


09/02/2017

### The method of Comparative Judgement for assessment and research

And data analysis in R

San Verhavert  
Sven De Maeyer  
Vincent Donche



[www.d-pac.be](http://www.d-pac.be)    [san.verhavert@uantwerpen.be](mailto:san.verhavert@uantwerpen.be)

---

---

---

---

---

---

---

---

2

### The method of Comparative Judgement for assessment and research


What?

- What is D-PAC?
- What is Comparative Judgement (CJ)?

How?

- Analysis in R
- Applications

- Research in D-PAC: in short



---

---

---

---

---

---

---

---

3

### What is D-PAC NOT!





---

---

---

---

---

---

---


---

4

### What is D-PAC?

- Digital Platform for the Assessment of Competences

1. Creating awareness
2. Developing a tool
3. Providing feedback



---

---

---

---

---


---

---

---

5

- What is D-PAC?
- What is Comparative Judgement (CJ)?
  - Terminology
  - Theoretical background
- Analysis in R
- Applications
- Research in D-PAC: in short



---

---

---

---

---

---


---

---

6

### Terminology

- Comparative Judgement or CJ
- Assessment
- Competence
- Assessor, judge: individual who assesses
- Assessee: individual who is assessed
- Representation [of competence]: product that is assessed



---

---

---

---

---

---

---

---


7

### Comparative Judgement (CJ)

- How do we assess?
- Assessment is difficult!
  - Consistent: over time, assessors, ...
- Making an absolute judgement is difficult and impossible (Laming, 1990)
- Every judgement is a comparison (Laming, 2003)

Rate the presenter:

7/20	20
16	9?
12	



---

---

---

---

---

---

---

---

---

---

8


### Comparative Judgement

- Every judgement is a comparison (Laming, 2003)
- Thurstone (1927): The Law of Comparative Judgement

$$X_{AB} = v_A - v_B$$

$$p(A > B | v_a, v_b, \sigma_{ab}^2) = p(x_{AB} = 1 | v_a, v_b, \sigma_{ab}^2)$$

$$= \frac{1}{\sigma_{ij}\sqrt{2\pi}} \int_0^\infty \exp\left\{-\frac{(X - \mu)^2}{2\sigma^2}\right\} dt$$



---

---

---

---

---

---

---

---

---

---


9

### Comparative Judgement

- Every judgement is a comparison (Laming, 2003)
- Thurstone (1927): The Law of Comparative Judgement
- Bradley-Terry-Luce model ( BTL model) (Bradley & Terry, 1952; Luce, 1959)
- Rasch model (Rasch, 1960)

$$p(A > B | v_a, v_b) = p(x_{AB} = 1 | v_a, v_b) = \frac{\exp(S_A - S_B)}{1 + \exp(S_A - S_B)}$$

$$p(x_{vj} = 1 | \alpha_j, \tau_v) = \frac{\exp(\alpha_j - \tau_v)}{1 + \exp(\alpha_j - \tau_v)}$$



(Andrich, 1978)

---

---

---

---

---

---

---


---

---

---

### Comparative Judgement

- Writing and IQ testing: Thurstone, 1925; 1927
- Language assessment: Pollitt & Murray 1995
- Alternative assessment method (Pollitt, 2004; 2009)
- Facilitating evaluation of open-ended tasks (Jones, Swan & Pollitt, 2015)
- Inter-rater reliability  $r = 0,86$  (Jones & Inglis, 2015)
- Efficiency



