Generalizability Theory; Understanding Variance in Research

Casperhp12345@gmail.com

2269064h@student.gla.ac.uk

A Brief History

- In 1951 Coefficient Alpha is formalized (Cronbach, 1951)
- In 1963 Generalizability Theory is conceived (Cronbach, Rajaratnam & Gleser, 1963)
- In 1972 it is developed into a comprehensive framework (Cronbach, 1972)
- In 2004 Lee Cronbach publishes his thoughts on Coefficient Alpha for its 50th anniversary (Cronbach & Shavelson, 2004)
- Let's see a few of those thoughts...

A Brief History

- "I doubt whether coefficient alpha is the best way of judging the reliability of the instrument to which it is applied."
- "My 1951 article made no clear distinction between results for the sample and results for the population."
- "I no longer regard the alpha formula as the most appropriate way to examine most data."

So why Gtheory, and what does it offer?

- An inclusion of multiple facets
- A consideration of interaction effects
- The ability to generate multiple specific generalizability coefficients
- The ability to estimate generalizability in similar, hypothetical experiments with different levels of facet occurence
- An application for crossed or nested designs

Walking through: G-study

- Package " gtheory"
- Long format data
- Each column a facet or score

Subject	block	question	score
1000	1	ANX1	1
1000	2	ANX1	2
1000	3	ANX1	1
1000	4	ANX1	3
1000	5	ANX1	3

```
func <- score ~ (1|Subject)+(1|question)+
(1|block)+
(1|Subject:question)+(1|Subject:block)+
(1|block:question)
```

```
g_stud_state <- gstudy(Gstate, formula = func)
g_stud_state
## $components
##              source         var percent n
## 1 Subject:question 0.109338184    15.7 1
## 2    Subject:block 0.083970644    12.0 1
## 3   block:question 0.002798998     0.4 1
## 4          Subject 0.083750955    12.0 1
## 5         question 0.135828015    19.5 1
## 6            block 0.001524688     0.2 1
## 7         Residual 0.280135471    40.2 1
##
## attr(,"class")
## [1] "gstudy" "list"
```

Walking through: D-study

- A function exists – "dstudy" – but it is temperamental
- Luckily the maths is easy: D-variance = G-variance/n, where n = number of facet occurrences

Walking through: D-study

	facets	gvar	percent	levels	dvar	Dpercent
1	Subject:question	0.109338184	15.7	20	5.466909e-03	4.71538620
2	Subject:block	0.083970644	12.0	5	1.679413e-02	14.48547999
3	block:question	0.002798998	0.4	100	2.798998e-05	0.02414226
4	Subject	0.083750955	12.0	1	8.375096e-02	72.23791085
5	question	0.135828015	19.5	20	6.791401e-03	5.85780306
6	block	0.001524688	0.2	5	3.049375e-04	0.26301848
7	Residual	0.280135471	40.2	100	2.801355e-03	2.41625915

```
occasions <- 5
questions <- 20


gtable <- g_stud_state$components %>%
  as.vector() %>%
  select(-n) %>%
  rename(facets = "source", gvar = "var") %>%
  mutate(levels = c(questions, occasions, questions*occasions, 1, questions,
occasions, questions*occasions)) %>%
  group_by(facets) %>%
  mutate(dvar = gvar/levels) %>% ungroup
```

Walking through: Facet distinctions

- Facet of differentiation: the object of measurement
- Fixed facet: we don’t want to generalize beyond the included number
- Random facet: we want to generalize to an infinite number of occurences
- Stratification facet: we don’t want to generalize between levels

Walking through: Coefficients

- $Relative\ G = E\rho^2 = \frac{\tau}{\tau + \delta} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2}$
- Where σ_p^2 is person-based variance and σ_δ^2 is the relative error variance, which includes interactions between facets, but not main effects. σ_δ^2 will change depending on which variance factors are included; for example, in a Person x Occasion x Item design, when calculating an overall G coefficient, $\sigma_\delta^2 = \sigma_{po}^2 + \sigma_{pi}^2 + \sigma_{poi}^2$

Walking through: Coefficients

- $Absolute\ G = \phi = \frac{\tau}{\tau + \Delta} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2}$
- Whilst similar to the above formula, Δ also includes main effects, and so for an overall G coefficient, $\sigma_\Delta^2 = \sigma_o^2 + \sigma_i^2 + \sigma_{oi}^2 + \sigma_{po}^2 + \sigma_{pi}^2 + \sigma_{poi}^2$

Big Caveat!

