The advantages of item-level assessment across research questions in personality psychology

Kendall Mather, University of Oregon





Overview of dissertation chapters

Chapter 1

Review paper:

The advantages of item-level assessment across research questions.

Goals of this chapter:

- Review previous item-level findings
- discuss barriers to reporting item-level findings
- Explore the advantages of item-level assessment for studies across the broad goals of *Prediction*, *Description*, and *Explanation*.

Chapter 2

Empirical study:

Demonstrating the Dowd Test Statistic (DTS) for item-level analysis.

Goals of this chapter:

- Address the limited statistical options for item-level research questions.
- Describe one option, the DTS test, and its advantages over familiar options (*t*-tests, *F*-tests).
- Demonstrate DTS test using data from the World Values Survey

Chapter 3

Empirical study:

Applying the DTS test to 2,623 items from the International Personality Item Pool (IPIP).

Goals of this chapter:

- Apply DTS test to compare responding on the 2,623 IPIP items across nations.
- Identify items demonstrating the most and fewest differences across nations.
- Identify nation pairs demonstrating the most and fewest differences across items

Chapter 4

Conclusions and takeaways

Goals of this chapter:

- Discuss remaining challenges and future directions for item-focused work.

Chapter 1
The advantages of item-level assessment across research questions

Personality psychology: the last refuge of the generalist?



"Exploring the frontiers of personality science poses major challenges. Perhaps the most central of these is the enduring task of personality psychology of **providing the integrative center for psychology**" (Little, 2010, p.1)



"Personality psychology is the last refuge of the generalist in psychology" (Revelle, 2008, p.2)

<u>A few of the many research questions personality psychology addresses:</u>

- What is the structure of personality?
- How does it develop and change?
- How does it vary across cultures?
- How well does it predict life outcomes?
- Can we detect it in our life narratives, digital footprints, and physical spaces?
- When is is healthy and when does it become maladaptive?
- Where does personality variation come from?

However, current methodological recommendations (e.g., use of broad trait measures, latent variable modeling) do not accommodate the diversity in topics addressed by the field.

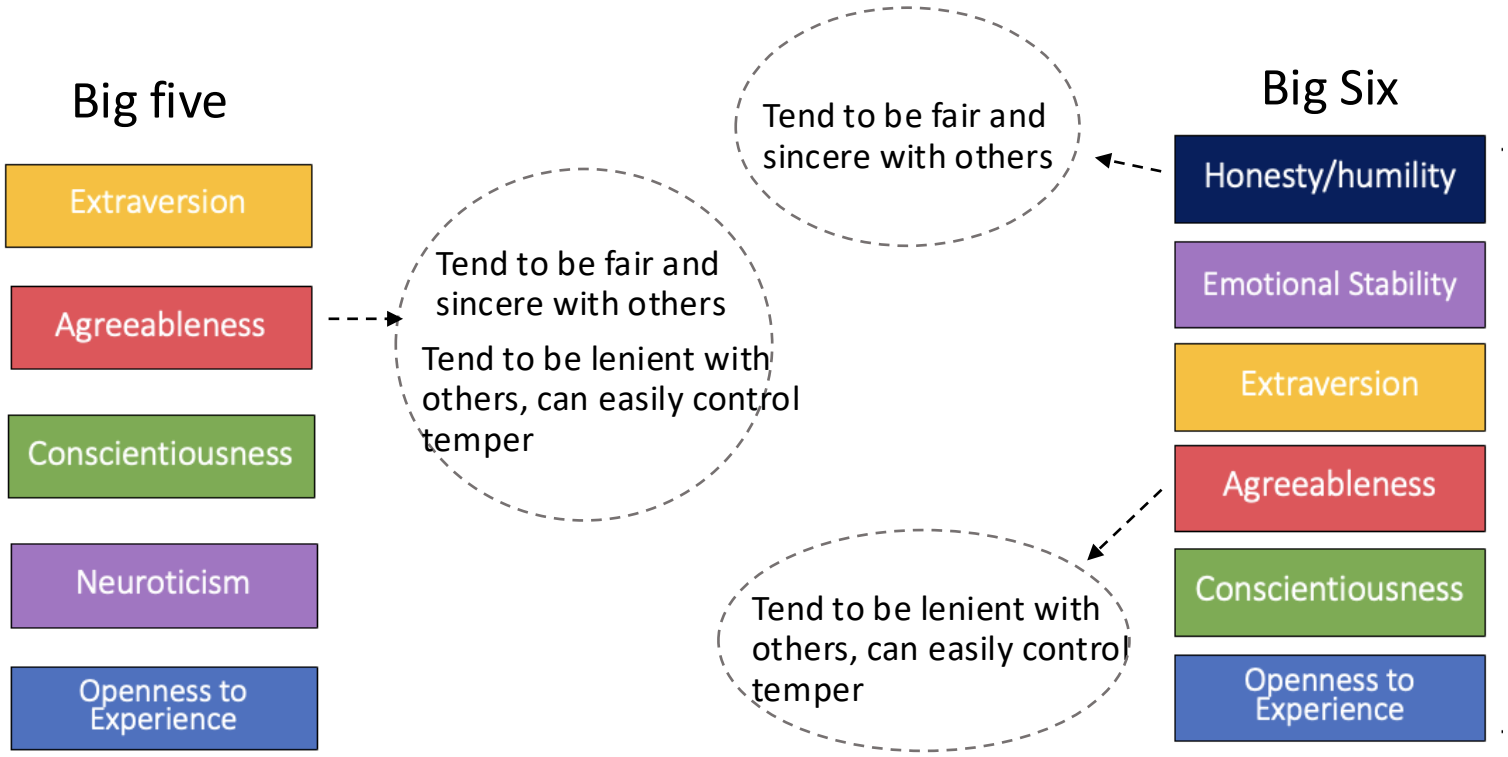
Studies differ according to their measurement needs

How much space is available to administer personality items?

DIW SOEP
The Socio-Economic Panel

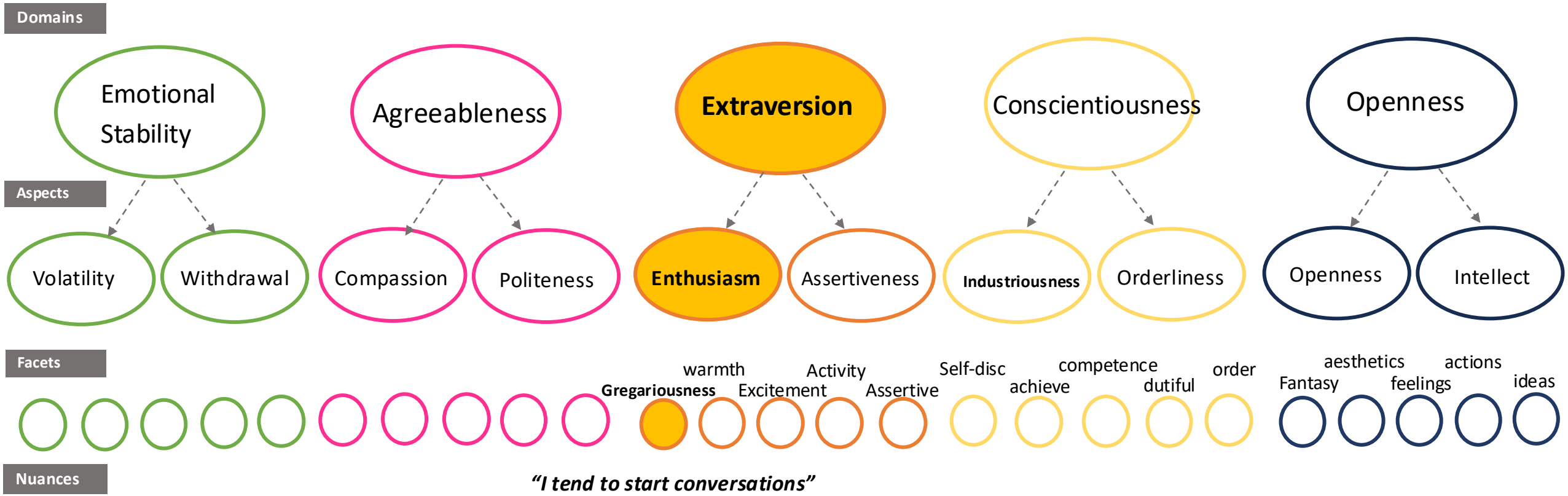
- Longitudinal survey of households in Germany (1984-2023)
- Core research questions : economic issues, work, health, well-being
- Uses 15-item Big Five measure (BFI-S; Lang et al., 2011). 3 items per trait.

What kinds of traits should be assessed? Which model best aligns with a particular theoretical perspective?



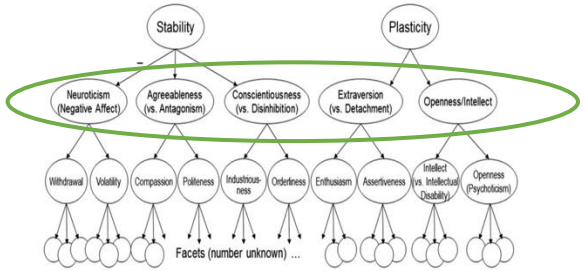
Studies differ according to their measurement needs

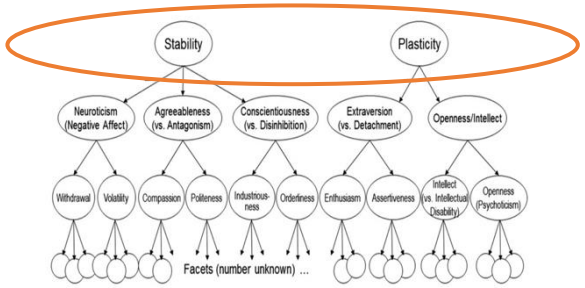
A common concern: which level of abstraction is optimal?

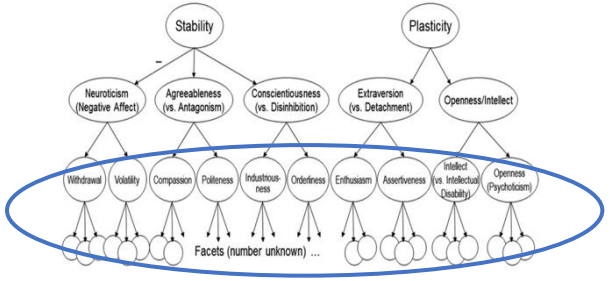


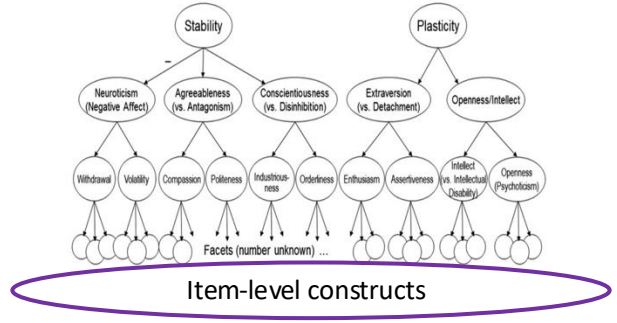
Trait models and levels of abstraction

Domains	Meta-traits	Aspects/facets	Items
Big five/Five Factor Model Big six/HEXACO Big seven, big eight	Big two Big three (or Pancultural three)	Big Five Facets Big Five Aspects	Individual items









- Most well-known
- Balances parsimony and comprehensiveness
- Portrayed as more generalizable than evidence suggests

- Relatively strong cross-cultural generalizability
- Strong overlap with frameworks from other fields
- Too broad to be useful in all research contexts

- More fine-grained than domain-level, more tractable than nuance level
- Subjectivity in determining facet content and number of facets

Trait models and levels of abstraction

Domains	Meta-traits	Aspects/facets	Items

Personality items have trait-like properties…

- Stability over time (Mõttus et al., 2017)
- Cross-rater agreement (Mõttus et al., 2014, 2019)
- Heritability (Mõttus et al., 2017)
- Unique developmental trajectories and associations with outcomes (Mõttus et al., 2019)

It remains somewhat uncommon to focus on the item-level.

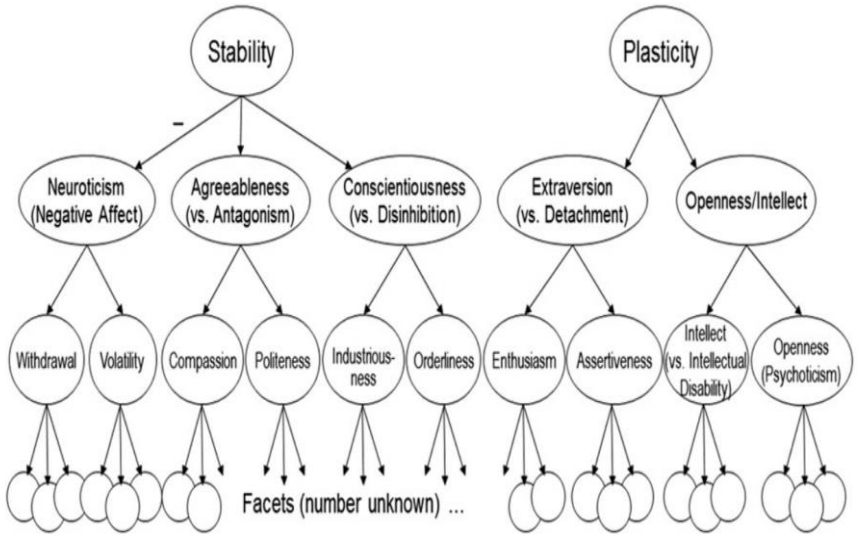
- Items are viewed mostly in terms of their relationship to broader constructs (i.e., indicators capturing a mixture of “true score” on a given trait and unique item variance)

This could be explained by the way personality frameworks are typically developed and the statistical techniques recommended for modeling and measuring traits.

How are trait models derived?

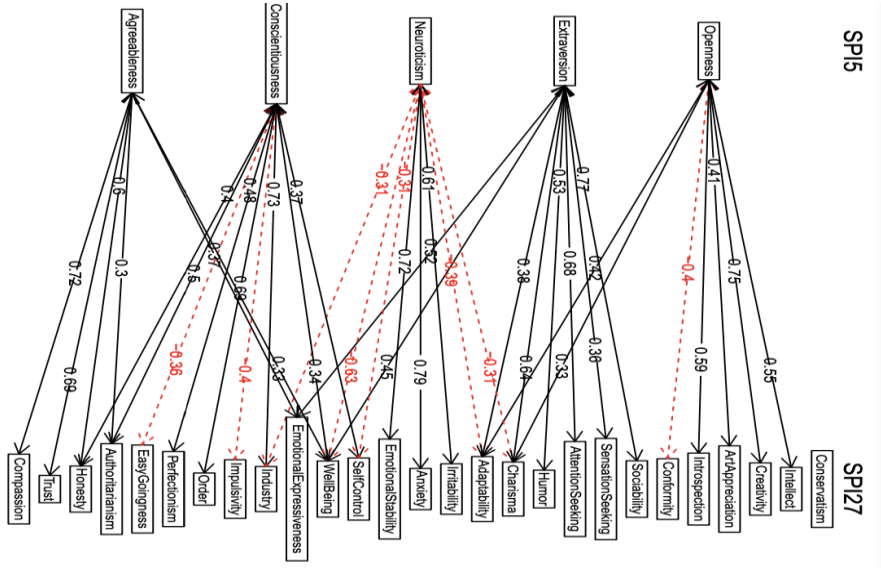
Top-down approaches

- Begin by identifying traits at highest level of interest.
- Lower levels traits reflect a portion of the content from just one trait at the level above (i.e., they are strictly nested within a higher-level trait)
- Lower-levels don't reflect optimal solution at that level.



Bottom-up approaches

- Each level is independent from all others—lower-levels reflect optimal partitioning of variance at that level.
- Lower-level traits (nuances, facets) are building blocks for levels above, form multiple traits at higher levels.
- Avoids missing interstitial traits.

