UME Bristo University of the West of England

Test-retest reliability for common tasks in vision science Kait Clark¹, Charlotte R. Pennington¹, Craig Hedge², Joshua Lee¹, & Austin Petrie¹ ¹University of the West of England, ²Cardiff University

Reliability

Reliability:

Consistency in results produced by a measure

Test-retest reliability: Correlation of scores taken at 2+ points in time

How well can a measure consistently distinguish between individuals who have high/low scores?

Repeatability \neq reliability

Highly robust tasks are often unreliable (e.g., Stroop) (Hedge, Powell, & Sumner, 2018)

Test-retest reliability is required to assess individual differences

Intraclass correlation coefficient

Variance between individuals

Variance between individuals + *error variance* + variance between sessions

ICC	Classification
0.8-1	Excellent
0.6-0.8	Good
0.4-0.6	Moderate
0.2-0.4	Fair
0-0.2	Poor

Low ICC could represent high error or the fact that individuals are very similar*

*cognitive tasks often designed to minimise variance

Participants and testing

Participants: 165 undergraduate psychology students

Testing sessions:

- Two sessions, separated by 1-3 weeks
- Each session two hours in length:
- 1 hour social cognition tasks (Pennington et al., in prep)
- 1 hour perceptual tasks (current poster)

Tasks

Useful field of view Motion coherence [MoCo] (UFOV) 90 ms 200 m Indicate direction of motion Fixate in the centre 400 dots, random motion Central number / peripheral dot Starting trial: 24% coherent flash for 90 ms 3 staircases, 1-up/1-down [%] Report number and dot location

Meal /meeane	таа	Dhe	
Task/measure			Spoormon's
			Speamans
Threshold	0.60	0.57	 Accounts for
UFOV			• Tends to be
Number Accuracy	0.47	0.46	lenient mea
Inner Accuracy	0.31	0.48	 Alians with I
Middle Accuracy	0.60	0.65	
Outer Accuracy	0.74	0.75	
MOT			UFOV - Inner Ri
Max Items	0.41	0.36	1
Threshold	0.36	0.31	0.8
VWM			
Capacity	0.75	0.77	0.0 SSIO
*similar to Xu et	al., 20	18	• 0.4
			0.2
Other tasks (from He	dge et a	1., 2018)	0.2 0.4
Stroop			Sess
RT Cost	0.66	0.70	
Accuracy Cost	0.44	0.44	MOT - Max Ite
Flanker			8 -
RT Cost	0.57	0.62	\sim 7
Accuracy Cost	0.72	0.63	• <u>-</u>
Posner cueing			S S S - •
Cueing effect	0.70	0.64	



Test-retest reliabilities





Research with Action in Mind

Psychological Sciences Research Group



- Higher ICCs reflected by higher betweensubject variance
- Minimal between-subject variance allows for robust, replicable effects
- (but difficult to assess individual differences) • Side note: Very little between-session variance
- (minimal practice effects)

Recommendations

Consider test-retest reliability before assessing individual differences

- Many cognitive tasks not designed to discriminate between individuals
- Unlikely to find IDs for tasks with low ICCs
- If a measure does not correlate with itself,
- it's unlikely to correlate with anything else

References

- Hedge, C., Powell, G., & Sumner, P. (2018). Why robust cognitive tasks do not produce reliable individual differences. Behavior Research Methods, 50(3), 1166-1186.
- Pennington, C. R., Clark, K., Shaw, D.J., Ploszajski, M., & Hedge, C. (in prep). Evaluating the test re-test reliability of tasks measuring social cognition.
- Xu, Z., Adam, K. C. S, Fang, X., & Vogel, E. K. (2018). The reliability and stability of visual working memory capacity. Behavior Research Methods, 50(2), 576-588.