- The elements of previous formulae are constructed differently depending on how the facets are defined. The aforementioned ones would be if all facets were considered random
- See Brennan (2001) or Bloch and Norman (2012) for rules on how to calculate coefficients for different designs and variables
- For good measure, I'll show how I calculated mine...

$$Relative\ G = E\rho^2 = \frac{\tau}{\tau + \delta} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2}$$

$$Absolute\ G = \phi = \frac{\tau}{\tau + \Delta} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2}$$

- Items were considered fixed, Occasions were considered random
- $\tau = \sigma_p^2 = P + P{:}I$
- τ is created by adding together D-study variance components including the facet of differentiation (in this case Person) with all its fixed facet interactions excluding any that feature random facets (in this case the Person and Item interaction).

$$Relative\ G = Ep^2 = \frac{\tau}{\tau + \delta} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2}$$

$$Absolute\ G = \phi = \frac{\tau}{\tau + \Delta} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2}$$

- Items were considered fixed, Occasions were considered random
- $\delta = \sigma_\delta^2 = P{:}O + P{:}O{:}I$
- δ is created by adding together every D-study variance component including interaction between the facet of differentiation (Person) and any random facet (in this case Occasion).

$$Relative\ G = Ep^2 = \frac{\tau}{\tau + \delta} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2}$$

$$Absolute\ G = \phi = \frac{\tau}{\tau + \Delta} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2}$$

- Items were considered fixed, Occasions were considered random
- $\Delta = {\Delta^2}_\delta = O + P{:}O + O{:}I + P{:}O{:}I$
- Δ is created by adding together every D-study variance component which include random facets (Occasion).

Walking through: Coefficients

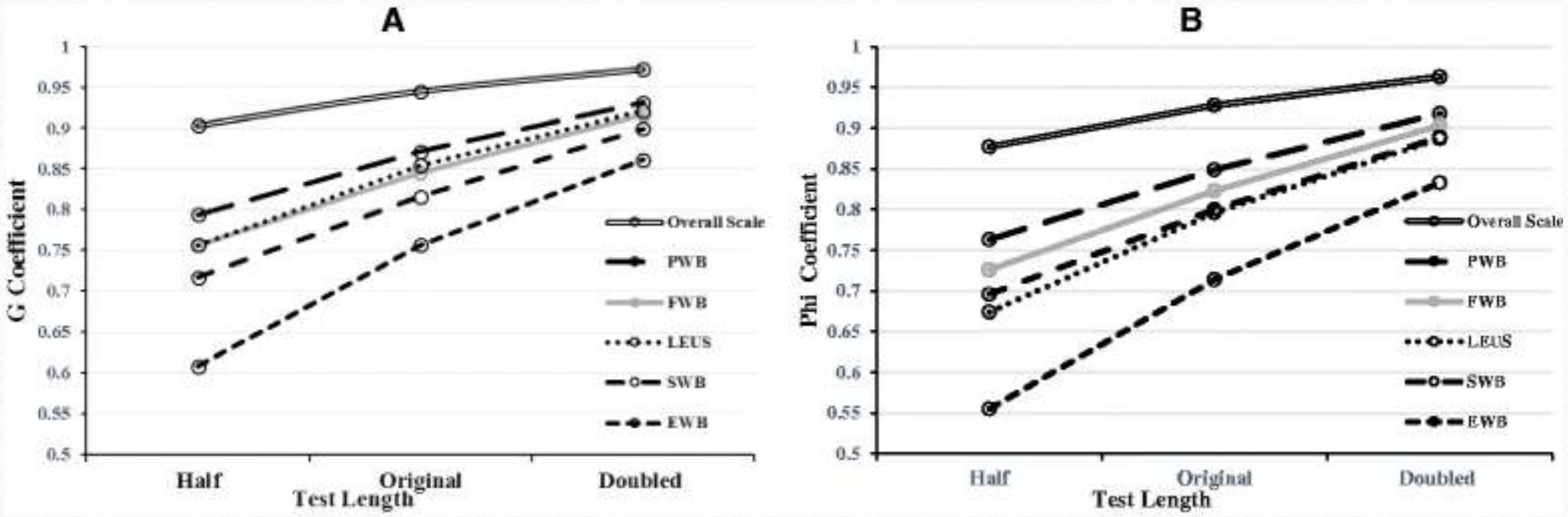
- TCI and SCI are calculated from G-study variances (Medvedev et al. 2017)
- $SCI = \dfrac{\sigma_s^2}{\sigma_s^2 + \sigma_t^2}$
- $TCI = \dfrac{\sigma_t^2}{\sigma_t^2 + \sigma_s^2}$
- Where $\sigma_t^2 = \sigma_p^2$ and $\sigma_s^2 = \sigma_{po}^2$
- Mathematically indistinguishable