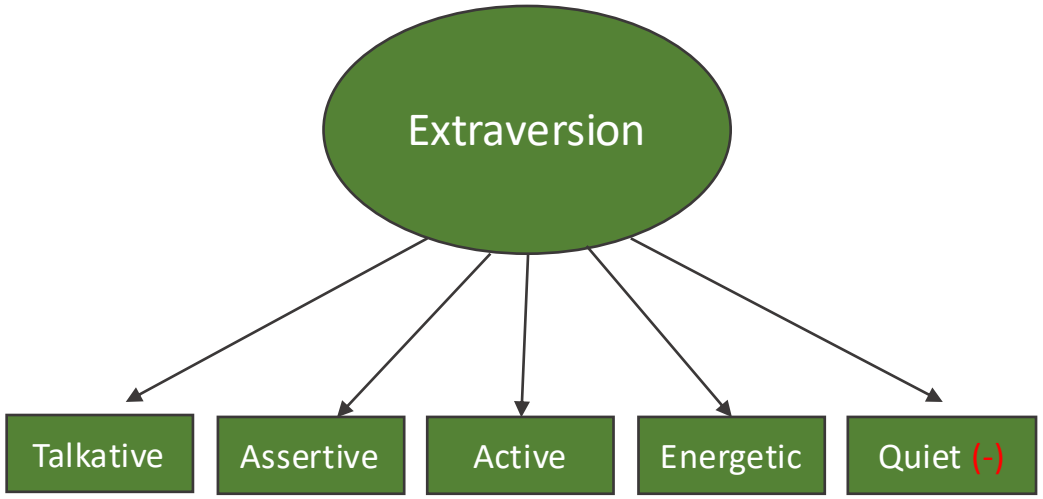


Personality structure is mostly considered from a top-down perspective—and the statistical techniques this approach involves reinforce the idea that broad constructs should be prioritized

Latent variable modeling as a default approach

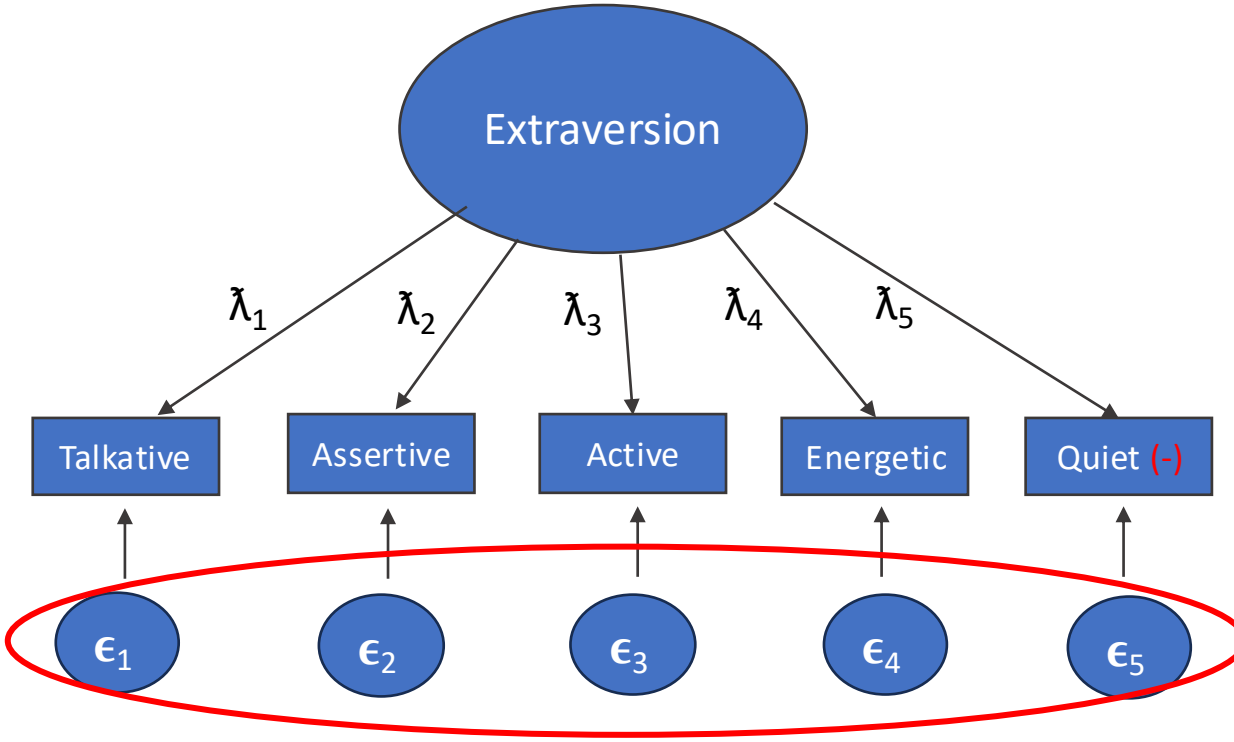
- Popular for representing and measuring personality attributes.
- The assumptions of latent variable models might not be warranted.

Theoretical assumptions



- Construct is a common cause of indicators
- Indicators are manifestations of the construct
- Indicators are interchangeable

Statistical assumptions

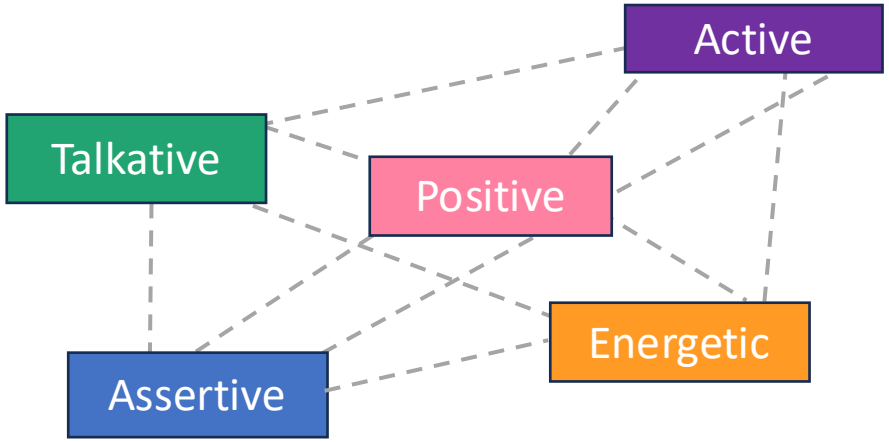


- Item-specific variance ("unique variance") is irrelevant and considered measurement error. It should be removed so only "true score" variance is captured.

Model choice influences how individual items are viewed

Do people show stable levels of extraversion because…

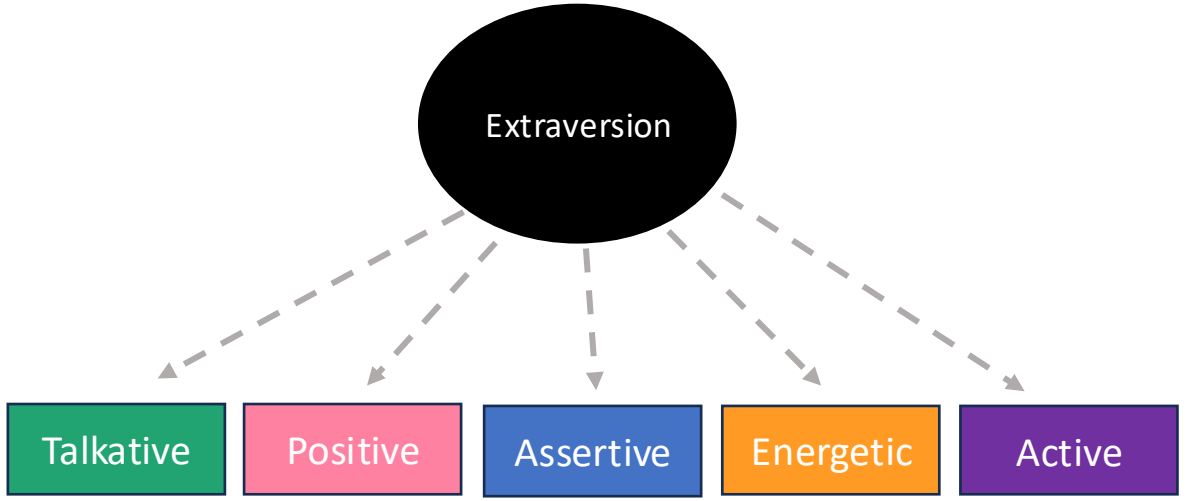
Its indicators are mutually reinforcing?



- **Individual items matter more**—causally relevant units that capture unique information about the system.

Cramer et al. (2012)

It is a proxy for stable neurophysiological activity?



- Range of personality tendencies associated with extraversion explained by a common cause, such as relatively stable neurophysiological activity (e.g., behavioral activation/inhibition systems).
- **Items may be seen as more interchangeable—**viewed in terms of relation to the broader latent construct.

Corr (2004)

Looking to the item-level to navigate model selection

- Establishing a stronger understanding of individual scale items—particularly their causal relations with each other (or lack thereof) and other variables—can help guide model selection (Rhemtulla et al., 2021)
  - Do items represent causally relevant parts of a dynamic system, or does it make more sense to treat them as interchangeable?

- Emerging perspectives on test validity also recommend a stronger focus on item-response processes and scale construction (Borsboom et al., [2003])
  - Current approaches emphasize collecting evidence for construct validity (Cronbach & Meehl, 1955)

Current methodological recommendations do not support a focus on the items

- Growing awareness of the consequences of measurement error (Cole & Preacher, 2014)
- Across many research areas, it is common to criticize any departure from LVM framework
- Psychology characterized by a "Latent-Variable Theory Psychological Testing Complex" (Condon & Mottus, 2021)
- LVM helped to address technical limitations (e.g., lack of statistical software) that no longer pose much of a challenge

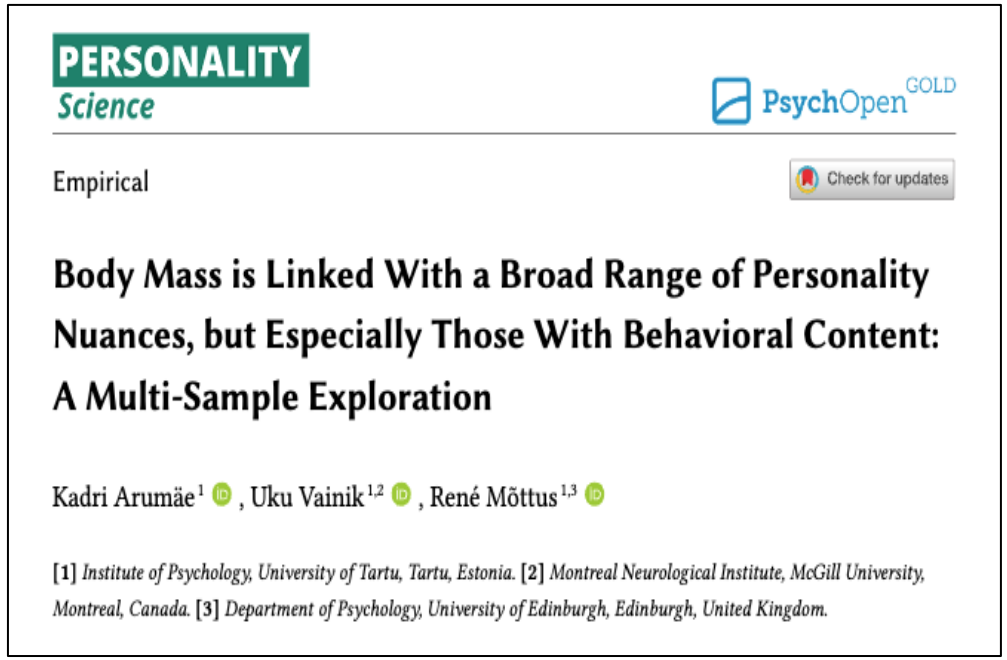
Item level assessment across research questions

Prediction	Description	Explanation

Item level assessment across research questions

Prediction	Description	Explanation

- More granular constructs tend to be stronger predictors of outcomes even after removing the variance associated with the broader domain (i.e., using *only* their unique variance; Mõttus et al., [2022]; Seeboth & Mõttus [2018]).



PERSONALITY Science

PsychOpen GOLD

Empirical

Check for updates

Body Mass is Linked With a Broad Range of Personality Nuances, but Especially Those With Behavioral Content: A Multi-Sample Exploration

Kadri Arumäe[1], Uku Vainik[1,2], René Mõttus[1,3]

[1] *Institute of Psychology, University of Tartu, Tartu, Estonia.* [2] *Montreal Neurological Institute, McGill University, Montreal, Canada.* [3] *Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom.*

- Administered NEO-PI-3, IPIP-NEO, and 100NP
- BMI showed strongest correlations with individual items relating to…
  - Orderliness (-)
  - Conservatism (+)
  - Impulsivity (+)
  - Yelling at others (+)
  - Talkativeness (+)
- Associations with broad domains were generally small and varied across inventory.

Item level assessment across research questions

Prediction	Description	Explanation

Item level assessment across research questions

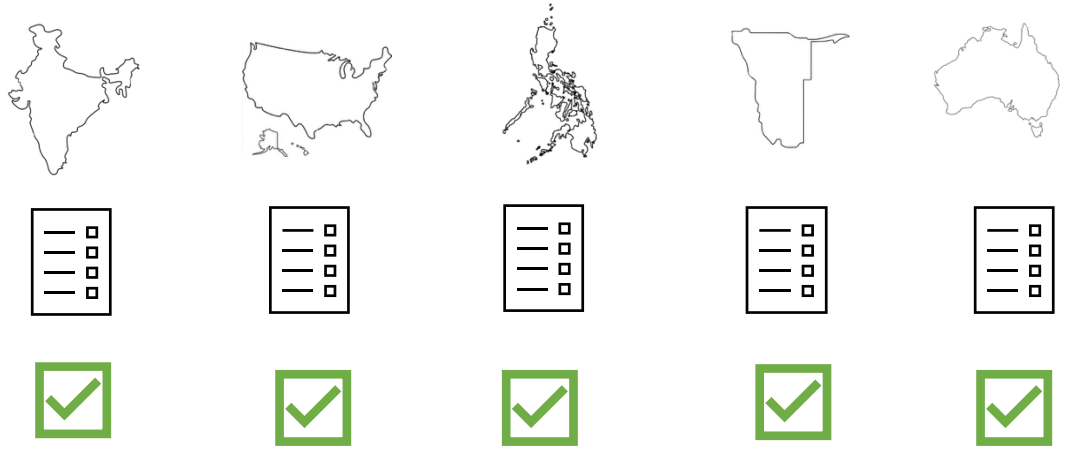
Prediction	Description	Explanation

- Descriptive findings provide the foundation for more prediction- and explanation-focused studies and should ideally be published in as much detail as possible (Mottus et al. 2020).
- Many studies rely on domain-level models, which lack of support for generalizability across nations, cultural contexts, and other groups (creates potential for misleading or uninformative findings).
- Evidence presented in prior work (taking an **etic approach**) does not justify claims of universality.

1. Identify a set of culturally diverse nations.

2. Administer big five or other measure developed in a western context and translated if applicable.

3. Similar factor structures across nations interpreted as evidence for generalizability.



- **The big five do not emerge spontaneously in lexical (emic) studies**
  - Thalmayer et al. (2021), Cheung et al. (2017), Wood et al. (2020

Item level assessment across research questions

Prediction	Description	Explanation

- Descriptive findings provide the foundation for more prediction- and explanation-focused studies and should ideally be published in as much detail as possible (Mottus et al. 2020).
- Many studies rely on domain-level models, which lack of support for generalizability across nations, cultural contexts, and other groups (creates potential for misleading or uninformative findings).
- Evidence presented in prior work (taking an etic approach) does not justify claims of universality.



"Big five has obtained the characteristics of an icon, with features that transcend the empirical facts"
– Boele de Raad (2010, p.2)

- Not always possible or ideal to take an emic approach
- Big two (model with the strongest generalizability) constructs might be more broad than ideal

Adopting a stronger item-level focus could be a strong etic strategy.

Item level assessment across research questions

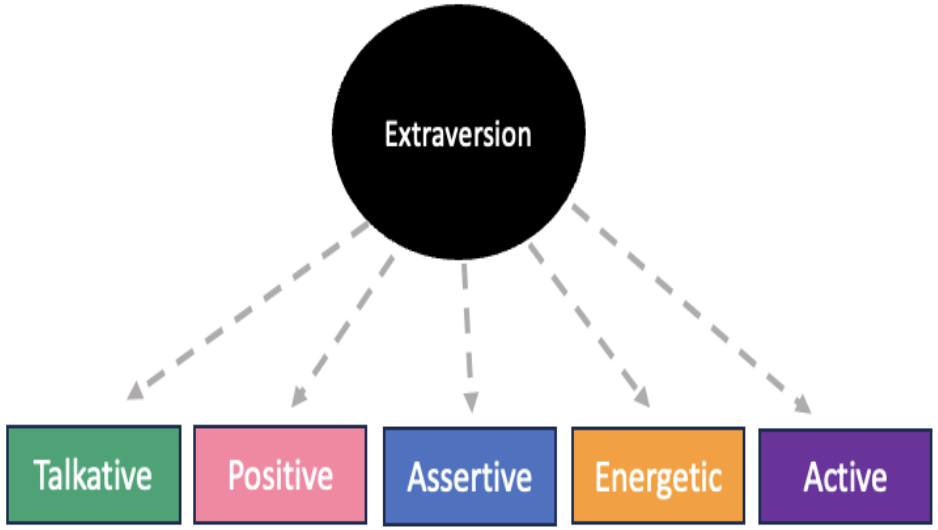
Prediction	Description	Explanation

Item level assessment across research questions

Prediction	Description	Explanation

- Explanation described as the main task of personality psychology in coming decades
  - Asendorpf, (2016); Baumert et al., (2017, 2019)
- Studies increasingly focus on personality "dynamics", "functioning", "processes", and "mechanisms".

Researchers generally agree that factorial traits are not explanations, but phenomena that need to be explained…



Extraversion used as explanation for attributes like talkativeness, assertiveness, high-energy levels, positive emotions, and their covariation.

