



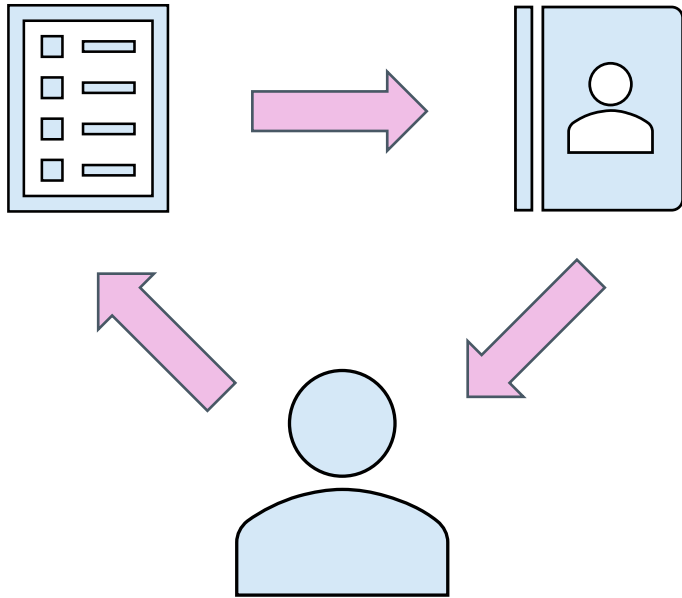
# **Modelling approaches to combining and comparing independent adaptive comparative judgement ranks**

Dr Jeff Buckley, Dr Niall Seery, & Prof. Richard Kimbell  
*Technological University of the Shannon: Midlands Midwest, Ireland*  
*Goldsmiths, University of London, UK*

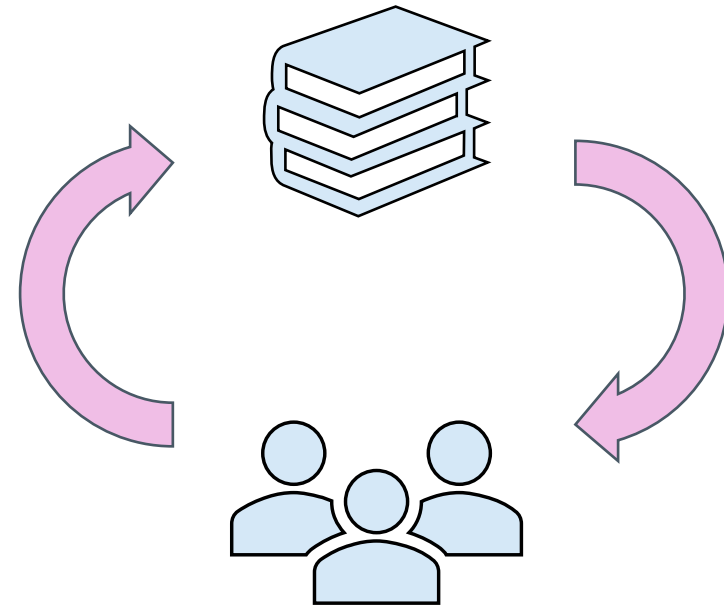


# Comparative judgement

## Criterion Assessment



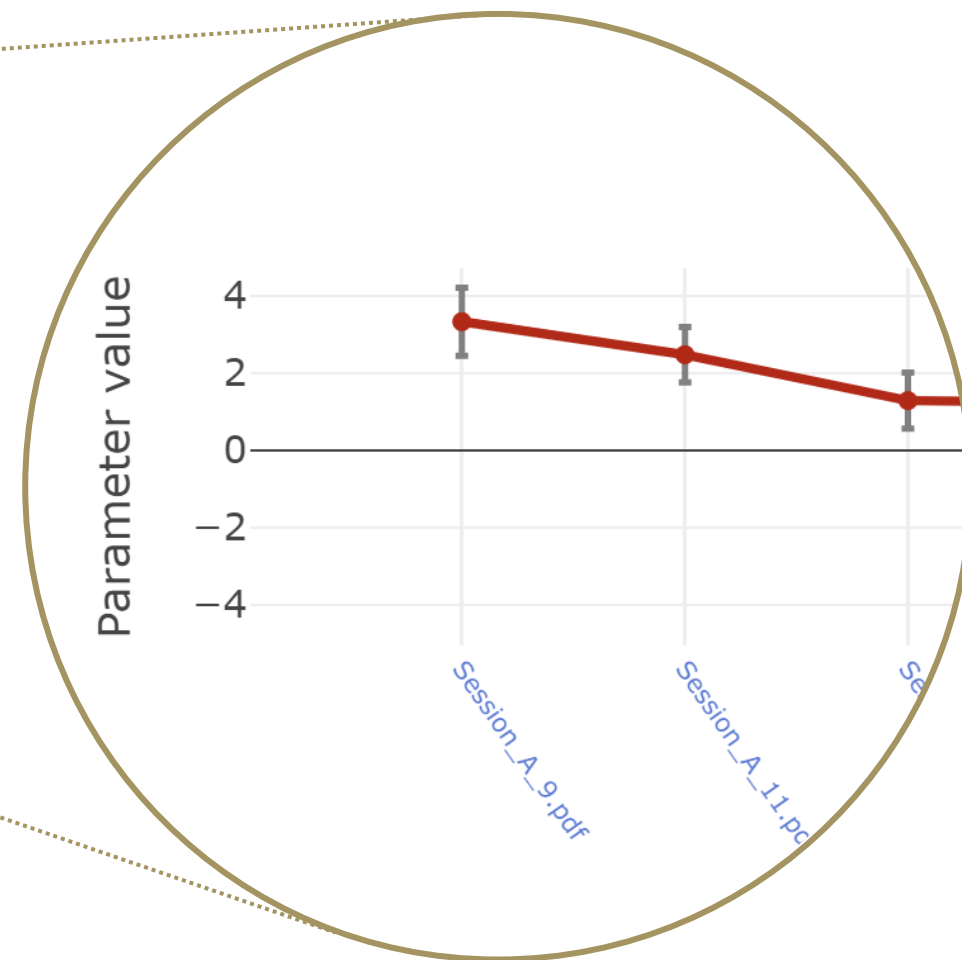
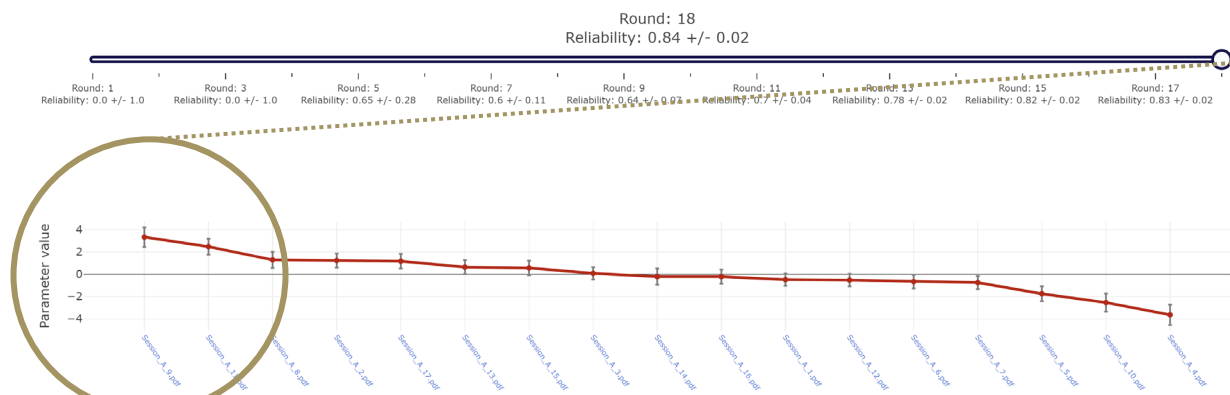
## Comparative Judgement