A few examples: D-study, Big Five

Items	Extraversion				Agreeableness				Conscientiousness			
	Measurement Occasions				Measurement Occasions				Measurement Occasions			
	1	2	3	5	1	2	3	5	1	2	3	5
1	.45	.52	.55	.58	.32	.39	.42	.45	.31	.37	.40	.43
2	.61	.68	.70	.73	.47	.55	.59	.62	.47	.54	.57	.60
3	.69	.75	.78	.80	.56	.64	.68	.70	.56	.63	.66	.69
5	.77	.83	.85	.86	.67	.74	.77	.79	.67	.73	.76	.78
7	.81	.86	.88	.90	.73	.79	.82	.84	.73	.79	.81	.83
10	.85	.89	.91	.92	.78	.84	.86	.88	.78	.83	.85	.87
12	.86	.91	.92	.93	.80	.86	.88	.90	.80	.85	.87	.89
15	.88	.92	.93	.94	.82	.87	.90	.91	.83	.87	.89	.91

Arterberry, Martens, Cadigan and Rohrer (2014)

A few examples: FACT-Leu



Meng et al. (2017)

A few examples: Blood Pressure Reliability

Study 1

Table 1

Variance component estimates for person × replication design

Source	Variance Component Estimates: Systolic Blood Pressure	Variance Component Estimates: Diastolic Blood Pressure
Person (P)	125.33	64.21
Replication (R)	.89	.06
P×R	22.57	13.66

Table 2

G–coefficients using $\sigma^2(\Delta)$ as error variance for P×R design with replications taken in the laboratory on the same day*

Number of Replications (n′(r))	G*–Coefficients: Systolic Blood Pressure	G*–Coefficients: Diastolic Blood Pressure
1	.84	.82
2	.91	.90
3	.94	.93
4	.96	.95
5	.96	.96

Llabre et al. (1988)

A few examples: Blood Pressure Reliability

Study 2

Table 3

Variance component estimates for P×R:D design

Source	Variance Component Estimates: Systolic Blood Pressure	Diastolic Blood Pressure
Person (P)	83.41	36.49
Day (D)	.73	0[a]
Replication (R:D)	.30	0[a]
P×D	24.91	21.69
P×R:D	9.91	7.99

Table 4

G-coefficients using σ^2 (Δ) as error for P×R:D design with measures taken in the laboratory*

Number of Days n′(d)	G*-Coefficients: Systolic Blood Pressure: 1 Replication	Systolic Blood Pressure: 2 Replications	Diastolic Blood Pressure: 1 Replication	Diastolic Blood Pressure: 2 Replications
1	.70	.73	.55	.59
2	.82	.84	.71	.74
3	.87	.89	.79	.81
4	.90	.92	.83	.85
5	.92	.93	.86	.88

Number of Replications n′(r)	Systolic Blood Pressure Same Day	Diastolic Blood Pressure Same Day
1	.89	.82
2	.94	.90
3	.96	.93
4	.97	.95
5	.98	.96

Llabre et al. (1988)

A few examples: Blood Pressure Reliability

Study 3

Table 5

Variance component estimates for P×R design both at home and at work

	Variance Component Estimates			
	Home		Work	
Source	Systolic Blood Pressure	Diastolic Blood Pressure	Systolic Blood Pressure	Diastolic Blood Pressure
Persons (P)	143.51	49.33	150.07	57.62
Replications (R)	6.06	1.55	0[a]	2.92
P×R	228.84	111.41	166.82	80.29

Table 6

G-coefficients using $\sigma^2(\Delta)$ as error for P×R design both at home and at work*

	G*-Coefficients			
	Home		Work	
Number of Replications	Systolic Blood Pressure	Diastolic Blood Pressure	Systolic Blood Pressure	Diastolic Blood Pressure
1	.38	.30	.47	.41
2	.55	.47	.64	.58
3	.65	.57	.73	.68
4	.71	.64	.78	.73
5	.75	.69	.82	.78
6	.79	.72	.84	.81
10	.86	.81	.90	.87

Llabre et al. (1988)

A few examples: Blood Pressure Reliability

Study 4

Table 7

Variance component estimates for P×R:S design

Source	Variance Component Estimates: Systolic Blood Pressure	Variance Component Estimates: Diastolic Blood Pressure
Persons (P)	100.63	37.47
Setting (S)	17.59	4.77
Replications (R:S)	2.15	1.45
P×S	46.87	20.38
P×R:S	139.41	83.92

Table 8

G-Coefficients for P×R:S design*

Number of Settings	G*-Coefficients: Systolic Blood Pressure 1[a]	5	10	20	Diastolic Blood Pressure 1	5	10	20
1	.33	.52	.56	.61	.25	.47	.53	.56
2	.49	.68	.72	.74	.40	.64	.69	.72
3	.59	.76	.79	.81	.50	.73	.77	.79

[a]Replications/setting.

