Pairwise Comparison Method

Presentation outline
• Method overview
• Empirical evaluation outcomes
• Operational deployment advice
ACER created LPSs for reading and mathematics using items from a range of assessments.

Items were located on the scale using a pairwise comparison approach, with judges rating which item within a pair of items is more difficult.

The International Standard Setting Exercise determined the location of the SDG4.1.1 Minimum Proficiency Levels on the scale using the Bookmark method.
Step 1: a self-assessment to determine whether the assessment instrument is of sufficient validity to be suitable for SDG 4.1.1 reporting.

Step 2: a pairwise comparisons exercise to place items from the assessment instrument onto a Learning Progression Scale (LPS).

Step 3: statistical, common items, linking to place MPL cut-scores on the assessment instrument reporting scale.
PCM Advantages

**Cheaper and faster** than other statistical linking methods.

**Panellist training is simple** and does not require extensive preparation.

**Can be implemented consistently** using an online application.

**New assessment items can be added to LPS** to build an invaluable resource to support capacity development and strengthen the LPS and SDG 4.1.1 reporting.
Evaluating the PCM core postulate

• items from ACER’s Progressive Achievement Tests (PAT) and items used in the initial learning progressions.

• reading 60 + 50 items and mathematics 61 + 60 items

• 2780 reading and 2994 mathematics item pairs

• mean item exposure = 42, max = 50, min = 30

• 21 reading judges and 20 mathematics did on average 133 and 157 comparisons respectively

• judges received extensive learning progressions training
Correlation PAT and LPS locations

**PAT vs. LPS location**

- **M**
  - $R = 0.93$, $p < 2.2e^{-16}$

- **R**
  - $R = 0.96$, $p < 2.2e^{-16}$
PCM replication and robustness

- same PAT items
- different judges (16 reading and 15 mathematics) did an average of 270 comparison
- judges had no experience with PAT items and received remote training
- in-person training on the comparative judgement task involved extensive learning progressions training

LPS correlation for PAT items between two studies: $r = 0.97$ for reading and $r = 0.94$ for mathematics.
PCM operational deployment pilot

- 40 reading and 62 mathematics items from Pacific Islands Literacy and Numeracy Assessment (PILNA)
- 11 reading and 13 mathematics judges did an average of 290 comparisons
- all training online – EQAP staff provided logistical assistance on the ground

PAT items original and LPS location correlation: $r = 0.86$ for reading and $r = 0.85$ for mathematics.
Correlation PILNA and LPS locations – across studies

PILNA vs. LPS location

**M**

\[ R = 0.85, p < 2.2e-16 \]

**R**

\[ R = 0.77, p = 6.1e-09 \]
PCM operational deployment indication

• PCM provides reliable item location estimates enabling robust statistical linking of assessment instruments and MPLs cut-scores via LPSs

• item exposure rate should be set to at least 40

• at least 15 judges should participate in an exercise, more is desirable

• a number of comparisons each judge did seem to have had no impact on the reliability of item location estimates

• the exercise can be successfully done remotely
Next steps

• ACER is developing a toolkit to enable the consistent implementation of the PCM

• ACER is looking for jurisdictions in which the approach can be implemented.

• At present, the LPS is only available for use with assessment items in English; however, ACER would be interested to implement the approach with a bilingual panel to determine if it is possible to link items in another language to the same LPS or whether separate LPSs are required for each language.
Thank you