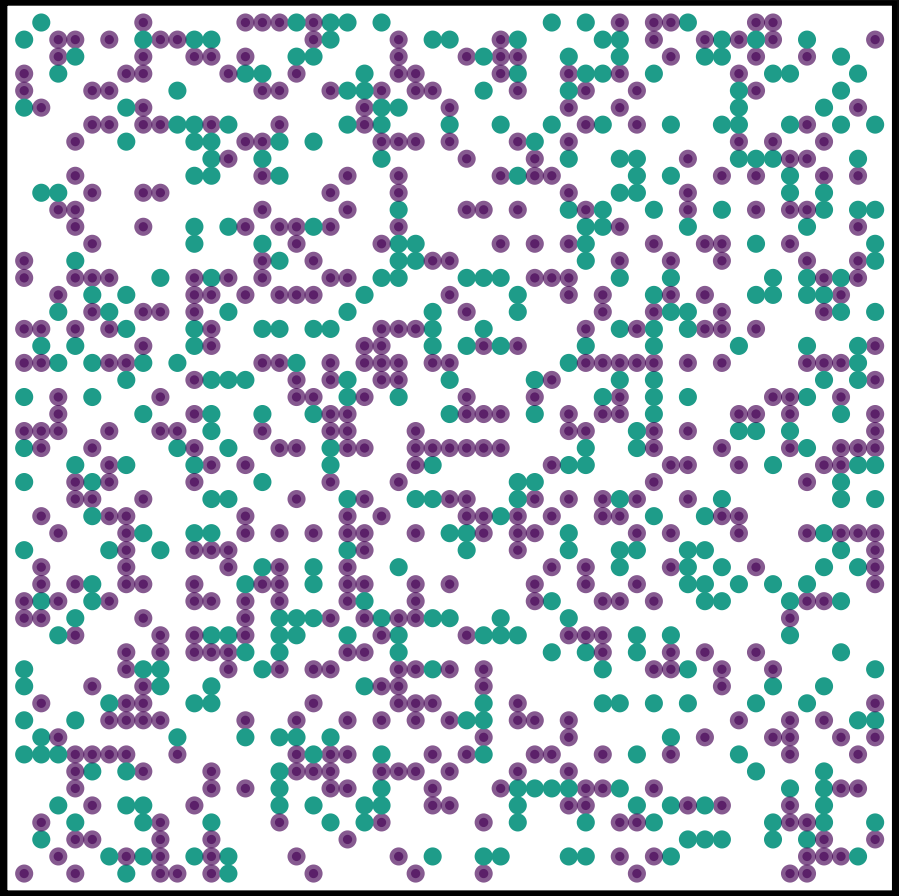
The Psychometrics of Interval responses

Matthias Kloft

DAGSTAT 2025, Berlin

Motivating Example

What is the percentage of purple dots?

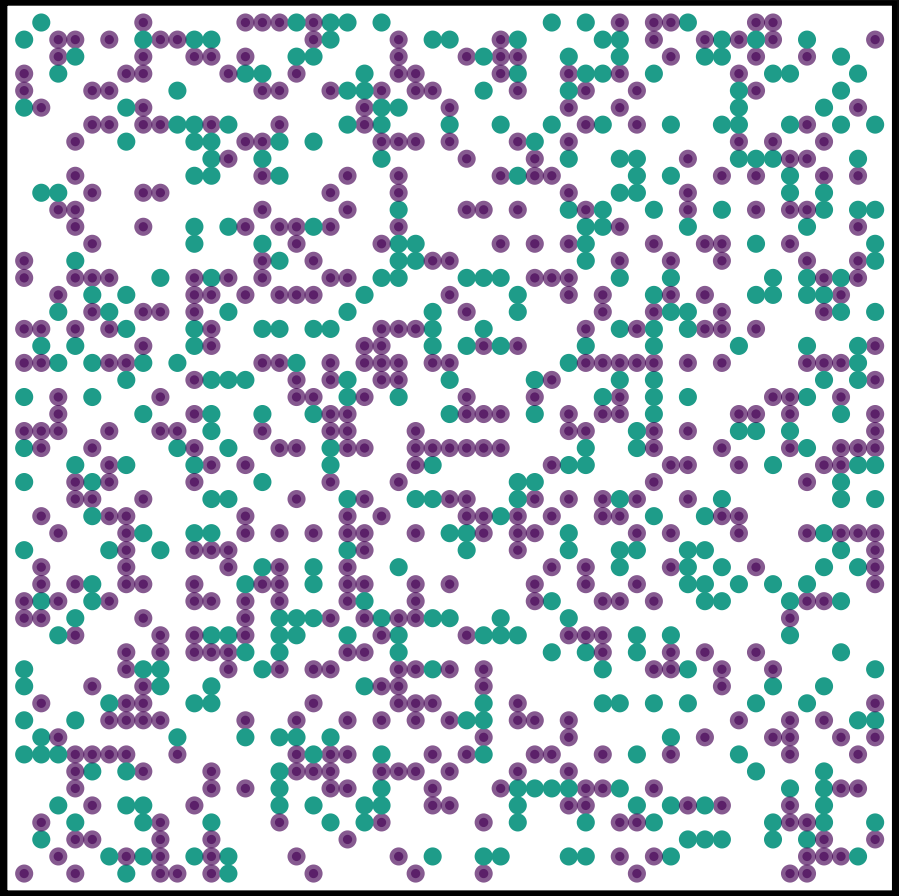


Think of a:

- Best guess
- Lower bound
- Upper bound

Motivating Example

What is the percentage of purple dots?



True percentage = 60%

Interval Responses

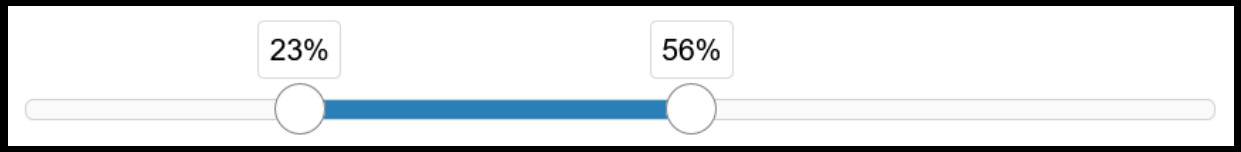
Single-range slider / visual analog scale (**VAS**)

Dual-range slider (**DRS**)

Interval Responses

Single-range slider / visual analog scale (**VAS**)

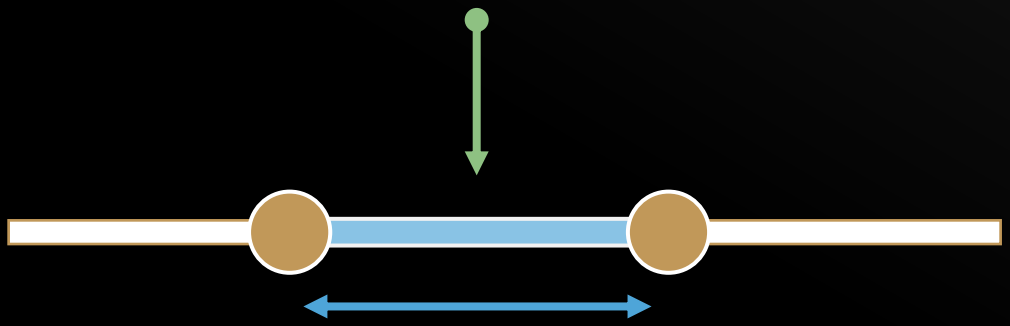
Dual-range slider (**DRS**)



noUISlider JavaScript range slider (Gersen, 2024)

Interval Responses: Descriptive Values

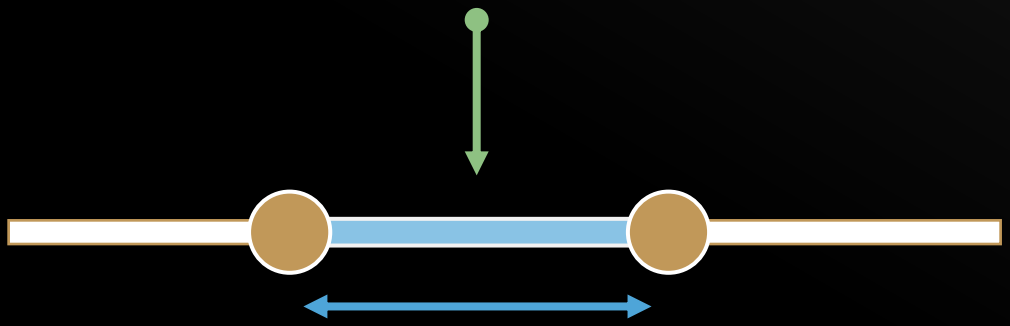
DRS **Location**: $\frac{y^{(L)}+y^{(U)}}{2}$



DRS **Width**: $y^{(U)} - y^{(L)}$

Interval Responses: Descriptive Values

DRS **Location**: $\frac{y^{(L)}+y^{(U)}}{2}$



DRS **Width**: $y^{(U)} - y^{(L)}$

There are better representations for modeling!

Interval Responses

Variability:

- Self-ratings, stimuli

Uncertainty:

- Influenced by expertise and confidence
- Estimation (e.g., forecasting)

Plausibility:

- No clear-cut true answer (e.g., verbal quantifiers like “seldom” or “likely”)
- Meta-uncertainty: plausible risks

Topics of the Talk

1. Convergent validity: DRS and VAS
2. Test-retest reliability of the DRS
3. Discriminant validity regarding the DRS Width

Convergent Validity: VAS & DRS Location

Kloft et al. (2023), Kloft et al. (2024)

Kloft et al., 2023: Cross-sectional Design

- 2 Response Formats: VAS & DRS
- 2 Extraversion Personality Scales, One per Response Format

Kloft et al., 2023: Cross-Sectional Design

- Modeling of extraversion scores via joint IRT models
  - **VAS**: Beta Response Model (Noel & Dauvier, 2007)
  - **DRS**: Dirichlet Dual Response Model
  - (Kloft et al., 2023)
  - **Correlation** of latent extraversion scores for VAS and DRS measures via Bayesian hierarchical model with MVN prior

Kloft et al., 2023

Correlation between latent extraversion score measured by VAS and DRS:

$\rho = .87$

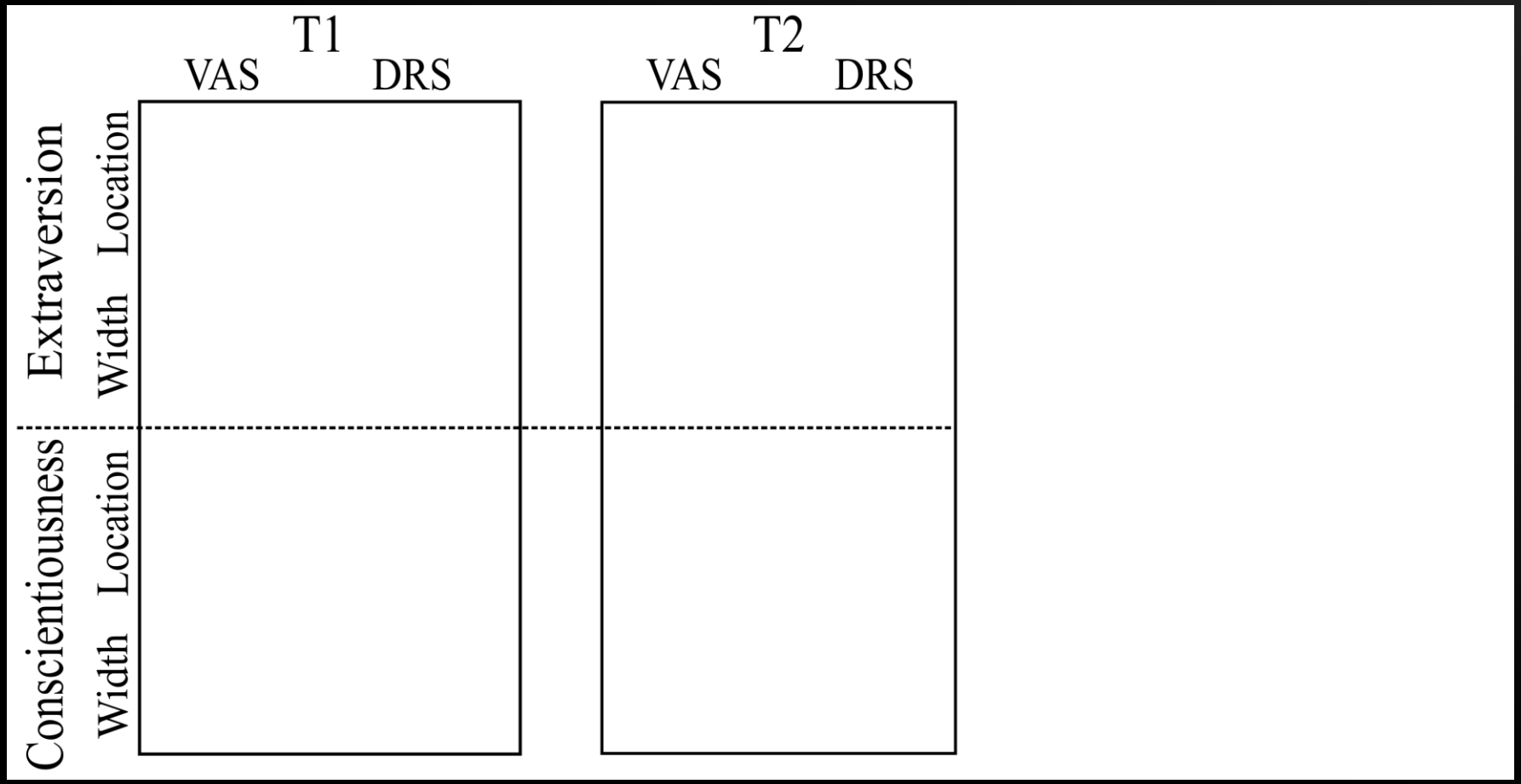
Kloft et al., 2024: MTMM Study Design

- 2 Time points (6-8 weeks apart)
- 2 Response Formats: VAS & DRS
- 2 Personality Scales: Extraversion & Conscientiousness person-descriptive adjectives
- Random split and assignment of items to response formats

Kloft et al., 2024: MTMM Study Design

- Modeling of personality scores via joint IRT models
  - **VAS**: Beta Response Model (Noel & Dauvier, 2007)
  - **DRS**: Dirichlet Dual Response Model
  - (Kloft et al., 2023)
  - **Correlations** of latent person scores via Bayesian hierarchical model with MVN prior

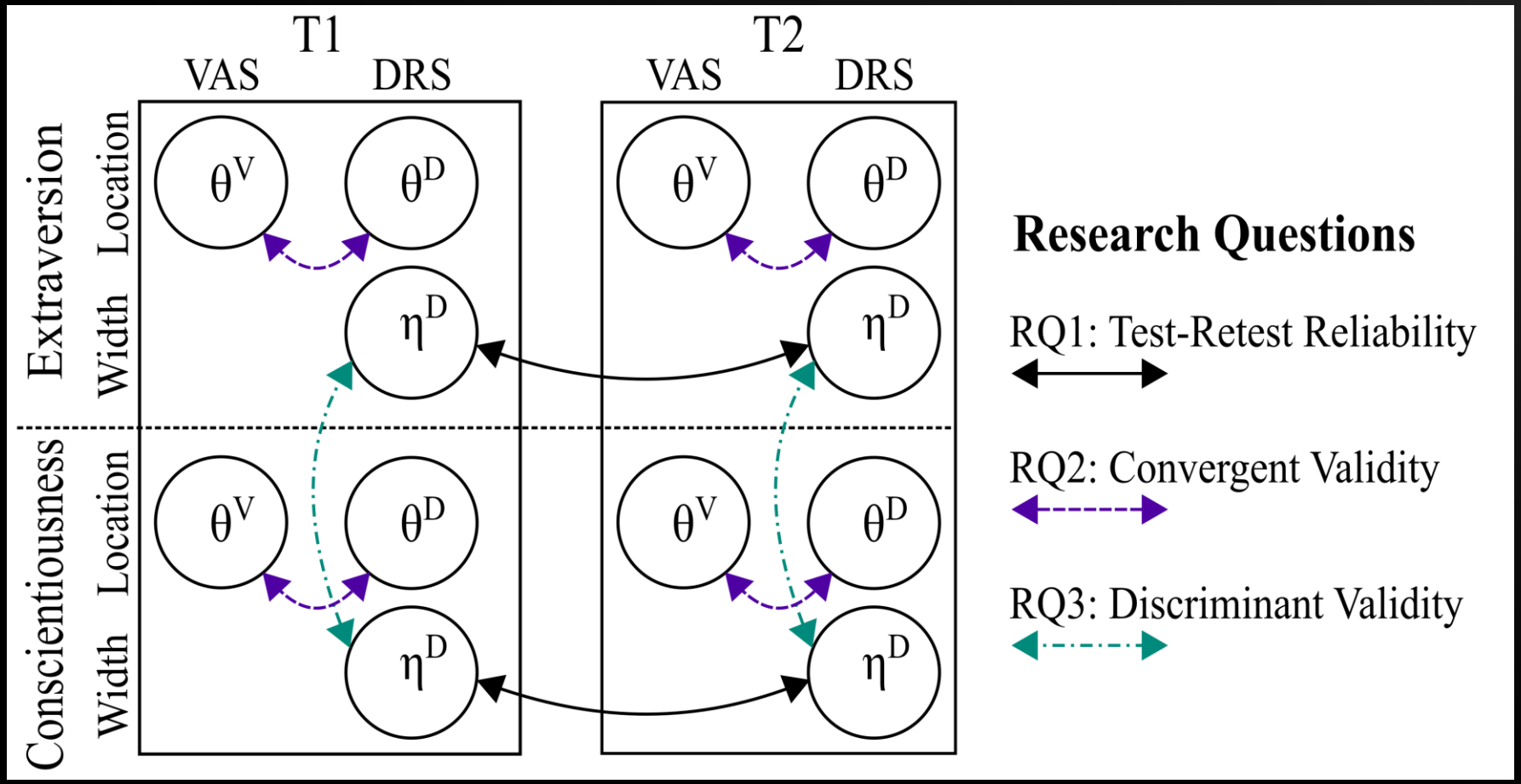
Kloft et al., 2024: MTMM Study Design



2 Time points, 2 Response Formats (VAS, DRS),

2 Scales (Extraversion & Conscientiousness)

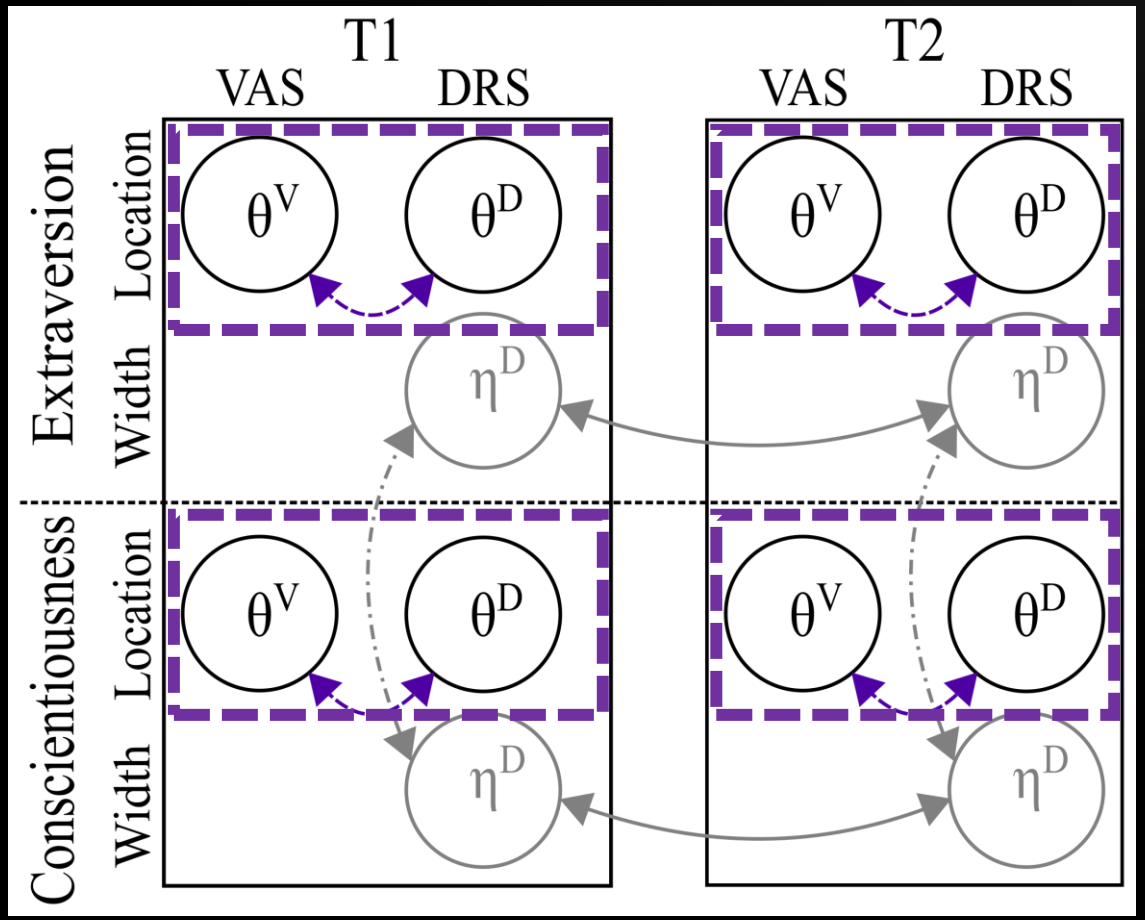
Kloft et al., 2024: MTMM Study Design



θ: scores for central tendency of personality

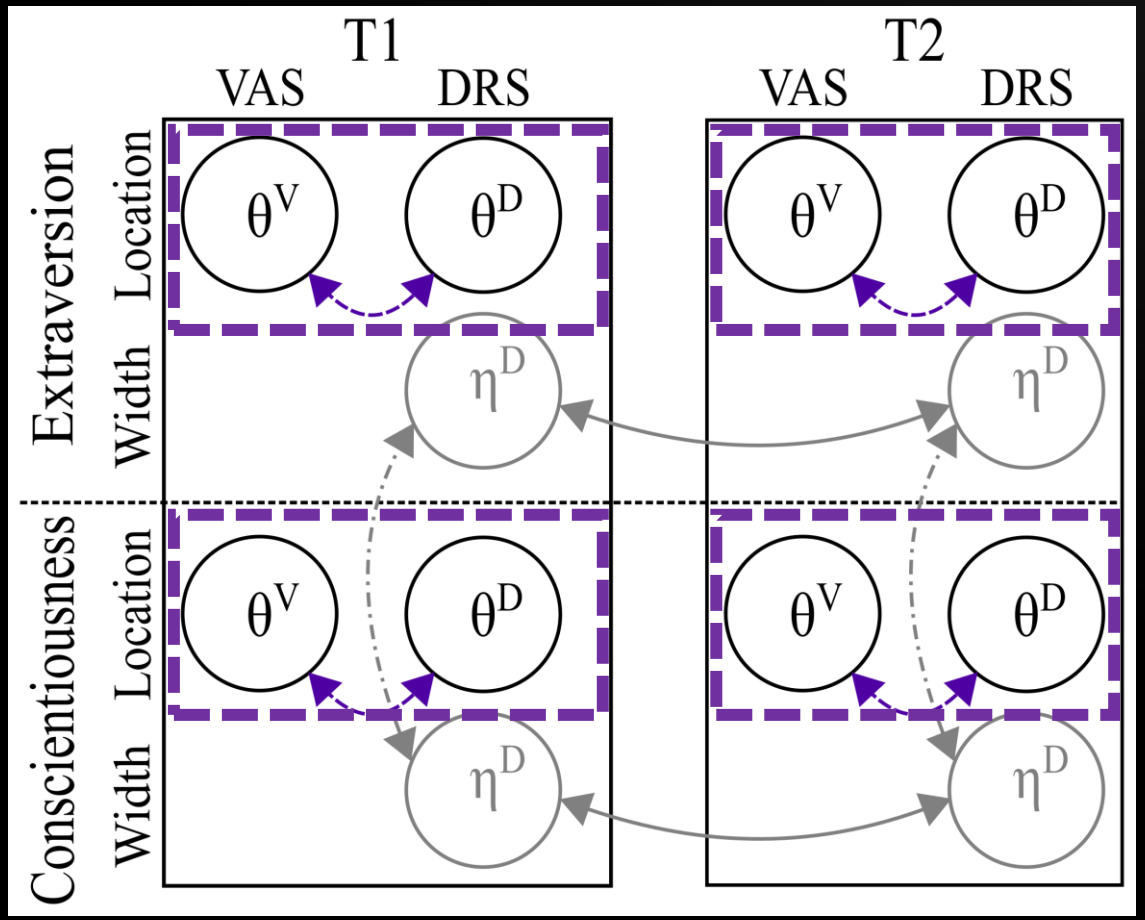
η: scores for perceived variability of personality