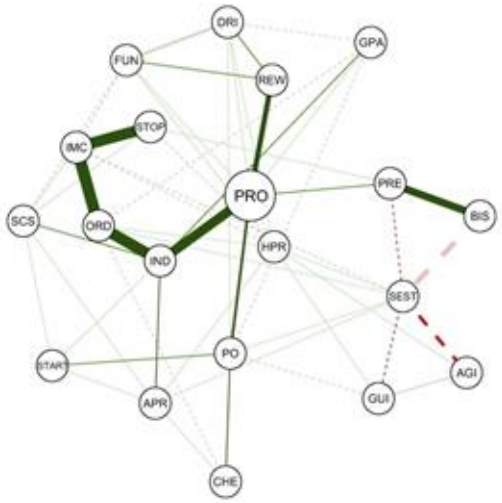
Talkativeness, assertiveness, high energy levels, and positive emotionality are **attributes used to define Extraversion.**

Item level assessment across research questions

Prediction	Description	Explanation

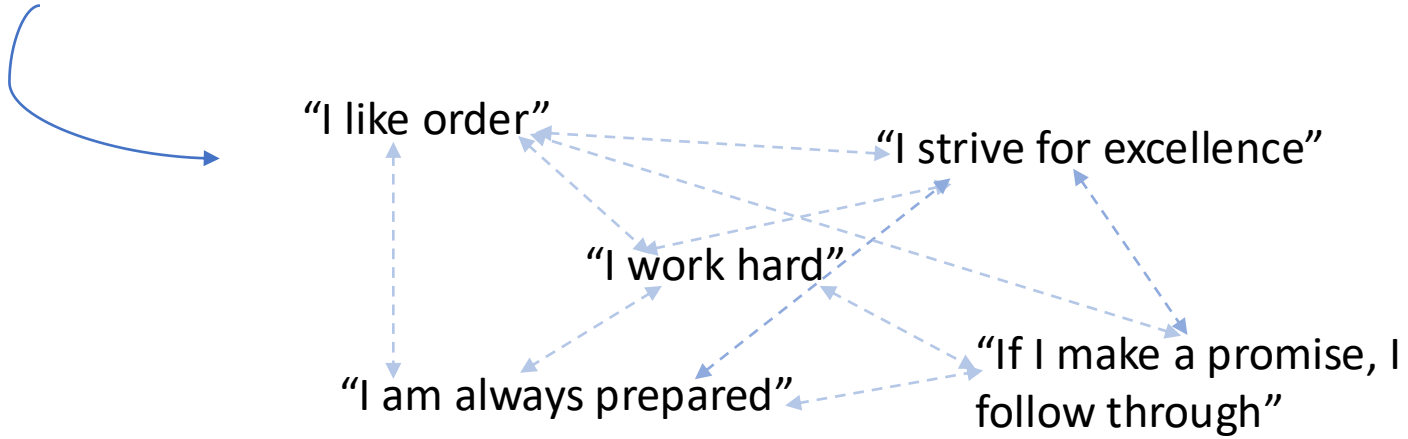
There is little agreement on the relationship between the nature of the relationship between factorial traits and personality processes.

- <u>One common approach</u>: Considering domain-level constructs as dynamic processes



Is conscientiousness a network of mutually reinforcing indicators?

- Avoids circularity issue
- reasonable considering trait scales are made up of affective, behavioral, cognitive, and goal-related item content (which may be mutually reinforcing).



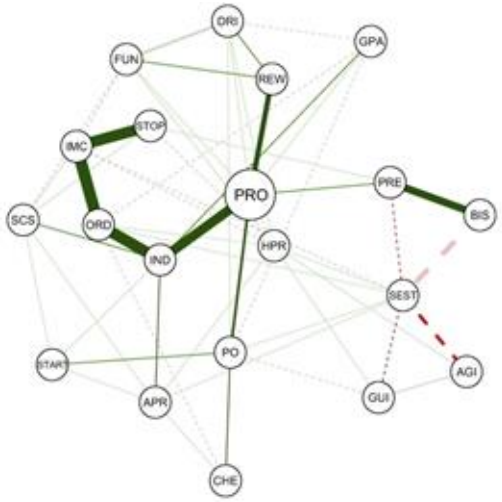
Costantini & Perugini (2016)

Item level assessment across research questions

Prediction	Description	Explanation

There is little agreement on the relationship between the nature of the relationship between factorial traits and personality processes.

- <u>One common approach</u>: Considering domain-level constructs as dynamic processes



Is conscientiousness a network of mutually reinforcing indicators?

- Avoids circularity issue
- reasonable considering trait scales are made up of affective, behavioral, cognitive, and goal-related item content (which may be mutually reinforcing).

However, this approach…

- Assumes all relevant content is captured in broad trait
- Proposed dynamic system based on population-level trait (Personality structure at the individual level differs from structure at the population level).

Starting from the item level could be an alternative approach for exploring within person processes

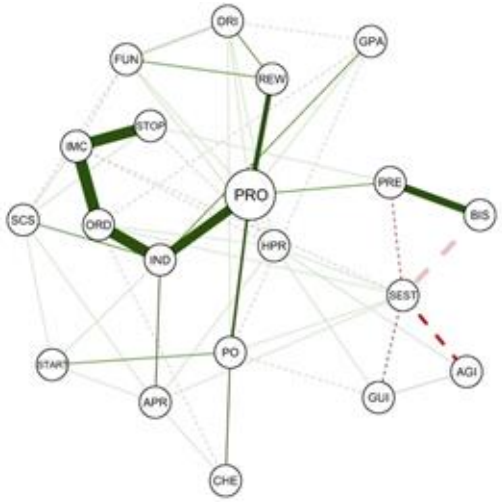
Costantini & Perugini (2016)

Item level assessment across research questions

Prediction	Description	Explanation

There is little agreement on the relationship between the nature of the relationship between factorial traits and personality processes.

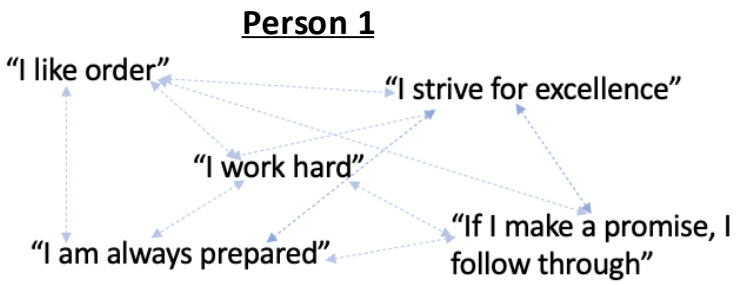
- <u>One common approach</u>: Considering domain-level constructs as dynamic processes

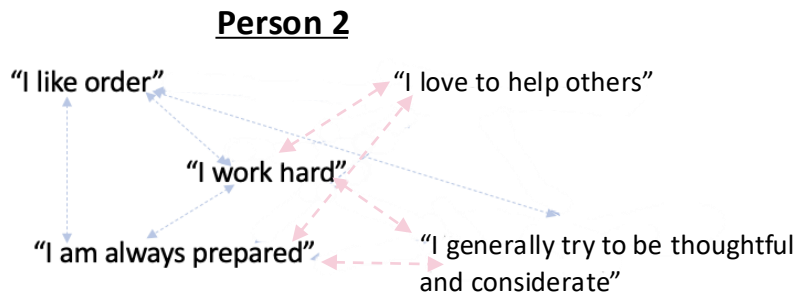


Costantini & Perugini (2016)

Is conscientiousness a network of mutually reinforcing indicators?

- Avoids circularity issue
- reasonable considering trait scales are made up of affective, behavioral, cognitive, and goal-related item content (which may be mutually reinforcing).





Starting from the item level could be an alternative approach for exploring within person processes

Chapter 2

Demonstrating the use of the Dowd Test Statistic for addressing item-level research questions

Statistical approaches for focusing on the item-level

Personality items have many trait-like properties…

- Stability over time (Mõttus et al., 2017)
- Cross-rater agreement (Mõttus et al., 2014, 2019)
- Heritability (Mõttus et al., 2017)
- Unique developmental trajectories and associations with outcomes (Mõttus et al., 2019)

However, It remains somewhat uncommon to focus on the item-level...

- Items are viewed mostly in terms of their relationship to broader constructs
- Most popular statistical techniques focus on the scale-level––unique item variance regarded as measurement error/irrelevant
- Few statistical options suitable for item-level assessment beyond descriptive statistics

Statistical approaches for focusing on the item-level

Traditional approaches fall short for item-focused research examining generalizability…

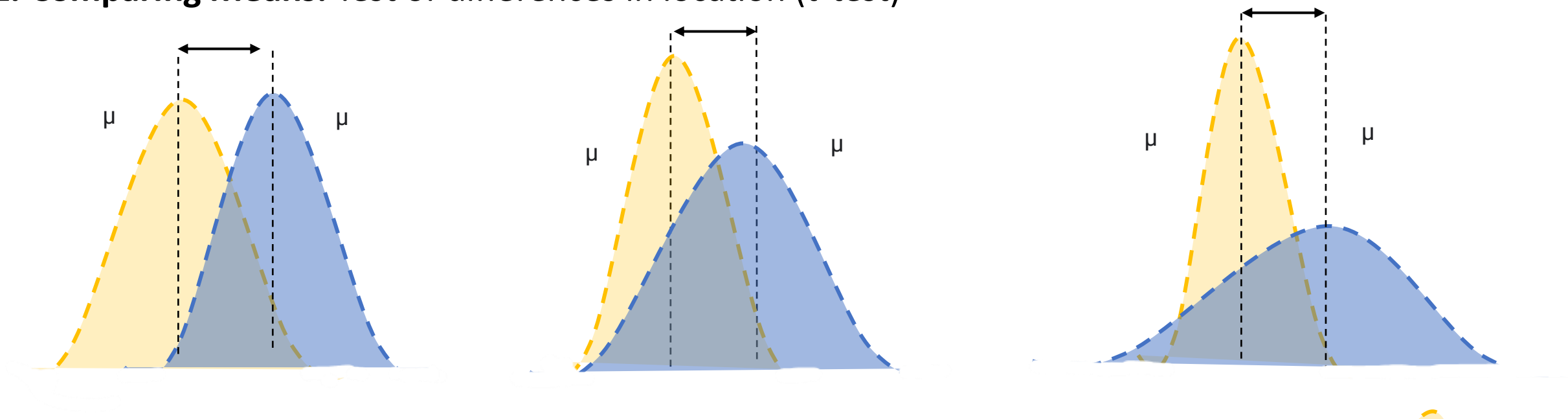
- These approaches assume a common, latent construct.
  - **Multigroup confirmatory factor analysis (CFA)**
    - **Configural invariance** establishes similar factor structure (i.e., number/similarity of constructs)
    - **Metric invariance** establishes that factor loadings are equal across groups
    - **Scalar invariance** establishes equal item intercepts across groups
  - **Differential Item Functioning (DIF)** aims primarily to identify items that function differently across groups––in reference to the underlying latent construct.

Chapter 2 demonstrates an approach for comparing groups at the item level

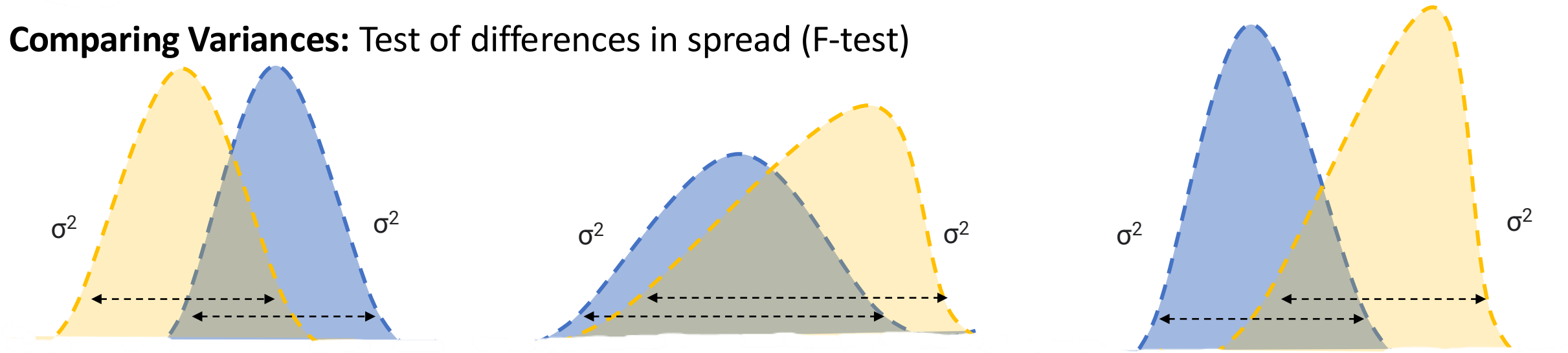
- Research question: How do nations compare on distinct but overlapping features of well-being?

Comparing item-level responding across groups: two well-known options

1. Comparing Means: Test of differences in location (t-test)



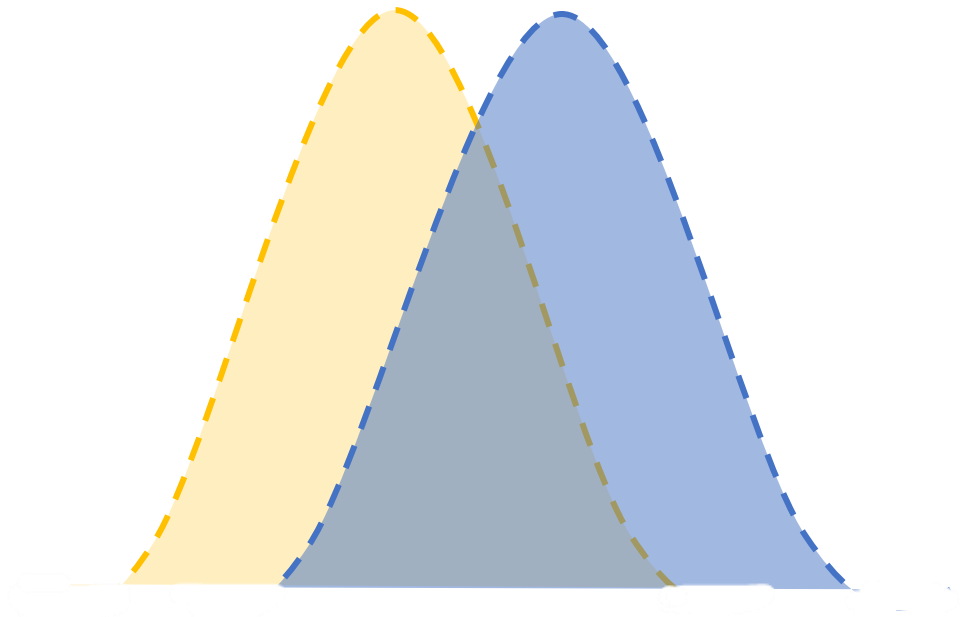
2. Comparing Variances: Test of differences in spread (F-test)



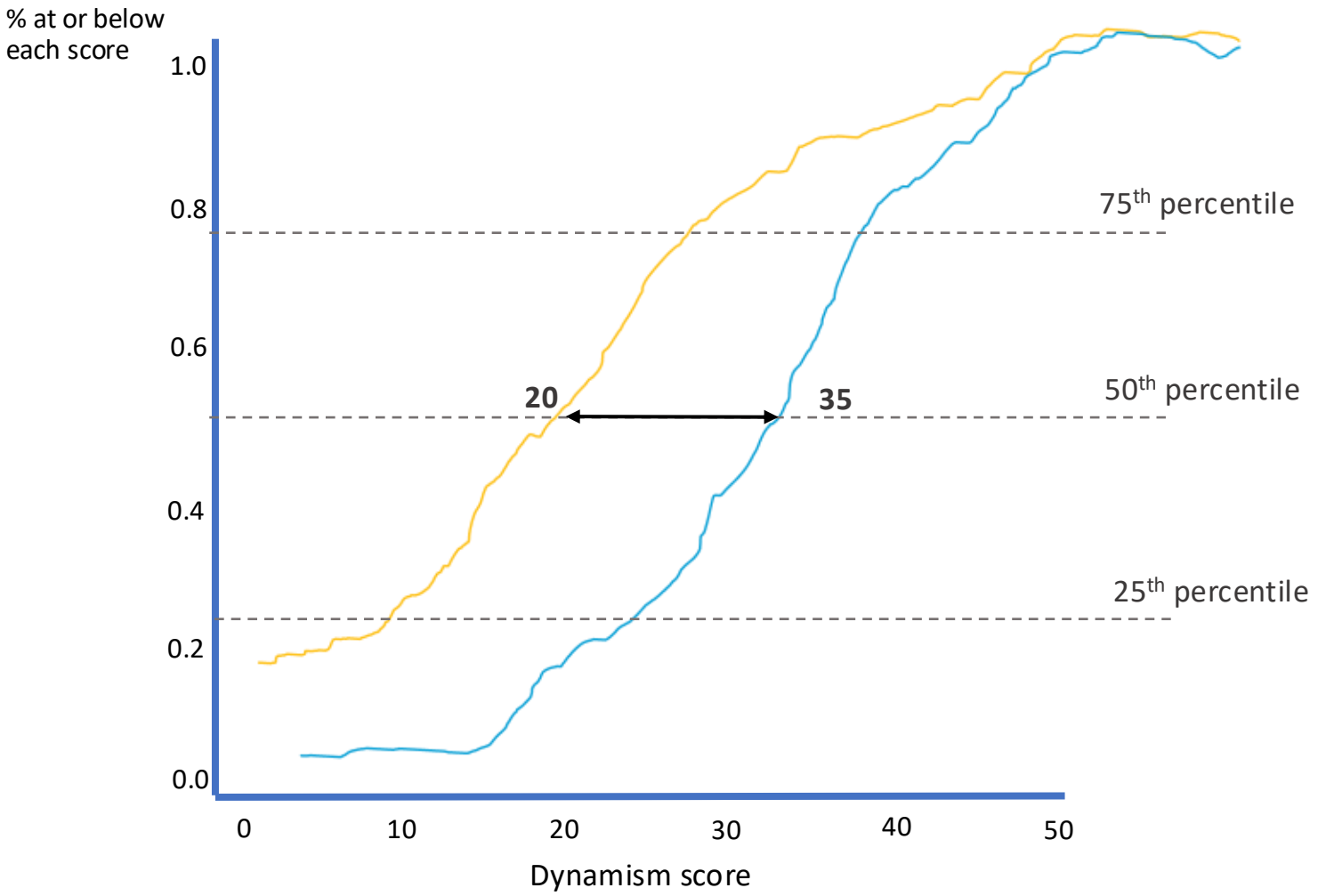
Comparing entire distributions using a single test

Dynamism: A tendency to be energetic, enthusiastic, confident, and socially bold.