---

---

---

---


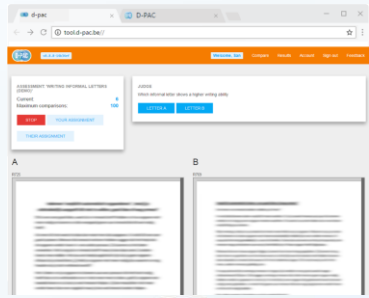
---

---

---

---

### CJ now



---

---

---

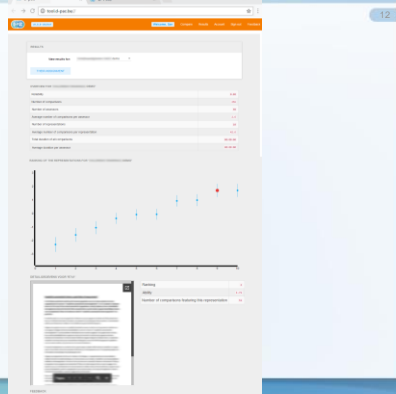
---

---

---

---

---



---

---

---

---

---


---

---

---

13

- What is D-PAC?
- What is Comparative Judgement (CJ)?
- Analysis in R
  - The data structure
  - Estimating the BTL model in R
  - A simulation and estimation results
- Applications
- Research in D-PAC: in short




---

---

---

---

---

---

---

---

---

---

---

14

### Analysis in R

comparison	assessment	assessor	representation A	representation B	selected representation	selected at	completed	Select base duration	Select base SD
3167	Papers	dpacl18@mevls.be	Paper 1	Paper 12	Paper 12	10/12/2015 13:07	1	292	5
3169	Papers	dpacl18@mevls.be	Paper 13	Paper 15	Paper 13	10/12/2015 13:15	1	178	6
3173	Papers	dpacl18@mevls.be	Paper 22	Paper 8	Paper 22				7
3177	Papers	dpacl18@mevls.be	Paper 20	Paper 16	Paper 20				7
3179	Papers	dpacl18@mevls.be	Paper 5	Paper 9	Paper 5				7
3183	Papers	dpacl18@mevls.be	Paper 7	Paper 10	Paper 7				7
3184	Papers	dpacl18@mevls.be	Paper 3	Paper 10	NA				7
3232	Papers	dpacl39@mevls.be	Paper 12	Paper 6	NA				7
3185	Papers	dpacl18@mevls.be	Paper 17	Paper 21	Paper 17				7
3188	Papers	dpacl39@mevls.be	Paper 2	Paper 14	Paper 2				7
3188	Papers	dpacl39@mevls.be	Paper 11	Paper 1	Paper 11				7
3187	Papers	dpacl18@mevls.be	Paper 6	Paper 20	Paper 6				7
3190	Papers	dpacl39@mevls.be	Paper 16	Paper 14	Paper 16				7
3198	Papers	dpacl39@mevls.be	Paper 7	Paper 20	Paper 7				7
3199	Papers	dpacl39@mevls.be	Paper 17	Paper 6	NA				7
3020	Papers	dpacl39@mevls.be	Paper 4	Paper 23	Paper 4				7
3191	Papers	dpacl39@mevls.be	Paper 12	Paper 13	Paper 12				7
3192	Papers	dpacl39@mevls.be	Paper 8	Paper 13	Paper 8				7
3194	Papers	dpacl39@mevls.be	Paper 11	Paper 13	Paper 11				7
3195	Papers	dpacl39@mevls.be	Paper 9	Paper 17	Paper 9	10/10/2015 16:35	1	293	7
3199	Papers	dpacl39@mevls.be	Paper 5	Paper 3	Paper 5	10/10/2015 16:38	1	184	6

comparison number  
assessment name  
assessor e-mail address  
representation A  
representation B  
selected representation  
when selected  
comparison completed  
selection duration  
selection perceived difficulty  
...

---

---

---

---

---

---

---

---

---

---

---

15

```


52 ## calculate score ##
53 #list
54 score <- numeric()
55
56 for ( i in 1: length( data_int_raw[representation.A] ) )
57 {
58   if ( data_int_raw[representation.A][i] == data_int_raw[selected.representation(i)] )
59     score <- append( score, 1 )
60   else if ( data_int_raw[representation.B][i] == data_int_raw[selected.representation(i)] )
61     score <- append( score, 0 )
62 } else
63   score <- append( score, na )
64 }
65 score <- append( score, na )
66 }
67 }
68
69 rm(i)
70
71 data_int <- data.frame( assessment = data_int_raw[assessment],
72                       assessor = data_int_raw[assessor],
73                       script1 = data_int_raw[representation.A],
74                       script2 = data_int_raw[representation.B],
75                       score = score )
76
77 rm( score, data_int_raw )
78
79