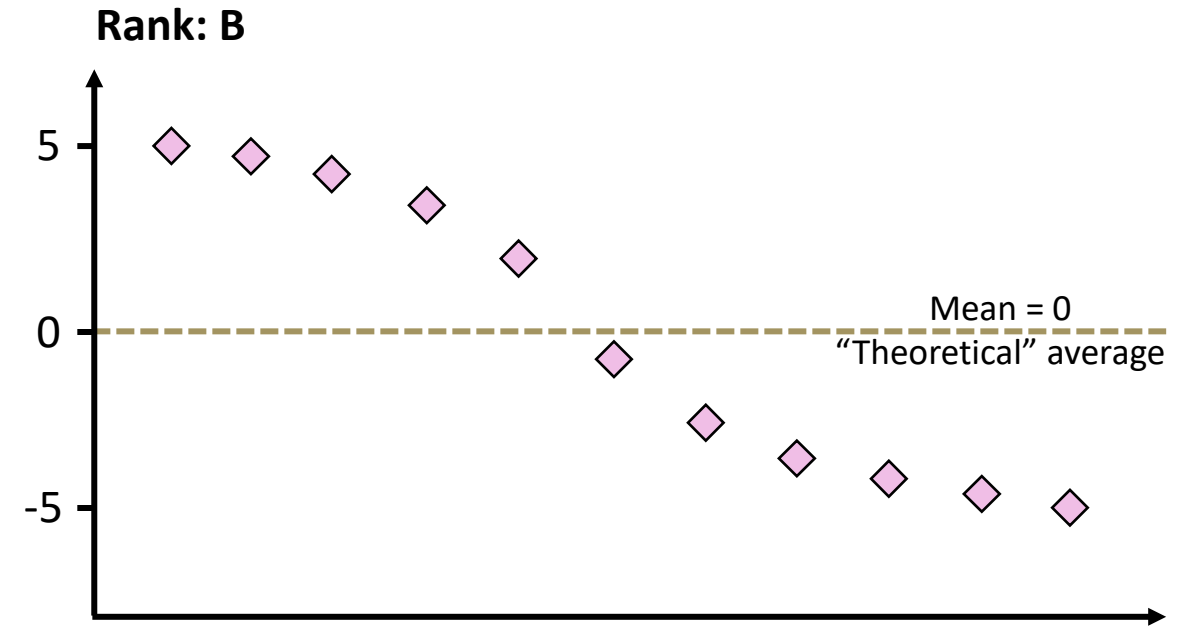
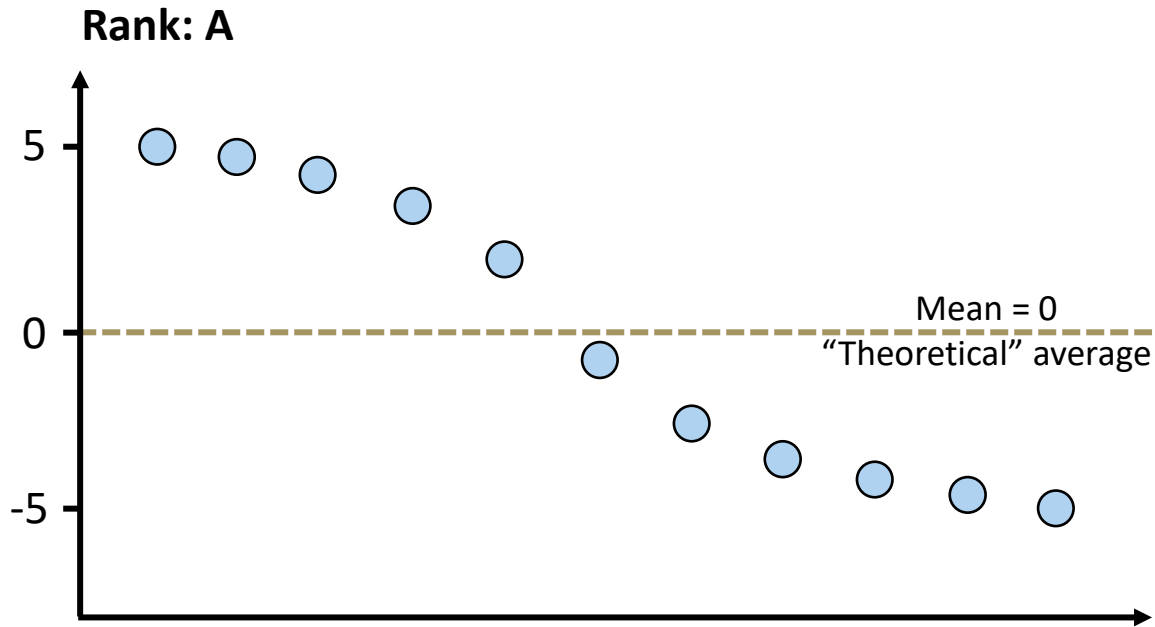
# Comparative judgement

The screenshot displays a web-based interface for comparative judgement. At the top, it shows 'My Judgement Sessions / (1/30)' and a 'View holistic statement' dropdown. The interface is split into two panels, A and B, each with a '1 of 1' indicator and navigation icons. Panel A contains the text 'Portfolio A presented here' and Panel B contains 'Portfolio B presented here'. A 'Current view' selector at the top right shows 'A', 'A&B', and 'B' options, with 'A&B' currently selected.

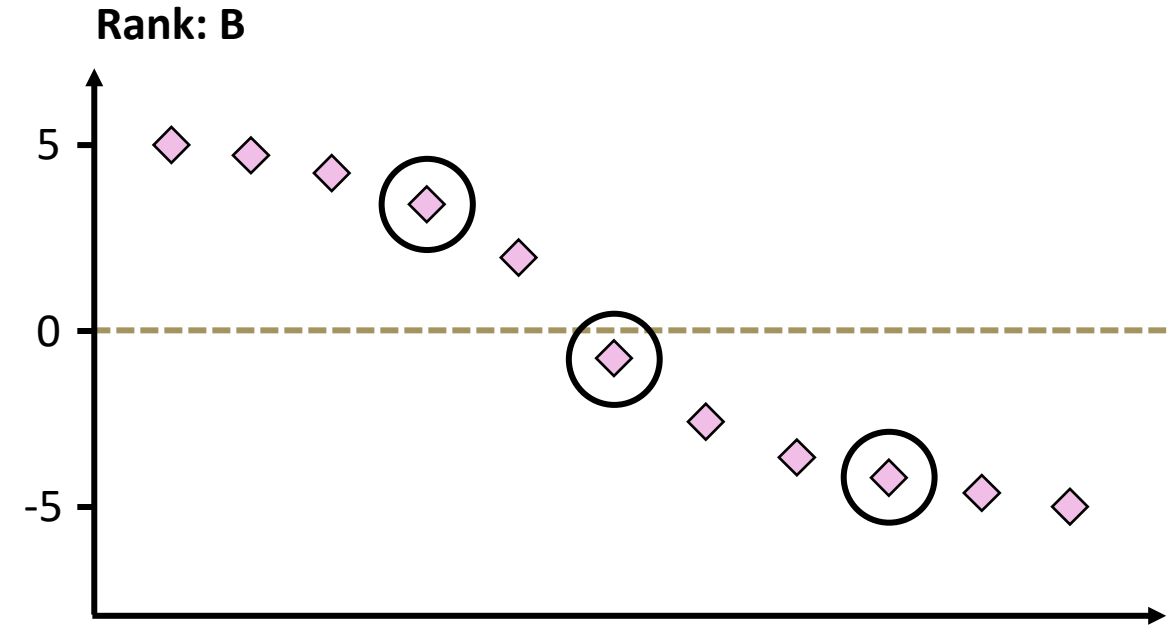
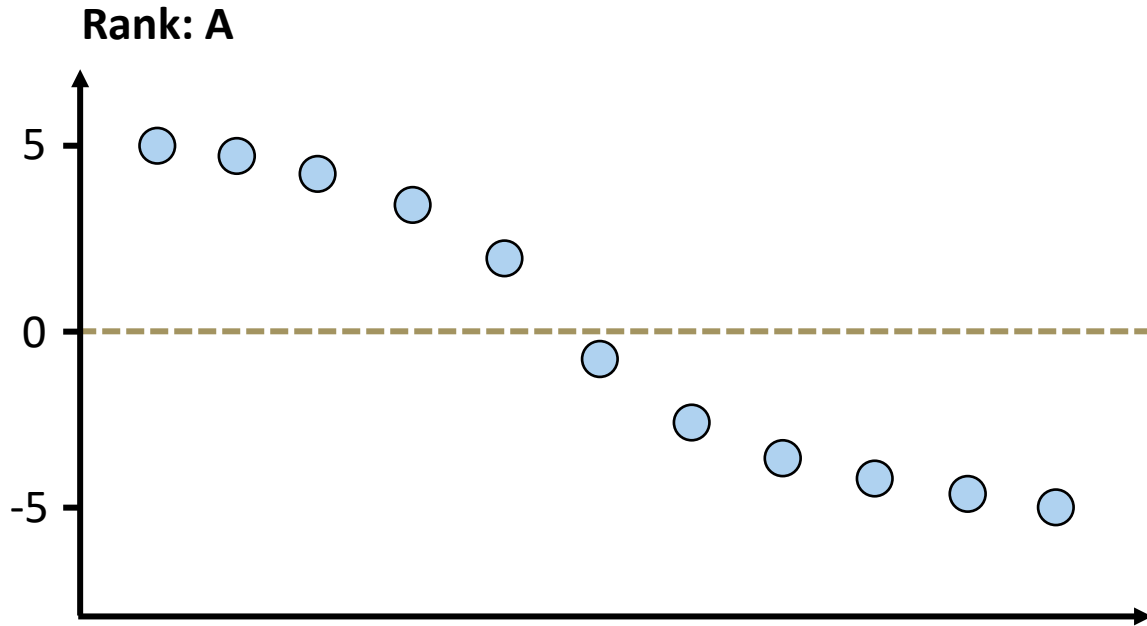
# Comparative judgement output