Distribution of Dynamism scores for two samples



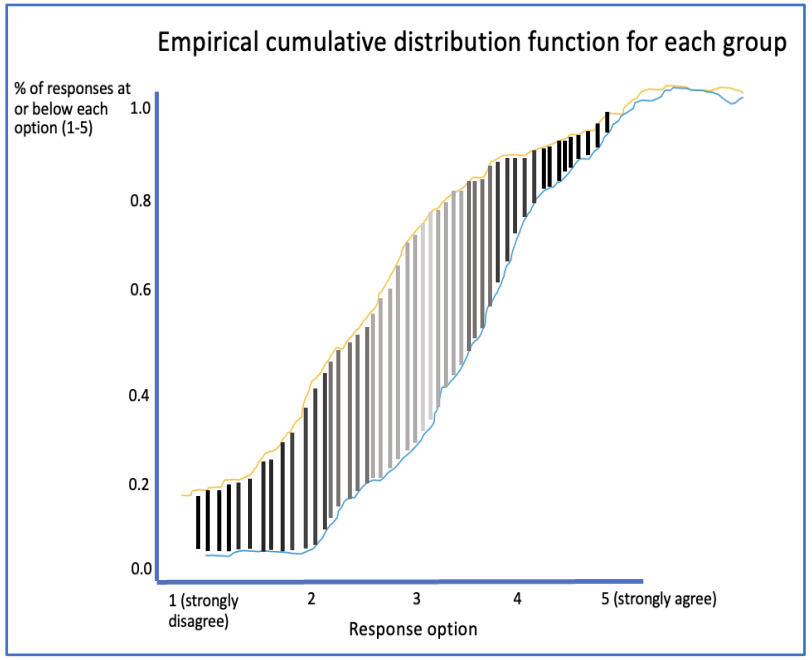
Empirical cumulative distribution function (ECDF) for each group



Comparing entire distributions using a single test: Several statistical options for comparing ECDFs

Significant *p*-value (*p*<.05) indicates that the difference between the ECDFs is fairly extreme if we assume that that samples are drawn from the same population

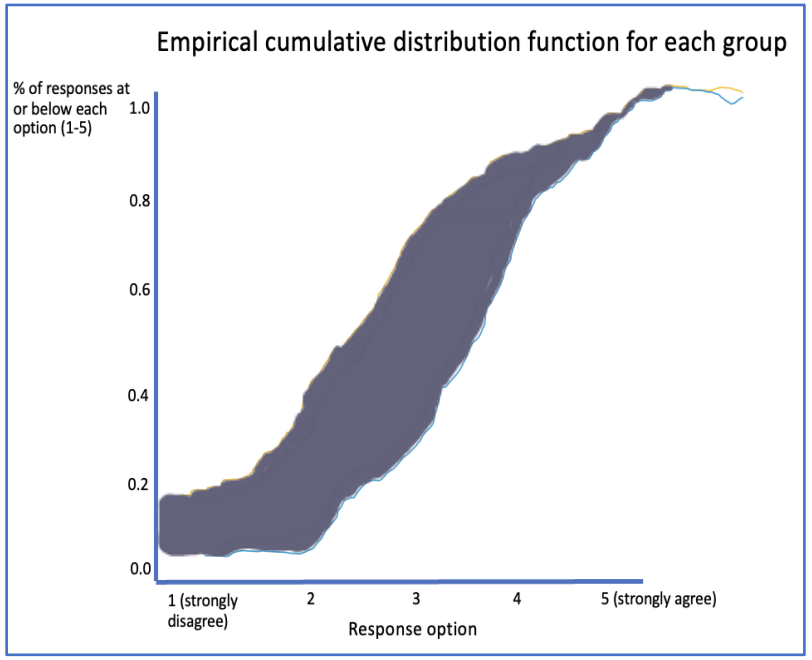
Anderson-Darling (AD) statistic



The weighted sum of all observed differences (Anderson & Darling, 1952)

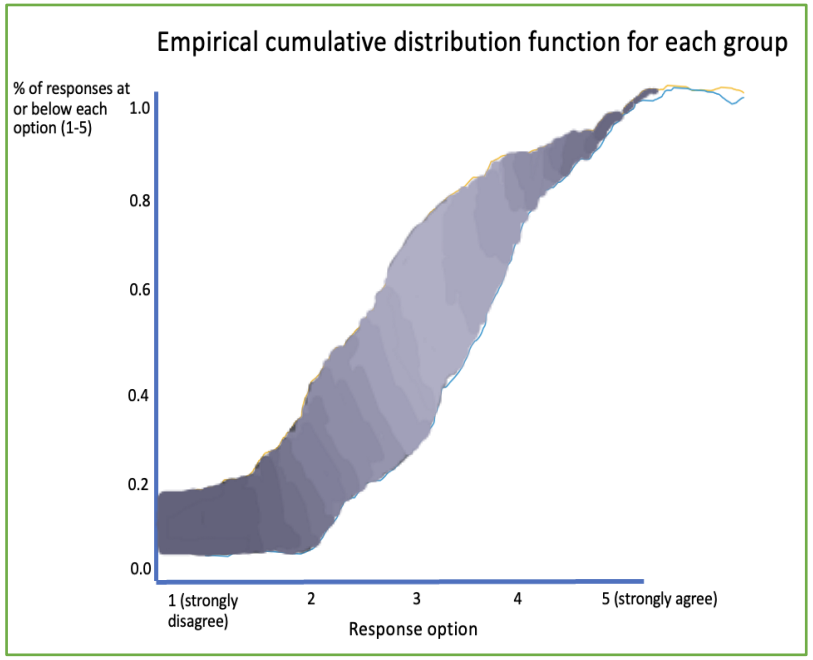
Putting less weight on observations in the middle of the distribution is done to improve the power of the test.

Wasserstein statistic



The area between the ECDFs (Wasserstein, 1969)

Dowd Test Statistic ☑



Synthesis of AD and Wasserstein: The weighted integral of the difference between the ECDFs (Dowd, 2020)

The Dowd Test Statistic (DTS)

$$DTS = \int_{-\infty}^{\infty} \frac{|\hat{F}(x) - \hat{E}(x)|}{\hat{D}(x)(1 - \hat{D}(x))} dx$$

The Dowd Test Statistic (DTS)

ECDF for sample 1
(e.g., Philippines)

ECDF for sample 2
(e.g., United States)

$$DTS = \int_{-\infty}^{\infty} \frac{|\hat{F}(x) - \hat{E}(x)|}{\hat{D}(x)(1-\hat{D}(x))} dx$$

Combined ECDF

- Used to estimate variance at each point, x

Test uses a domain classifier that predicts whether each response option, x––e.g., 1 (strongly disagree) through 6 (strongly agree)––came from one sample or the other

The Dowd Test Statistic (DTS)

ECDF for sample 1
(e.g., Philippines)

ECDF for sample 2
(e.g., United States)

$$DTS = \int_{-\infty}^{\infty} \frac{|\hat{F}(x) - \hat{E}(x)|}{\hat{D}(x)(1 - \hat{D}(x))} dx$$

Combined ECDF

- Used to estimate variance at each point, x

Example: for the item *I am quiet around strangers*—rated from 1 (strongly disagree) to 6 (strongly agree)….

- If D(3) = .9, model predicts the response of 3 came from the Philippines
- If D(3) = .1, model predicts the response of 3 came from the United States
- If D(3) = around .5, the model is uncertain which sample the response came from

The Dowd Test Statistic (DTS)

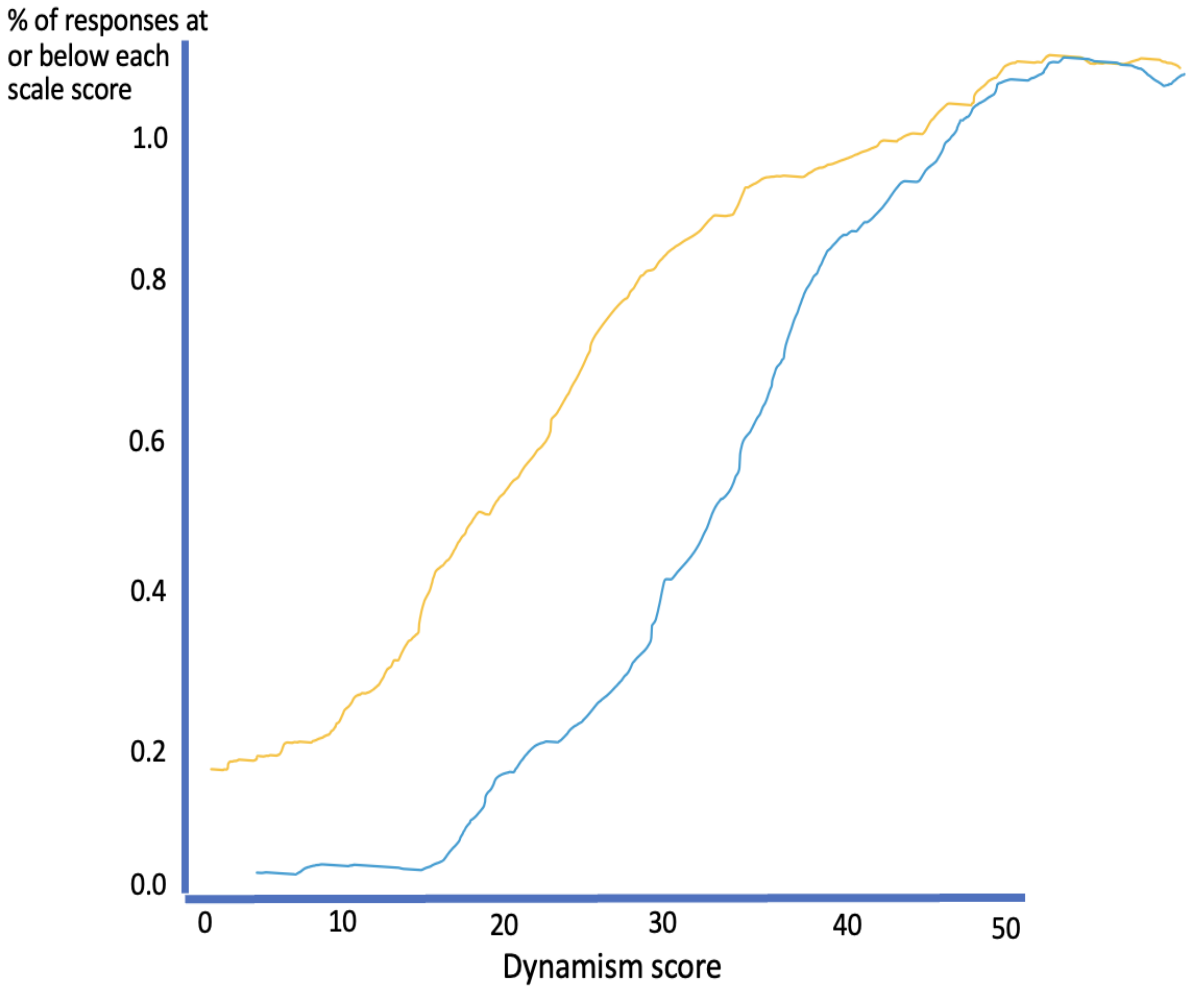
$$DTS = \int_{-\infty}^{\infty} \frac{|\hat{F}(x) - \hat{E}(x)|}{\hat{D}(x)(1-\hat{D}(x))} dx$$

When the classifier is confident about which group the response came from (i.e., closer to .1 or .9), the denominator is smaller, which contributes to a larger DTS statistic.

ECDFs for scale scores vs. item scores

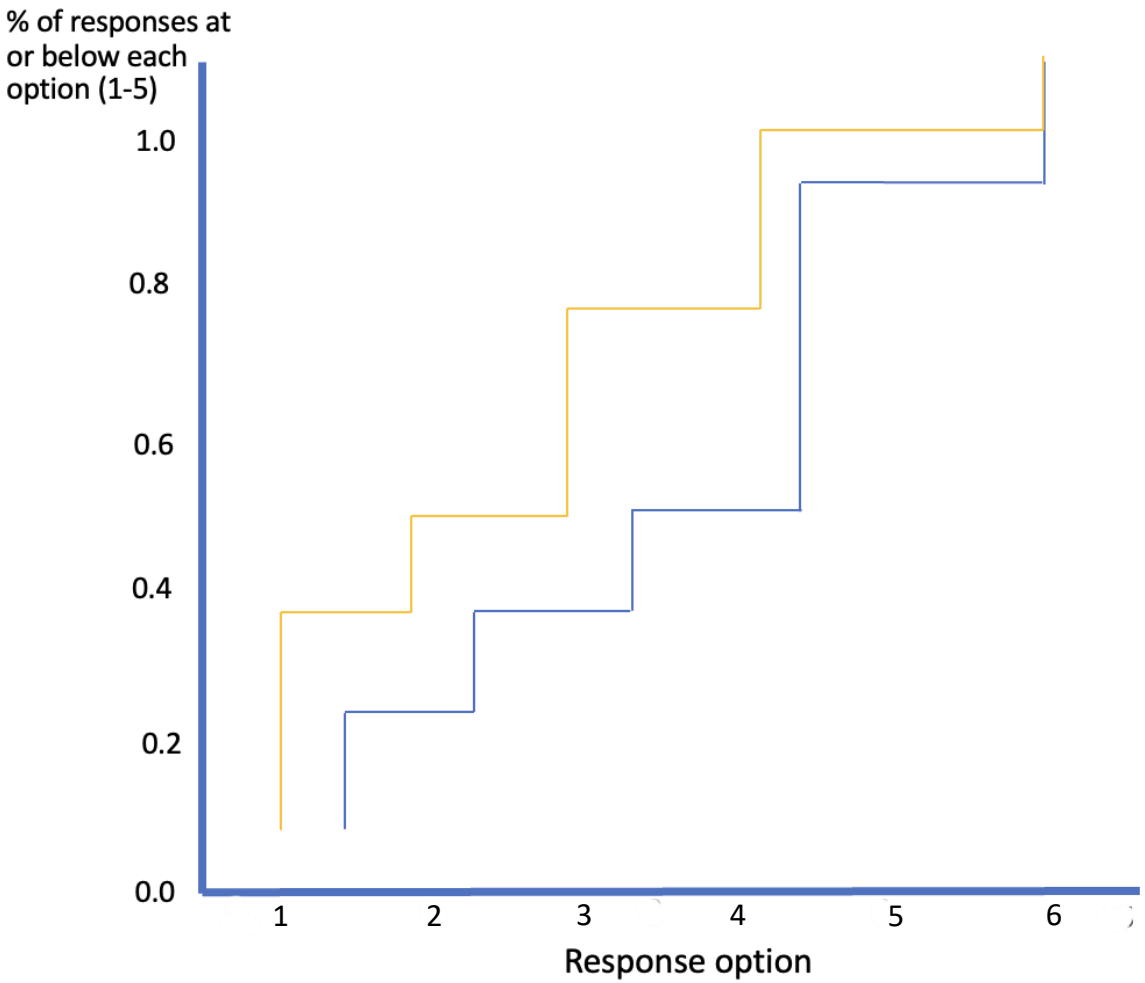
Empirical cumulative distribution function for each group

Dynamism scale score



Empirical cumulative distribution function for each group

Dynamism item: "Tend to be confident"

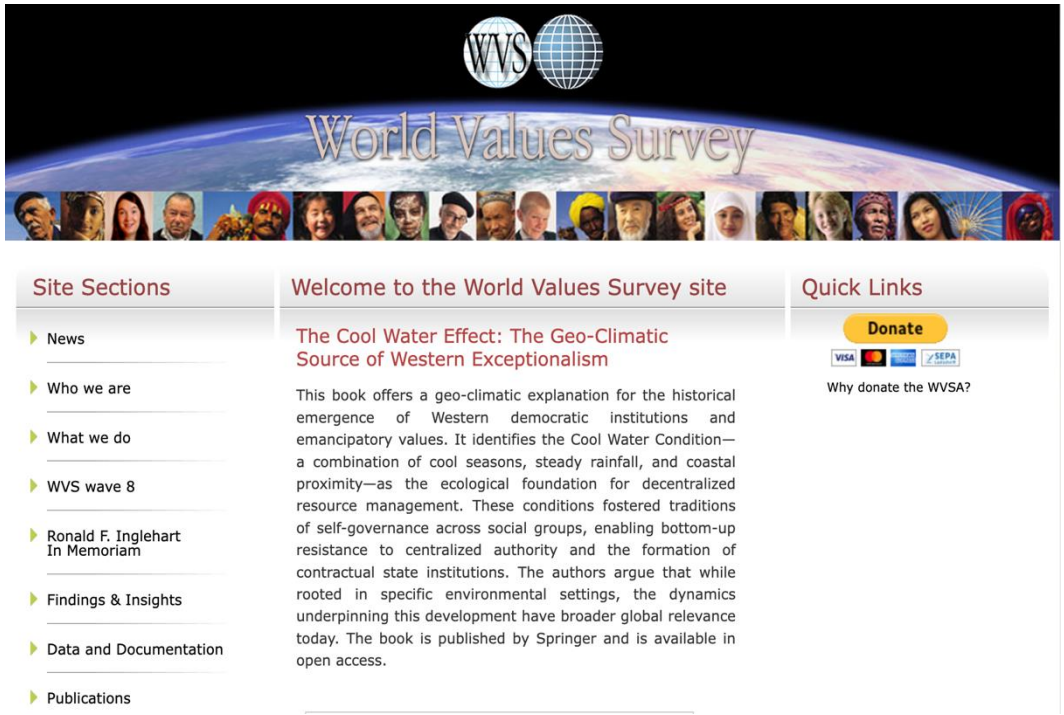


Participants: World Values Survey

Research question 1: How do nations differ on various features of well-being?