Llabre et al. (1988)

A few examples: Blood Pressure Reliability

Study 5

Table 9

Variance component estimates for P×R:I design

Source	Variance Component Estimates: Systolic Blood Pressure	Diastolic Blood Pressure
Persons (P)	210.36	109.14
Instrument (I)	19.71	.51
Replication (R:I)	0[a]	.20
P×I	34.05	29.50
P×R:I	20.74	14.18

Table 10

G-Coefficients using $\sigma^2(\Delta)$ as error for P×R:I design—all replications taken in the laboratory*

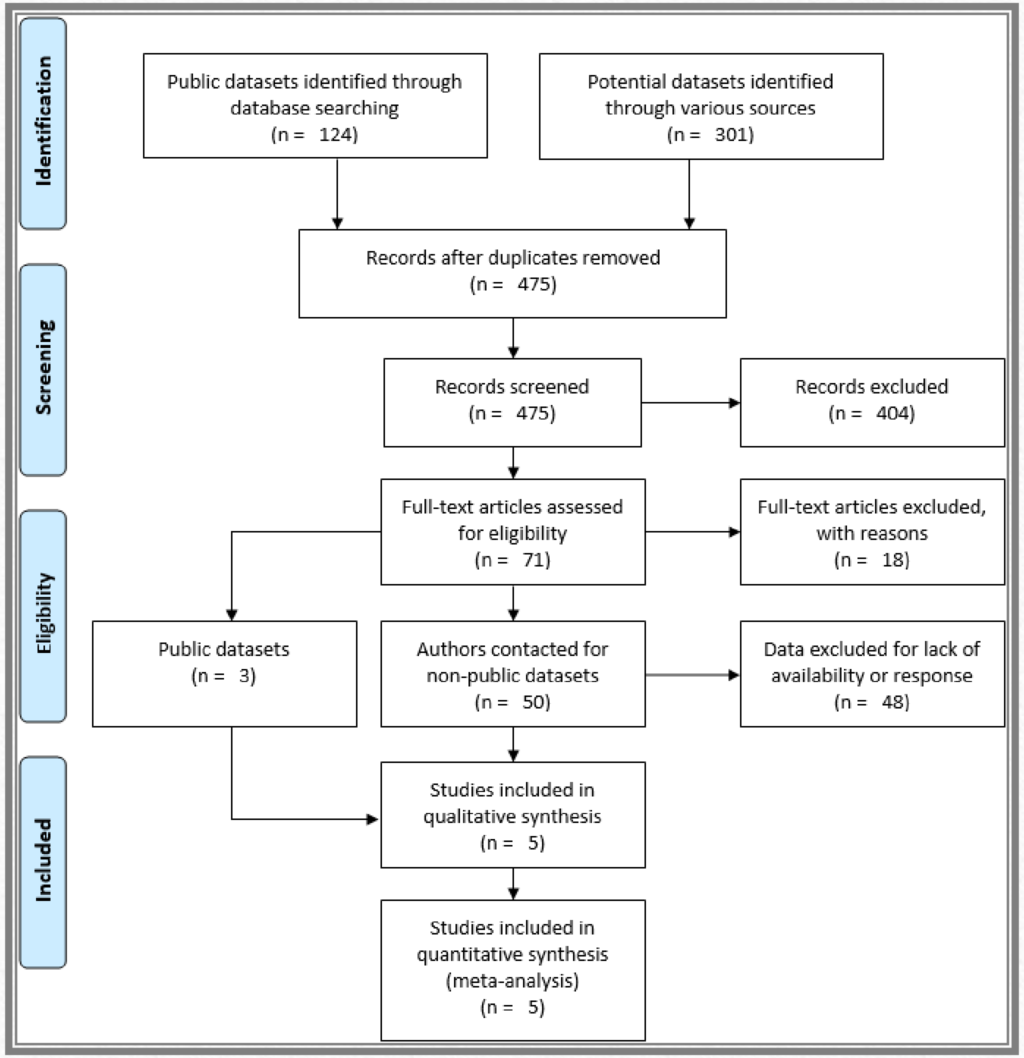
	G*-Coefficients									
	Systolic Blood Pressure					Diastolic Blood Pressure				
Number of Instruments	1[a]	2	3	4	5	1	2	3	4	5
1	.74	.77	.78	.78	.78	.71	.75	.76	.77	.77
2	.85	.87	.87	.88	.88	.83	.85	.86	.87	.87
3	.89	.91	.91	.92	.92	.88	.90	.90	.91	.91

