open-ended task

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

Submit and peer assess

Choose Left

File submissions

Equations

An equation is a mathematical expression that often contains some algebra. An algebraic expression contains an unknown variable such as X or Y that you need to solve to find the value. Foe example, in the expression X + 7 = 14 you solve for X. To solve this type of expression you need to balance it by doing the same thing to one side as you do to the other. In the above example you subtract 7 from the left and right side. This leaves x = 7.

These equations can become much more complex. For example, you may see an equation like 7x + 2y – 1 = 8. To solve this, we would still need to perform the same action on both sides. The final answer for this equation is x = 1 and y=1. This type of equation asks you to solve for 2 unknown variables.

You may also be asked to solve simultaneous equations. This is when two or more different equations contain the same variables that have the same unknown variable. For example, x - y = 10 and 2x – y = 18.

This can be solved through elimination or substitution. Below is an example of elimination.

x - y –(2x-y)= 10 – 18.

x - 2x = -8

x = 8

We then put the found value into the equation.

8-y=10

8 - -2 = 10

Y=-2

Equations can be useful as it can be used to solve some where not every value is known. Equations are also useful in the real world. Equations are applied frequently in scientific formula and even in other fields such as economics and finance.

Choose Right

File submissions

An equation is a mathematical formula used to find out an Unknown Answer.

for example "Algebra" → "x"

You can use an equation to help rearrange the unknown with letters and then adapt accordingly with BIDMAS

This can be useful in all aspects of life. → finding Angles → Building Buildings → Bridges

An Equation may look like this:

$$\frac{100}{5} = x$$

x would = 20.

Discussion in groups

Choose Left

File submissions

Equations

An equation is a mathematical expression that often contains some algebra. An algebraic expression contains an unknown variable such as X or Y that you need to solve to find the value. Foe example, in the expression X + 7 = 14 you solve for X. To solve this type of expression you need to balance it by doing the same thing to one side as you do to the other. In the above example you subtract 7 from the left and right side. This leaves x = 7.

These equations can become much more complex. For example, you may see an equation like 7x + 2y – 1 = 8. To solve this, we would still need to perform the same action on both sides. The final answer for this equation is x = 1 and y=1. This type of equation asks you to solve for 2 unknown variables.

You may also be asked to solve simultaneous equations. This is when two or more different equations contain the same variables that have the same unknown variable. For example, x - y = 10 and 2x – y = 18.

This can be solved through elimination or substitution. Below is an example of elimination.

[illegible] 18

[illegible] and value into the equation.

8-y=10

8 - -2 = 10

Y=-2

Equations can be useful as it can be used to solve some where not every value is known. Equations are also useful in the real world. [illegible] applied frequently in scientific formula and even in other fields such as econ[illegible]

Choose Right

File submissions

An equation is a mathematical formula used to find out an unknown answer.

for example "Algebra" —→ "x"

you can use an equation to help rearrang the unknown with letters and then adapt accordingly with BIDMAS

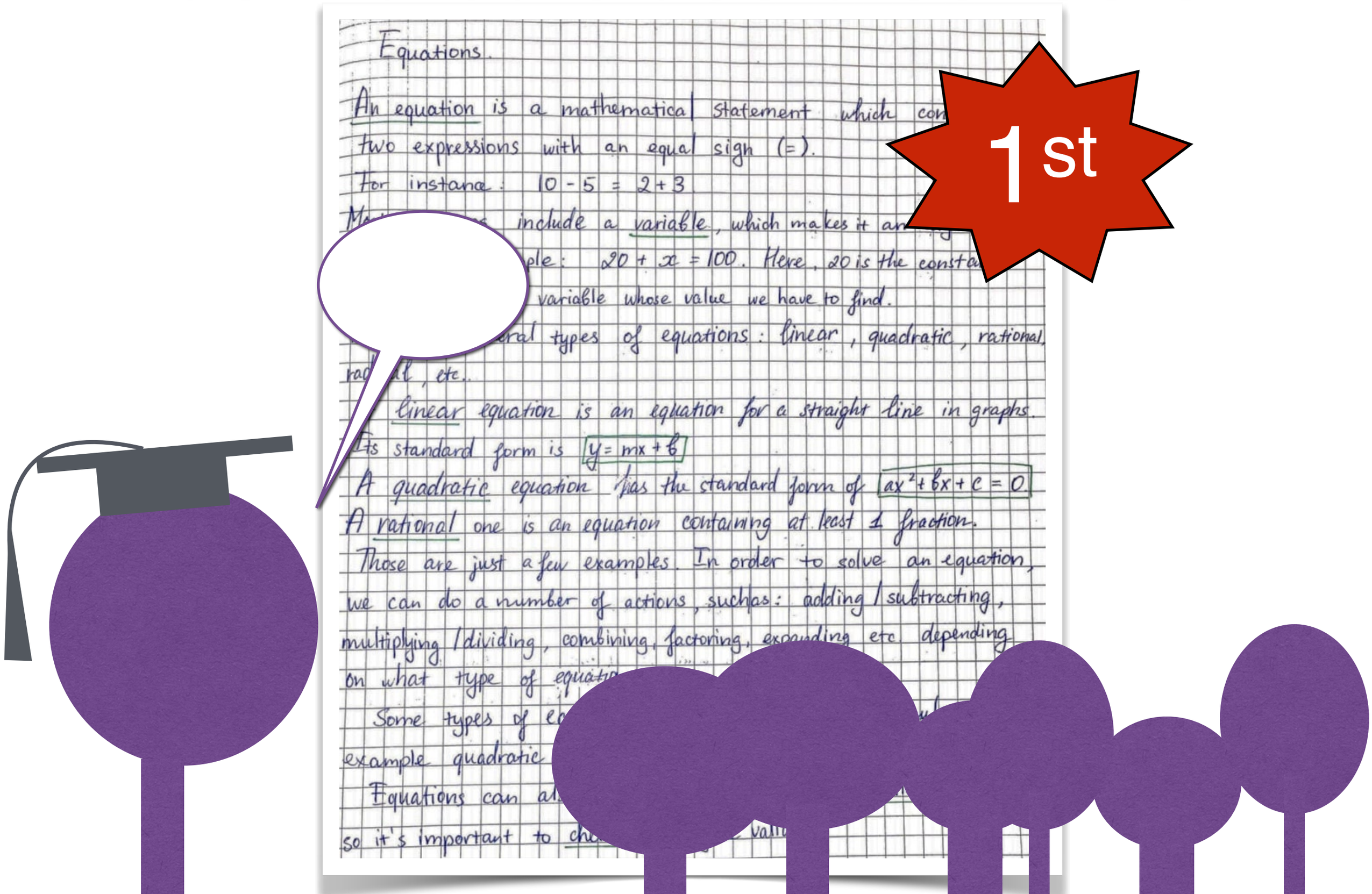
This can be useful in all aspects [illegible] life. —→ [illegible] angles [illegible] dings [illegible] edges