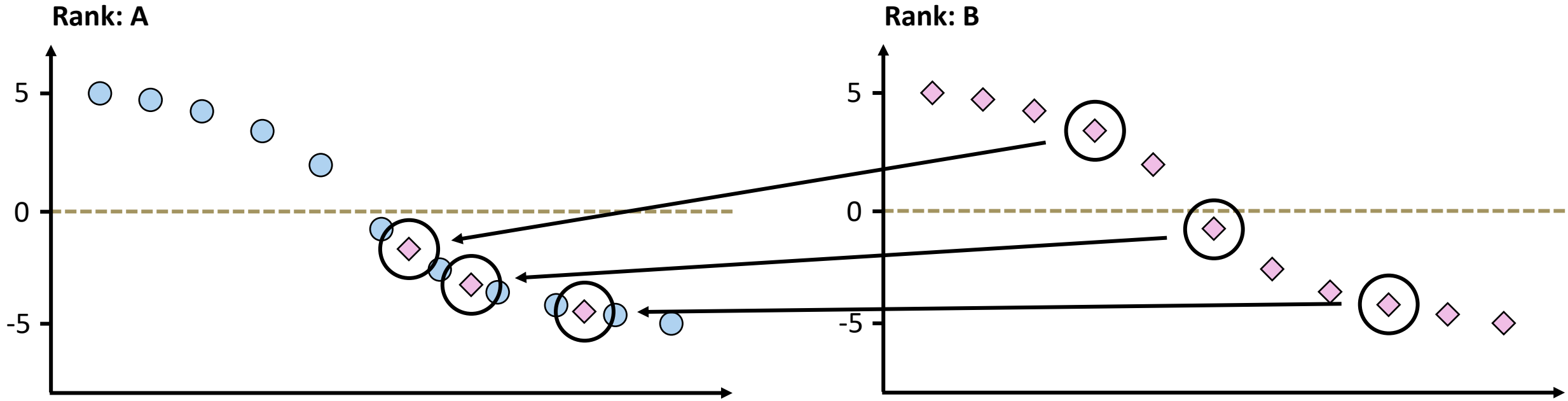
# The limitation



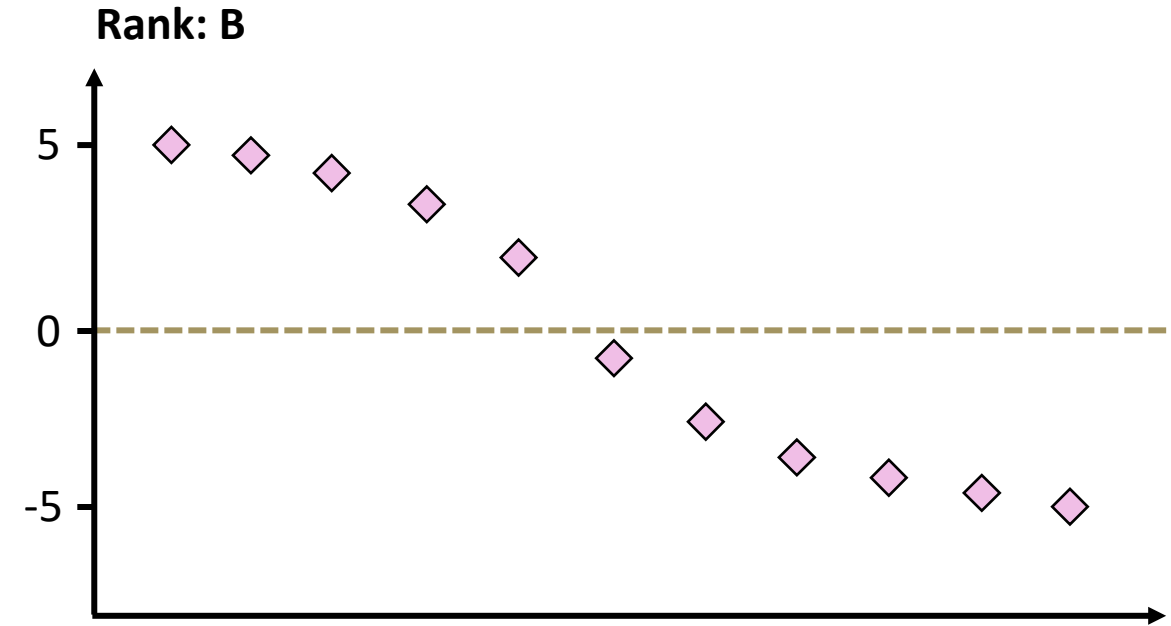
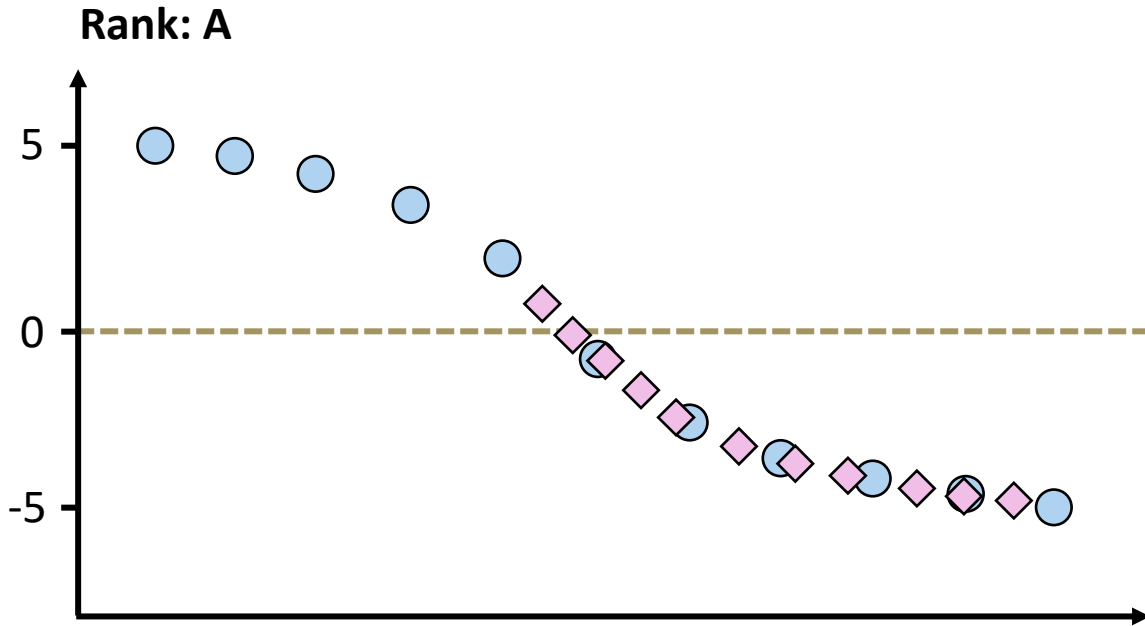
# A solution



# A solution



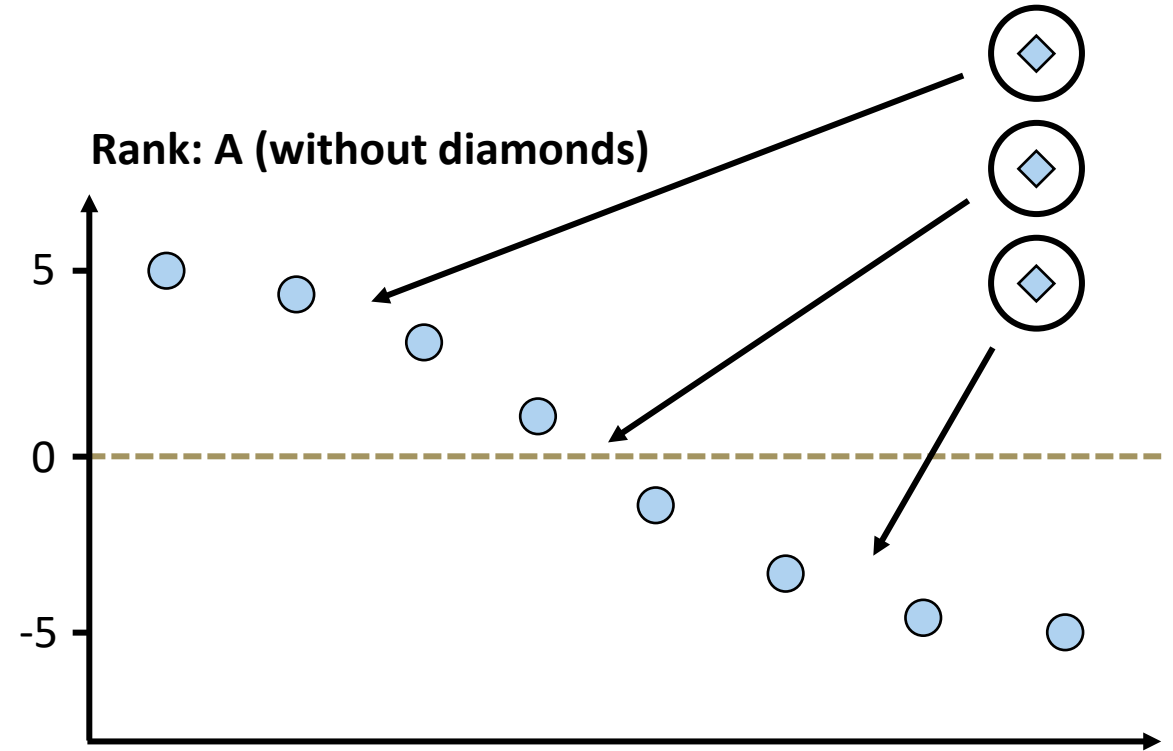
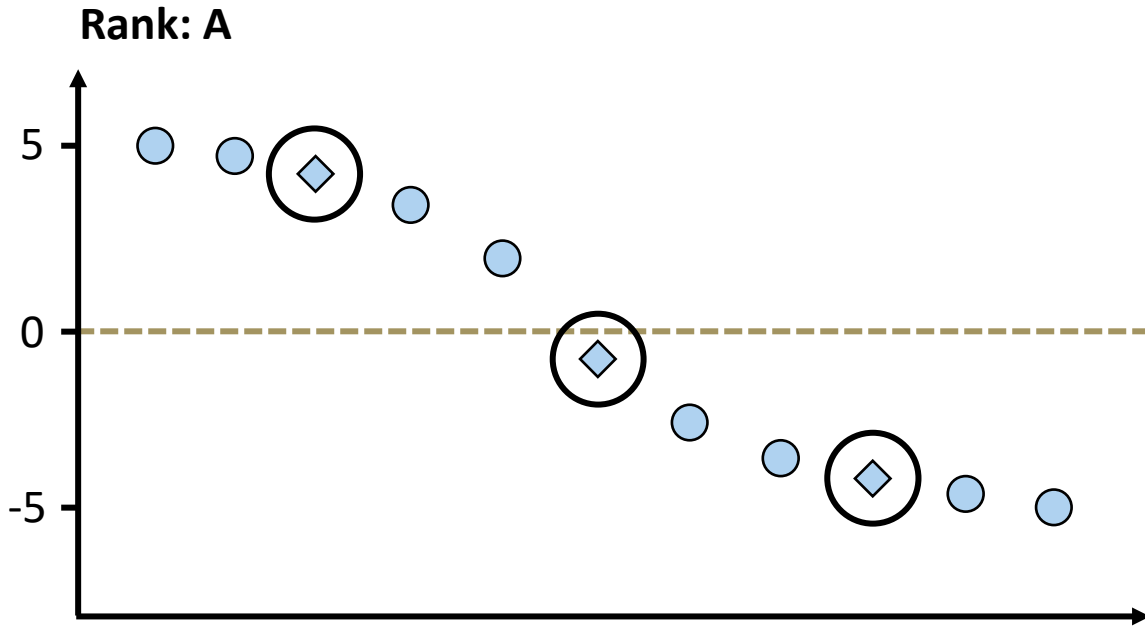
# A solution





# Last year

Buckley and Canty (2022)



# Timeline

## Last Year

Start with an original rank  
Remove “some” portfolios  
Recompute original rank without portfolios  
Place portfolios back into the rank



## This Year

Take two independent ranks, A and B  
Select portfolios from A and judge into B  
Use these A portfolios to rescale A rank  
Merge scaled A and original B ranks

# Underpinning model

The table shows how many wins each portfolio has against another. Here, we can see that **portfolio A won against portfolio B three times**.