Llabre et al. (1988)

Dissertation: Background

- Meta-analysis
- Secondary data
- Exploratory analysis
- State-Trait Anxiety Inventory
- An evaluation of full scale and individual items

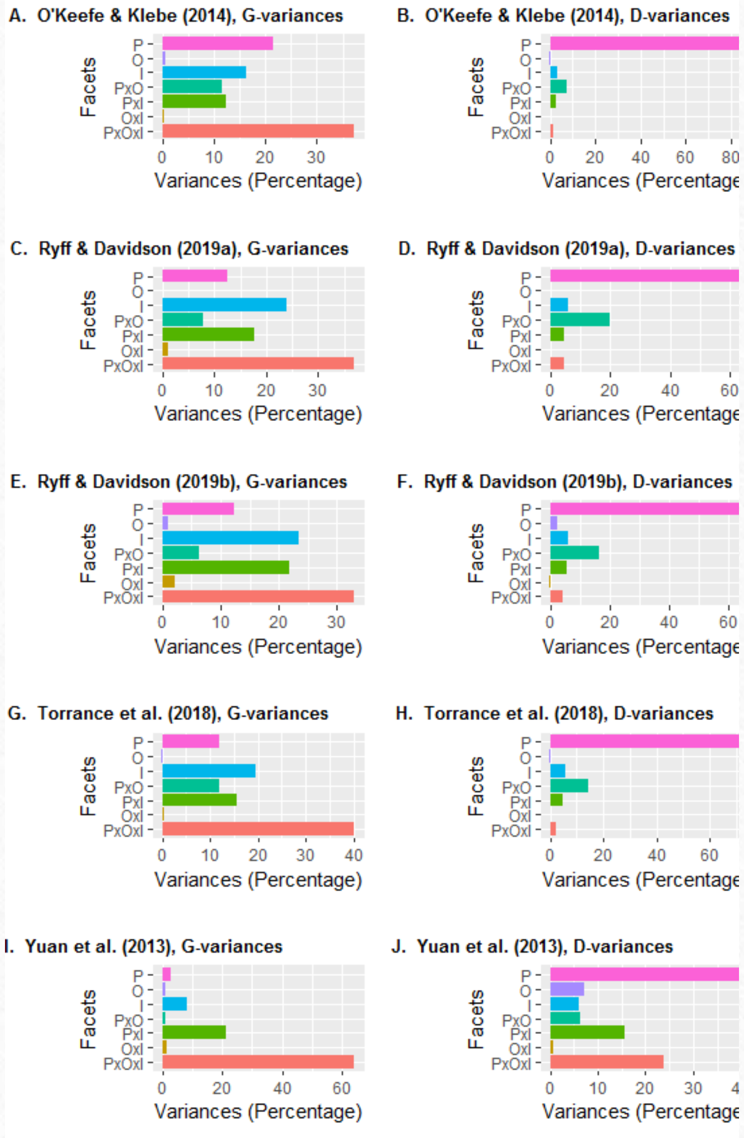
Dissertation: Background



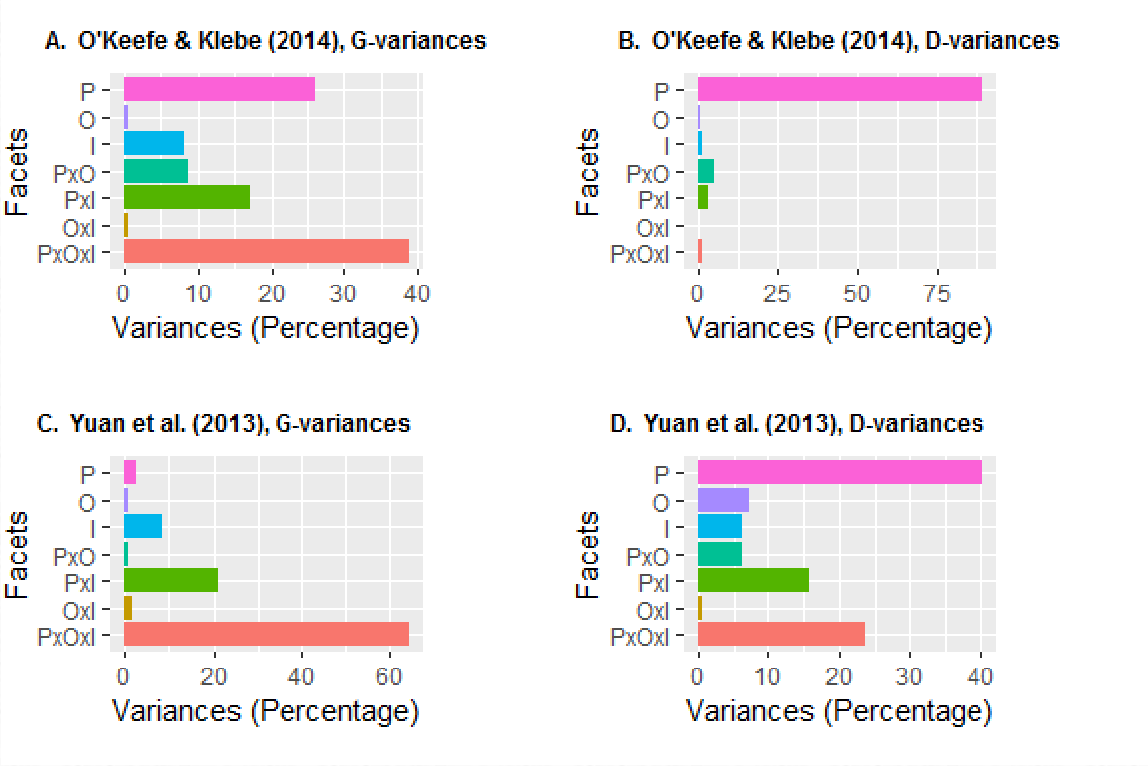
Dissertation: Coefficients

Study	Participants	Subscale	Alpha	ϕ	Ep^2	TCI
O'Keefe & Klebe (2014)	100	state	0.93	0.91	0.91	0.65
O'Keefe & Klebe (2014)	100	trait	0.93	0.94	0.94	0.95
Ryff & Davidson (2019a)	318	state	0.88	0.73	0.73	0.61
Ryff & Davidson (2019a)	318	trait	0.88	-	-	-
Ryff & Davidson (2019b)	122	state	0.87	0.75	0.77	0.66
Ryff & Davidson (2019b)	122	trait	0.9	-	-	-
Yuan et al. (2013)	52	state	0.89	0.6	0.65	0.76
Yuan et al. (2013)	52	trait	0.85	0.6	0.61	0.68
Torrance et al. (2018)	24	state	0.87	0.82	0.82	0.5
Torrance et al. (2018)	24	trait	-	-	-	-

