

Supplemental Material S2. Test-retest reliability studies: number of participants, children's age, interval and the correlations between the 1st and 2nd session. The table is listed by the interval between 1st and 2nd sessions.

| Test–retest study | Number of participants | Children’s age | Interval between the 1st and 2nd sessions | Test–retest reliability; correlations or percentage of agreements between the 1st and 2nd session |
|--|------------------------|----------------|---|---|
| Kenyan (Alcock et al., 2015) | 17 | WG | 7 days | C: 0.69 P: 0.54 |
| German (Szagun et al., 2009; Table 10) | 57 | WS | 8–10 days | P: 0.87* Co: 0.94* |
| British (Alcock et al., 2020) ¹ | 31 | 11–15M | 3–15 days | scores on each subscale between 1st and 2nd sessions did not differ significantly relatively low correlations between 1st and 2nd session on production (no more info) |
| Canadian French (Trudeau et al., 2008, Table 2, Table 3) | 30 | 8–10M | 2–3 weeks | C: 0.89 P: 0.79 G: 0.79 |
| | 29 | 13–15M | | C: 0.91 P: 0.92 G: 0.93 |
| | 30 | 19–21M | | P: 0.98 Co: 0.97 |
| | 32 | 26–28M | | P: 0.95 Co: 0.90 |

| Test–retest study | Number of participants | Children’s age | Interval between the 1st and 2nd sessions | Test–retest reliability; correlations or percentage of agreements between the 1st and 2nd session |
|---|------------------------|----------------|---|---|
| Peninsular Spanish (López Ornat et al., 2005, Table 4.45, Table 4.46) | 67 | WG | 2–3 weeks | C: 0.98* P: 0.99* G: 0.94* |
| | | WS | | P: 0.99* Co: 0.98* |
| Mexican Spanish (Jackson-Maldonado et al., 1993, Table 6, Table 7) | 16 | WG | 1 month | C: 0.66* P: 0.53* |
| | 20 | WS | | P: 0.62* |
| Chinese Cantonese (Tardif & Fletcher, 2008, Table 4.6, Table 4.7) | 48 | WG | 1 month | C: 0.73* P: 0.77* G: 0.75* |
| | 94 | WS | | P: 0.90* Co: 0.87* |
| Chinese Mandarin (Tardif & Fletcher, 2008, Table 4.6, Table 4.7) | 90 | WG | 4–6 weeks | C: 0.76* P: 0.91* G: 0.64* |
| | 150 | WS | | P: 0.89* Co: 0.86* |
| Basque (García et al., 2008, Table 6) | 20 | WG | 6 weeks | C: 0.92 P: 0.83 G: 0.83 |

| Test–retest study | Number of participants | Children’s age | Interval between the 1st and 2nd sessions | Test–retest reliability; correlations or percentage of agreements between the 1st and 2nd session |
|---|------------------------|----------------|---|---|
| | 28 | WS | | P: 0.98 Co: 0.99 |
| Norwegian (Simonsen et al., 2014, Table 2, Table 3) | 329 | WG | Average 46 days | C: 0.62–0.95 P: 0.47–0.97 G: 0.39–0.95 |
| | 397 | WS | | P: 0.40–1.00 |
| Polish (Smoczyńska et al., 2015) | 89 | WG | 45 days at maximum | C: 0.90 P: 0.85 G: 0.94 |
| | 64 | WS | 45 days at maximum | P: 0.97 |
| Cypriot Greek (Taxitari et al., 2016) | N/A | WS | 1 month | N/A |
| Estonian (Schults, 2016) | 10 | 14M | 2 months | C: 0.74 P: 0.93 |
| Swedish (Berglund & Eriksson, 2000a, Table 5) | 54 | WG | 2 months | C: 82–93% P: 59–77% G: 91–97% |
| | 60 | WS | 3 months | P: 86–94% Co: 93–94% |

Note. C = comprehended words; P = produced words; G = gestures; Co = section Complexity; * = partial correlation.

¹ The test–retest was assessed also as interform reliability—the caregivers first completed the paper form and during the second session the online form or vice-versa.