On-line Supplementary Materials

Table S1.

*Interrater Agreement for EIT Items across 3 Studies*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Son Dissertation(r = 2, n = 93) |  | Isbell Dissertation(r = 3, n = 20) |  | Isbell Pilot(r = 2, n = 5) |
| Item | Agreement (%) | ICC2a |  | Agreement (%)\* | ICC2a |  | Agreement (%) | ICC2a |
| I01 | 84 | 0.82 |  | 75 | 0.62 |  | 80 | 0.88 |
| I02 | 83 | 0.96 |  | 45 | 0.71 |  | 80 | 0.93 |
| I03 | 86 | 0.93 |  | 40 | 0.70 |  | 60 | 0.83 |
| I04 | 79 | 0.91 |  | 65 | 0.62 |  | 80 | 0.93 |
| I05 | 73 | 0.89 |  | 65 | 0.89 |  | 100 | 1.00 |
| I06 | 85 | 0.96 |  | 65 | 0.84 |  | 60 | 0.83 |
| I07 | 81 | 0.95 |  | 40 | 0.78 |  | 80 | 0.88 |
| I08 | 77 | 0.93 |  | 60 | 0.89 |  | 60 | 0.69 |
| I09 | 87 | 0.97 |  | 60 | 0.81 |  | 80 | 0.96 |
| I10 | 78 | 0.95 |  | 75 | 0.86 |  | 80 | 0.96 |
| I11 | 80 | 0.94 |  | 60 | 0.93 |  | 80 | 0.60 |
| I12 | 82 | 0.96 |  | 40 | 0.86 |  | 100 | 1.00 |
| I13 | 81 | 0.94 |  | 65 | 0.93 |  | 40 | 0.75 |
| I14 | 81 | 0.93 |  | 65 | 0.94 |  | 40 | 0.75 |
| I15 | 75 | 0.88 |  | 55 | 0.90 |  | 40 | 0.40 |
| I16 | 84 | 0.93 |  | 60 | 0.89 |  | 40 | 0.50 |
| I17 | 80 | 0.93 |  | 65 | 0.90 |  | 100 | 1.00 |
| I18 | 76 | 0.91 |  | 55 | 0.89 |  | 80 | 0.87 |
| I19 | 80 | 0.95 |  | 50 | 0.85 |  | 60 | 0.71 |
| I20 | 85 | 0.97 |  | 50 | 0.86 |  | 80 | 0.60 |
| I21 | 84 | 0.95 |  | 50 | 0.90 |  | 60 | 0.80 |
| I22 | 81 | 0.90 |  | 50 | 0.83 |  | 100 | 1.00 |
| I23 | 80 | 0.96 |  | 70 | 0.96 |  | 80 | 0.80 |
| I24 | 76 | 0.88 |  | 55 | 0.89 |  | 100 | 1.00 |
| I25 | 82 | 0.93 |  | 70 | 0.95 |  | 40 | -0.20 |
| I26 | 77 | 0.90 |  | 70 | 0.94 |  | 80 | 0.71 |
| I27 | 80 | 0.94 |  | 75 | 0.95 |  | 80 | 0.00 |
| I28 | 81 | 0.92 |  | 55 | 0.92 |  | 80 | 0.87 |
| I29 | 74 | 0.89 |  | 60 | 0.90 |  | 60 | 0.81 |
| I30 | 75 | 0.91 |  | 60 | 0.91 |  | 60 | 0.64 |
|  |  |  |  |  |  |  |  |  |
| Mean | 80 | 0.93 |  | 59 | 0.86 |  | 72 | 0.75 |

\*Exact agreement was reached for all 3 scorers.

Table S2.

*Measurement Summary of 3-Facet RSM (Integrated Data, All Scorers)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Facet | N | Measure M(SD) | Measure Range | Infit Range | Outfit Range | Sep. Strata | Rel. |
| Persons | 318 | 0.09 (1.57) | -4.90 – 4.16 | 0.39 – 2.25 | 0.24 – 7.89 | 6.91 | 0.98 |
| Scorers | 5 | 0.00 (0.08) | -0.10 – 0.11 | 0.76 – 1.12 | 0.94 – 1.33 | 2.32 | 0.84 |
| Items | 30 | 0.00 (0.99) | -2.36 – 1.46 | 0.70 – 2.08 | 0.66 – 4.98 | 17.14 | 1.00 |

Table S3

*Category Statistics of 3-Facet RSM (Integrated Data, All Scorers)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Score | N | Average Measure | Expected Measure | Outfit | Rasch-Andrich Threshold (SE) |
| 0 | 3437 | -2.20 | -2.18 | 1.0 | - |
| 1 | 2054 | -0.99 | -0.91 | 0.9 | -0.98 (.03) |
| 2 | 2446 | 0.17 | -0.02 | 1.7 | -0.63 (.03) |
| 3 | 2617 | 0.82 | 0.88 | 0.9 | 0.35 (.03) |
| 4 | 3124 | 2.06 | 2.08 | 1.0 | 1.26 (.03) |

*Note.* The Average Measure is the average Rasch person ability measure of all examinees who received a particular score on the rating scale, while the Expected Measure is model-predicted average Rasch person ability measure of examinees who receive a particular score. Outfit indicates the degree to which unexpected observations are present for a score; values near 1.0 are ideal. Rasch-Andrich thresholds indicate the point in the examinee ability continuum at which two adjacent categories are equally probable.

Table S4.

*Measurement Summary of 2-Facet RSM (Integrated Data, Single Scorer)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Facet | N | Measure M(SD) | Measure Range | Infit Range | Outfit Range | Sep. Strata | Rel. |
| Persons | 318 | 0.09 (1.57) | -4.90 – 4.16 | 0.36 – 2.48 | 0.40 – 8.17 | 6.91 | 0.98 |
| Items | 30 | 0.00 (0.99) | -2.36 – 1.46 | 0.70 – 2.03 | 0.66 – 4.23 | 17.14 | 1.00 |

*Note.* Rating scale thresholds: 0/1 = -1.06 logits, 1/2 = -0.59 logits, 2/3 = 0.38 logits, 3/4 = 1.27 logits.

Table S5.

*Linear regression EIT item averages on select linguistic features following Kim et al. (2016)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | SE | ß | *p* |
| Intercept | 3.87 | 0.40 |  |  |
| Syllables | -0.07 | 0.04 | -0.35 | 0.10 |
| Relative Clauses | -0.27 | 0.20 | -0.21 | 0.18 |
| Vocabulary Score | -0.09 | 0.05 | -0.34 | 0.08 |

F(3,26) = 9.55, p = 0.0002, Multiple R2 = 0.54, Adjusted R2 = 0.47.



*Figure S1.* Wright Map summarizing the 3-facet model with all raters.



*Figure S2.* Category probability curves for the 3-facet RSM.



*Figure S3.* Wright Map summarizing a 3-facet hybrid model with rating scale category threshold estimated separately for each rater.