Dissertation: State variance profiles



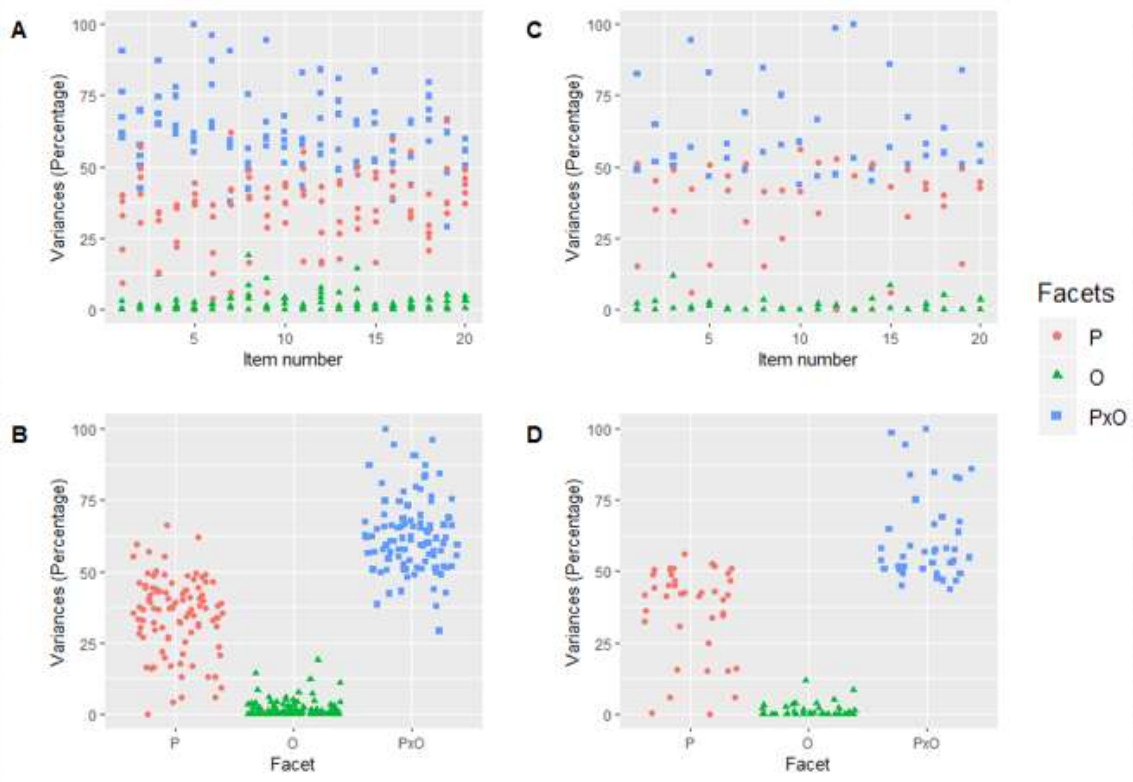
Dissertation: Trait variance profiles



Dissertation: Item-level variance profiles

Left Side:

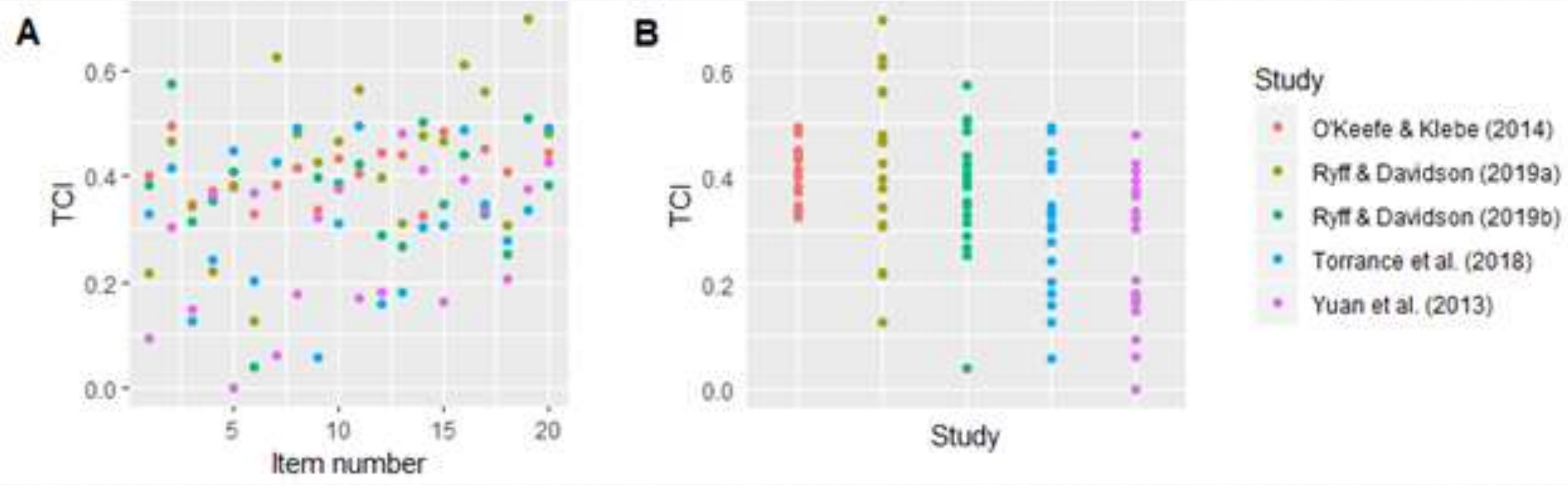
State subscale



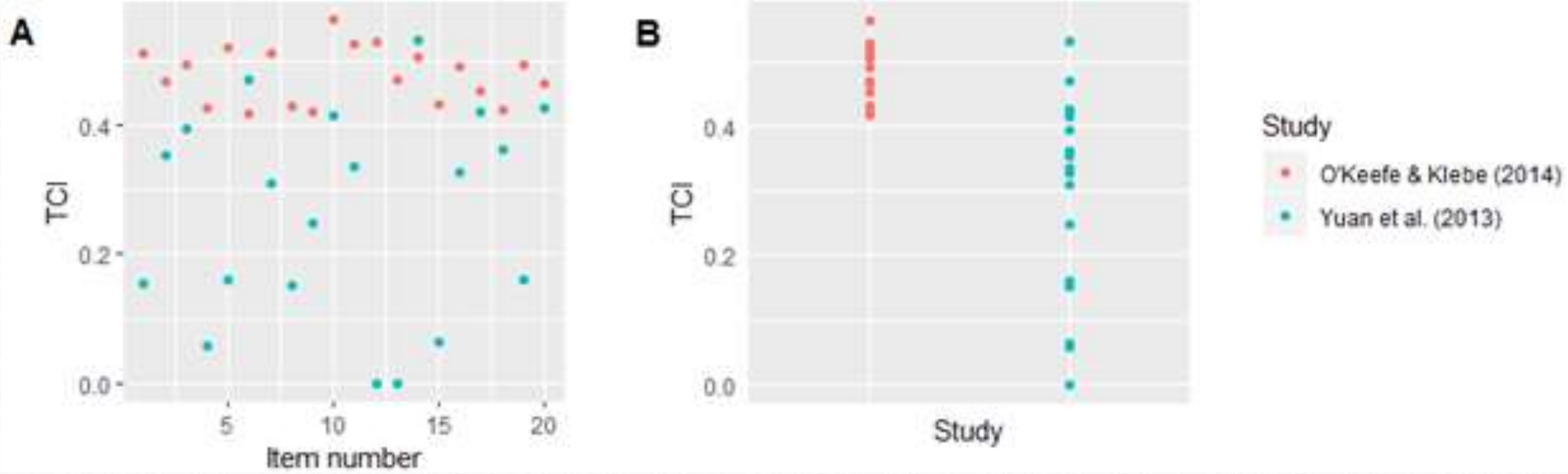
Right Side:

Trait subscale

Dissertation: State item-level TCI distributions



Dissertation: Trait item-level TCI distributions



Dissertation: Item-level TCI, descriptive statistics

Study	State TCI Mean	State TCI SD	State TCI Range	Trait TCI Mean	Trait TCI SD	Trait TCI Range
O'Keefe & Klebe (2014)	0.41	0.05	0.33 - 0.49	0.48	0.04	0.42 - 0.56
Ryff & Davidson (2019a)	0.43	0.15	0.13 - 0.7	-	-	-
Ryff & Davidson (2019b)	0.37	0.11	0.04 - 0.57	-	-	-
Torrance et al. (2018)	0.32	0.13	0.06 - 0.49	-	-	-
Yuan et al. (2013)	0.27	0.14	0 - 0.48	0.27	0.16	0 - 0.53

Discussion

- https://osf.io/29vjp/
- Human error; Gtheory is only as good as the facets we decide to include
- Current lack of guidelines, visualization techniques, rules of thumb
- State and trait separation is mathematically and conceptually rigid
- Analysis of individual items are noisy
- Steyer, Mayer, Geiser and Cole (2015)

What's next?

- Situated measures
- Collecting data through virtual agents
- Passive data collection
- Building individual user profiles
- A comparison of Latent State-Trait Theory and Gtheory
- Stepwise Gtheory

Conclusion

- Help improving Gtheory and alternative methods
- Do justice to your data; consider distributions, individual differences and generalizability within and outside your sample.
- Variance has meaning!

References

- Arterberry, B. J., Martens, M. P., Cadigan, J. M., & Rohrer, D. (2014). Application of Generalizability Theory to the Big Five Inventory. *Personality and Individual Differences, 69*, 98 - 103.