- **Total *N* = 13,627** participants from Wave 7 (data collection began in 2017, but only used data from 2021-2023)
- Nations: Armenia (*n* = 1223), Czechia (*n* = 1200), India (*n* = 1692), Kenya (*n* = 1259), Libya (*n* = 1185), Maldives (*n* = 1036), Morocco (*n* = 1200), Netherlands (*n* = 2145), Slovakia (*n* = 1200), Great Britain (*n* = 2469), Uruguay (*n* = 1000), Uzbekistan (*n* = 1249), Venezuela (*n* = 1190)

78 nation pairs (across 13 nations) x 5 items = 390 DTS tests



Research question 2: How have different features of well-being changed within nations over time?

- **Total N = 22,578** (from Waves 6 and 7. Wave 6 data collection occurred from 2010-2014)
- Nations: Armenia (*n* = 2,323), India (*n* = 5,743), Libya (*n* = 3,316), Morocco (*n* = 2,400), Netherlands (*n* = 4,047), Uruguay (*n* = 2,000), and Uzbekistan (*n* = 2,749).

7 nations x 5 items = 35 DTS tests

World values survey collects data on topics like…

- Social values, attitudes
- **Happiness and well-being**
- Economic values
- Science and technology
- National security
- Ethical norms

World Values Survey: Happiness and Well-being Items

HAPPINESS AND WELL-BEING

Q46. Taking all things together, would you say you are (*read out and code one answer*):

1 Very happy
2 Rather happy
3 Not very happy
4 Not at all happy

Q47. All in all, how would you describe your state of health these days? Would you say it is… (*read out*):

1 Very good
2 Good
3 Fair
4 Poor
5 Very poor

Q48. Some people feel they have completely free choice and control over their lives, while other people feel that what they do has no real effect on what happens to them. Please use this scale where 1 means "no choice at all" and 10 means "a great deal of choice" to indicate how much freedom of choice and control you feel you have over the way your life turns out (*code one number*):

No choice at all									A great deal of choice
1	2	3	4	5	6	7	8	9	10

(SHOW CARD 5)

Q49. All things considered, how satisfied are you with your life as a whole these days? Using this card on which 1 means you are "completely dissatisfied" and 10 means you are "completely satisfied" where would you put your satisfaction with your life as a whole? (*Code one number*):

Completely dissatisfied									Completely satisfied
1	2	3	4	5	6	7	8	9	10

(SHOW CARD 6)

Q50. How satisfied are you with the financial situation of your household? Please use this card again to help with your answer (*code one number*):

Completely dissatisfied									Completely satisfied
1	2	3	4	5	6	7	8	9	10

- Reverse scored Happiness and Health items for consistent interpretation

- Some items could reasonably be included in a measure of the same broader construct (e.g., happiness and overall life satisfaction)

- Other items have less conceptual overlap with the others (e.g., satisfaction with one's financial situation)

Results

RQ1: Cross-national comparisons

RQ2: Within-nation comparisons

Results

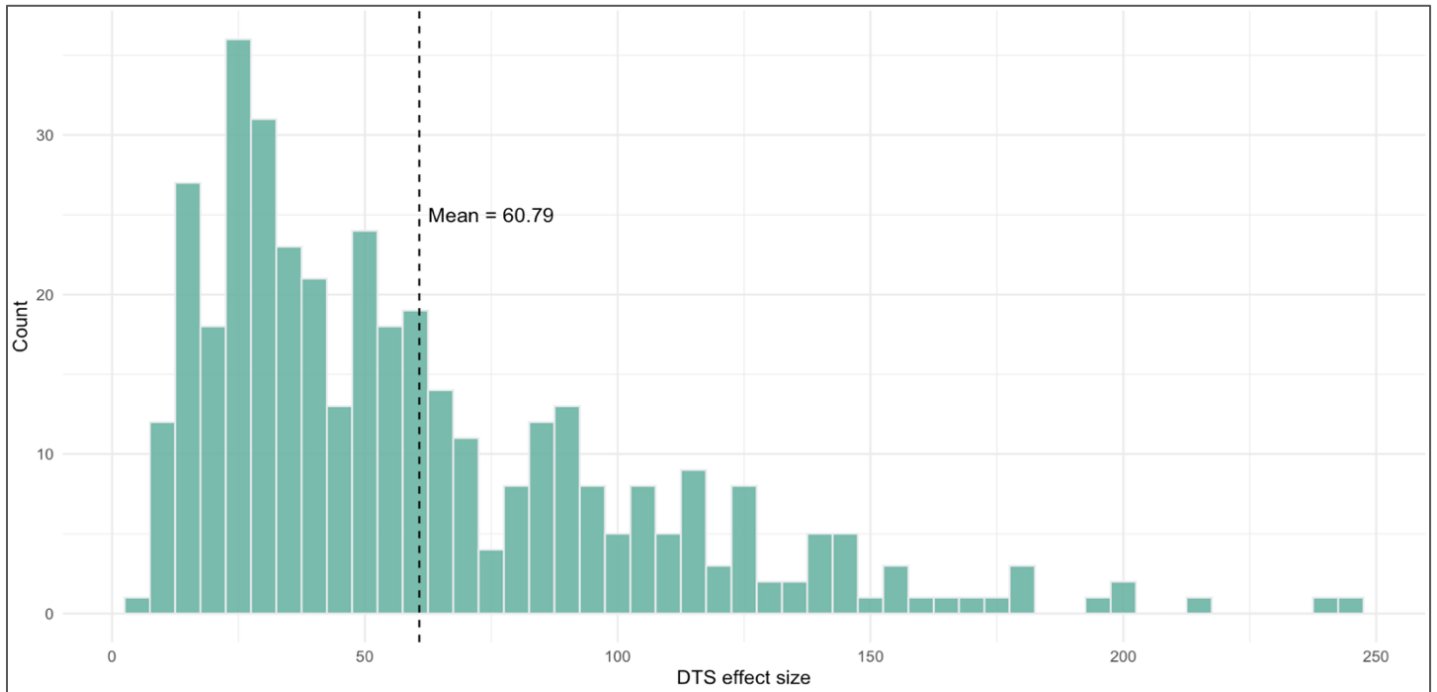
RQ1: Cross-national comparisons

RQ2: Within-nation comparisons

383 of the 390 DTS tests (98%) were statistically significant ($p < .05$)….

- DTS Effect sizes ranged from 7.36 to 244.56 ($M = 60.79$; $SD = 43.51$)

Histogram of statistically significant DTS effect sizes



DTS Effect sizes are interpreted in a relative way (i.e., no "objective" benchmarks). Here….

- effect sizes from 0 to 10 might be thought of as small
- Effect sizes over 100 might be thought of as large

Results

RQ1: Cross-national comparisons

RQ2: Within-nation comparisons

383 of the 390 DTS tests (98%) were statistically significant ($p < .05$)….

- DTS Effect sizes ranged from 7.36 to 244.56 ($M = 60.79$; $SD = 43.51$)

Non-significant DTS comparisons—indicating no difference between the nations' distributions

Item	Content	Nation 1	Nation 2	DTS effect size	P-value
Q46	Happiness	Czech Rep	Morocco	4.11	0.27
Q47	Health	Kenya	Uzbekistan	7.39	0.07
Q47	Health	Uruguay	Venezuela	6.21	0.19
Q47	Health	Czech Rep	Maldives	5.05	0.33
Q48	Freedom	Czech Rep	Slovakia	10.42	0.40
Q49	Life Satisfaction	Uzbekistan	Venezuela	18.81	0.05
Q49	Life Satisfaction	India	Venezuela	16.74	0.07
Q50	Financial Situation	Maldives	Uruguay	13.80	0.17

Results

RQ1: Cross-national comparisons

Happiness

Taking all things together, would you say you are…
1 (Not at all happy) to 4 (Very happy)

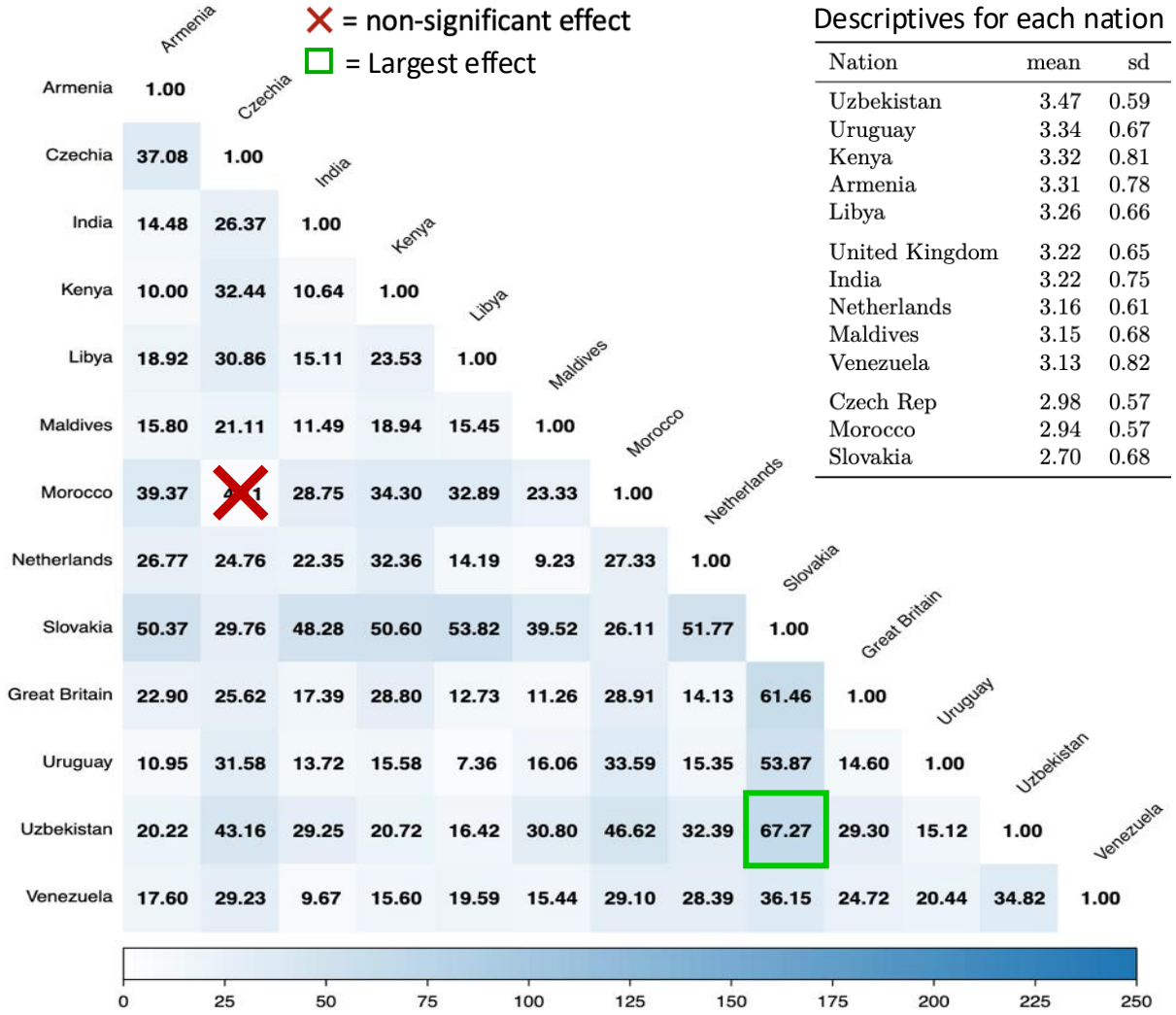
- Overall: *M* = 3.17, *SD* = 0.70
- Highest mean happiness reported by **Uzbekistan** (*M* = 3.47; *SD* = 0.59)
- Lowest mean happiness reported by **Slovakia** (*M* = 2.70; *SD* = 0.68)

Most DTS effect sizes were relatively small

- Ranged from 4.11 to 67.27 (*M* = 26.15, *SD* = 0.70)
- All tests significant except for Czech Republic and Morocco

Consistent with previous findings establishing limited variability in happiness across contexts…

RQ2: Within-nation comparisons



Descriptives for each nation

Nation	mean	sd
Uzbekistan	3.47	0.59
Uruguay	3.34	0.67
Kenya	3.32	0.81
Armenia	3.31	0.78
Libya	3.26	0.66
United Kingdom	3.22	0.65
India	3.22	0.75
Netherlands	3.16	0.61
Maldives	3.15	0.68
Venezuela	3.13	0.82
Czech Rep	2.98	0.57
Morocco	2.94	0.57
Slovakia	2.70	0.68

Results

RQ1: Cross-national comparisons

Health

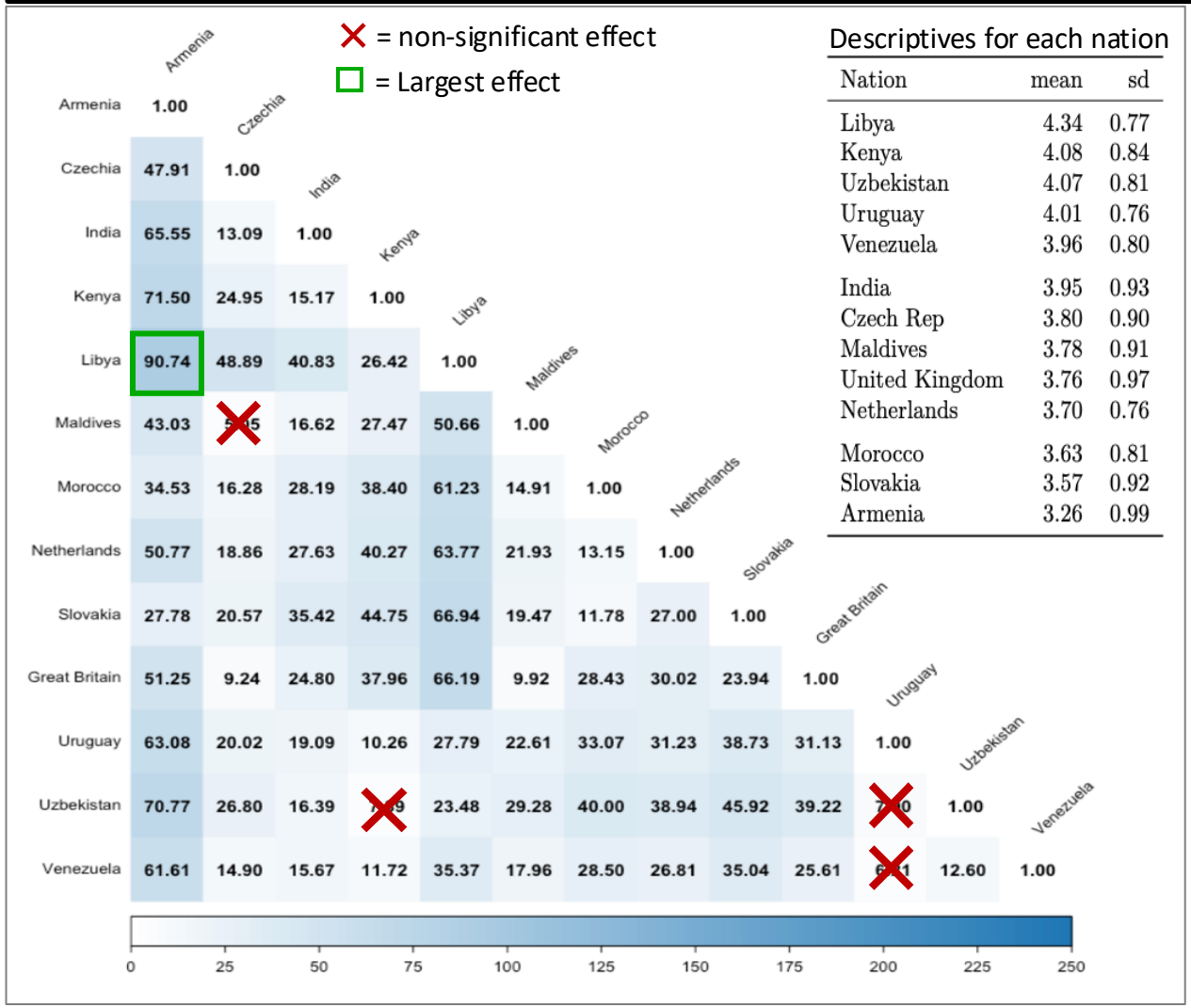
All in all, how would you describe your state of health these days?....
From 1 (Very poor) to 5 (Very good)

- Overall: *M* = 3.83, *SD* = 0.90
- Highest mean health reported by **Libya** (*M* = 4.43; *SD* = 0.77)
- Lowest mean health reported by **Armenia** (*M* = 3.26; *SD* = 0.99)

Most DTS effect sizes were relatively small, but slightly more varied....

