Teachers' ratings of children's personal, social and emotional development at the start of school: Is there evidence of differential item functioning between groups?

Katharine Bailey & Christine Merrell

Centre for Evaluation and Monitoring, Durham University

Background

- PSED

- A concern over valid interpretations of assessment data for children from diverse language and cultural backgrounds

- Assumption of comparability

- Bias associated with teachers' ratings

Assessing young children's personal, social and emotional development

- Focus is children at the start of school

- Appropriate for teachers to rate development on the basis of their observations of children in the school setting

- Analysis of PIPS Baseline Assessment data

PIPS PSED

- Comfort
- Independence
- Confidence
- Concentration on teacher-directed activities
- Concentration on self-directed activities
- Actions
- Relationships with adults
- Relationships with peers
- Rules
- Cultural awareness
- Communication

Example: Confidence

1. Very hesitant. Does not join in group activities and rarely talks.

2. Fairly hesitant. Reluctant to participate in group activities or talk.

3. Will join in group activities or talks when prompted.

4. Quite confident. Keen to join in group activities or talk within the school setting.

5. Very confident. Keen to participate in group activities within the school.

Sample

- PSED scores for 1162 children

- Started school in England in 2011

- 597 boys (51.4%) and 564 girls (48.5%)
  1 case with missing data (0.1%)

- 119 with EAL (10.2%) 792 not EAL (68.2%)
 251 cases with missing data (21.6%)

Mean scores

	EAL	Not EAL
	Mean	Mean
Comfort	3.16	3.20
Independence	2.97	3.11
Confidence	2.86	3.10
Concentration – T	2.82	2.90
Concentration - S	2.95	3.05
Actions	2.88	3.05
Relationships – Adults	2.76	2.98
Relationships - Peers	2.88	3.19
Rules	3.27	3.40
Cultural awareness	2.12	2.27
Communication	2.50	3.07