Kloft et al., 2024: MTMM Study Design



- Correlation between personality scores
- across response formats,
- for both scales,
- within each time point

Kloft et al., 2024: Correlations



T1: T2:

$\hat{\rho} = .93$ $\hat{\rho} = .96$

$\hat{\rho} = .88$ $\hat{\rho} = .90$

Convergence: VAS & DRS Location

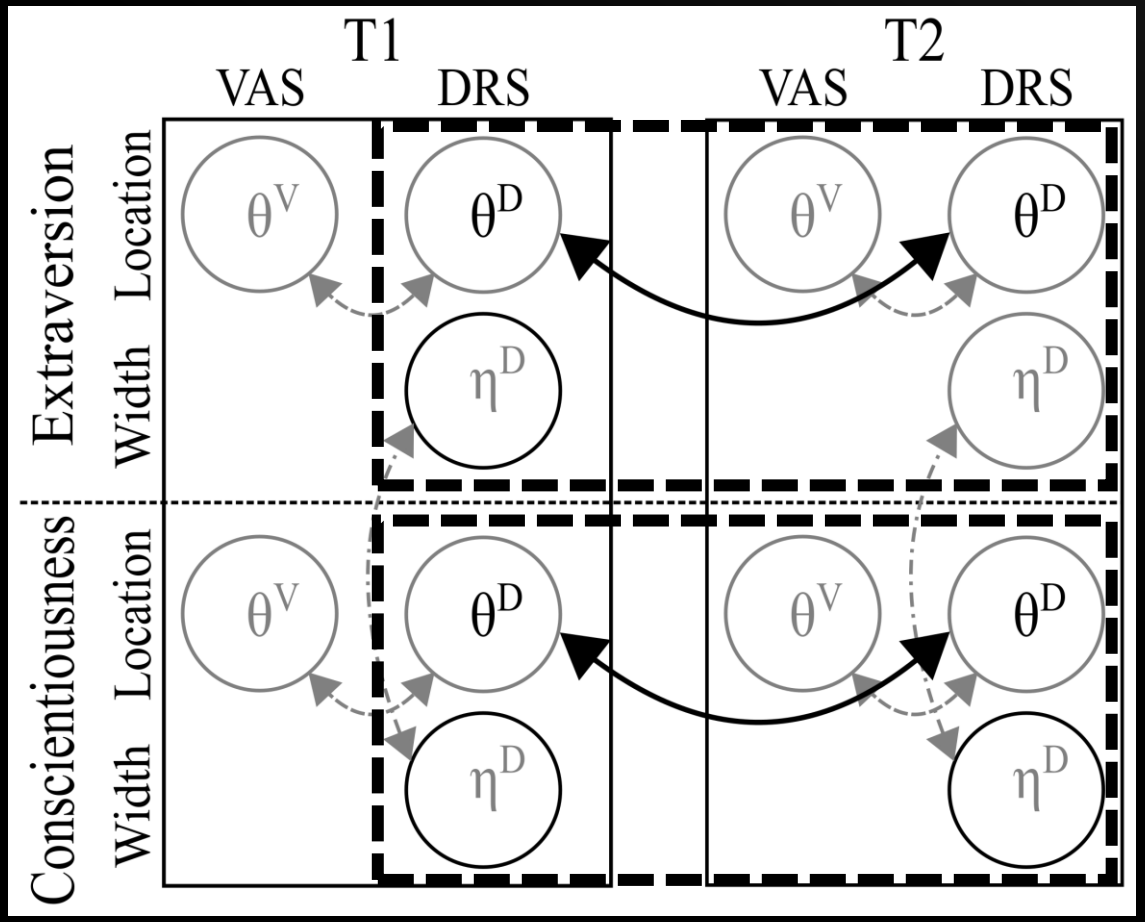
- Correlations between .87 and .93

- VAS and DRS Location measures may be used **equivalently** at least in this personality application

Test-Retest Reliability

Kloft et al. (2024)

Kloft et al., 2024: MTMM Study Design

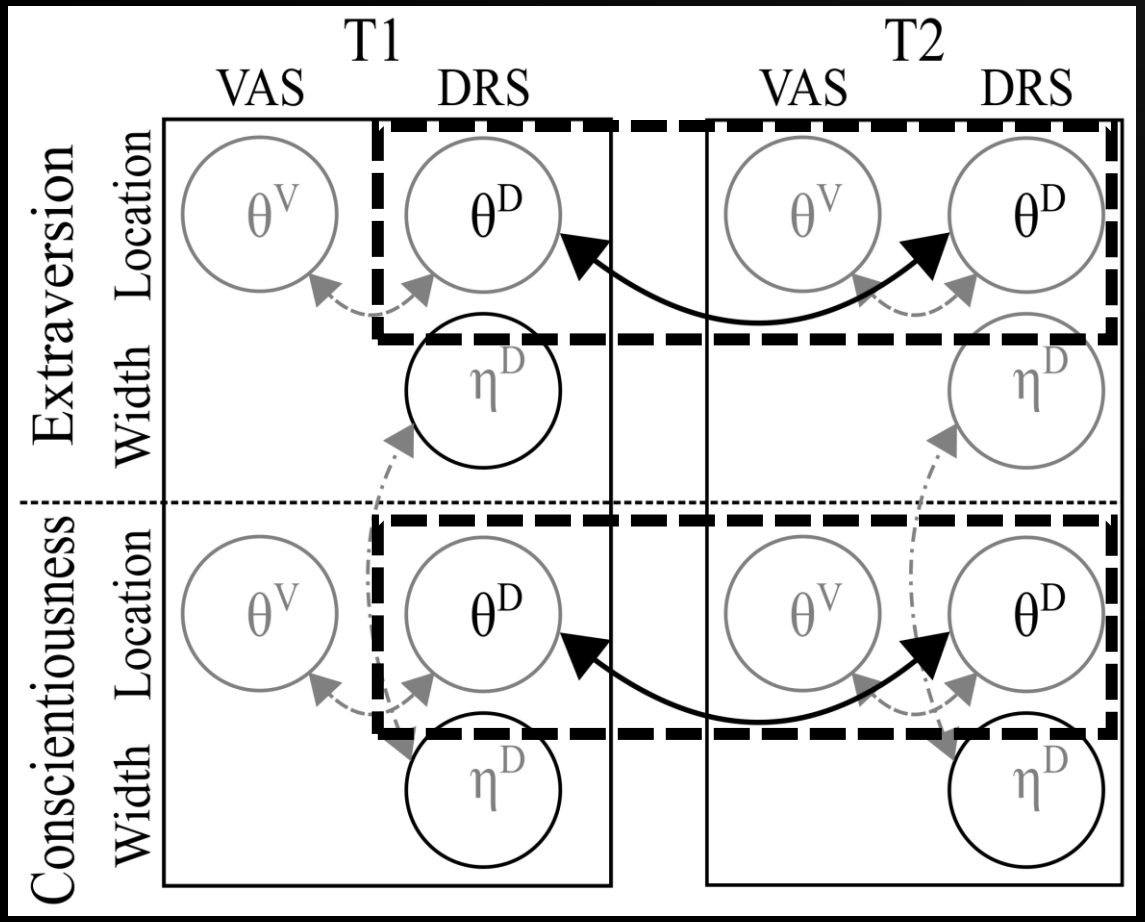


- Correlations of personality scores
- measured by DRS Location & Width
- across **time**