- Ranged from 5.05 to 90.74 (*M* = 31.90, *SD* = 18.30)
- All tests significant except for...
  - Uzbekistan and Uruguay, Uzbekistan and Kenya, Uruguay and Venezuela, Czech Republic and the Maldives

RQ2: Within-nation comparisons



Descriptives for each nation

Nation	mean	sd
Libya	4.34	0.77
Kenya	4.08	0.84
Uzbekistan	4.07	0.81
Uruguay	4.01	0.76
Venezuela	3.96	0.80
India	3.95	0.93
Czech Rep	3.80	0.90
Maldives	3.78	0.91
United Kingdom	3.76	0.97
Netherlands	3.70	0.76
Morocco	3.63	0.81
Slovakia	3.57	0.92
Armenia	3.26	0.99

Results

RQ1: Cross-national comparisons

Perceived freedom

….indicate how much freedom of choice and control you feel you have over the way your life turns out
From 1 (no choice at all) to 10 (a great deal of choice)

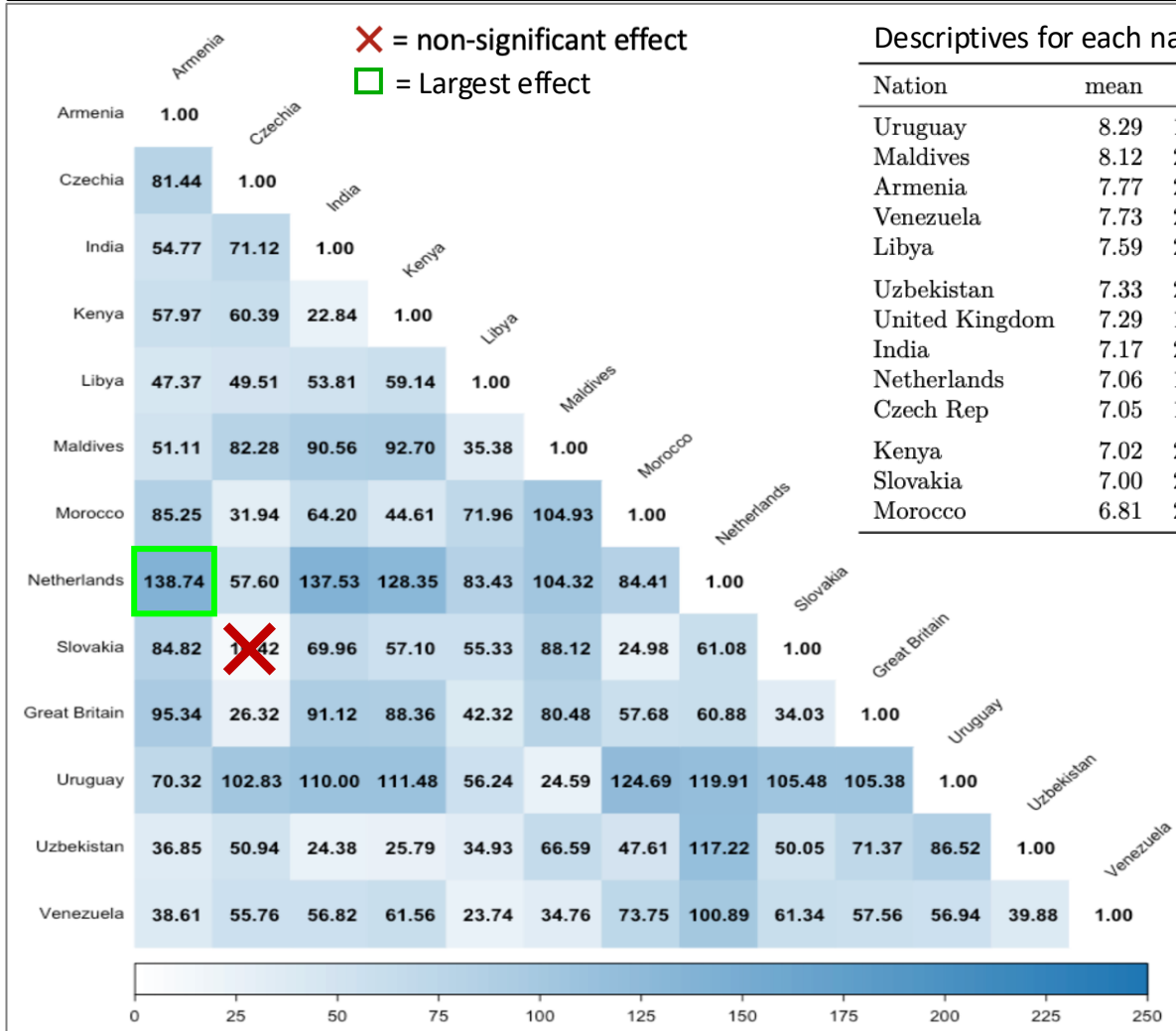
- Overall: *M* = 7.35, *SD* = 2.23
- Highest mean perceived freedom reported by **Uruguay** (*M* = 8.29; *SD* = 1.83)
- Lowest mean perceived freedom reported by **Morocco** (*M* = 6.81; *SD* = 2.16)

DTS effect sizes were larger, and more varied….

- Ranged from 10.42 to 138.74 (*M* = 67.70, *SD* = 29.84)
- All tests significant except for Czech Republic and Slovakia

Differences appear to be more driven by differences in variances than means

RQ2: Within-nation comparisons



Descriptives for each nation

Nation	mean	sd
Uruguay	8.29	1.83
Maldives	8.12	2.06
Armenia	7.77	2.62
Venezuela	7.73	2.25
Libya	7.59	2.21
Uzbekistan	7.33	2.57
United Kingdom	7.29	1.94
India	7.17	2.67
Netherlands	7.06	1.54
Czech Rep	7.05	1.98
Kenya	7.02	2.59
Slovakia	7.00	2.02
Morocco	6.81	2.16

Results

RQ1: Cross-national comparisons

RQ2: Within-nation comparisons

Overall life satisfaction

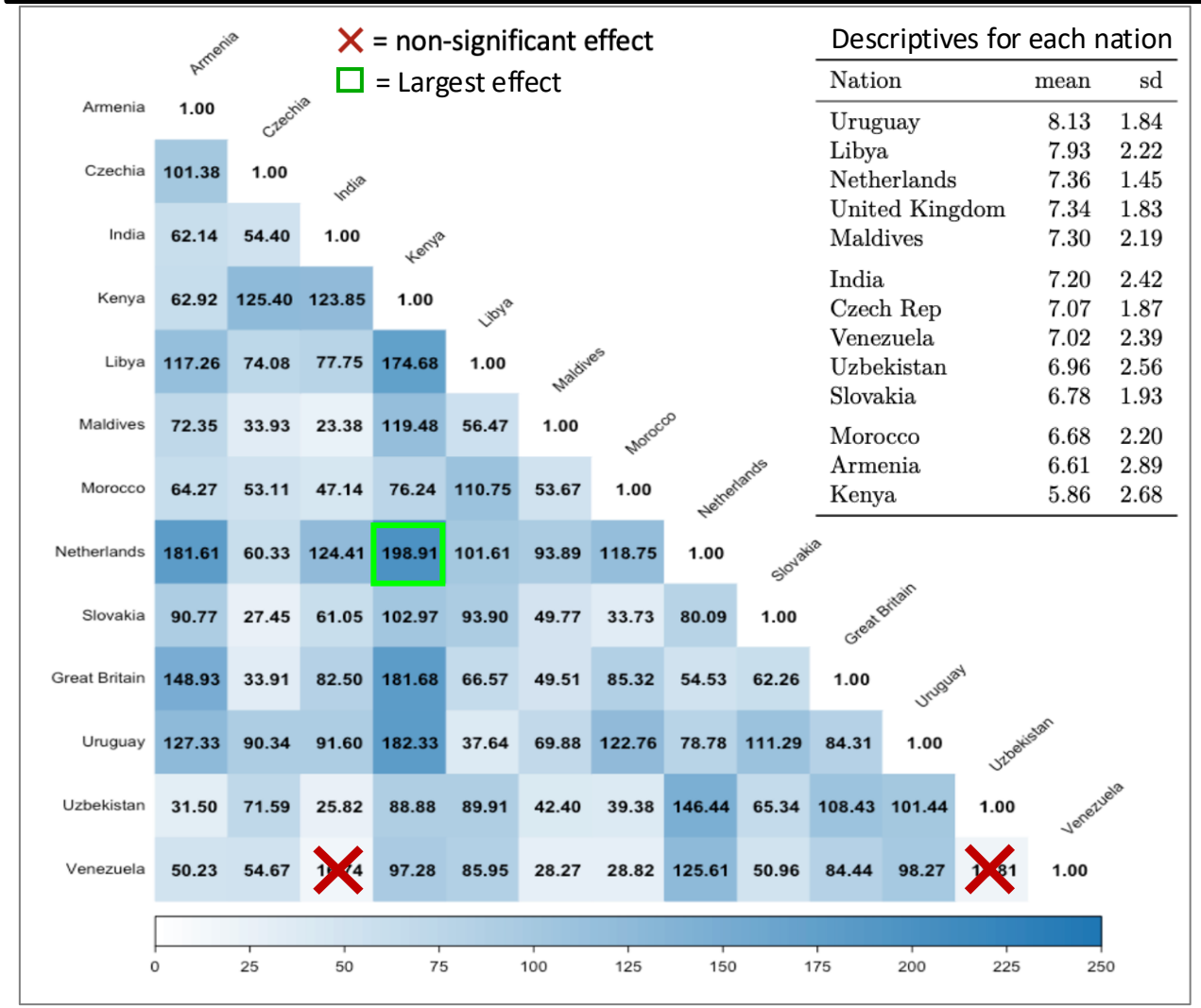
All things considered, how satisfied are you with your life as a whole these days?
From 1 (completely dissatisfied) to 10 (completely satisfied)

- Overall: *M* = 7.11, *SD* = 2.24
- Highest mean life satisfaction reported by **Uruguay** (*M* = 8.13; *SD* = 1.84)
- Lowest mean life satisfaction reported by **Kenya** (*M* = 7.28; *SD* = 1.69)

DTS effect sizes were larger, and more varied….

- Ranged from 16.74 to 198.91 (M = 82.26, SD = 41.49)
- All tests significant except for two for Venezuela with Uzbekistan and India

Nations with higher DTS effect sizes overall tended to show much higher or lower variance than other nations



Descriptives for each nation

Nation	mean	sd
Uruguay	8.13	1.84
Libya	7.93	2.22
Netherlands	7.36	1.45
United Kingdom	7.34	1.83
Maldives	7.30	2.19
India	7.20	2.42
Czech Rep	7.07	1.87
Venezuela	7.02	2.39
Uzbekistan	6.96	2.56
Slovakia	6.78	1.93
Morocco	6.68	2.20
Armenia	6.61	2.89
Kenya	5.86	2.68

Results

RQ1: Cross-national comparisons

RQ2: Within-nation comparisons

Financial satisfaction

How satisfied are you with the financial situation of your household?

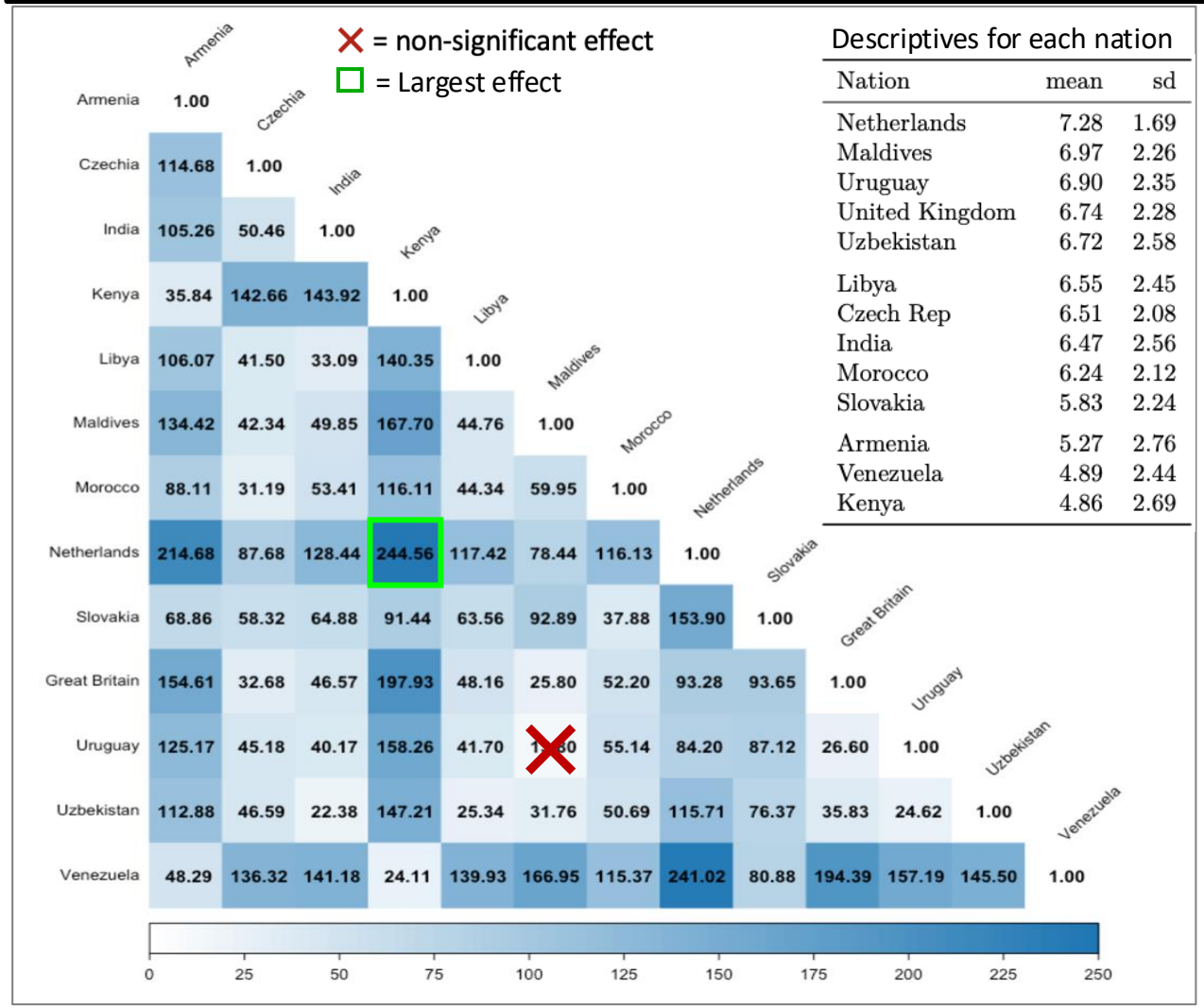
From 1 (completely dissatisfied) to 10 (completely satisfied)

- Overall: *M* = 6.32, *SD* = 2.45
- Highest mean financial satisfaction reported by the **Netherlands** (*M* = 7.28; *SD* = 1.69)
- Lowest mean financial satisfaction reported by **Kenya** (*M* = 4.86; *SD* = 2.69)

DTS effect sizes largest, most varied for this item….

