

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

① Meaningfulness

② The case in point

- Tower of London
- Scoring systems

③ Real data application

④ Final remarks

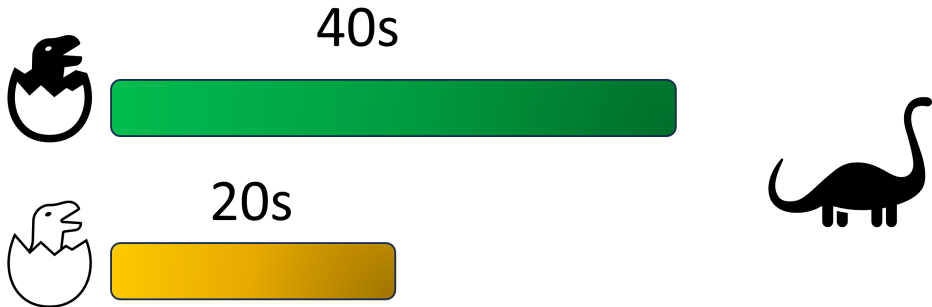
In the measurement of lengths, the ratio between the measures of two fixed lengths a and b is constant and independent of the measurement unit:

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

where φ and φ' are two different scales of measurement of length.

Meaningful comparisons

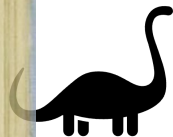
The comparison between a and b is meaningful if it is invariant under all the unit transformations.





40s

20s



1 Meaningfulness

2 The case in point

- Tower of London
- Scoring systems

3 Real data application

4 Final remarks

1 Meaningfulness

2 The case in point

- Tower of London
- Scoring systems

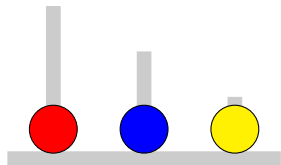
3 Real data application

4 Final remarks

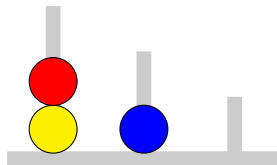
Meaningfulness

└ The case in point

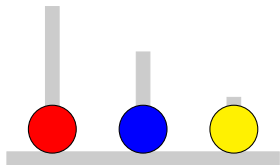
└ Tower of London



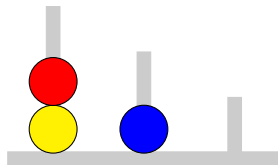
Starting configuration



Goal configuration



Starting configuration



Goal configuration

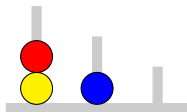
Item difficulty influenced by:

- 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



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

1 Meaningfulness

2 The case in point

- Tower of London
- Scoring systems

3 Real data application

4 Final remarks

Meaningfulness

└ The case in point

└ Scoring systems

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

Anderson et al. – AN

For each of the 12 items:

Assign	if time is
3	< 15 s
2	< 30 s
1	< 60 s
0	≥ 60 s

For each of the 12 items:

Assign	if time is
9	< 6 s
8	6 – 10 s
7	11 – 20 s
6	21 – 40 s
5	41 – 60 s
0	> 60 s

Subtract the number of unsuccessful attempts

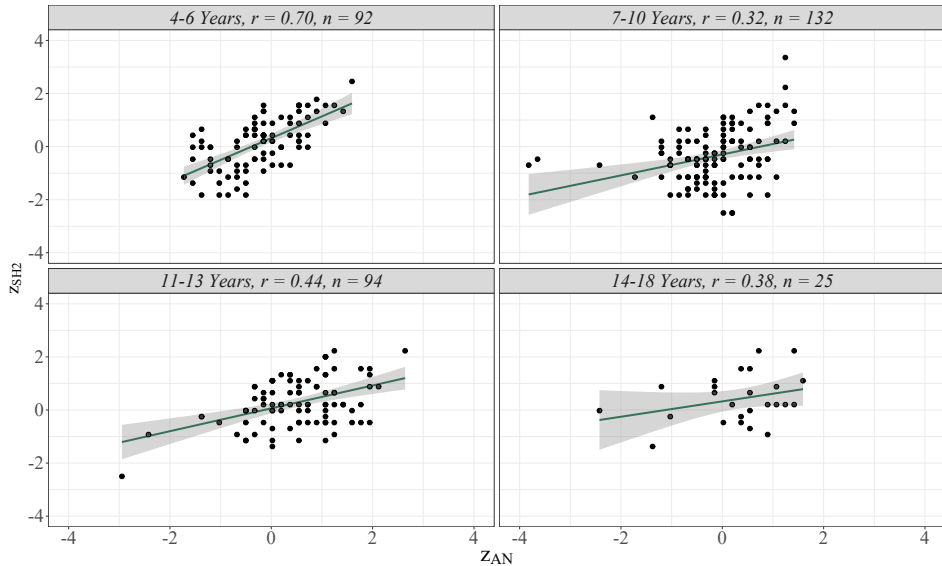
① Meaningfulness

② The case in point

- Tower of London
- Scoring systems

③ Real data application

④ Final remarks



Is it really bad...?

