Le misure in psicologia sono significanti? Il caso del test della Torre di Londra

Ottavia M. Epifania, Luca Stefanutti, Pasquale Anselmi, Andrea Brancaccio, Debora de Chiusole

Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata,
Università di Padova

Convegno AIP-Sezione Sperimentale 2023
Simposio: Crisi di replicabilità o crisi di validità? L'importanza delle misure

19 Settembre 2023

The ratio between the measures of a and b is constant and independent of the measurement unit:

$$\frac{\varphi(a)}{\varphi(b)} = \frac{\varphi'(a)}{\varphi'(b)},$$

where φ and φ' are two different scales of measurement of the same variable.

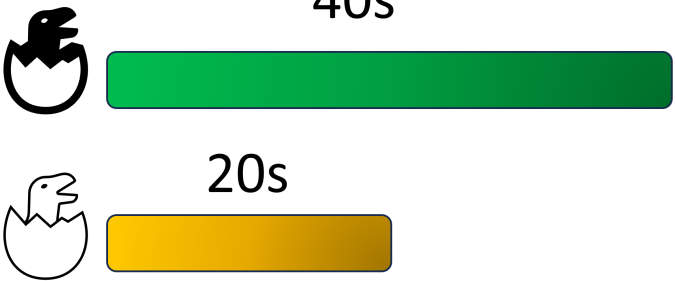
The ratio between the measures of a and b is constant and independent of the measurement unit:

$$\frac{\varphi(a)}{\varphi(b)} = \frac{\varphi'(a)}{\varphi'(b)},$$

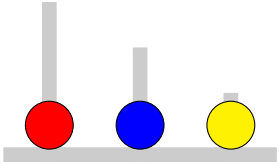
where φ and φ' are two different scales of measurement of the same variable.

Meaningful comparisons

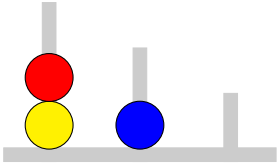
The comparison between a and b is meaningful if it is invariant under all the unit transformations.


40s
20s

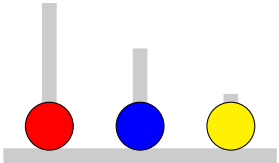

40s
20s



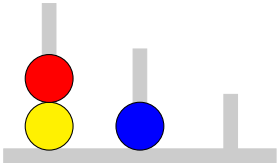
Starting configuration



Goal configuration



Starting configuration



Goal configuration

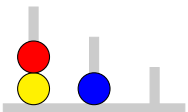
Item difficulty influenced by:

- Number of moves
- Number of alternative paths
- Hierarchy of the starting/goal configuration

The Tower of London Test (ToL Test)

Shallice (1982)

- 12 problems
- Same starting configuration
- More than one attempt per item



Problem	Minimum moves	Alternative paths
Example	2	1
1	2	1
2	2	1
3	3	2
4	3	1
5	4	2
6	4	1
7	4	1
8	4	1
9	5	2
10	5	1
11	5	1
12	5	2

Scoring	Attempts	Response times	Item score	Total score
Shallice 1	✓	✓	0-1	0-12
Shallice 2	×	✓	0-3	0-36
Anderson et al.	✓	✓	0-9	0-108
Kirkorian et al.	✓	×	0-3	0-36

Scoring	Attempts	Response times	Item score	Total score
Shallice 1	✓	✓	0-1	0-12
Shallice 2	×	✓	0-3	0-36
Anderson et al.	✓	✓	0-9	0-108
Kirkorian et al.	✓	×	0-3	0-36

Shallice 2 – SH2

For each of the 12 items:

Assign	if time is
3	≤ 15 s
2	$15 \dashv 30$ s
1	$30 \dashv 60$ s
0	> 60 s

Anderson et al. – AN

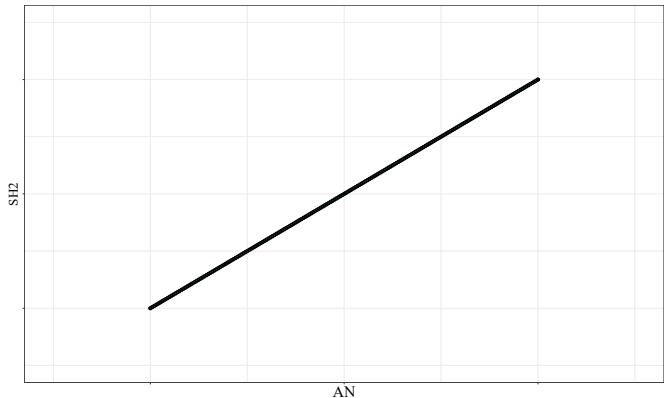
For each of the 12 items:

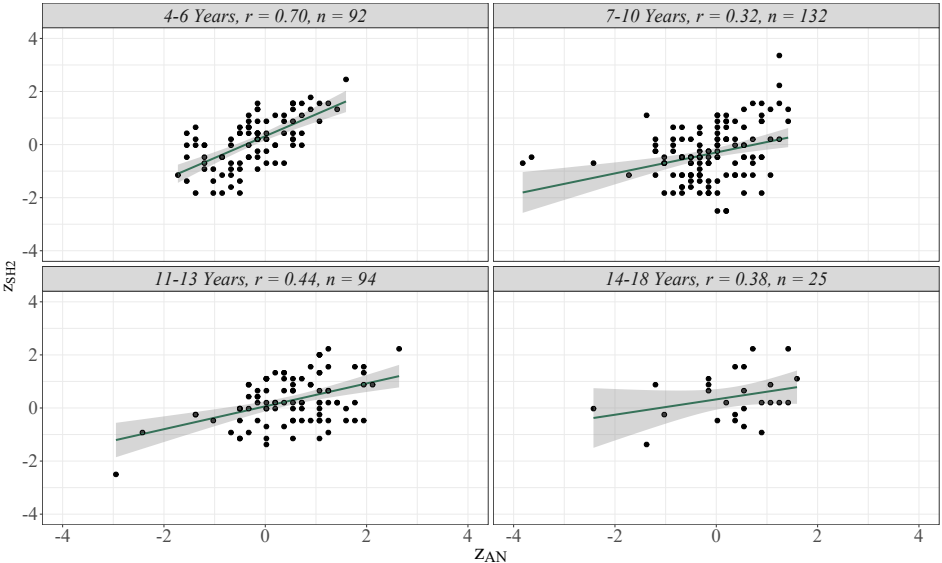
Assign	if time is
9	≤ 6 s
8	$6 \dashv 10$ s
7	$11 \dashv 20$ s
6	$21 \dashv 40$ s
5	$41 \dashv 60$ s
0	> 60 s

Subtract the number of unsuccessful attempts

Both scorings are based on the discretization of the response times → There should not be differences in the **order** of the total score of the respondents according to the scoring method

Both scorings are based on the discretization of the response times → There should not be differences in the **order** of the total score of the respondents according to the scoring method




4-6 Years, r = 0.70, n = 92
7-10 Years, r = 0.32, n = 132
11-13 Years, r = 0.44, n = 94
14-18 Years, r = 0.38, n = 25
zSH2
zAN
4
2
0
-2
-4

Is it really bad...?