- Ranged from 13.80 to 244.56 (M = 90.61, SD = 55.21)
- All tests significant except for test between the Maldives and Uruguay

Satisfaction with one's financial situation is likely the most dependent on external circumstances



Descriptives for each nation

Nation	mean	sd
Netherlands	7.28	1.69
Maldives	6.97	2.26
Uruguay	6.90	2.35
United Kingdom	6.74	2.28
Uzbekistan	6.72	2.58
Libya	6.55	2.45
Czech Rep	6.51	2.08
India	6.47	2.56
Morocco	6.24	2.12
Slovakia	5.83	2.24
Armenia	5.27	2.76
Venezuela	4.89	2.44
Kenya	4.86	2.69

Results

RQ1: Cross-national comparisons

RQ2: Within-nation comparisons

Did DTS detect cross-national differences missed by other tests?

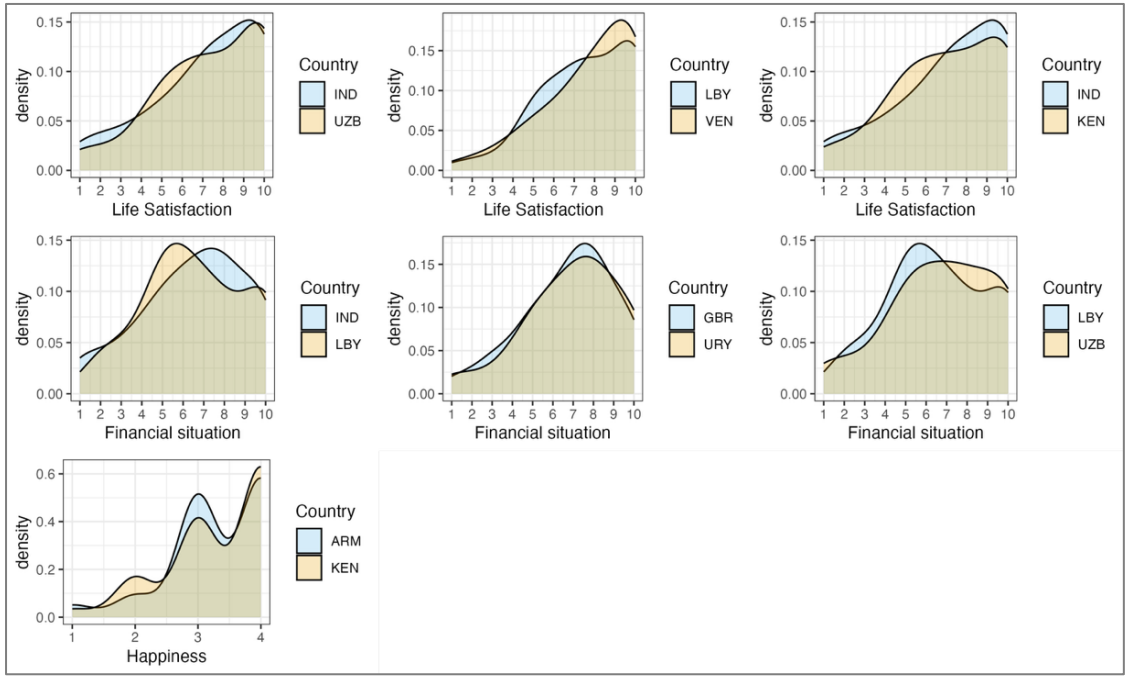
….Yes, but not many.

Number of significant comparisons identified per test:

- **DTS**: 383/390 (98.2%)
- ***t*-test**: 332/390 (85.1%)
- ***F*-tes**t: 309/390 (79.2%)

- Many comparisons were detected as statistically significant by DTS and *either* the *t*-test or the *F*-test….
- **DTS uniquely detected only 7 differences** between nations (which appear minimal upon visual inspection)

Comparisons detected by DTS, but missed by t and F tests



Results

RQ1: Cross-national comparisons

RQ2: Within-nation comparisons

Results

RQ1: Cross-national comparisons

RQ2: Within-nation comparisons

Among the 35 DTS tests, 34 were statistically significant ($p < .05$)….

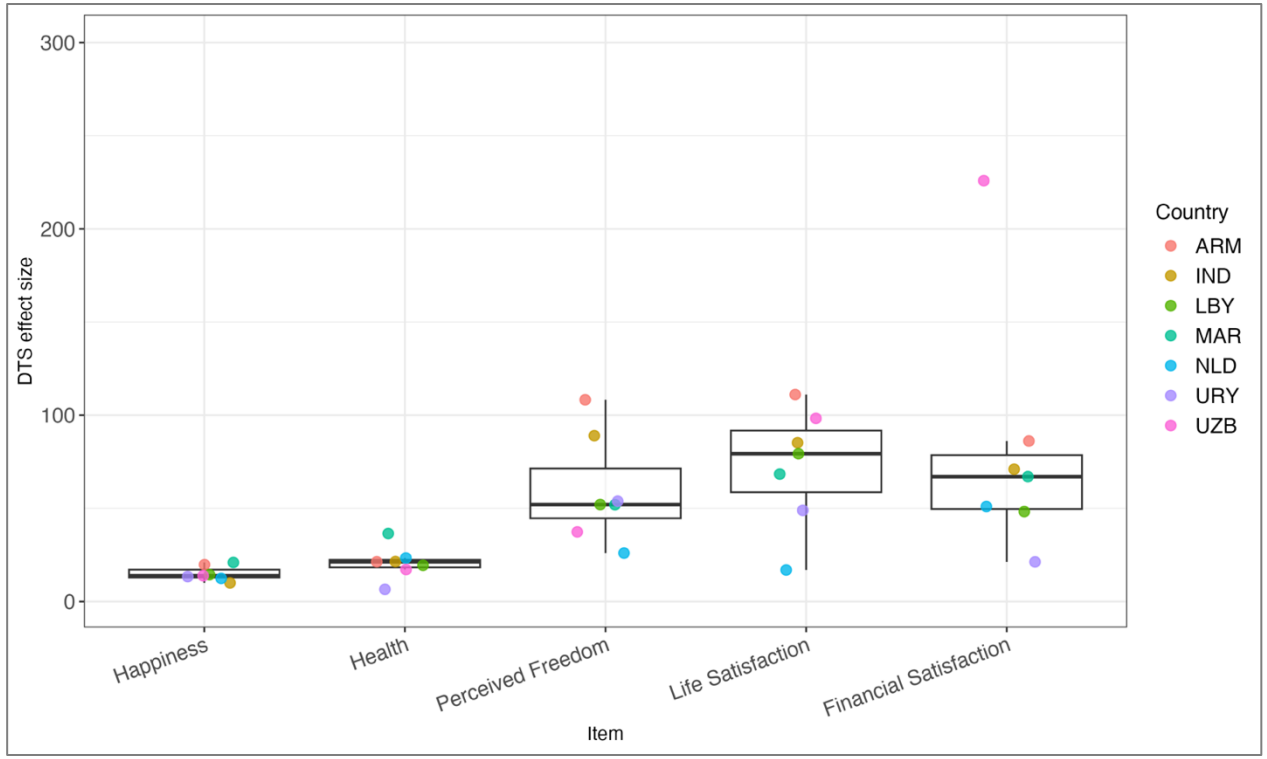
- DTS Effect sizes ranged from 9.98 to 225.87 (M = 51.19; SD = 43.44)
- Only the comparison for Uruguay on self-reported health was non-significant.

Most within-nation effect sizes were smaller than between-nation effect sizes. Some exceptions:

- Uzbekistan: financial satisfaction (DTS = 225.97, $p < .001$).
- Armenia: life satisfaction (DTS = 111.06, $p < .001$) and perceived freedom and control (DTS = 108.26, $p < .001$).

Findings mirrored cross-national comparisons with larger and more varied effect sizes for freedom, life satisfaction, and financial satisfaction.

DTS Effect Sizes Across Waves by Nation and Well-Being Item



Results

RQ1: Cross-national comparisons | RQ2: Within-nation comparisons

Did DTS detect within-national differences missed by other tests?

….No comparisons were uniquely identified by DTS

- 13 of the 35 comparisons were missed by either the *t*-test or *F*- test, but none were missed by both.

Chapter 2 conclusions

Some findings highlight the value of item-level assessment

- Cross-national differences in responding to the life satisfaction item were considerably larger and more varied than differences in responding to the happiness item (though this is potentially due to differences in the number of item response options).

Advantages of DTS over familiar alternatives remain unclear

- DTS appears to offer only a marginal advantage over more familiar approaches in detecting differences
- Useful as a single omnibus test—alternative to conducting multiple types of tests
- DTS could be more useful for detecting other types of individual differences if applied using a more comprehensive set of items

Chapter 3

Exploring Generalizability from the Bottom up: Applying DTS to demonstrate an item-level approach to personality assessment

Note: Data for this study have been collected and analyzed. A manuscript based on findings from this work is currently In-Press at *Current Research in Ecological and Social Psychology (CRESP).*

Applying DTS to the 2,623 IPIP items

Objective: Compare responding across pairs of 13 diverse nations on the 2,623 IPIP items

- Identify items that elicit the most and fewest differences across nations.
- Identify nations/nation pairs that demonstrate the most and fewest differences across items.

- IPIP items assess a variety of individual differences including (but not limited to) the Big Five domains and facets

International Personality Item Pool

A Scientific Collaboratory for the Development of Advanced Measures of Personality and Other Individual Differences

Site Overview

This is the official website for the International Personality Item Pool (IPIP). The site includes over 3,000 items and over 250 scales that have been constructed from the items. New items and scales are developed on an irregular basis. The items and scales are in the public domain, which means that one can copy, edit, translate, or use them for any purpose without asking permission and without paying a fee. However, the grant that supported the creation of this website has expired, so if you find the IPIP website useful, we ask you to consider making a donation through the link below.

Support the IPIP Website with a Voluntary Donation

Warning about the nature of this site

For persons wandering into this site who have not completed a university course or two in psychological assessment, BEWARE: This site includes highly technical scientific information, and **NO PERSONALITY TESTS ARE ADMINISTERED HERE**. If you are looking to take an on-line personality questionnaire based on IPIP items, browse to our page IPIP Implementations around the Web and read about online IPIP-based questionnaires under the heading "High-Quality Implementations of IPIP Measures."

If you are looking for software to administer personality questionnaires based on IPIP items on your own computer or the Web, browse to our page Creating Your Own Online Assessment with the IPIP:

The 3,320 IPIP Items in Alphabetical Order (plus their Survey and Item Number[1])

[1]All but the 654 T items, 127 W items, and 1 Y item from the 3,320 IPIP items have been administered to the Eugene-Springfield Community Sample in 12 of the 30 separate surveys described in the January 2016 ORI Technical Report. The survey numbers and letter codes of IPIP items included in the survey are as follows: (2) H; (8) X; (10) E; (14) N, B; (15) P; (21) A, C; (24) D; (25) Q; (26) R; (27) M; (28) V, (29) S. The T items are from the CAT Personality Disorder Scales (Simms, et al., 2011), the W items are new items added to the VIA measure of character strengths (du Plessis & de Bruin, 2015), and the Y item was created for the Big-Five Aspect Scales (DeYoung, Quilty, & Peterson, 2007).

Abuse people's confidences.	H1131
Accept apologies easily.	H133
Accept challenging tasks.	H515
Accept little from others.	H615
Accept others' weaknesses.	H139
Accept people as they are.	H106
Accept the consequences of my actions.	H266
Accept the first thing that comes along.	H1384
Accept what others say.	H962
Accomplish a lot of work.	H554
Accomplish my work on time.	H230
Acknowledge others' accomplishments.	H152
Acquire skills quickly.	H1338
Act according to my conscience.	H178
Act according to my own feelings.	H1367
Act as a leader.	Q176
Act as I please.	H795
Act as if I'm somebody else and completely identify myself with the part.	A18
Act as if I'm constantly on stage.	T17
Act as if some laws do not apply to me.	T19

Alphabetical Index of 274 Labels for 463 IPIP Scales

Each construct listed below is followed by links to those IPIP scales that seem to measure that same construct. The abbreviations for the link labels are explained in the footnote at the bottom of this page.

Achievement-striving (NEO: C4), (TCI: P3), (MPQ: AC), (6FPQ: IT1), (HPI: Competitive HIC)
Activity-level (NEO: E4), (JPI: Enl)
Adaptability (6FPQ: AG3)
ADHD (ADHD: Span, et al., 2002)
Adventurousness (NEO: O4), (CPI: Sp), (6FPQ: OP1), (ORVIS: Adventure)
Aesthetic Appreciation/Artistic Interests (NEO: O2), (HEX: O-AesA), (VIA: App), (HPI: Culture HIC), (BFAS: Openness)
Affective Lability (CAT-PD: Affective Lability)
Agreeableness (NEO Domain), (Big-Five Domain), (Big-7: 525), (6FPQ: AG), (BFAS: Agreeableness)
Altruism (NEO: A3), (ORVIS: Altruism)
Ambition/Drive (BAS-Drive: Carver & White, 1994)
Amiability (CPI: Ami)
Anger (CAT-PD: Anger), (NEO: N2), (CAT-PD: Anger), (BFAS: Volatility)
Anhedonia (CAT-PD: Anhedonia)
Anxiety (CAT-PD: Anxiousness), (NEO: N1), (JPI: Axy), (16PF: O), (HEX: E-Anxi), (BIS-Anxiety: Carver & White, 1994), (BFAS: Withdrawal)
Appearance-Consciousness (HPI: Appearance HIC), (ORAIS: Fashion)
Artistic Interests/Aesthetic Appreciation (NEO: O2), (HEX: O-AesA), (VIA: App), (HPI: Culture HIC), (BFAS: Openness)
Assertiveness (AB5C: I+/III+), (NEO: E3), (CPI: Do), (16PF: E), (MPQ: SP), (BFAS: Assertiveness)
Attention to Emotions (Barchard, 2001)
Attractiveness (Big-7: 525)
Authenticity/Integrity/Honesty (VIA: Int)
Behavioral Inhibition/Activation System (BIS/BAS: Carver & White, 1994)
Belligerence (MPQ: AG)

Multi-Construct IPIP Inventories

Lexical Big Five Inventories
- Goldberg's (1992) Big-Five Factor Markers
- Donnellan et al.'s (2006) highly abridged B5 factor markers ("Mini-IPIP")
- Sibley et al.'s (2011) extension of the "Mini-IPIP" to include HEXACO Factor 6, the "Mini-IPIP6"
- Saucier's (1997) 7-factor scales:
- Hofstee, de Raad and Goldberg's (1992) 45 AB5C facets:
- DeYoung, Quilty, and Peterson's (2007) 10 Big-Five Aspects:
- Markey and Markey's (2009) 8 IPIP Interpersonal Circumplex Scales:

Inventories Based on the Five Factors:
- Goldberg's (1999) IPIP representation of Costa and McCrae's (1992) NEO-PI-R Domains:
- Goldberg's (1999) IPIP representation of Costa and McCrae's (1992) NEO-PI-R Facets:
- Johnson's (2014) IPIP-NEO-120 representation of Costa and McCrae's (1992) NEO-PI-R Facets:
- Maples et al. (2014) IPIP-NEO-120 representation of Costa and McCrae's (1992) NEO-PI-R Facets:
- Maples-Keller et al. (2019) IPIP-NEO-60 representation of Costa and McCrae's (1992) NEO-PI-R Domains:

Inventories Derived from the Big Five with More Factors
- Hogan and Hogan's (1992) Hogan Personality Inventory (HPI):
- Jackson, Paunonen, and Tremblay's (2000) Six Factor Personality Questionnaire (6FPQ):
- Lee and Ashton's (2004) HEXACO-PI:

Other Popular Inventories:
- Cattell's 16 Personality Factor Questionnaire (16PF; Conn & Rieke, 1994):
- California Psychological Inventory (CPI; Gough, 1996):

Dataset and participants

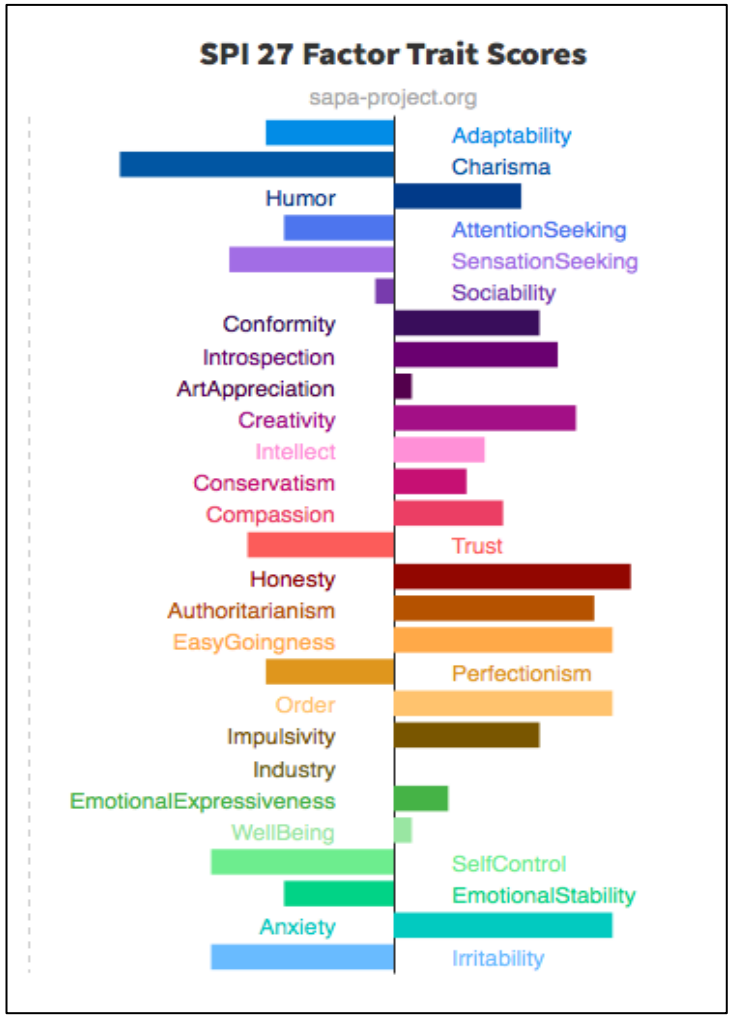
Dataset: The SAPA Project

- Online data collection platform that uses a planned missingness design (about 90% missingness)
  - Participants are administered a random subset of items
- Total sample = 1,025,954
  - Participants indicated that they were 18 or older, and speak English "Well" or "Very well/Fluent"

Nations: India, Pakistan, United States, Nigeria, the Philippines, South Africa, Bangladesh, Brazil, United Kingdom, Zimbabwe, Argentina, Mexico, and Egypt.