θ: scores for central tendency of personality

η: scores for perceived variability of personality

Autocorrelation: DRS Location

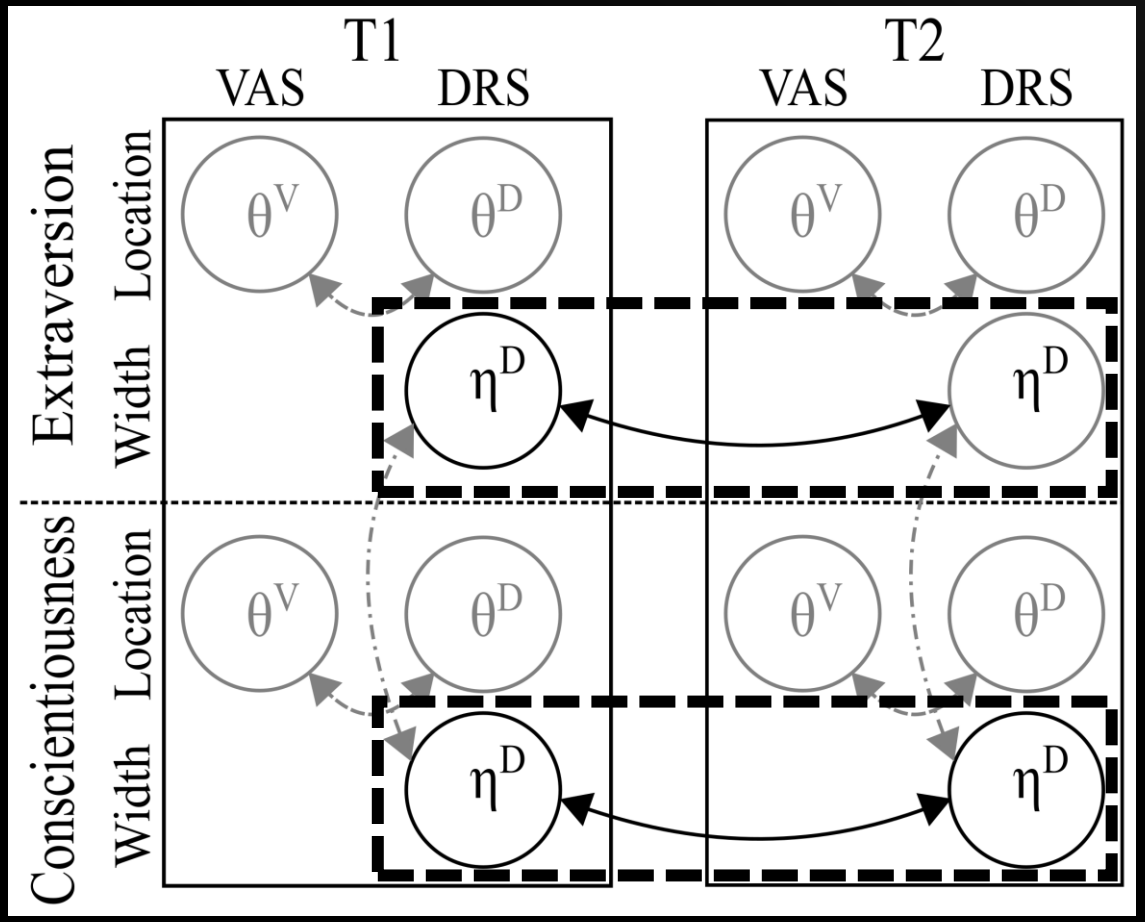


Location:
$\hat{\rho} = .92$

$\hat{\rho} = .87$

High reliability for DRS **Location** scores

Autocorrelation: DRS Width



Location:
$\hat{\rho} = .92$

Width:
$\hat{\rho} = .81$

$\hat{\rho} = .87$

$\hat{\rho} = .73$

Good reliability for DRS **Width** scores

Reliability: DRS Location & Width

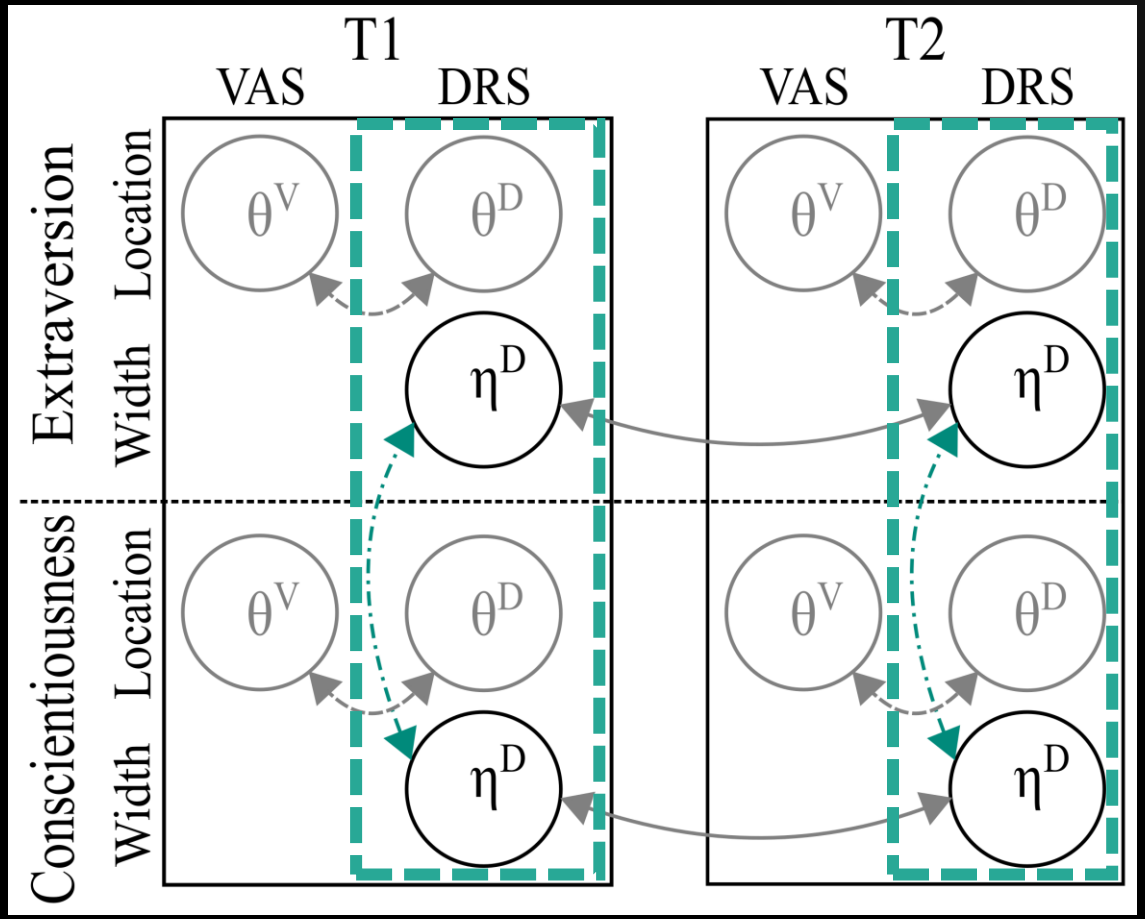
- Test-retest reliability:
  - Good for DRS Location scores
  - Seems okay for DRS Width scores

- Caveat: this might be due to strong response biases
  - ➢ Investigate discriminant validity between different constructs

Discriminant Validity: DRS Width

Kloft et al. (2024), Kloft & Heck (2024)

Kloft et al., 2024: MTMM Study Design

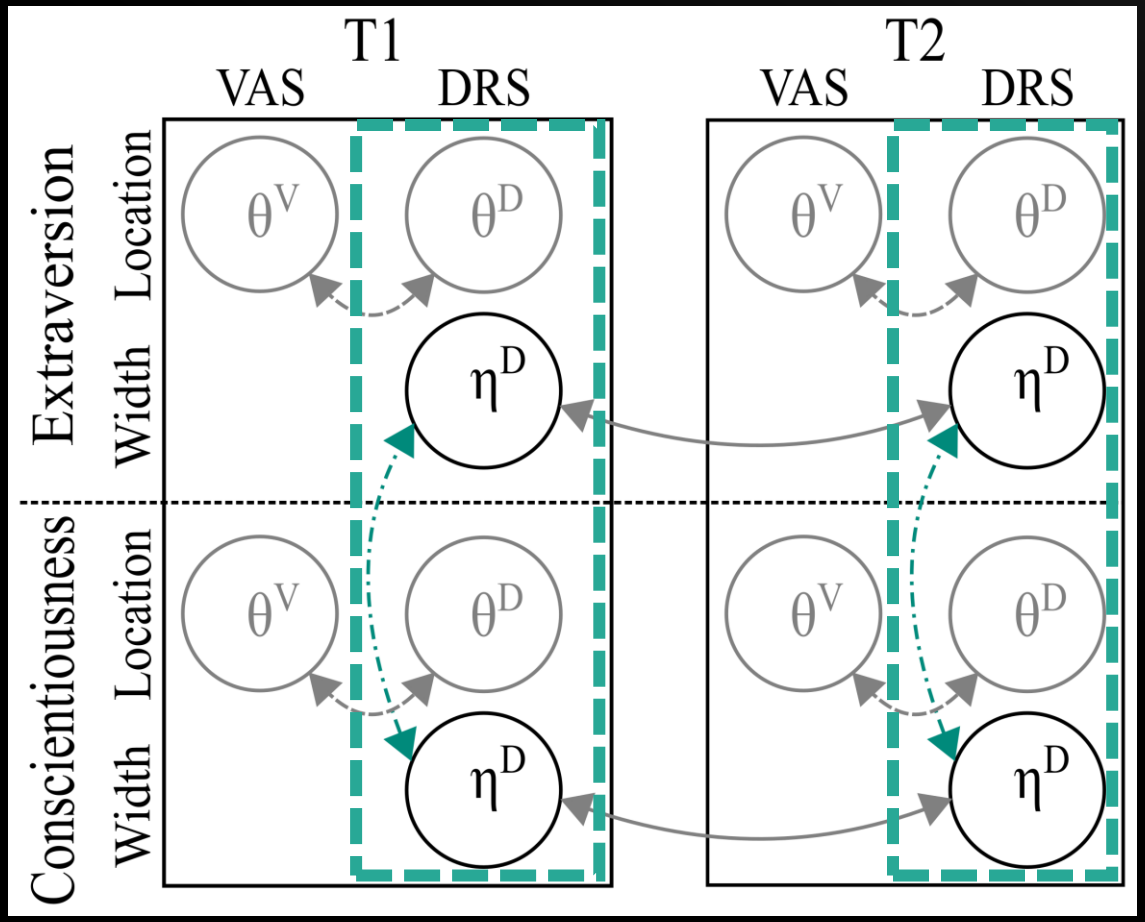


- Correlations of personality scores
- measured by DRS Location & Width
- across **personality scales**

θ: scores for central tendency of personality

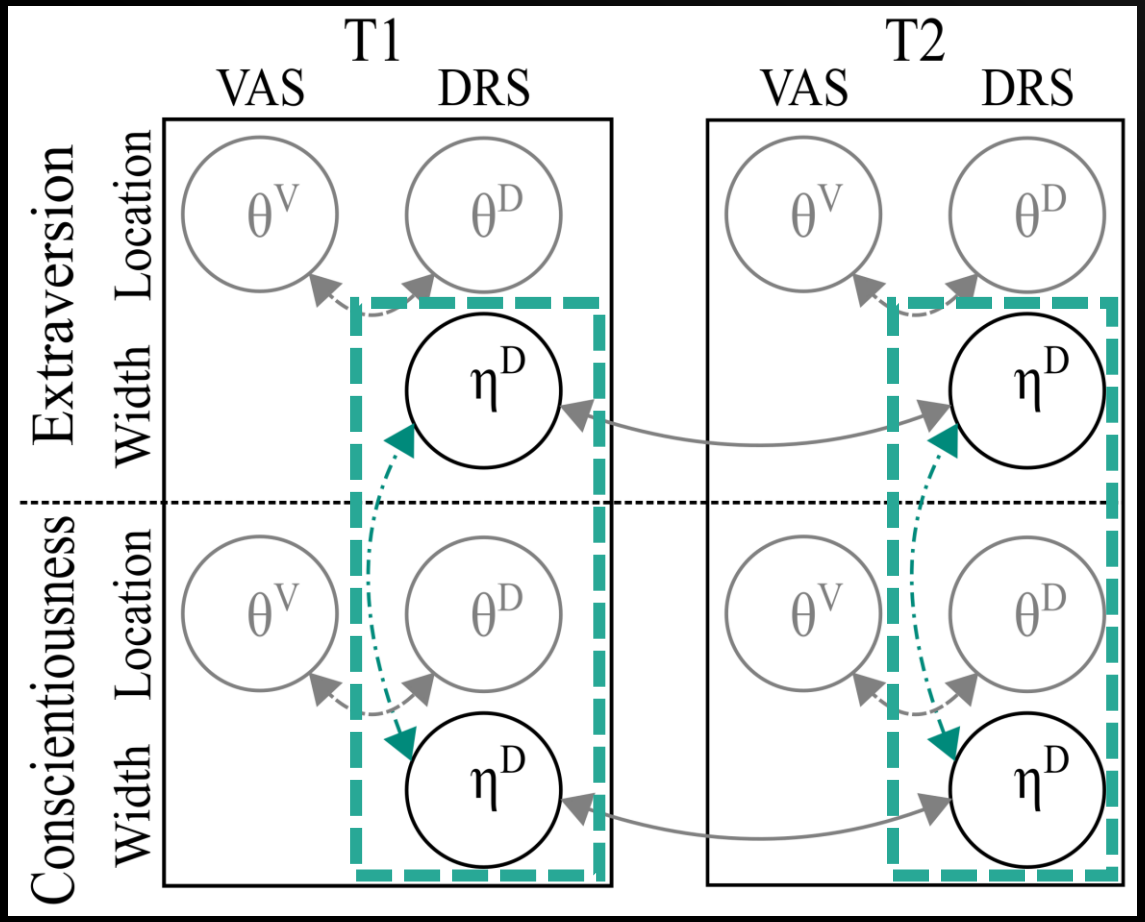
η: scores for perceived variability of personality

Kloft et al., 2024: MTMM Study Design



Different perceived variability for **different** traits?