Portfolio	A	B	C	D
A	-	3	0	1
B	4	-	5	0
C	0	2	-	1
D	5	0	2	-

## Bradley-Terry-Luce model

$$\alpha_i = \frac{W_i}{\sum_{j \neq i} \frac{w_{ij} + w_{ji}}{\alpha_i + \alpha_j}}$$

$$\text{Portfolio A parameter value} = \frac{\text{Sum of Portfolio A wins}}{\text{For all unique portfolios } \dots \frac{\text{Portfolio A wins} + \text{Comparator portfolio wins}}{\text{Portfolio A parameter value} + \text{Comparator portfolio parameter value}}}$$

$$\text{Portfolio A parameter value} = \frac{3 + 0 + 1}{\frac{3 + 4}{1 + 1} + \frac{0 + 0}{1 + 1} + \frac{1 + 5}{1 + 1}} = 0.615$$

# Methodology

## ACJ Session A

17 portfolios from  
School A

Coded portfolio.aX

To be scaled rank

## ACJ Session B

18 portfolios from  
School B

Coded portfolio.bX

Fixed rank

## ACJ Session C

35 portfolios from  
School A and B

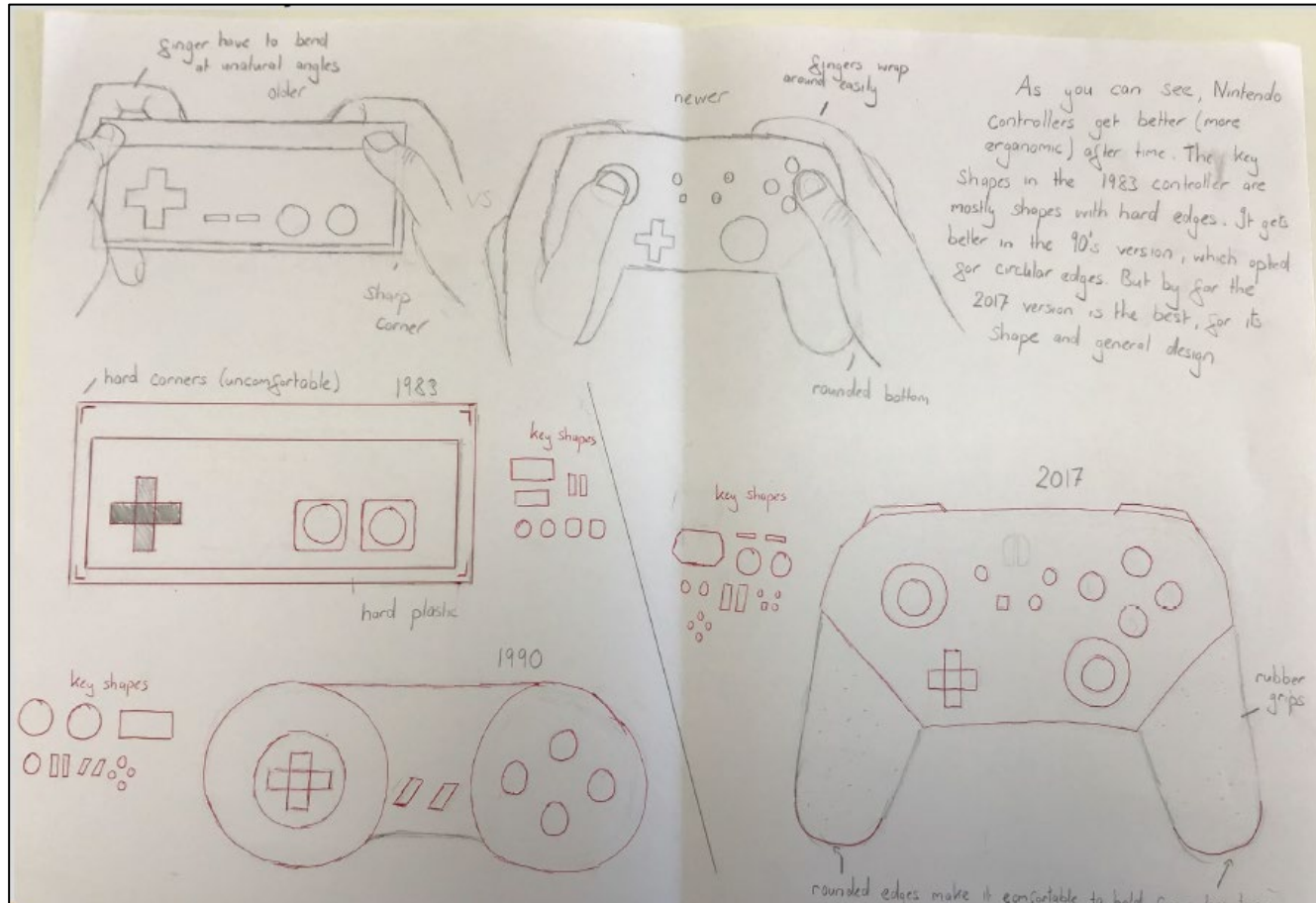
“Goal state”

## CJ Session D

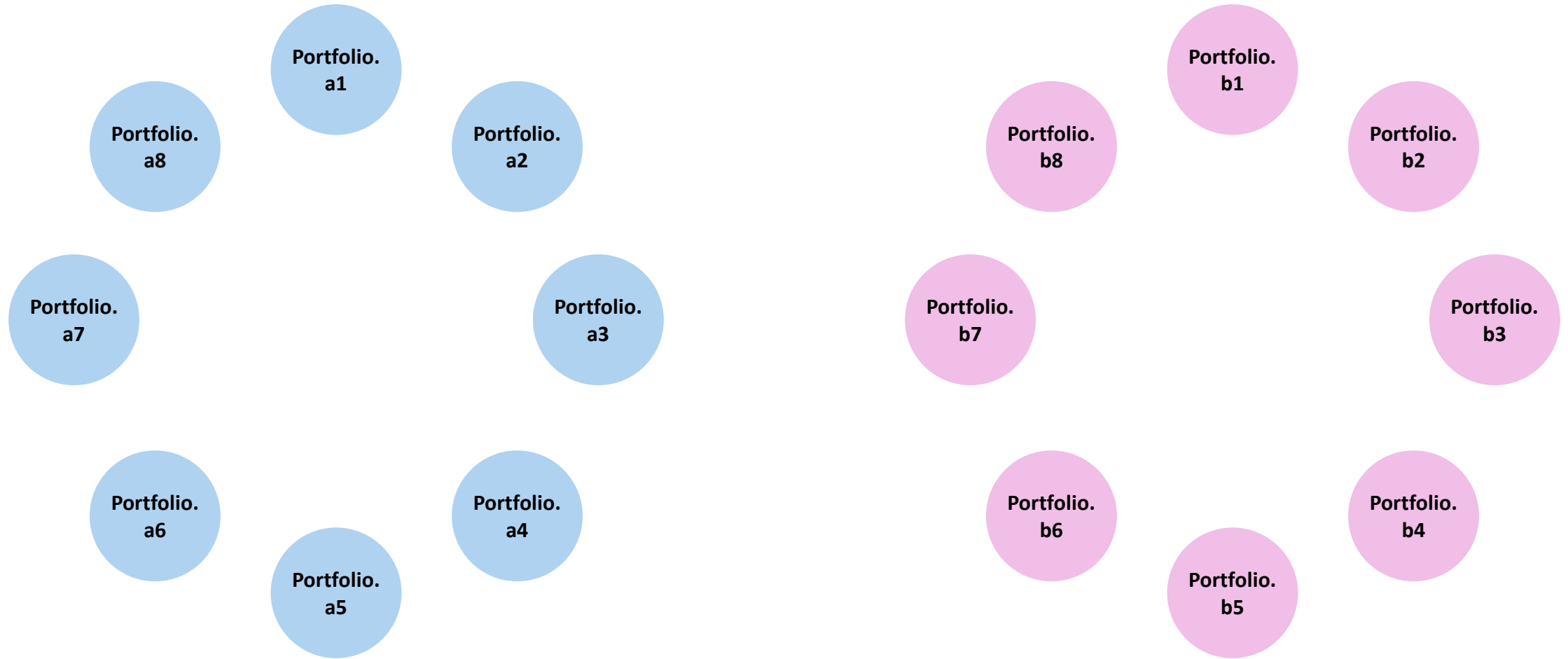
Purposeful section of  
School A and B portfolios  
with pre-determined  
judgements