An Equation [illegible] like this:

$$\frac{100}{5} = x$$

[illegible]d = 20.

Facilitated class discussion

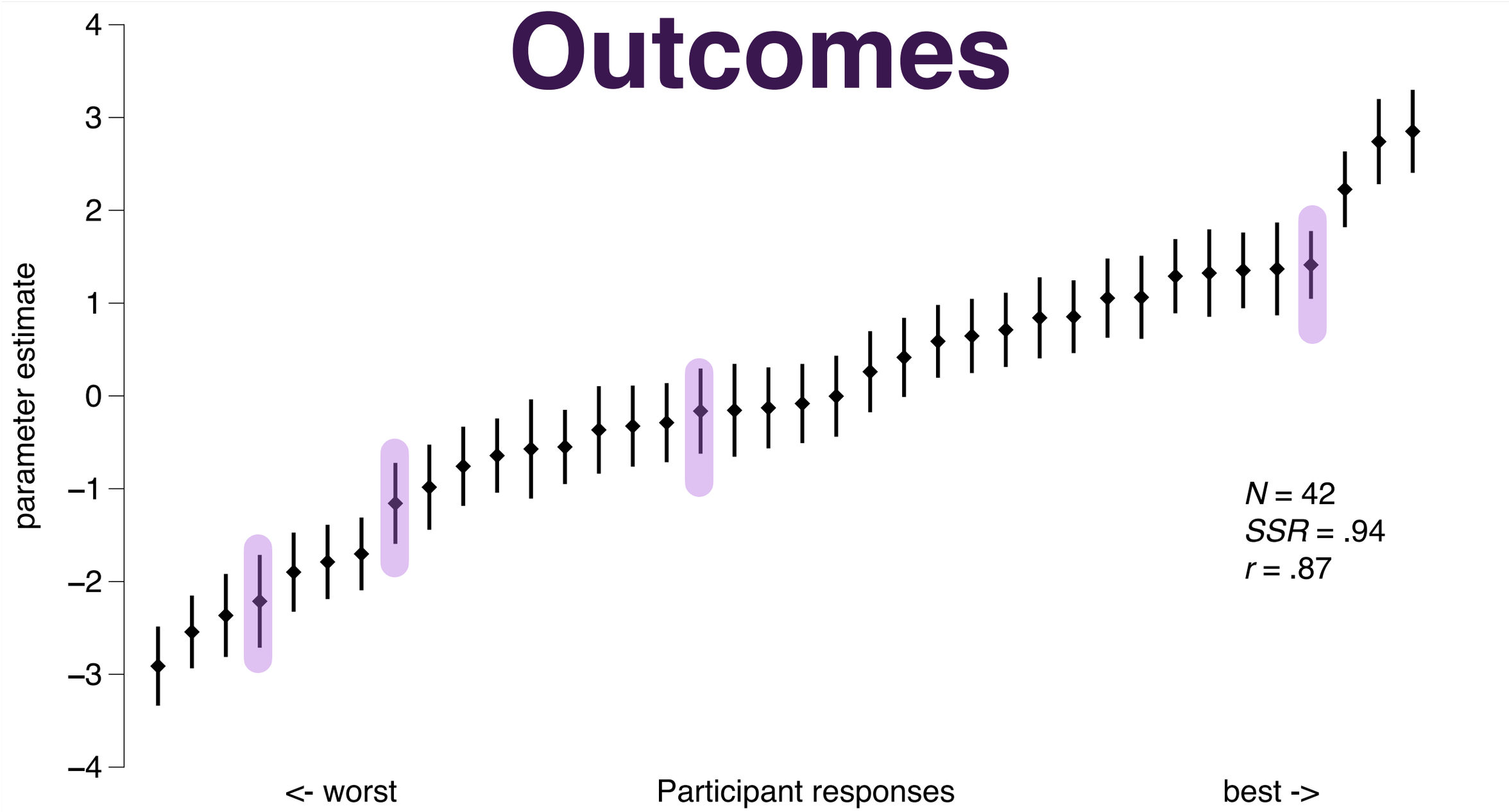


Questions and issues

Isn’t it norm-referenced?

- No, there are various methods for benchmarking comparative judgement outcomes
  (e.g. Jones & Alcock 2014; Hunter & Jones 2018)
- In fact benchmarking and standard setting is a particular strength
  (e.g. Jones et al. 2016; Heldsigner & Humphry 2013)
- You can compare unlike objects
  (e.g. Jones et al. 2016; Hunter & Jones 2018)

Outcomes



- *SSR* is Scale Separation Reliability (analogous to Cronbach's alpha)
- *r* is inter-rater reliability (split-halves technique, 20 iterations)

Bisson et al. (2016)

Isn’t it opaque?

- Validity resides in the collective judgement of experts
- Varied approaches to validity reported
  - Convergent and divergent
    (e.g. Jones & Inglis 2015*;* Jones et al. 2013)
  - Content analysis
    (e.g. Hunter & Jones 2018)
  - Interviews and surveys of judges
    (e.g. Jones & Alcock 2014*;* van Daal et al. 2019)
  - Expert vs. novice vs. non-expert judges
    (e.g. Jones & Alcock 2014*;* Jones & Sirl 2017)

What about feedback?

- Students don’t get red ink on their work
- Some CJ engines allow judges to make comments
- We can see a rich and high-quality form of feedback when students
  - compare their peers' work
  - discuss their peers' work in small groups
  - engage in lecture-facilitated large group discussions
  - reflect on which submissions come top

Thank you

Ian Jones
I.Jones@lboro.ac.uk

Thanks to Tom Bramley for the Shakespeare metaphor.

Try comparative judgement at
www.nomoremarking.com (stand alone)
or
https://github.com/ianjones/moodleCJ (Moodle plug-in)

Loughborough University
Mathematics Education Centre

Let’s do some judging

tinyurl.com/CJ090322

References

Bisson, M.-J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education, 2*, 141–164. 10.1007/s40753-016-0024-3

Evans, S., Jones, I., & Dawson, C. (2014). Do subject specialists produce more useful feedback than non-specialists when observing mathematics lessons? *Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 33–40). Vancouver, Canada: IGPME.