```

```

107 ## use "0@pac.be" off assessor name
108 data_ext[assessor] <- sub(data_ext[assessor], pattern = "0@pac.be",
109                          fixed = TRUE, replacement = "")
110

```




---

---

---

---

---

---

---

---

---

---

---



19

## Estimating Bradley-Terry-Luce Model in R

- Our own function
  - CML
- btm function from sirt (Robitzsch, 2016)
  - Minorization-Maximization procedure (Hunter, 2004)

D-PAC

---

---

---

---

---

---

---

---

20

## Estimating Bradley-Terry-Luce Model in R

- Our own function
  - CML
- btm function from sirt (Robitzsch, 2016)
- BTm function from BradleyTerry2 (Turner & Firth, 2012)
  - glm function
  - Least Squares Maximum Likelihood
- vglm and Brat functions from VGAM (Yee, 2010)
  - glm like function
  - Least Squares Maximum Likelihood

D-PAC

---

---

---

---

---

---

---

---

21

## vglm from VGAM

```
vglm( Brat( DataMat ) ~ 1,
      brat( refgp = refCatNo, refvalue = 1 ),
      trace = T )
```

	Paper 1	Paper 2	Paper 3	Paper 4
Paper 1	NA	5	1	0
Paper 2	2	NA	8	2
Paper 3	0	2	NA	10
Paper 4	10	8	0	NA

D-PAC

---

---

---

---

---

---

---


---

22

### Estimating Bradley-Terry-Luce Model in R

- Our own function
  - CML
- btm function from sirt (Robitzsch, 2016)
- BTm function from BradleyTerry2 (Turner & Firth, 2012)
  - glm function
  - Least Squares Maximum Likelihood
- vglm and Brat functions from VGAM (Yee, 2010)
  - glm like function
  - Least Squares Maximum Likelihood

lbtPC.fit function in prefmod, btmodel function in psychotools, ...




---

---

---

---

---

---

---


---

23

### Estimating Bradley-Terry-Luce Model in R

- Our own function
  - CML
- btm function from sirt (Robitzsch, 2016)
- BTm function from BradleyTerry2 (Turner & Firth, 2012)
  - Least Squares Maximum Likelihood
- vglm and brat functions from VGAM (Yee, 2010)
  - Least Squares Maximum Likelihood

Bias




---

---

---

---

---

---

---

---


24

### Estimating Bradley-Terry-Luce Model in R

- Our own function
  - $\epsilon$  bias correction
- btm function from sirt (Robitzsch, 2016)
  - $\epsilon$  bias correction
- BTm function from BradleyTerry2 (Turner & Firth, 2012)
  - bias reduction by function modification
- vglm and brat functions from VGAM (Yee, 2010)

Bias  
Unbound likelihood  
Scale '0'

(Bertoli Barsotti & Punzo, 2012)