To scale Session A  
portfolios in Session B

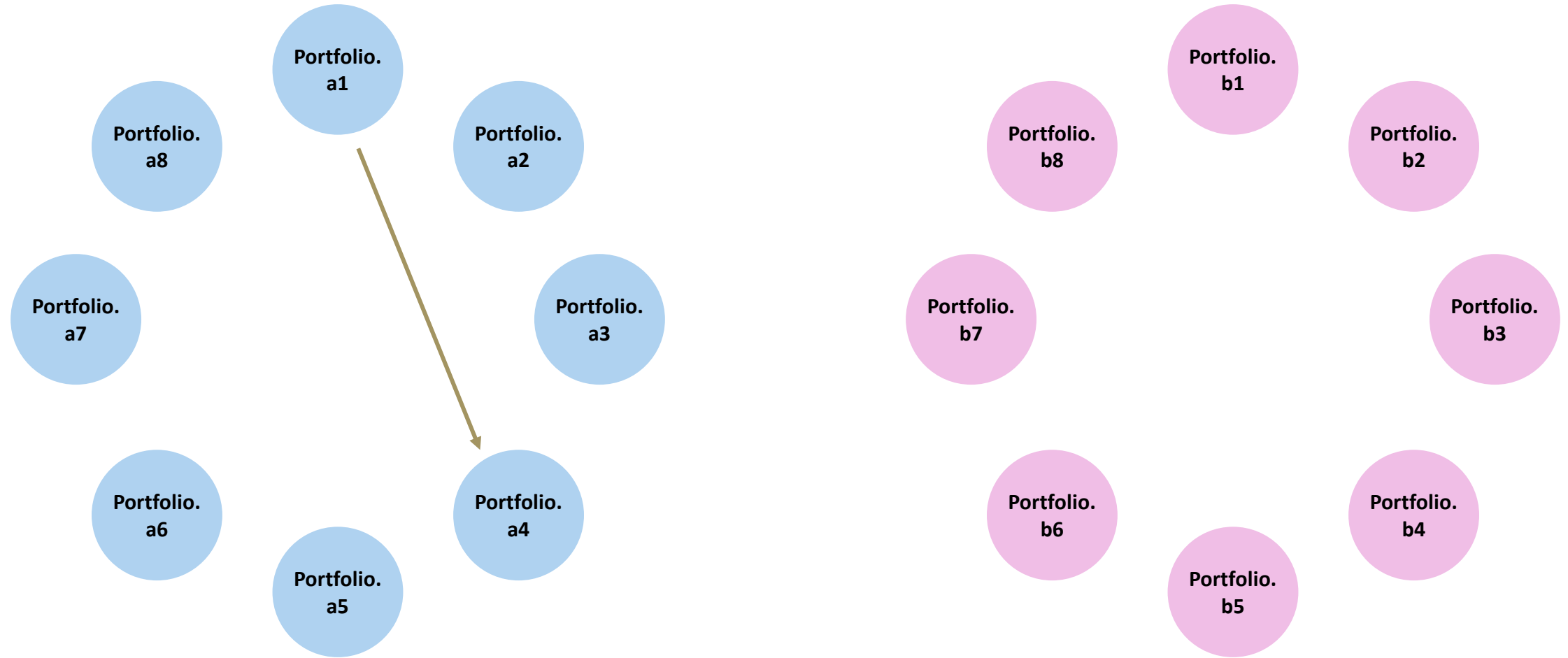
# Student activity



# Nature of data



# Nature of data

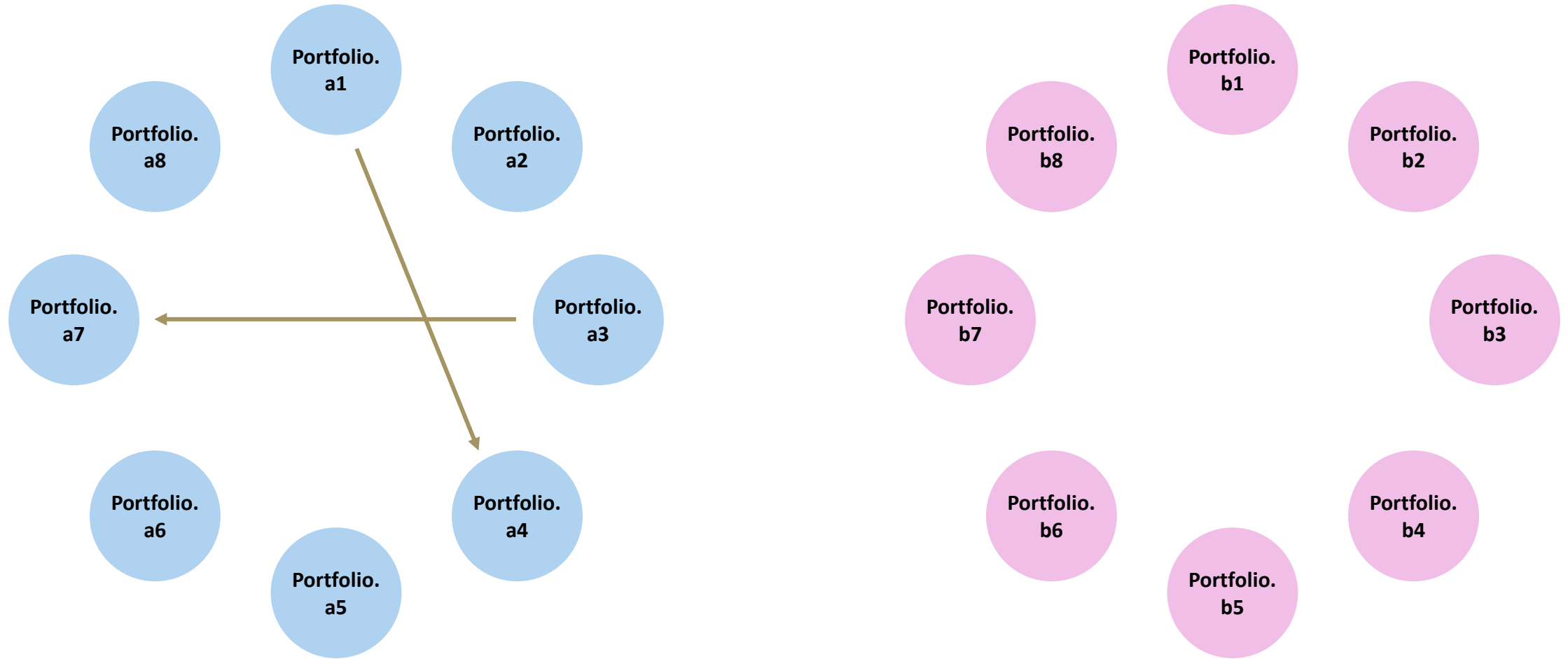


# Nature of data

Portfolio.1	Portfolio.2	Result	Judge
Portfolio.a1	Portfolio.a4	1	Judge.1



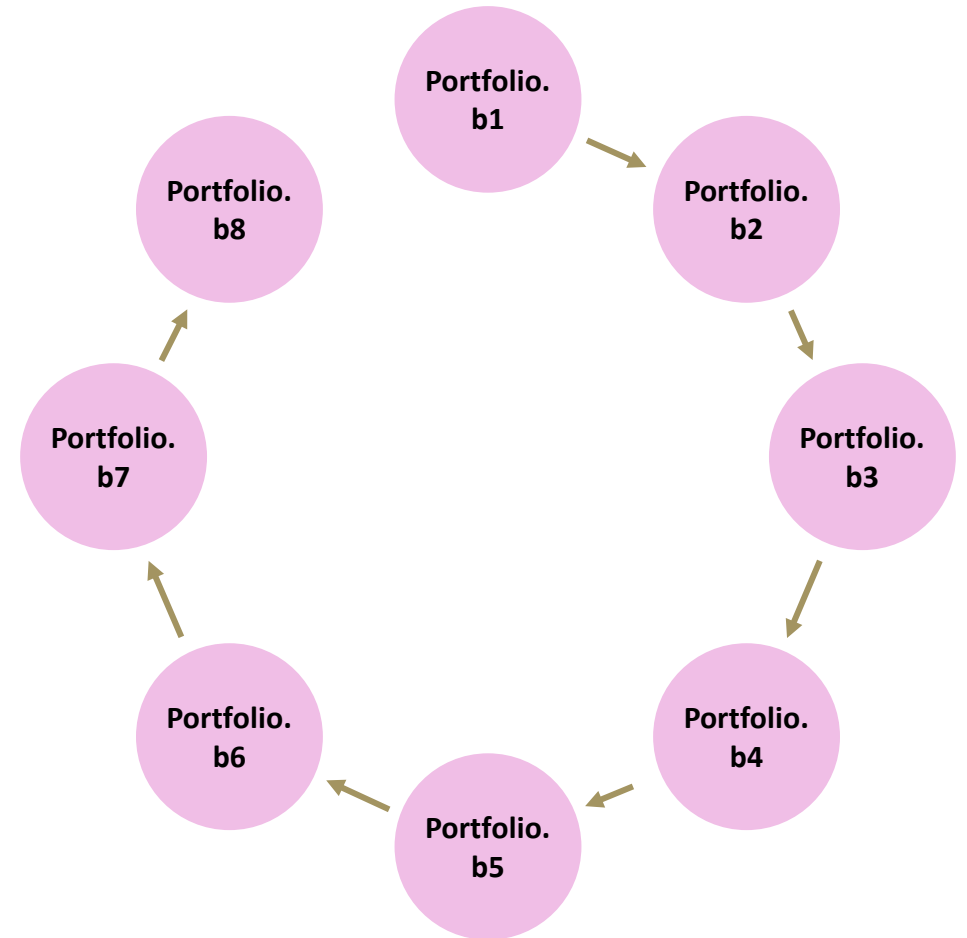
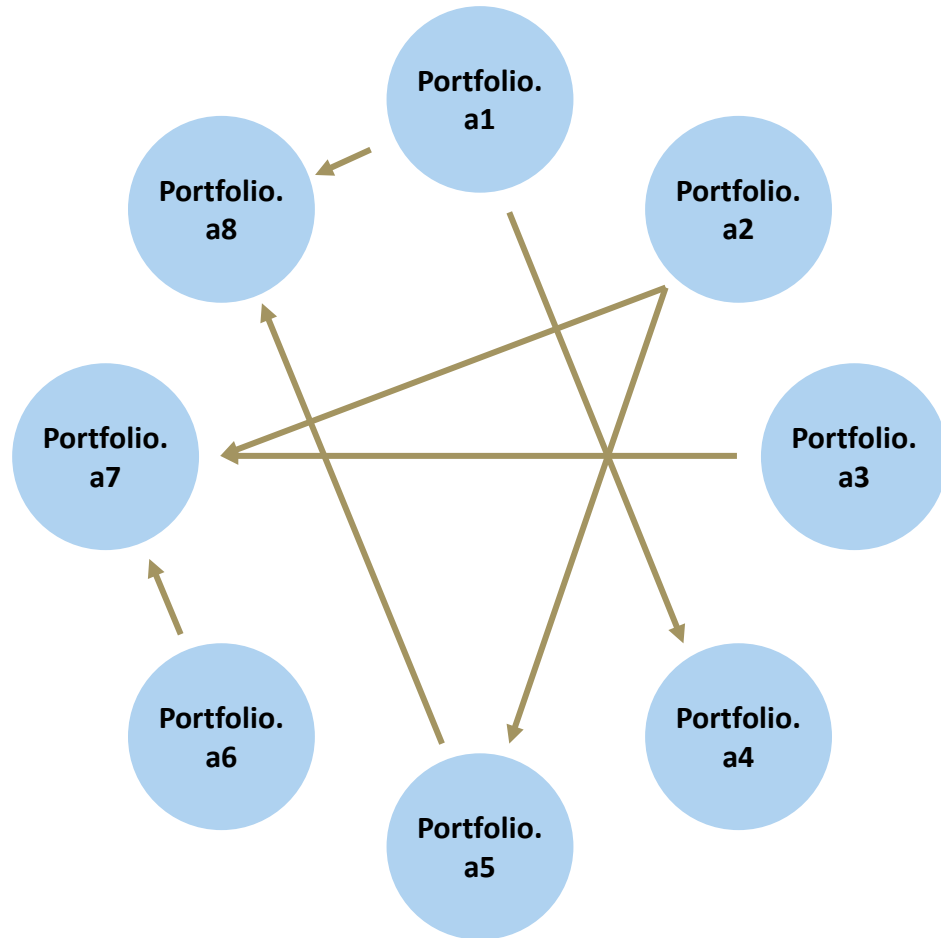
# Nature of data