Heldsinger, S. A., & Humphry, S. M. (2013). Using calibrated exemplars in the teacher-assessment of writing: An empirical study. *Educational Research, 55*, 219–235. 10.1080/00131881.2013.825159

Hunter, J., & Jones, I. (2018). Free-response tasks in primary mathematics: A window on students' thinking. *Proceedings of the 41st Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 41, pp. 400–407). Auckland, New Zealand: MERGA.

Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education, 39*, 1774–1787. 10.1080/03075079.2013.821974

Jones, I., Bisson, M., Gilmore, C., & Inglis, M. (2019). Measuring conceptual understanding in randomised controlled trials: Can comparative judgement help? *British Educational Research Journal, 45*, 662–680. 10.1002/berj.3519

Jones, I., & Inglis, M. (2015). The problem of assessing problem solving: Can comparative judgement help? *Educational Studies in Mathematics, 89*, 337–355. 10.1007/s10649-015-9607-1

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: IGPME.

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

Jones, I., Wheadon, C., Humphries, S., & Inglis, M. (2016). Fifty years of A-level mathematics: Have standards changed? *British Educational Research Journal, 42*, 543–560. 10.1002/berj.3224

Kimbell, R. (2012). Evolving project e-scape for national assessment. *International Journal of Technology and Design Education, 22*, 135–155. https://doi.org/10.1007/s10798-011-9190-4

McMahon, S., & Jones, I. (2015). A comparative judgement approach to teacher assessment. *Assessment in Education: Principles, Policy & Practice, 22*, 368–389. https://doi.org/10.1080/0969594X.2014.978839

van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V., & De Maeyer, S. (2019). 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, 26*, 59–74. 10.1080/0969594X.2016.1253542

Verhavert, S., Bouwer, R., Donche, V., & Maeyer, S. D. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice, 26*, 1–22. 10.1080/0969594X.2019.1602027