- Bloch, R., & Norman, G. R. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Medical Teacher, 34*(11), 960-992. Retrieved 10 13, 2019
- Brennan, R. L. (2001). *Generalizability theory*. Springer-Verlag.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297 - 334.
- Cronbach, L. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. Wiley.

- Cronbach, L. J., & Shavelson, R. J. (2004). My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educational and Psychological Measurement, 64*(3), 391 - 418.
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: a liberalization of reliability theory. *British Journal of Statistical Psychology, 16*(2), 137 - 163.
- Llabre, M., Ironson, G., Spitzer, S., Gellman, M., Weidler, D., & Schneiderman, N. (1988). How Many Blood Pressure Measurements are Enough?: An Application of Generalizability Theory to the Study of Blood Pressure Reliability. *Psychophysiology, 25*(1), 97 - 106.
- Medvedev, O. N., Krägeloh, C. U., Narayanan, A., & Siegert, R. J. (2017). Measuring Mindfulness: Applying Generalizability Theory to Distinguish between State and Trait. *Mindfulness, 8*(4), 1036 - 1046.
- Meng, Q., Yang, Z., Wu, Y., Xiao, Y., Gu, X., Zhang, M., . . . Li, X. (2017). Reliability analysis of the Chinese version of the Functional Assessment of Cancer Therapy – Leukemia (FACT-Leu) scale based on multivariate generalizability theory. *Health and quality of life outcomes, 15*(1), 93 - 112.
- Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A Theory of States and Traits-Revised. *Annual Review of Clinical Psychology, 11*(1), 71-94.