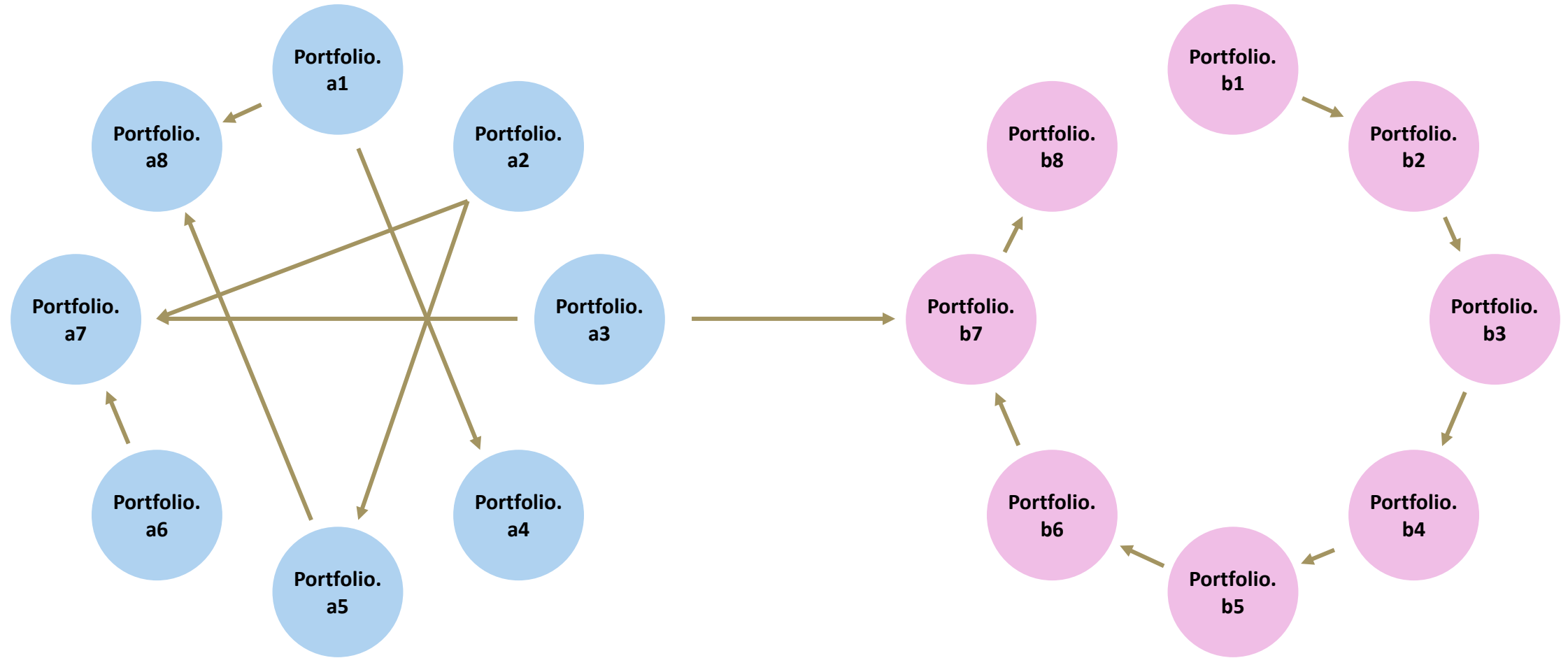
# Nature of data

Portfolio.1	Portfolio.2	Result	Judge
Portfolio.a1	Portfolio.a4	1	Judge.1
Portfolio.a3	Portfolio.a7	0	Judge.2