---

---

---

---

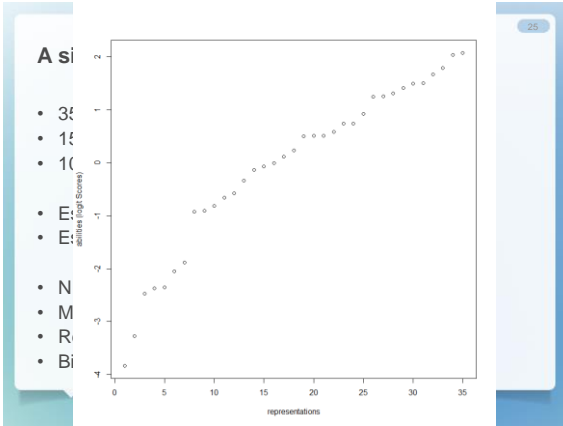
---

---

---

---






---

---

---

---

---

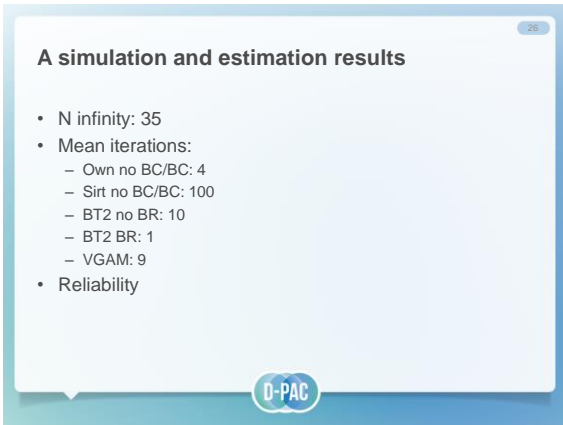
---

---

---

---

---




---

---

---

---

---

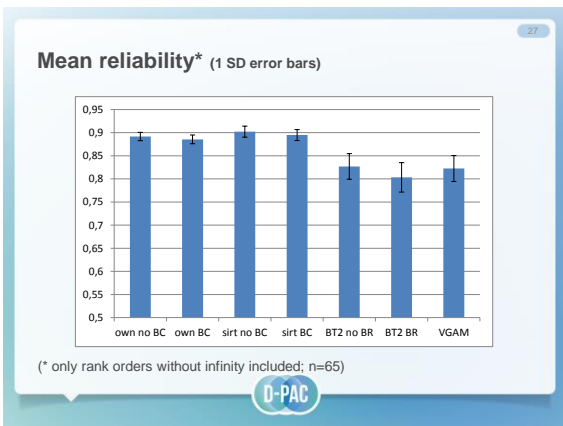
---

---

---

---

---




---

---

---

---

---

---

---

---

---

---

28

### A simulation and estimation results

- N infinity: 35
- Mean iterations:
  - Own no BC/BC: 4
  - Sirt no BC/BC: 100
  - BT2 no BR: 10
  - BT2 BR: 1
  - VGAM: 9
- Reliability
- Bias

**D-PAC**

---

---

---

---

---

---

---

---

29

### Mean bias (1 SD error bars)

Method	Mean Bias (approx.)	SE
own no BC	0.4	0.52
own BC	0.4	0.52
sirt no BC	0.6	0.54
sirt BC	0.4	0.54
BT2 no BR	0.8	0.69
BT2 BR	0.7	0.69
VGAM	0.75	?

**D-PAC**

---

---

---

---

---

---

---

---

30

### A simulation and estimation results

- Conclusion
  - Our own function
  - btm from sirt with  $\epsilon$  bias correction

**D-PAC**

---

---

---

---

---

---

---

---

31

- What is D-PAC?
- What is Comparative Judgement (CJ)?
- Analysis in R
- **Applications**
- Research in D-PAC: in short

D-PAC

---

---

---

---

---

---

---

---

32

### Applications

Within the D-PAC project

**Assessments**

- # writing skills
- Reporting and reasoning
- ER models
- Moodboards

**Jury's**

- Selection based on rank order
- Consultation/meeting based on rank order
- Accountability (FB)

**Research**  
26 try-outs

**Assessment tool**  
10 try-outs

**Learning tool**  
11 try-outs

**Professionalization**  
1 try-out

**Data collection and Scale development:**

- Input PhD's
- # writing skills
- Reading skills
- Visual skills
- Mathematical problem solving
- Speech (audiologist)
- Number estimation
- Delphi study

**Peer assessment**

- Learning effect of comparing
- Learning effect of FB

(Maarten Goossens)

D-PAC

---

---

---

---

---

---

---

---

33

- What is D-PAC?
- What is Comparative Judgement (CJ)?
- Analysis in R
- Applications
- **Research in D-PAC: in short**

D-PAC

---

---

---

---

---

---


---

---

34

### Research in D-PAC: in short

1. CJ method: reliability, efficiency and validity
2. Feedback to assessees
3. Feedback to organizations
4. Feedback to assessors
5. Design research: userfriendly?