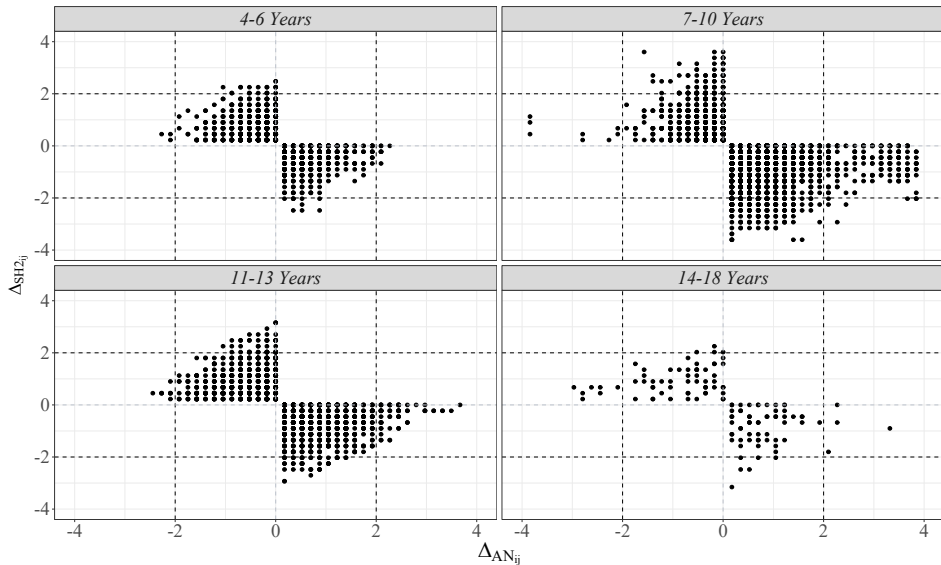
Respondent $i, j \in \{1, \dots, N\}$

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

$$\Delta_{AN_{ij}} = z_{AN_i} - z_{AN_j}$$

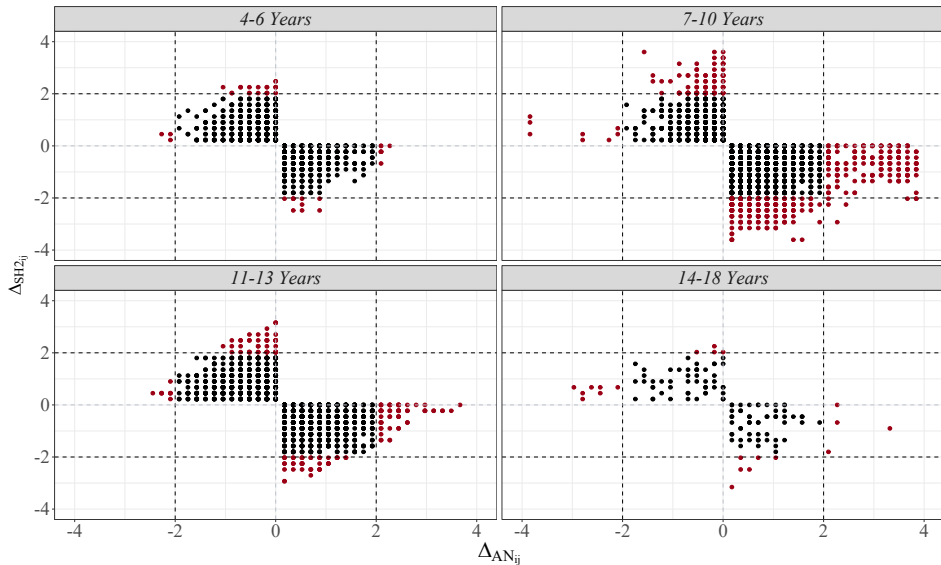
- SH2 Comparison (Δ_{SH2}): The standardized SH2 score of each subject i is compared against the standardized SH2 score of every other subject j

$$\Delta_{SH2_{ij}} = z_{SH2_i} - z_{SH2_j}$$

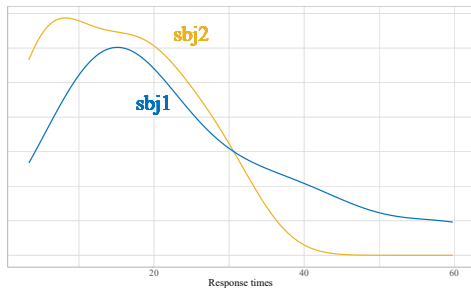


Meaningfulness

└ Real data application



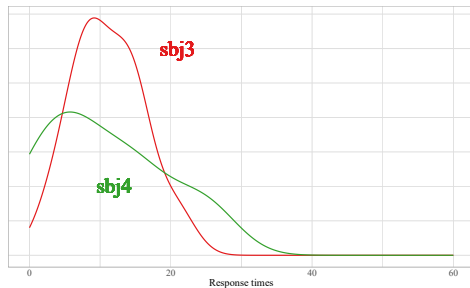
$$\Delta_{AN} > 2 \text{ \& } \Delta_{SH2} \approx 0$$



	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

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



	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

- ① Meaningfulness
- ② The case in point
 - Tower of London
 - Scoring systems
- ③ Real data application
- ④ **Final remarks**

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

Live long and prosper