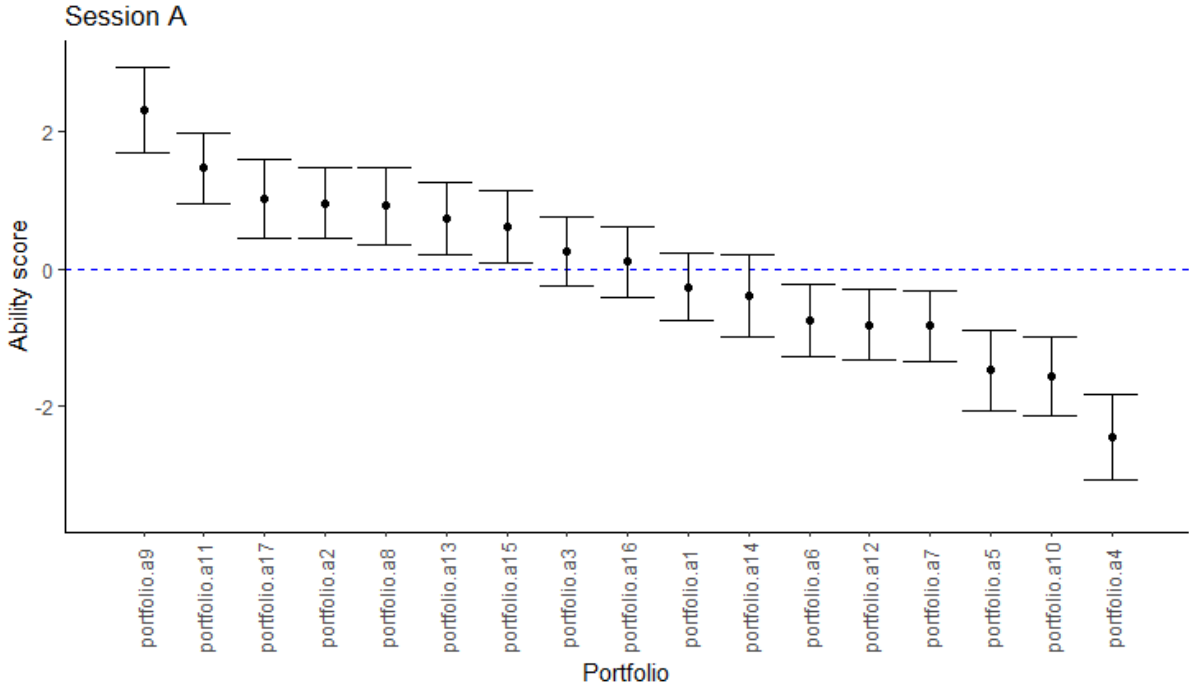
# Chaining



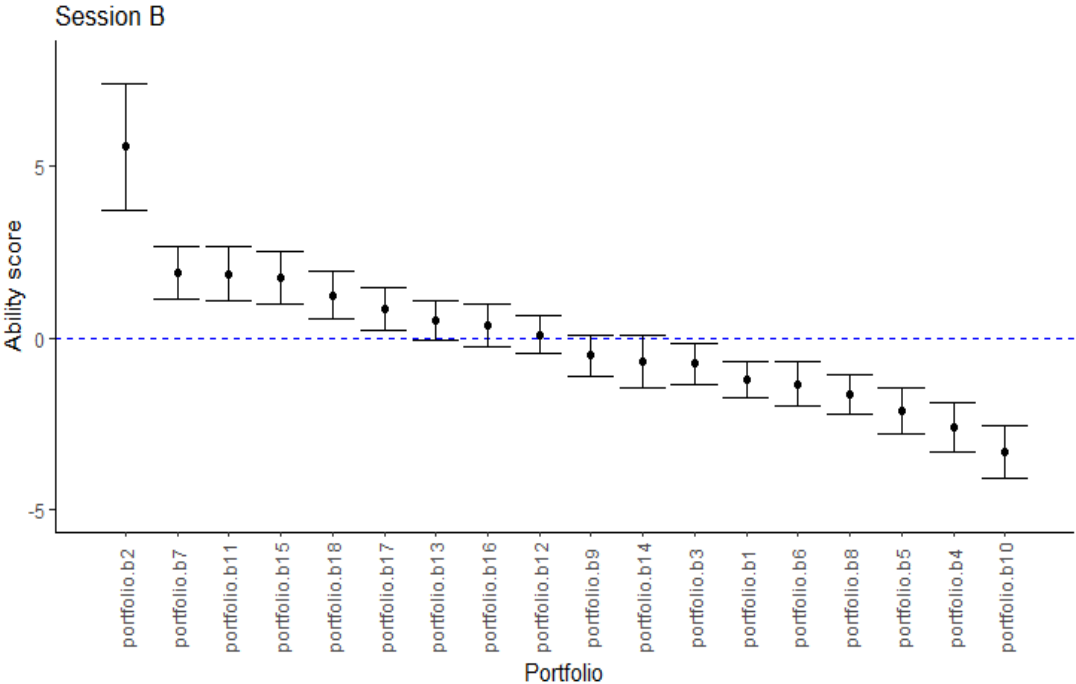
# Chaining



# Initial ranks

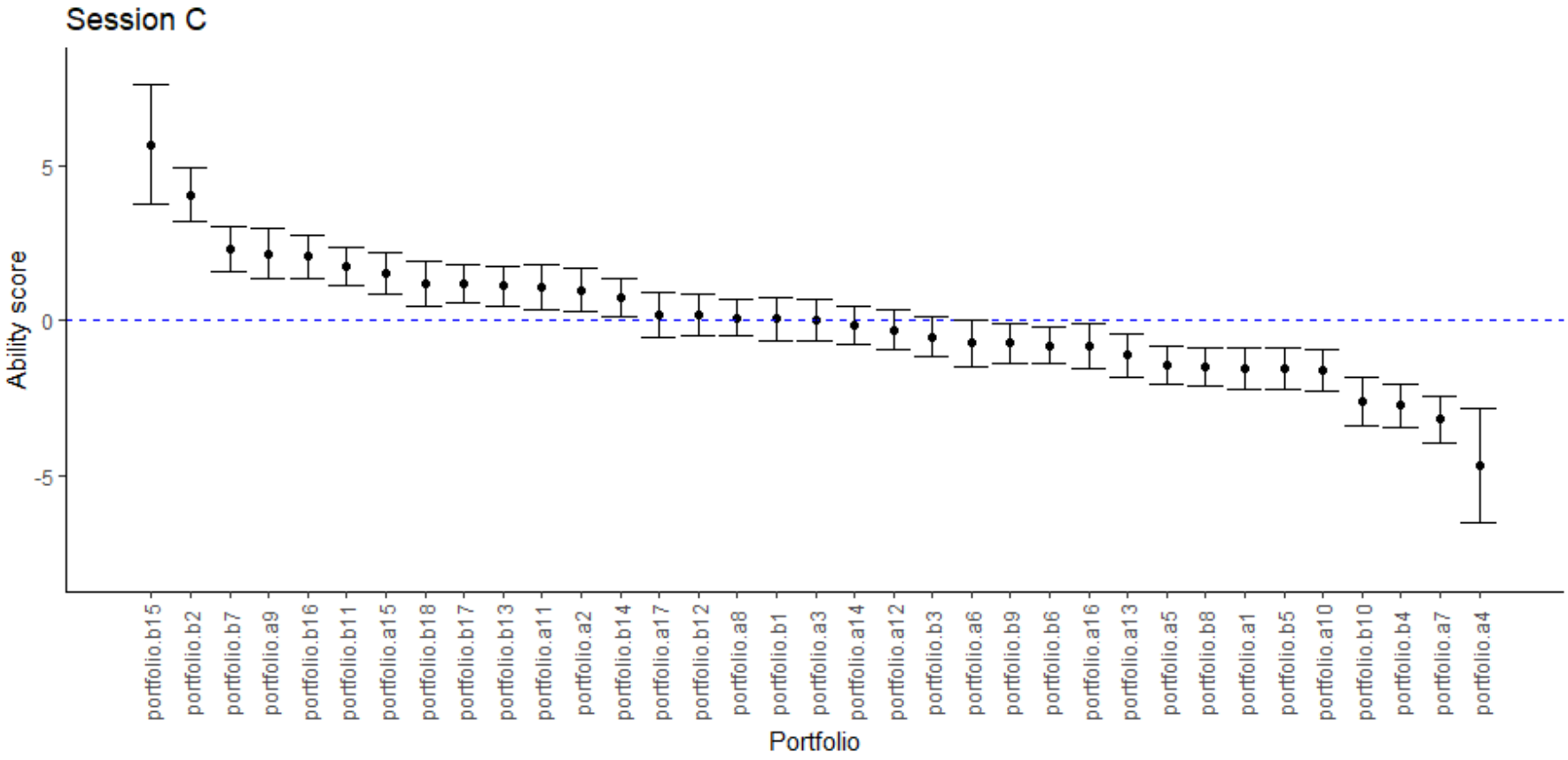


SSR: 0.8004



SSR: 0.8594

# Initial ranks



SSR: 0.8393

# Methodology

**Step 1** Select portfolios from rank A

**Step 2** Select portfolios from rank B

**Step 3** Select judges to make comparisons

**Step 4** Compute new rank A portfolios using BTL model.  
Rank B portfolio parameter values are fixed.

**Step 5** Recompute original rank A but fix the portfolio parameter values computed in Step 4

**Step 6** Merge recomputed rank A with original rank B

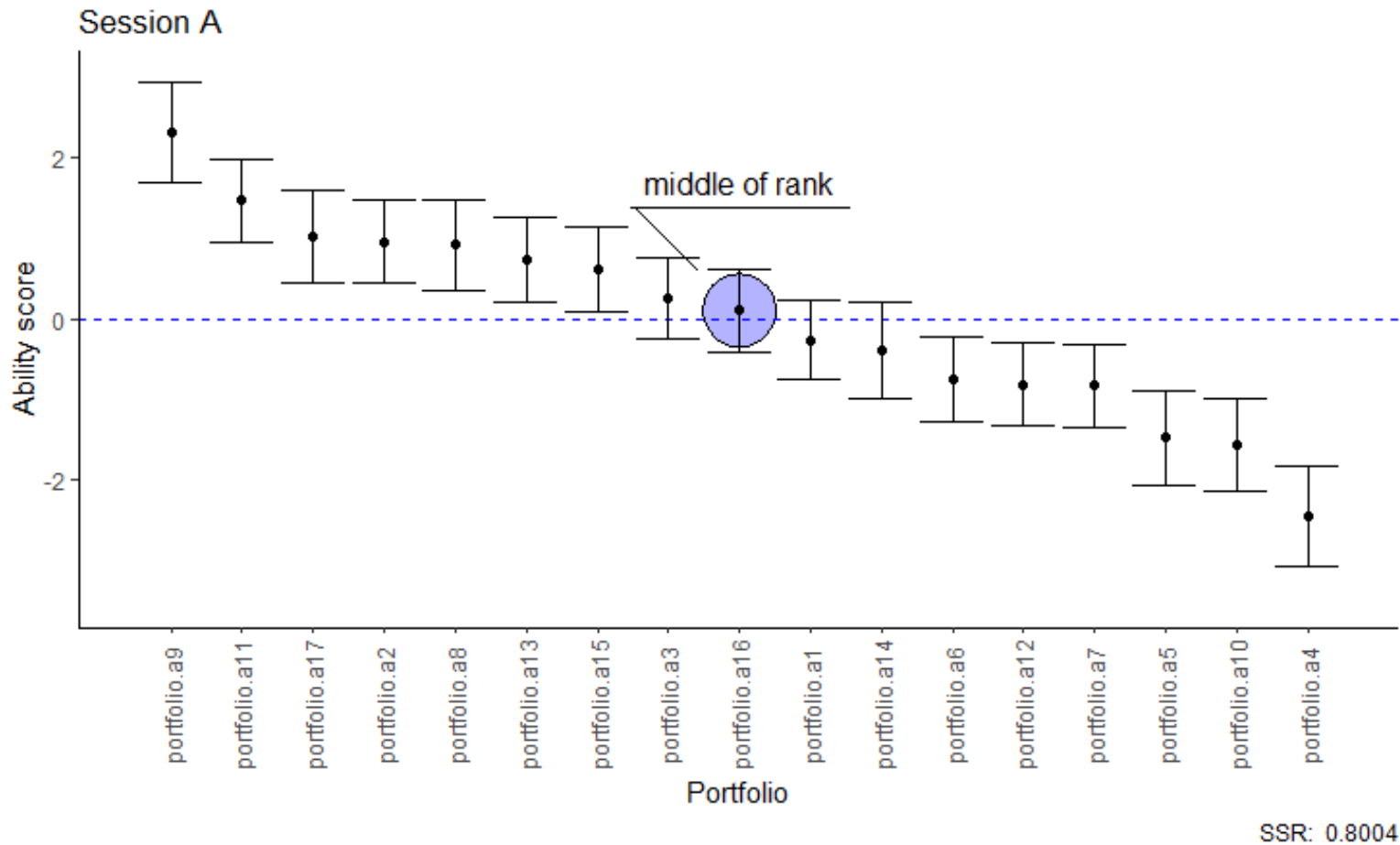
**Step 7** Compare merged ranks from Step 6 with original rank C (correlation analysis)

# Judge selection

Judge	Session A misfit	Session B misfit	Absolute difference between 1 and average misfit
judge.2	0.797566	1.253231	0.025398
judge.12	1.045002	1.093468	0.069235
judge.7	1.076206	1.112101	0.094153
judge.11	0.88119	1.342626	0.111908
judge.3	1.256328	1.006778	0.131552
judge.1	0.679718	0.819701	0.25029
judge.13	0.760399	1.794187	0.277293
judge.5	1.304007	0.138742	0.278625
judge.6	1.603682	0.972127	0.287904
judge.9	0.72778	0.676694	0.297763
judge.10	1.560905	1.419284	0.490095
judge.4	0.704051	0.18031	0.55782
judge.8	0.26236	0.473132	0.632254