Kloft et al., 2024: Correlations

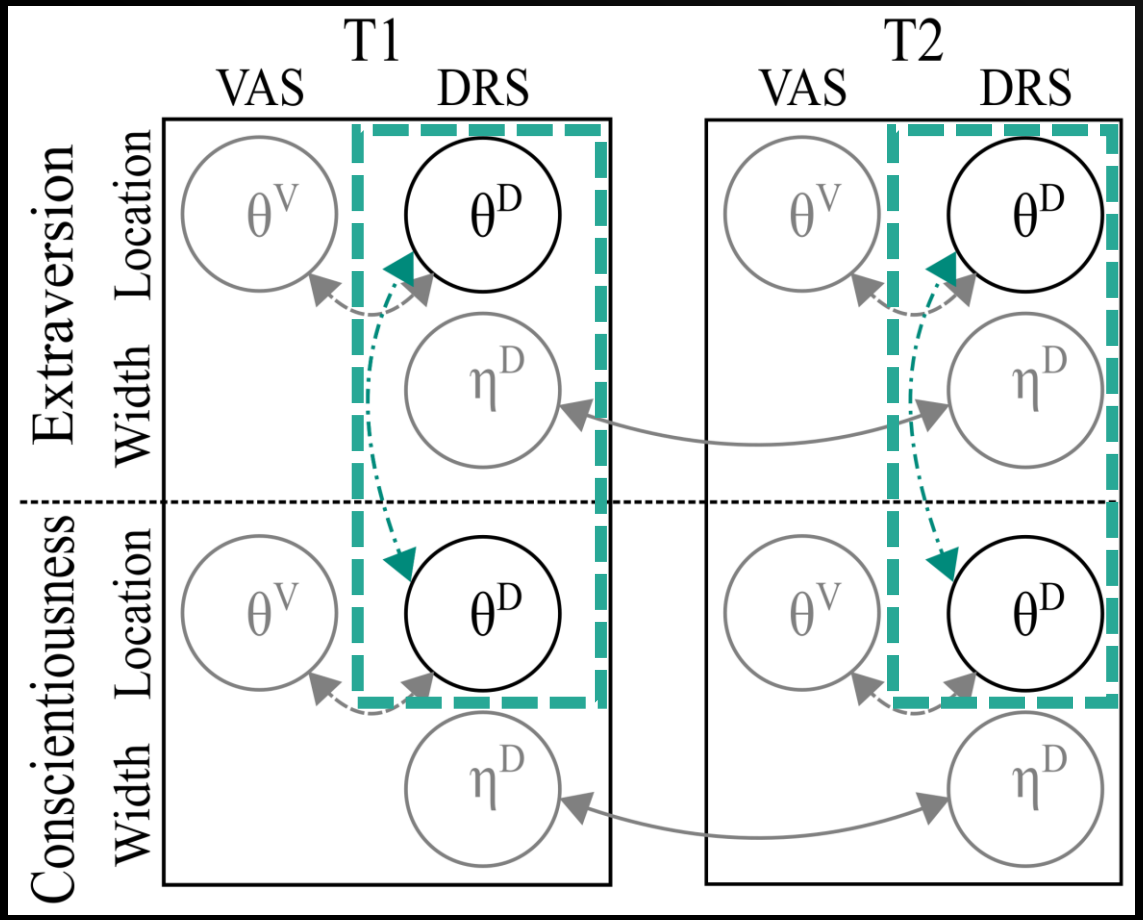


Width:

$\hat{\rho} = .94$ $\hat{\rho} = .96$

Poor discrimination of DRS **Width** scores

Kloft et al., 2024: Correlations



Location:

$\hat{\rho} = .30$ $\hat{\rho} = .31$

Width:

$\hat{\rho} = .94$ $\hat{\rho} = .96$

Reasonable correlation of DRS **Location** scores

Kloft et al., 2024: Conclusions

What do DRS **Widths** measure?

- Personal **preference** for a particular width
  - Preference response style
- **Global trait** of perceived variability in personality?
- Limitation: only two constructs, both from the personality domain
- Follow-up Study: Kloft & Heck, 2024

Kloft & Heck, 2024: Aim of Study

- Focus on DRS **Widths**:
  - **Sensitivity** to different tasks?
  - Just a preference response style?
  - Factorial **dimensionality** across different tasks?

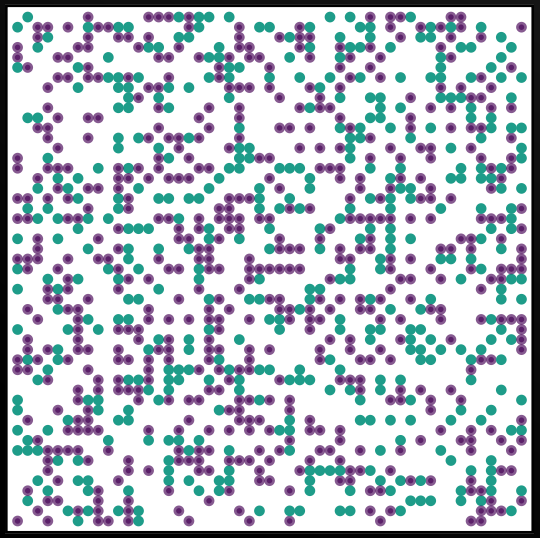
Kloft & Heck, 2024: Study Design

- Seven tasks of varying similarity
- Logit transformation of interval responses (Smithson & Broomel, 2024; see also additional slides)
- Factor Analyses: EFA, CFA
  - We use **only** the transformed **widths**
  - Transformation acts as a **link function**
    - Accounts for boundedness
    - Adjusts the width for the location

Kloft & Heck, 2024: Applications / Tasks

Extraversion & Conscientiousness:

Talkative $= [75\% - 93\%]$ applicable



Color Dot Estimation:

% purple dots $= [46\% - 73\%]$

(true $= 60\%$)

Election Forecasting: outcomes for 6 parties

- *Green party* $= [10\% - 19\%]$

Kloft & Heck, 2024: Applications / Tasks

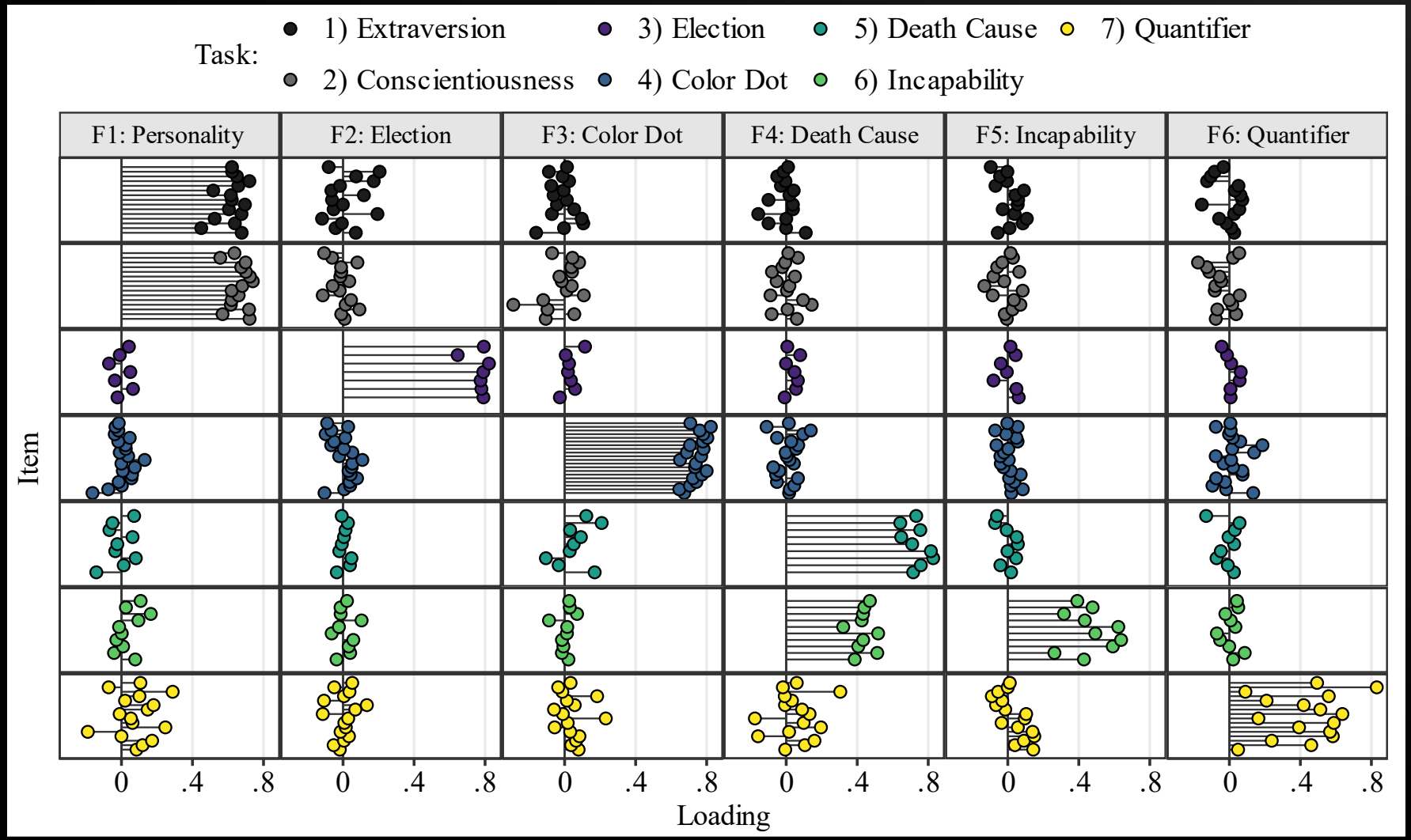
Estimation of percentages:

- Death Causes: *heart diseases*
- Reasons for Incapability for work: *mental health*

Judgment of verbal quantifiers as probabilities that a so described event would occur:

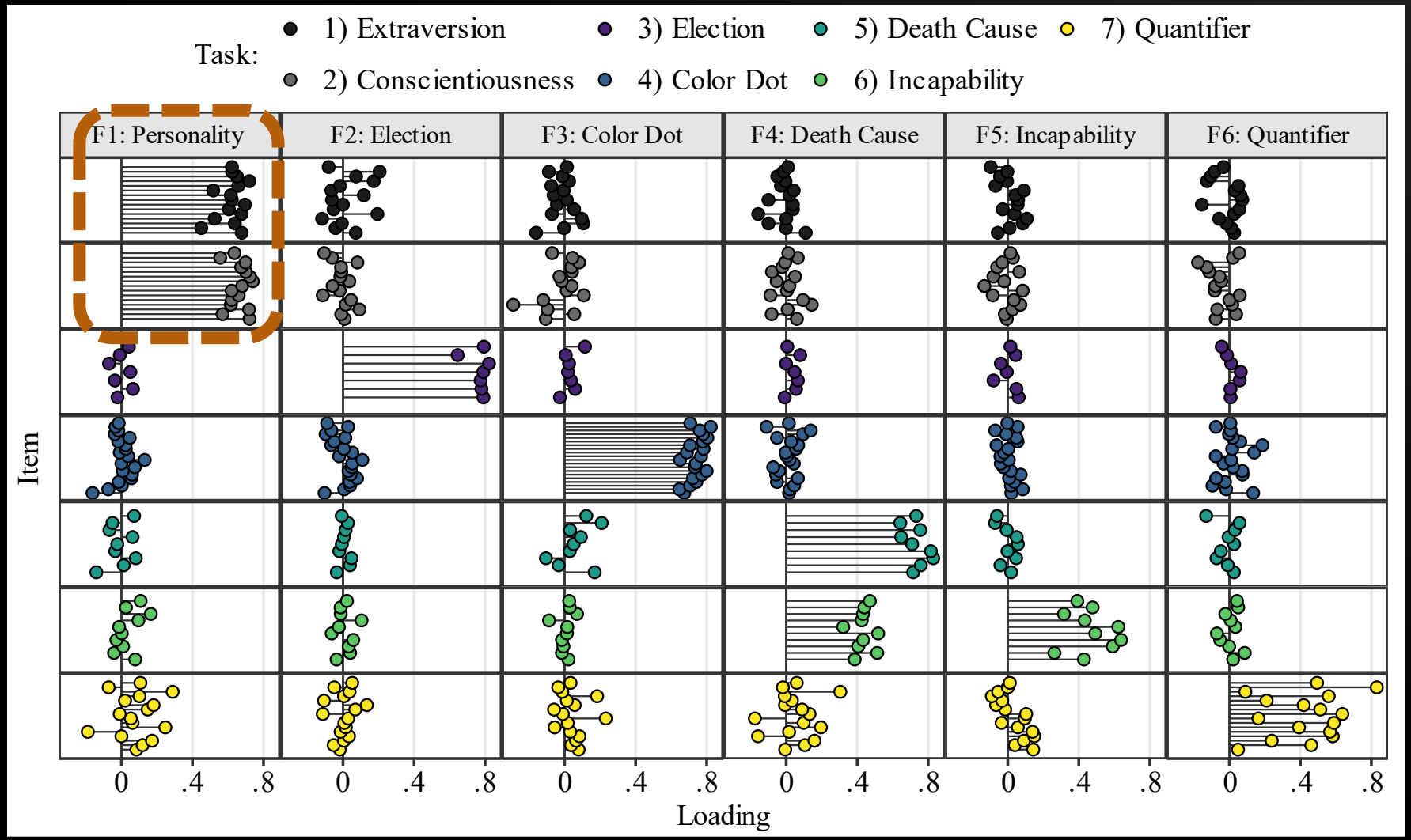
- *Seldom* = [5% – 30%]