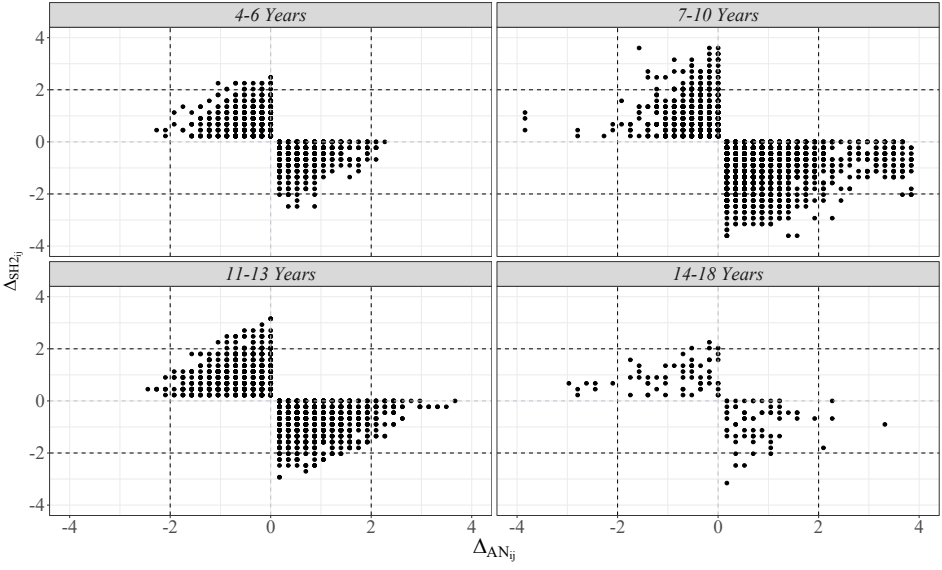
Respondents $i, j \in \{1, \dots N\}$

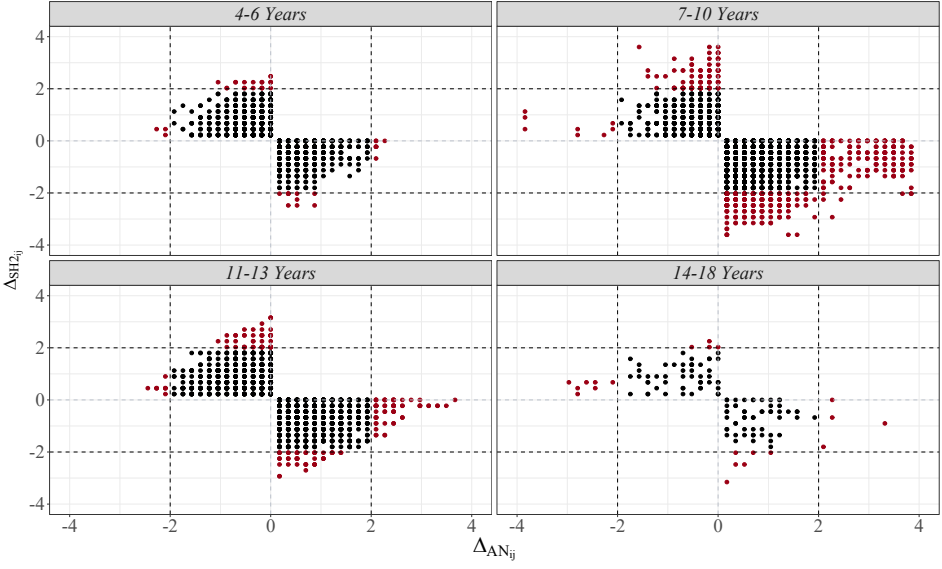
- AN Comparison (Δ_{AN}): The standardized AN score of each subject i is compared against the standardized AN score of every other subject j

$$\Delta_{\text{AN}_{ij}} = z_{\text{AN}_i} - z_{\text{AN}_j}$$

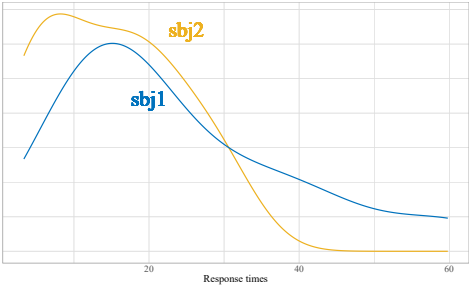
- SH2 Comparison (Δ_{SH2}): The standardized SH2 score of each subject i is compared against the standardized SH2 score of every other subject j

$$\Delta_{\text{SH2}_{ij}} = z_{\text{SH2}_i} - z_{\text{SH2}_j}$$


4-6 Years
7-10 Years
11-13 Years
14-18 Years
$\Delta_{SH2_{ij}}$
$\Delta_{AN_{ij}}$
4
2
0
-2
-4


4-6 Years
7-10 Years
11-13 Years
14-18 Years
$\Delta_{SH2_{ij}}$
$\Delta_{AN_{ij}}$
4
2
0
-2
-4

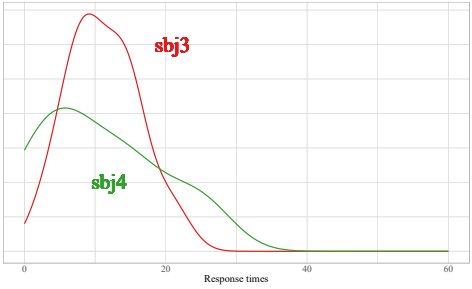
$\Delta_{\text{AN}} > 2$ & $\Delta_{\text{SH2}} \approx 0$



	z_{AN}	z_{SH2}	Accuracy	Time (sd)
sbj1	−1.55	0.43	0.75	24.10 (15.60)
sbj2	0.72	0.43	0.75	14.51 (9.22)

	Δ_{AN}	Δ_{SH2}
sbj1 − sbj2	2.27	0.00

$\Delta_{\text{AN}} \approx 0$ & $\Delta_{\text{SH2}} > 2$



	z_{AN}	z_{SH2}	Accuracy	Time (sd)
sbj3	−0.15	1.55	0.75	11.14 (4.96)
sbj4	0.20	−0.70	0.58	10.72 (8.60)

	Δ_{AN}	Δ_{SH2}
sbj3 − sbj4	−0.35	2.25

Highlights

- Different scoring systems → The focus is shifted: Fast and furious or slow and steady?
- Different scoring systems might favor a cognitive theory over a contrasting one (raising also replicability issues)

But

What if the performance of the respondents could suggest the most appropriate scoring system? Currently underway

Thank you!

ottavia.epifania@unipd.it