---

---

---

---

---

---

---

---

### Thank you!

San Verhavert  
san.verhavert@uantwerpen.be



---

---

---

---

---

---

---

---



*"Because human judgements are comparisons"*

www.d-pac.be



---

---

---

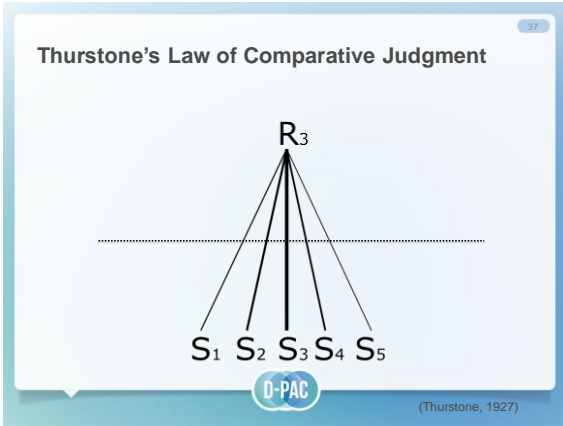
---

---

---

---

---



---

---

---

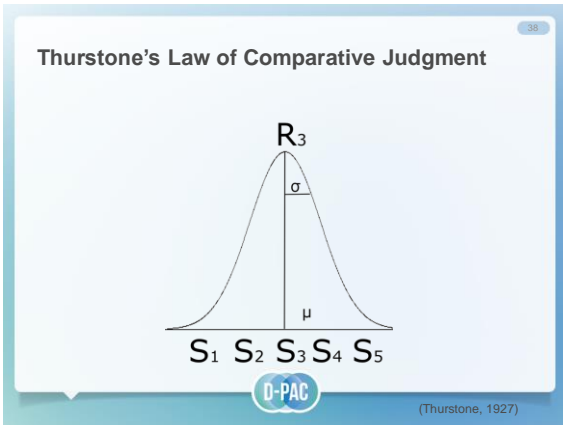
---

---

---

---

---



---

---

---

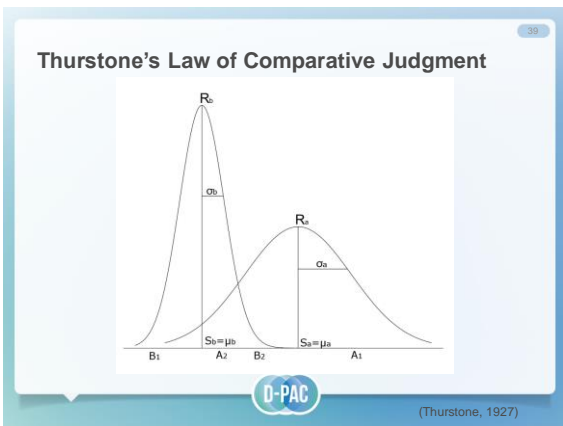
---

---

---

---

---



---

---

---

---

---

---

---

---

40

### Thurstone's Law of Comparative Judgment

$p(A > B | v_a, v_b, \sigma_{ab}^2) = p(x_{AB} = 1 | v_a, v_b, \sigma_{ab}^2)$   
 $= \frac{1}{\sigma_{ij}\sqrt{2\pi}} \int_0^\infty \exp\left\{-\frac{(X-\mu)^2}{2\sigma^2}\right\} dt$

$X_{AB} = \frac{v_A - v_B}{\sigma}$   
 $X_{AB} = v_A - v_B$

D-PAC

(Thurstone, 1927)

---

---

---

---

---

---

---

---

41

### Bradley-Terry-Luce Model (BTL)

$p(A > B | v_a, v_b) = p(x_{AB} = 1 | v_a, v_b) = \frac{\exp(S_A - S_B)}{1 + \exp(S_A - S_B)}$   
 $p(x_{vj} = 1 | \alpha_j, \tau_v) = \frac{\exp(\alpha_j - \tau_v)}{1 + \exp(\alpha_j - \tau_v)}$

$X_{AB} = v_A - v_B$

D-PAC

(Andrich, 1978)

---

---

---

---

---

---

---

---

42

### Difference between methods

Method	Value (approx.)
own v sirt no BC	0.3
own v sirt BC	0.15
own v BT2 no BC/BR	0.6
own v BT2 BC/BR	0.4
own v vgam	0.55
sirt v BT2 no BC/BR	0.55
sirt v BT2 BC/BR	0.4
sirt v vgam	0.55
BT2 v vgam	0.05

D-PAC

---

---

---

---

---

---

---

---