Interval Width: EFA Loadings



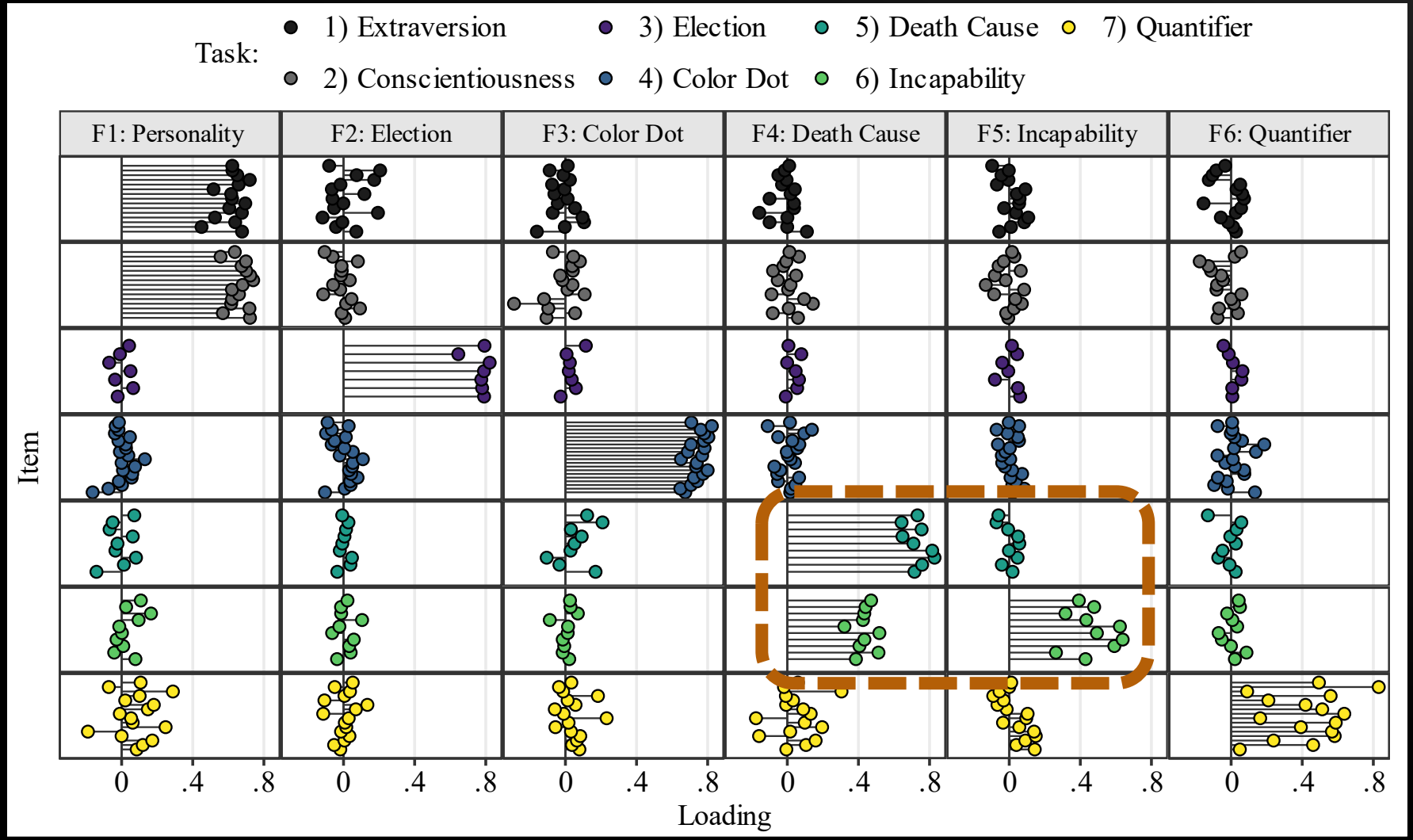
DRS **Width** only!

Interval Width: EFA Loadings



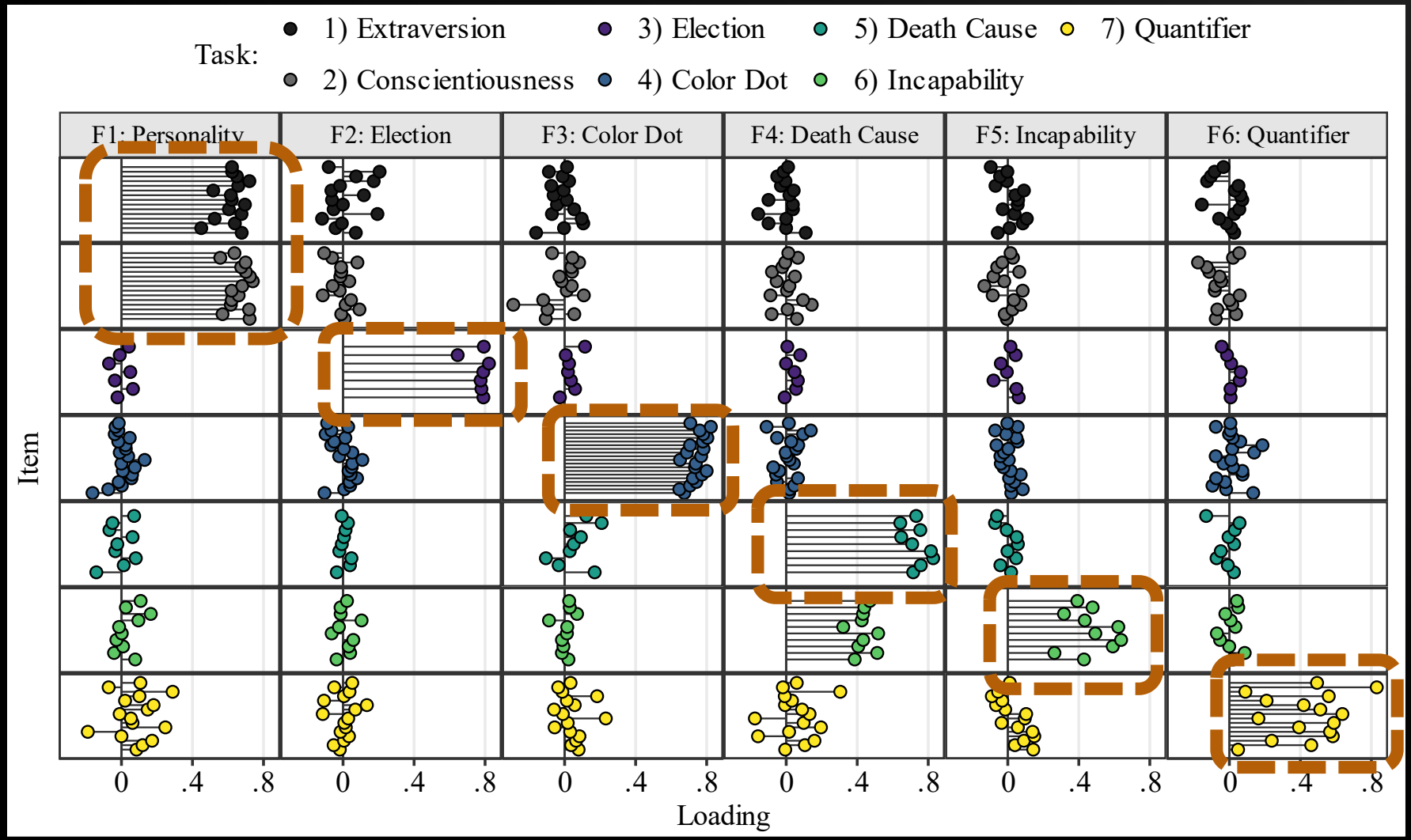
Common factor for Extraversion & Conscientiousness