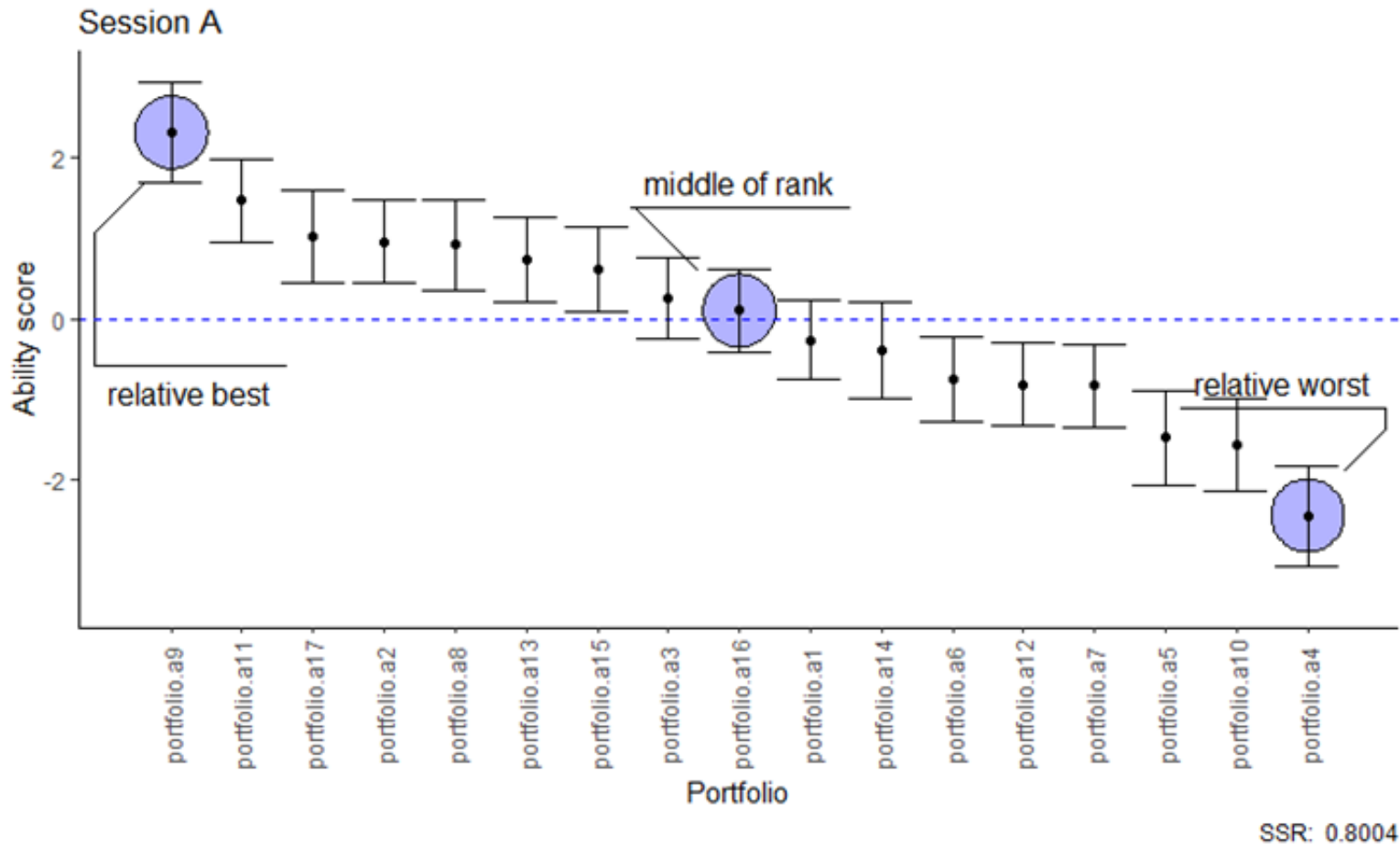
# Three approaches: Approach 1



Session B portfolio	Compared Session A portfolio by judge.2
portfolio.b5	portfolio.a16
portfolio.b18	portfolio.a16
portfolio.b6	portfolio.a16
portfolio.b12	portfolio.a16
portfolio.b13	portfolio.a16
portfolio.b1	portfolio.a16
portfolio.b14	portfolio.a16
portfolio.b15	portfolio.a16
portfolio.b10	portfolio.a16
portfolio.b3	portfolio.a16
portfolio.b17	portfolio.a16
portfolio.b2	portfolio.a16
portfolio.b11	portfolio.a16
portfolio.b16	portfolio.a16
portfolio.b4	portfolio.a16
portfolio.b7	portfolio.a16
portfolio.b9	portfolio.a16
portfolio.b8	portfolio.a16

**Correlation with original rank C**  
 $r = .79$  [95% CI; .62, .99],  $p < .001$   
 $\rho = .80$  [95% CI; .64, .90],  $p < .001$

# Three approaches: Approach 2



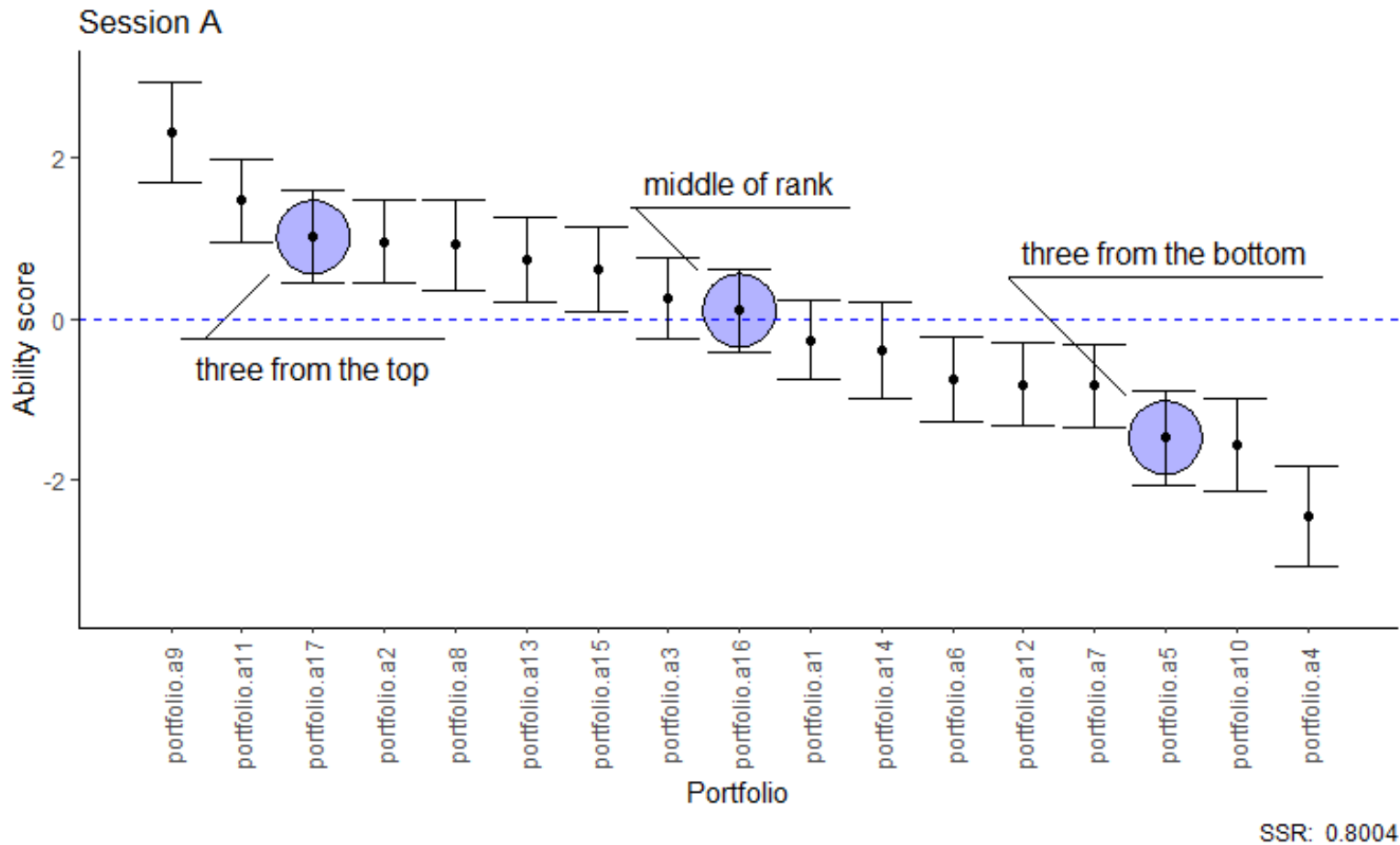
Session B portfolio	Compared Session A portfolio by judge.12	Compared Session A portfolio by judge.7	Compared Session A portfolio by judge.11
portfolio.b16	portfolio.a9	portfolio.a4	portfolio.a16
portfolio.b5	portfolio.a9	portfolio.a4	portfolio.a16
portfolio.b12	portfolio.a9	portfolio.a4	portfolio.a16
portfolio.b15	portfolio.a9	portfolio.a4	portfolio.a16
portfolio.b9	portfolio.a16	portfolio.a9	portfolio.a4
portfolio.b17	portfolio.a16	portfolio.a9	portfolio.a4
portfolio.b6	portfolio.a16	portfolio.a9	portfolio.a4
portfolio.b4	portfolio.a16	portfolio.a9	portfolio.a4
portfolio.b2	portfolio.a4	portfolio.a16	portfolio.a9
portfolio.b7	portfolio.a4	portfolio.a16	portfolio.a9
portfolio.b18	portfolio.a4	portfolio.a16	portfolio.a9
portfolio.b10	portfolio.a4	portfolio.a16	portfolio.a9

**Correlation with original rank C**

$r = .48$  [95%; .18, .70],  $p = .003$

$\rho = .47$  [95% CI; .15, .70],  $p = .005$

# Three approaches: Approach 3



Session B portfolio	Compared Session A portfolio by judge.3	Compared Session A portfolio by judge.1	Compared Session A portfolio by judge.13
portfolio.b16	portfolio.a17	portfolio.a5	portfolio.a16
portfolio.b5	portfolio.a17	portfolio.a5	portfolio.a16
portfolio.b12	portfolio.a17	portfolio.a5	portfolio.a16
portfolio.b15	portfolio.a17	portfolio.a5	portfolio.a16
portfolio.b9	portfolio.a16	portfolio.a17	portfolio.a5
portfolio.b17	portfolio.a16	portfolio.a17	portfolio.a5
portfolio.b6	portfolio.a16	portfolio.a17	portfolio.a5
portfolio.b4	portfolio.a16	portfolio.a17	portfolio.a5
portfolio.b2	portfolio.a5	portfolio.a16	portfolio.a17
portfolio.b7	portfolio.a5	portfolio.a16	portfolio.a17
portfolio.b18	portfolio.a5	portfolio.a16	portfolio.a17
portfolio.b10	portfolio.a5	portfolio.a16	portfolio.a17

## Correlation with original rank C

$r = .79$  [95% CI; .62, .99],  $p < .001$

$\rho = .80$  [95% CI; .64, .90],  $p < .001$

# Implications and future work

## Implications

Ranks can now be merged and directly compared.

Experimental research and large-scale application now possible.

## Further analysis

How should reliability of rank merging be considered?

## Further analysis

We had a lot of judgements.

Do the number of judgements impact how well ranks merge?

## Future work

Large scale confirmatory study.

# References

Buckley, J., & Canty, D. (2022). Assessing performance: The technical challenges in developing and validating “ACJ-Steady State”. *PATT2022: PATT on the edge: Technology, innovation, and education*. Newfoundland & Labrador, Canada: Memorial University.

Thank you!

Jeff Buckley

Email: [Jeffrey.buckley@tus.ie](mailto:Jeffrey.buckley@tus.ie)

 [@jeffbuckley92](https://twitter.com/jeffbuckley92)