Analysis of full scale using Winsteps

- Person reliability = 0.92

- Item reliability = 1.00

DIF measure – full scale
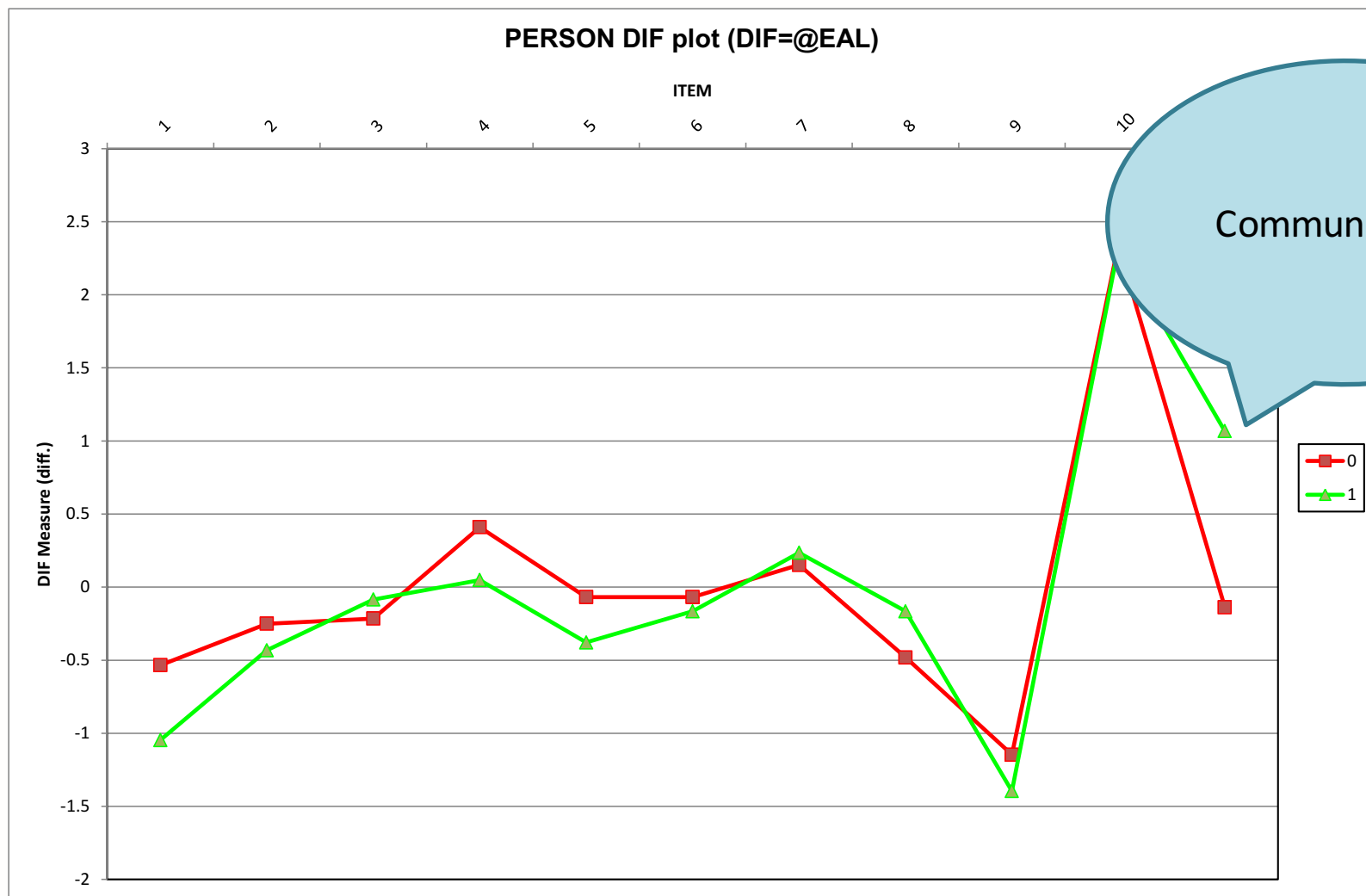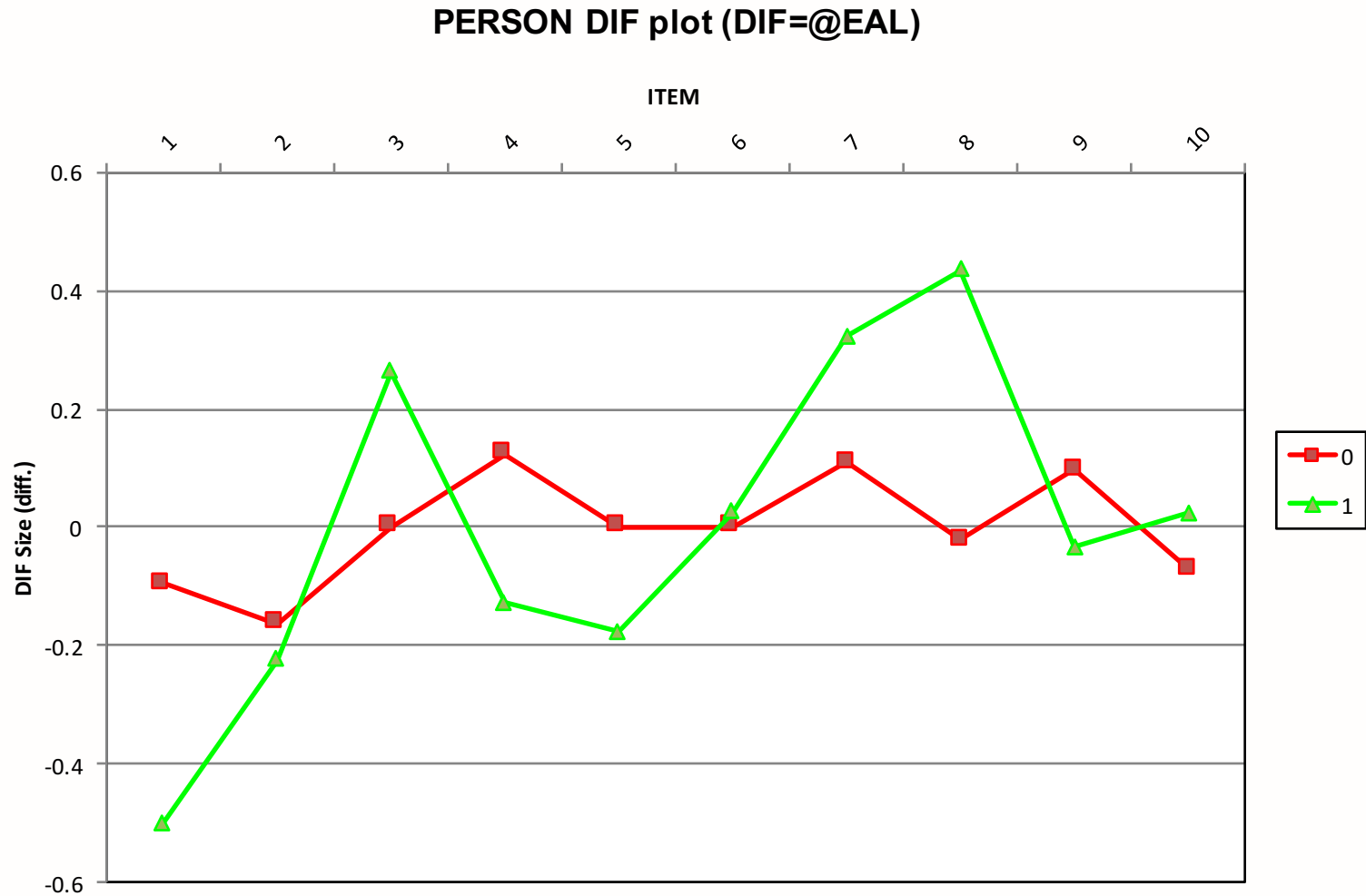
Reliability excluding Communication

- Person reliability = 0.91

- Item reliability = 1.00

DIF measure – communication removed

PERSON DIF plot (DIF=@EAL)

DIF size – communication removed

PERSON DIF plot (DIF=@EAL)

DIF t value – communication removed

PERSON DIF plot (DIF=@EAL)

Largest Differences

- Item 1 – Comfort

- Item 4 – Concentration on teacher-directed activities

- Item 8 – Relationships with peers

Implications

- Analysis of DIF suggests the assessment is appropriate for use with children from diverse language and cultural backgrounds

- … with the exception of Communication

- EAL children have lower scores overall

- More analysis is needed

Thank you