Interval Width: EFA Loadings



High cross loadings: similar tasks

Interval Width: EFA Loadings

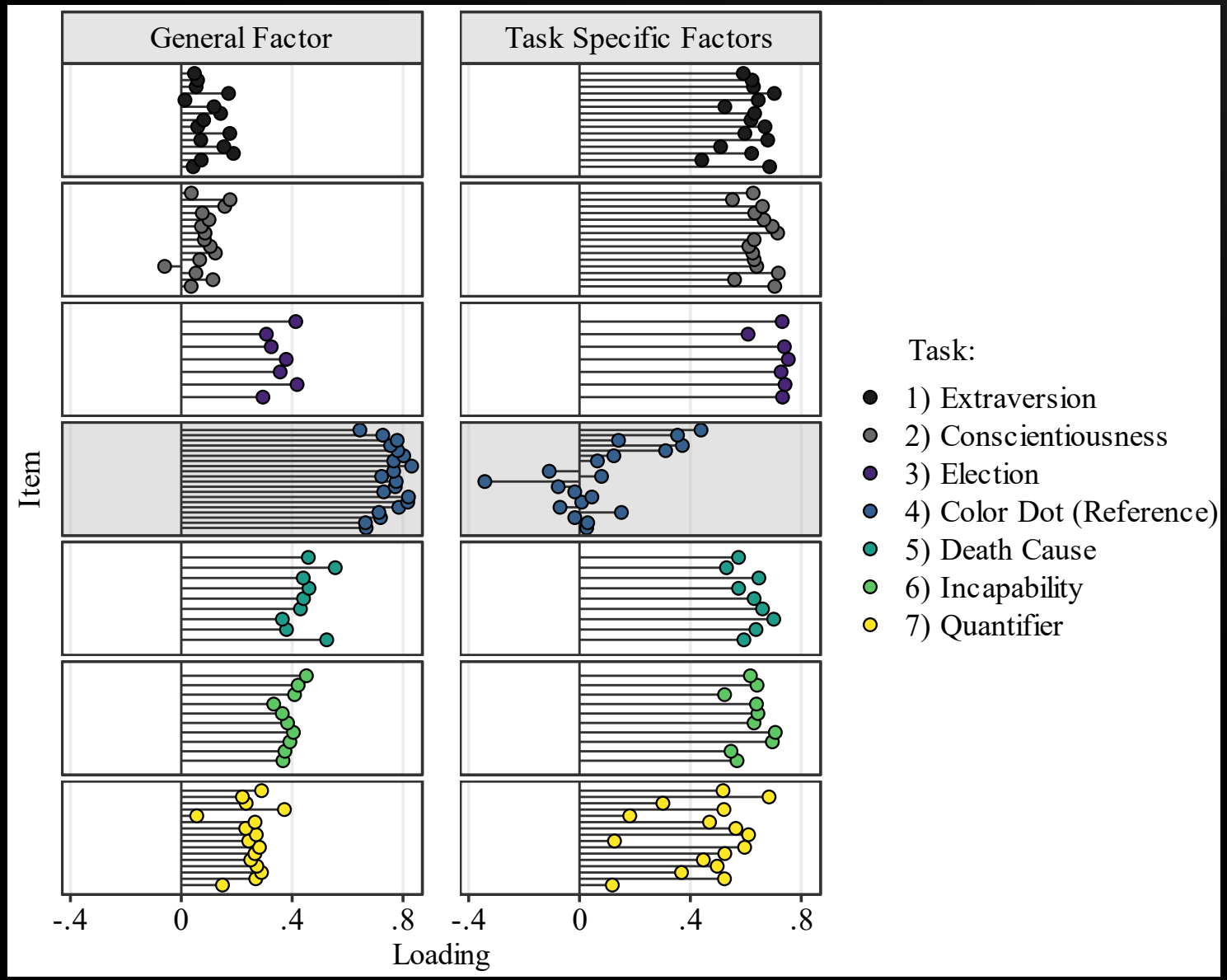


Factor structure **follows** task structure

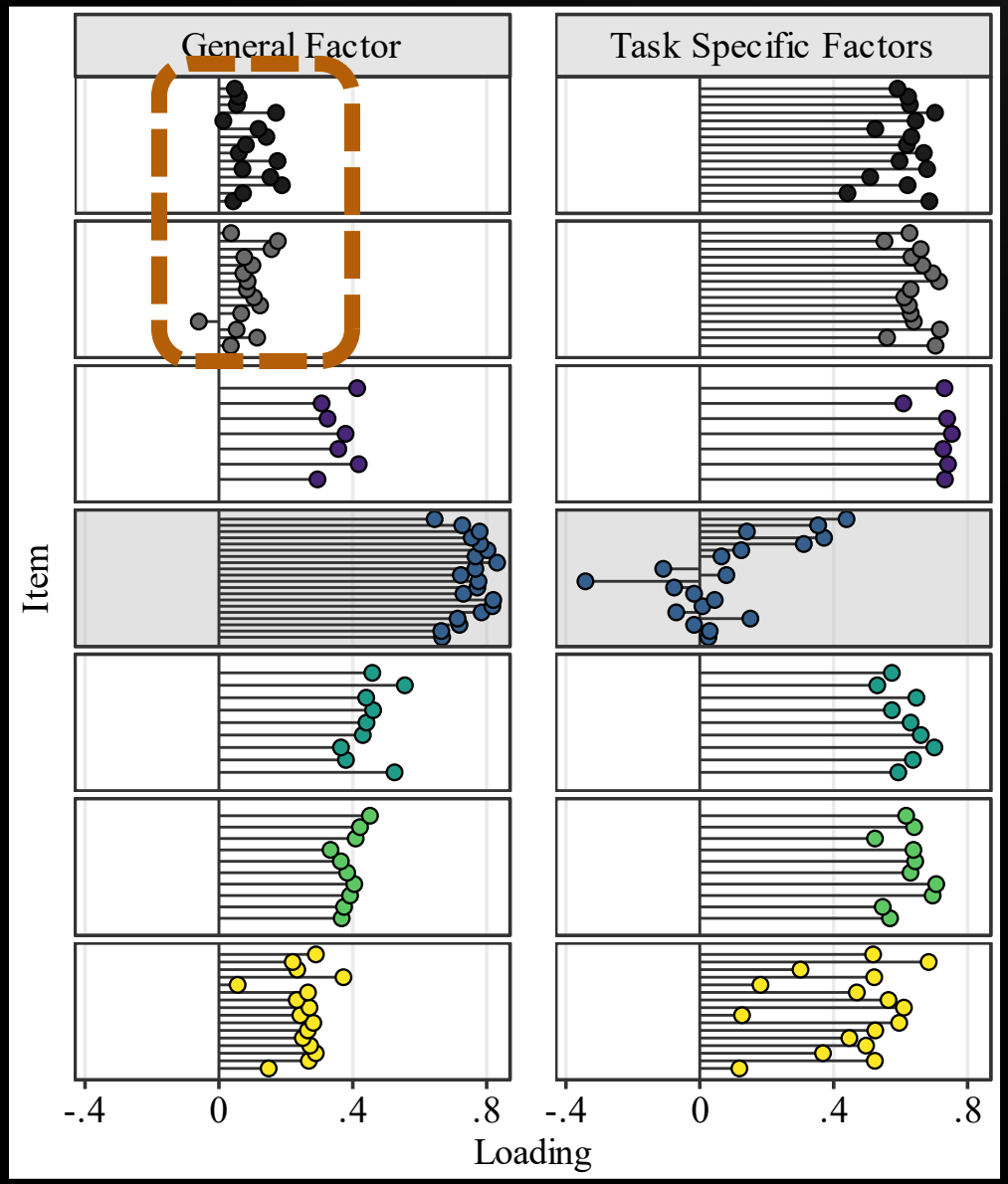
Kloft & Heck, 2024: Bifactor Model

- Again only transformed DRS **Width**
- Bifactor-($S.I-1$) model (Eid et al., 2017)
- 6 **Specific** Factors
  - Assignment by theoretical task
- **General** Factor: preferred width / response style
  - Color dot estimation as reference task

Loadings: General vs. Specific Factor(s)



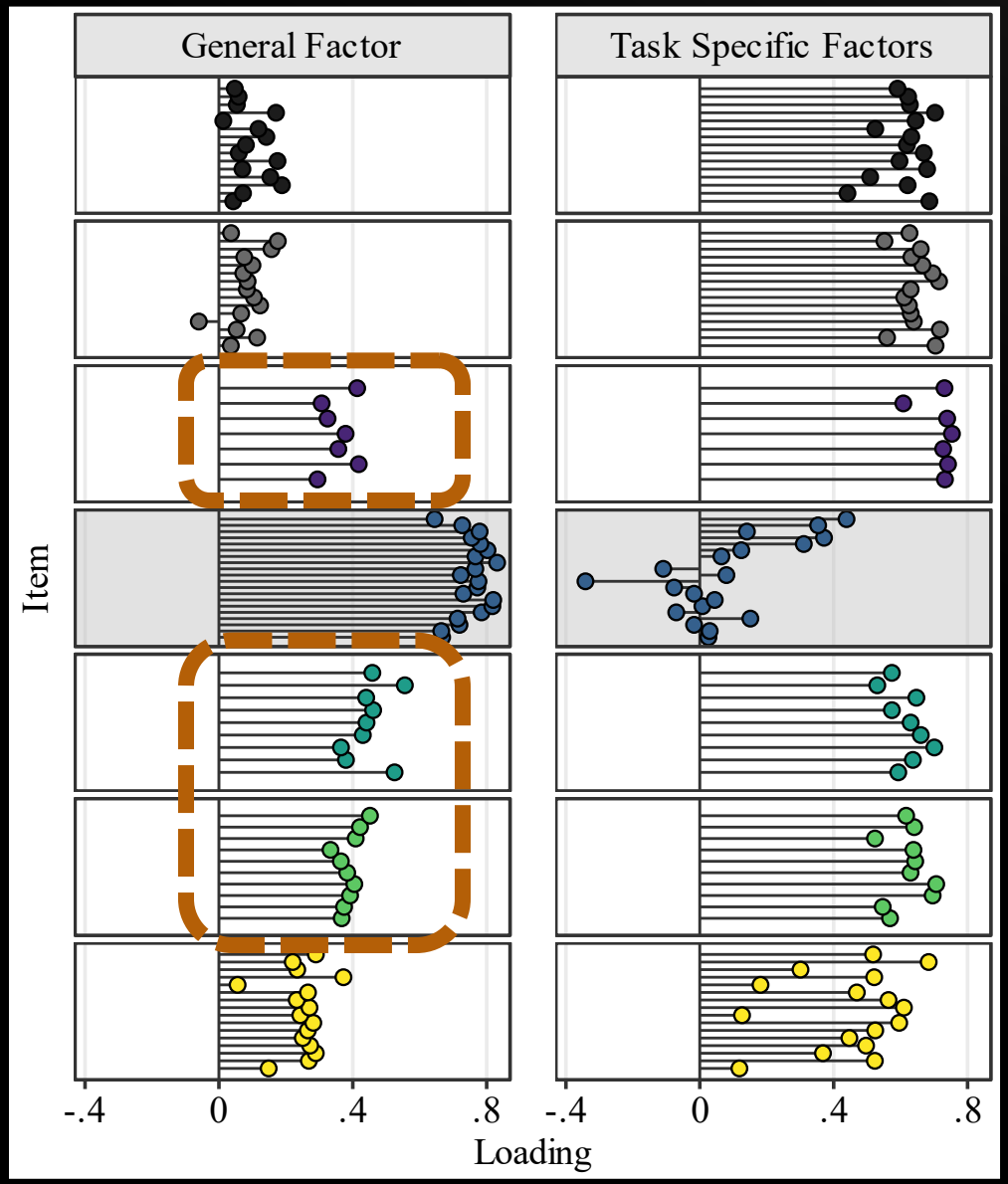
Loadings: General vs. Specific Factor(s)



Personality:

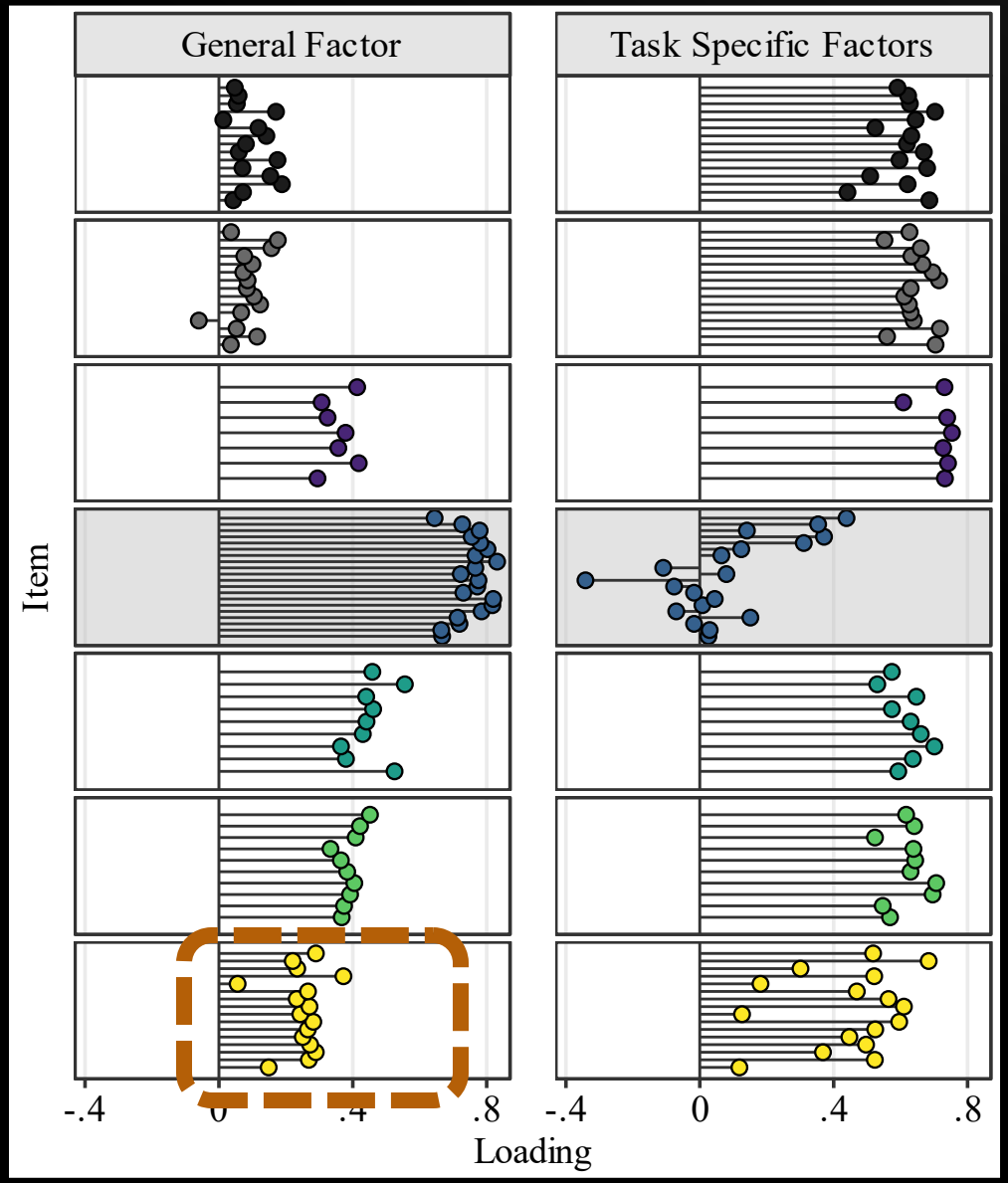
mostly small to zero loadings on General Factor

Loadings: General vs. Specific Factor(s)



Estimation tasks: mostly small to medium loadings on General Factor

Loadings: General vs. Specific Factor(s)



Verbal quantifiers judgment task: mostly small loadings on General Factor

Kloft & Heck, 2024: Bifactor Model

- Across all tasks: low influence of the general factor compared to the specific factor

Bifactor Model: Correlations Between Special Factors

	Personality	Election	Color Dot (Reference)	Death Cause	Incapability
Election	0.11 [-0.03,0.24]				
Color Dot (Reference)	-0.16 [-0.29,-0.02]	-0.15 [-0.28,-0.02]			
Death Cause	0.07 [-0.07,0.2]	0.26 [0.13,0.38]	-0.02 [-0.16,0.11]		
Incapability	0.1 [0.06,0.32]	0.24 [0.11,0.37]	-0.03 [-0.16,0.11]	0.66 [0.58,0.73]	
Quantifier	0.28 [0.15,0.4]	0.19 [0.06,0.32]	-0.06 [-0.2,0.07]	0.3 [0.18,0.42]	0.36 [0.24,0.48]

Bifactor Model: Correlations Between Special Factors

	Personality	Election	Color Dot (Reference)	Death Cause	Incapability
Election	0.11 [-0.03,0.24]				
Color Dot (Reference)	-0.16 [-0.29,-0.02]	-0.15 [-0.28,-0.02]			
Death Cause	0.07 [-0.07,0.2]	0.26 [0.13,0.38]	-0.02 [-0.16,0.11]		
Incapability	0.1 [0.06,0.32]	0.24 [0.11,0.37]	-0.03 [-0.16,0.11]	0.66 [0.58,0.73]	
Quantifier	0.28 [0.15,0.4]	0.19 [0.06,0.32]	-0.06 [-0.2,0.07]	0.3 [0.18,0.42]	0.36 [0.24,0.48]

Conclusions

- Similar tasks show less discriminant validity
  - Shared expertise vs. shared response bias?
- DRS Width: respondents seem to be sensitive to requirements of different tasks
- Perceived variability of personality: DRS measures will probably not be valid
- Central tendency: DRS yields similar measurements to the VAS

Future Research

- More work to disentangle expertise and uninformed confidence
- Test-retest reliability in domains other than personality
- Any ideas how to better model the data from different tasks?

Thanks to:

Prof. Dr. Daniel W. Heck

Björn Siepe

Dr. Jean-Paul Snijder

Dr. Raphael Hartmann

Prof. Dr. Andreas Voss

Contact:

kloft@uni-marburg.de/

Slides:

https://github.com/matthiaskloft

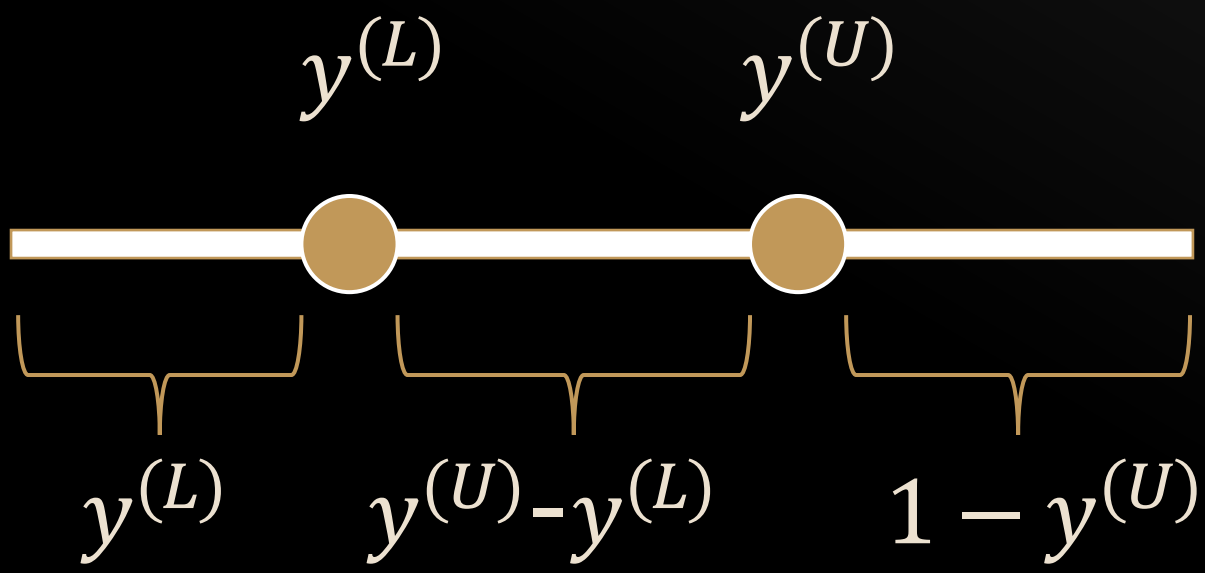
References

- Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in *G*-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541–562. https://doi.org/10.1037/met0000083
- **Kloft, M.**, Hartmann, R., Voss, A., & Heck, D. W. (2023). The Dirichlet dual response model: An item response model for continuous bounded interval responses. *Psychometrika*. https://doi.org/10.1007/s11336-023-09924-7
- **Kloft, M.**, Snijder, J.-P., & Heck, D.W. (2024). Measuring the variability of personality traits with interval responses: Psychometric properties of the dual-range slider response format. Behavior Research Methods, 56 (4), 3469–3486. https://doi.org/10.3758/s13428-024-02394-4
- **Kloft, M.**, & Heck, D.W. (2024). Discriminant validity of interval response formats: Investigating the dimensional structure of interval widths. Educational and Psychological Measurement 0 (0). https://doi.org/10.1177/00131644241283400
- Gersen, L. (2024). Leongersen/noUiSlider [Software]. https://github.com/leongersen/noUiSlider
- Smithson, M., & Broomell, S. B. (2024). Compositional data analysis tutorial: Psychological Methods. *Psychological Methods, 29* (2), 362–378. https://doi.org/10.1037/met0000464

Additional Slides

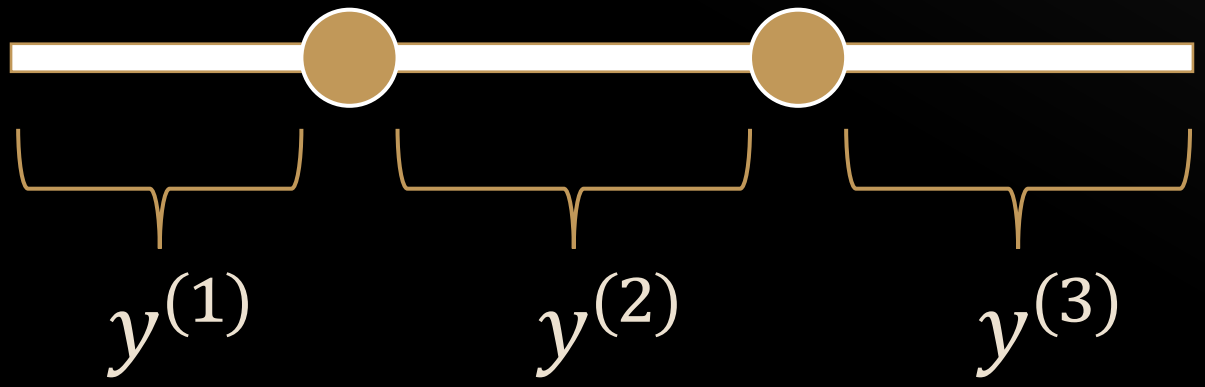
Compositional Data

- Components must sum to one: simplex



Compositional Data

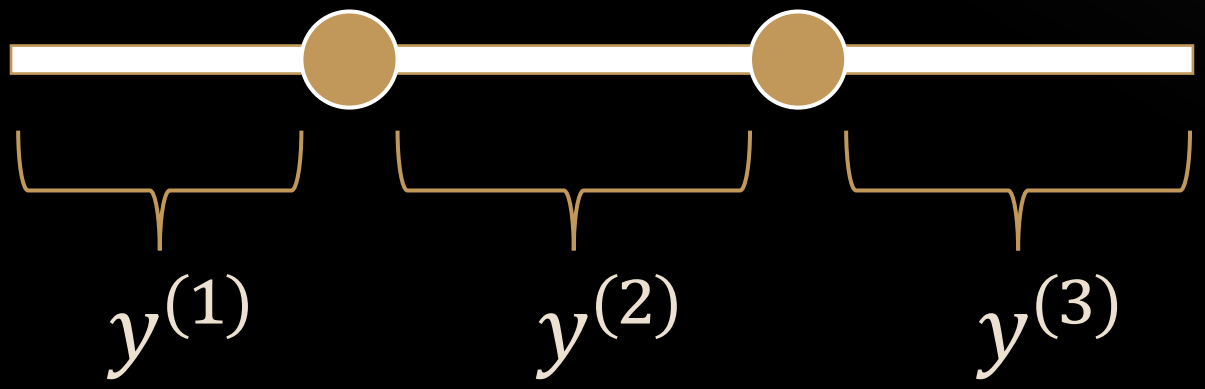
- Components must sum to one: simplex



Log-Ratios

Unbounded **Location**: $\log\left(\frac{y^{(1)}}{y^{(3)}}\right)$

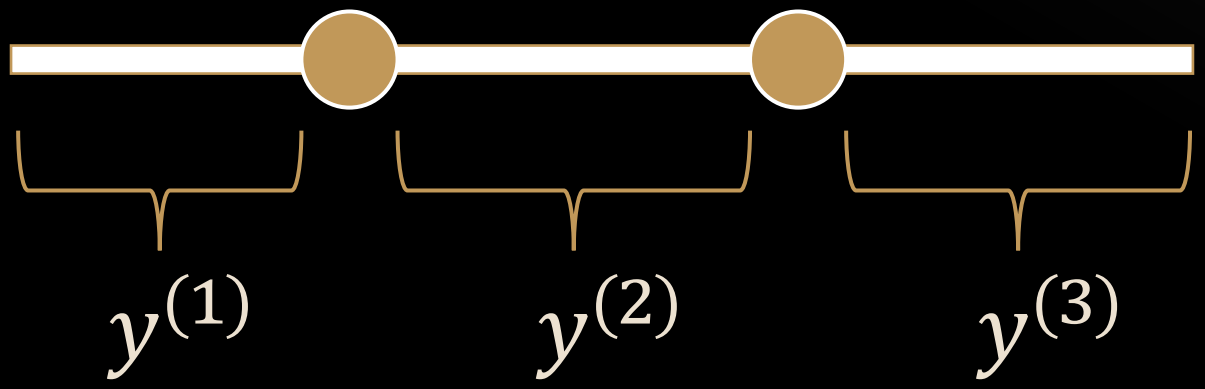
- Compares outer components



Log-Ratios

Unbounded **Width**: $\log\left(\frac{y^{(2)}}{\sqrt{y^{(1)} \times y^{(3)}}}\right)$

- Compares interval width to geometric mean of outer components



Isometric Log-Ratio Transformation

- Smithson & Broomel (2024)

$$\mathbf{z} = \begin{pmatrix} z^{loc} \\ z^{wid} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1}{2}} \log\left(\frac{y^{(1)}}{y^{(3)}}\right) \\ \sqrt{\frac{2}{3}} \log\left(\frac{y^{(2)}}{\sqrt{y^{(1)} \times y^{(3)}}}\right) \end{pmatrix}$$

Data Example

- More suitable for models using a normal distribution