- Nations included based on minimum of 100 responses for all IPIP items

78 total pairs across these 13 nations



Participants receive information about their personalities in exchange for participation

Analyses

Tests comparing nations on IPIP items:

78 Nation pairs **x** 2,623 IPIP items **=** 204,594 Total DTS tests

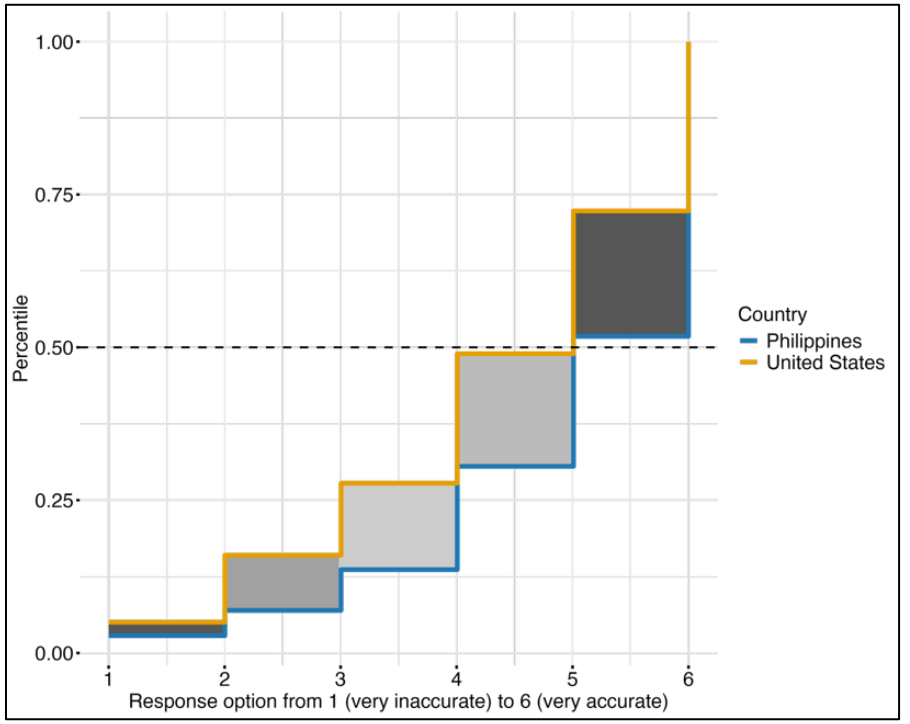
Across 13 nations…

Argentina
Bangladesh
Brazil
Egypt
United Kingdom
India
Mexico
Nigeria
Pakistan
Philippines
United States
South Africa
Zimbabwe

Example test: *I am quiet around strangers*

Nations: Philippines and United States

DTS effect size = 94.67, $p < .001$



Analyses will identify….

1. Nation pairs demonstrating the most differences (and fewest)
    - Might point to cultural differences, variation in response styles

2. Items eliciting the most differences in responding (and fewest)
    - Might point to generalizability (or lack thereof) of certain items
    - Could clarify whether any items assessing the same broader construct show different patterns of responding across nations

Note: these analyses are exploratory and not designed to test specific hypotheses, but some groups of items (e.g., Openness-related items) are expected to show more cross-national differences than others based on previous findings (Thalmayer & Schwaba, 2024).

Results

Results by item

Results by nation pair

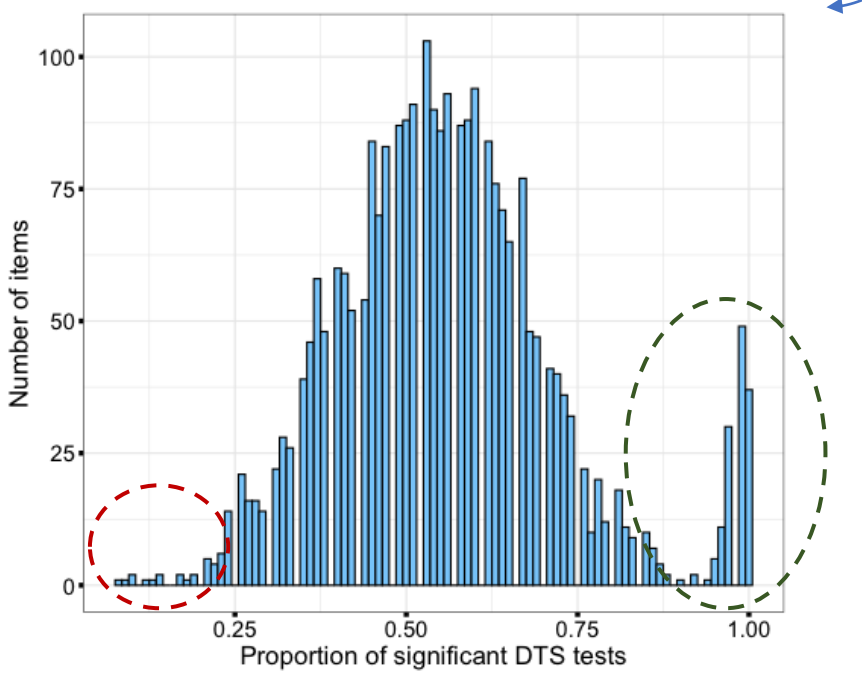
DTS contributions

Results

Results by item | Results by nation pair | DTS contributions

56% of the 204,594 DTS tests were statistically significant ($p < .05$)….

- Proportion of significant tests per item across the 2,623 items was roughly normally distributed (ranging from .08 to 1.00)



Some items had a notably high proportion of significant tests (~100%) across nation pairs.

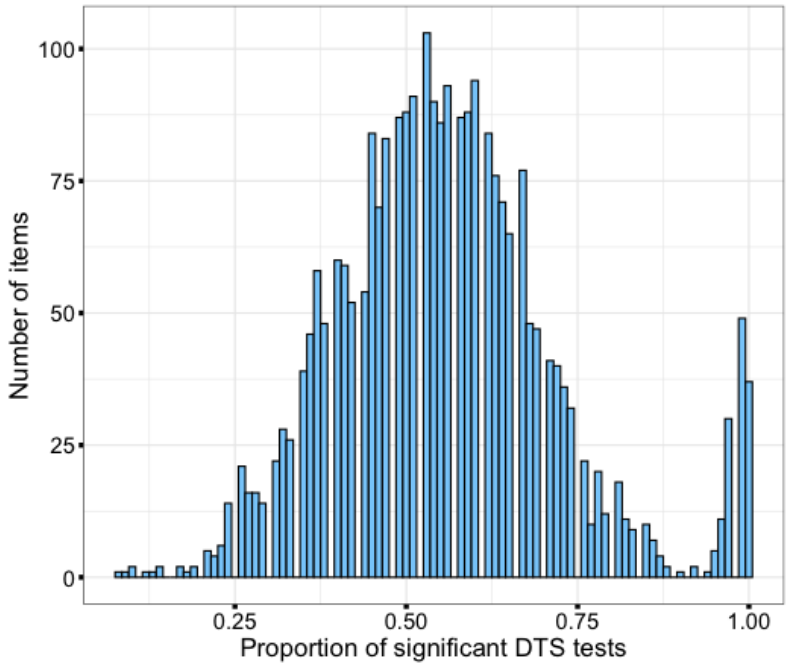
Others had a fairly low proportion of significant tests (<20%) across nation pairs.

Results

Results by item | Results by nation pair | DTS contributions

56% of the 204,594 DTS tests were statistically significant ($p < .05$)….

- Proportion of significant tests per item across the 2,623 items was roughly normally distributed (ranging from .08 to 1.00)



Note that DTS outcomes do not appear to be independent of overall item-variance. Item standard deviation was correlated with….

- Proportion of significant DTS tests ($r = .26$, $p < .001$)
- Mean DTS effect size ($r = .14$, $p < .001$)

Higher item-variance likely contributed to higher DTS effect sizes for some items in addition to substantive cultural differences

Results

Results by item | Results by nation pair | DTS contributions

Items with the highest proportion of significant tests: 37 items with 100% significant tests

Am able to come up with new and different ideas
Am full of ideas.
Am not bothered by disorder.
Am skilled in handling social situations.
Believe in one true religion.
Believe that people are basically moral.
Dislike myself.
Don't consider myself religious.
Feel comfortable with myself.
Get angry easily.
Have a natural talent for influencing people.
Have a slow pace to my life.
Jump into things without thinking.
Let things proceed at their own pace.
Like order.
Like to be thought of as a normal kind of person.
Like to stand during the national anthem.
Like to take it easy.
Love dangerous situations.
Never spend more than I can afford.
Never splurge.
Often forget to put things back in their proper place.
Often make decisions on the spur of the moment.
Panic easily.
Prefer variety to routine.
Put on a show to impress people.
Rarely get irritated.
Rarely overindulge.
Seek danger.
Start tasks right away.
Take risks.
Tend to vote for conservative political candidates.
Think quickly.
Think that my moods don't change more than most people's do.
Trust people to mainly tell the truth.
Trust what people say.
Use others for my own ends.

Results

Results by item | Results by nation pair | DTS contributions

Items with the highest proportion of significant tests: 37 items with 100% significant tests

Am able to come up with new and different ideas
Am full of ideas.
Am not bothered by disorder.
Am skilled in handling social situations.
Believe in one true religion.
Believe that people are basically moral.
Dislike myself.
Don't consider myself religious.
Feel comfortable with myself.
Get angry easily.
Have a natural talent for influencing people.
Have a slow pace to my life.
Jump into things without thinking.
Let things proceed at their own pace.
Like order.
Like to be thought of as a normal kind of person.
Like to stand during the national anthem.
Like to take it easy.
Love dangerous situations.
Never spend more than I can afford.
Never splurge.
Often forget to put things back in their proper place.
Often make decisions on the spur of the moment.
Panic easily.
Prefer variety to routine.
Put on a show to impress people.
Rarely get irritated.
Rarely overindulge.
Seek danger.
Start tasks right away.
Take risks.
Tend to vote for conservative political candidates.
Think quickly.
Think that my moods don't change more than most people's do.
Trust people to mainly tell the truth.
Trust what people say.
Use others for my own ends.

Impulsivity or sensation-seeking related items

Results

Results by item | Results by nation pair | DTS contributions

Items with the highest proportion of significant tests: 37 items with 100% significant tests

Am able to come up with new and different ideas
Am full of ideas.
Am not bothered by disorder.
Am skilled at handling social situations.
Believe in one true religion.
Believe that people are basically moral.
Dislike myself.
Don't consider myself religious.
Feel comfortable with myself.
Get angry easily.
Have a natural talent for influencing people.
Have a slow pace to my life.
Jump into things without thinking.
Let things proceed at their own pace.
Like order.
Like to be thought of as a normal kind of person.
Like to stand during the national anthem.
Like to take it easy.
Love dangerous situations.
Never spend more than I can afford.
Never splurge.
Often forget to put things back in their proper place.
Often make decisions on the spur of the moment.
Panic easily.
Prefer variety to routine.
Put on a show to impress people.
Rarely get irritated.
Rarely overindulge.
Seek danger.
Start tasks right away.
Take risks.
Tend to vote for conservative political candidates.
Think quickly.
Think that my moods don't change more than most people's do.
Trust people to mainly tell the truth.
Trust what people say.
Use others for my own ends.

Social influence or attention-seeking items

Results

Results by item | Results by nation pair | DTS contributions

Items with the highest proportion of significant tests: 37 items with 100% significant tests

Am able to come up with new and different ideas
Am full of ideas.
Am not bothered by disorder.
Am skilled at handling social situations.
Believe in one true religion.
Believe that people are basically moral.
Dislike myself.
Don't consider myself religious.
Feel comfortable with myself.
Get angry easily.
Have a natural talent for influencing people.
Have a slow pace to my life.
Jump into things without thinking.
Let things proceed at their own pace.
Like order.
Like to be thought of as a normal kind of person.
Like to stand during the national anthem.
Like to take it easy.
Love dangerous situations.
Never spend more than I can afford.
Never splurge.
Often forget to put things back in their proper place.
Often make decisions on the spur of the moment.
Panic easily.
Prefer variety to routine.
Put on a show to impress people.
Rarely get irritated.
Rarely overindulge.
Seek danger.
Start tasks right away.
Take risks.
Tend to vote for conservative political candidates.
Think quickly.
Think that my moods don't change more than most people's do.
Trust people to mainly tell the truth.
Trust what people say.
Use others for my own ends.

Conscientiousness-related items

Results

Results by item | Results by nation pair | DTS contributions

Items with the highest proportion of significant tests: 37 items with 100% significant tests

Am able to come up with new and different ideas
Am full of ideas.
Am not bothered by disorder.
Am skilled in handling social situations.
Believe in one true religion.
Believe that people are basically moral.
Dislike myself.
Don't consider myself religious.
Feel comfortable with myself.
Get angry easily.
Have a natural talent for influencing people.
Have a slow pace to my life.
Jump into things without thinking.
Let things proceed at their own pace.
Like order.
Like to be thought of as a normal kind of person.
Like to stand during the national anthem.
Like to take it easy.
Love dangerous situations.
Never spend more than I can afford.
Never splurge.
Often forget to put things back in their proper place.
Often make decisions on the spur of the moment.
Panic easily.
Prefer variety to routine.
Put on a show to impress people.
Rarely get irritated.
Rarely overindulge.
Seek danger.
Start tasks right away.
Take risks.
Tend to vote for conservative political candidates.
Think quickly.
Think that my moods don't change more than most people's do.
Trust people to mainly tell the truth.
Trust what people say.
Use others for my own ends.

openness-related items

Results

Results by item | Results by nation pair | DTS contributions

More Openness-related items with > 90% significant comparisons….

Spend time reflecting on things
Love to reflect on things
Tend to vote for liberal political candidates
Seek adventure
Learn things slowly
Enjoy examining myself and my life
Don't like the idea of change
Do not like poetry
Believe in the importance of art
Am quick to understand things
Can handle a lot of information

Try to understand myself
Am quick to understand things
Can handle a lot of information
Try to understand myself
Love to think up new ways of doing things
Have read the great literary classics
Have a vivid imagination
Catch on to things quickly
Am an original thinker
Do not enjoy going to art museums
Would hate to be considered odd or strange

Results

Results by item | Results by nation pair | DTS contributions

Items with the fewest significant tests (< 20%)

Believe that there are many sides to most issues	7.7%
Want to understand things	9.0%
Enjoy the beauty of nature	10.3%
Know how to fly a helicopter	10.3%
Have little patience for slackers	11.5%
Am committed to principles of justice and equality	12.8%
Am able to see how other peoples' emotions can affect their behaviors	14.1%
Don't have a soft side	14.1%
Have lived in all 50 states in the U.S.A.	16.7%
Love to chat	16.7%
Look at the facts	17.9%
Think a lot before complaining in a restaurant if my meal is not well-prepared	19.2%
Want to know the reasons why	19.2%

Results

Results by item | Results by nation pair | DTS contributions

Items with the fewest significant tests (< 20%)

Believe that there are many sides to most issues	7.7%
Want to understand things	9.0%
Enjoy the beauty of nature	10.3%
Know how to fly a helicopter	10.3%
Have little patience for slackers	11.5%
Am committed to principles of justice and equality	12.8%
Am able to see how other peoples' emotions can affect their behaviors	14.1%
Don't have a soft side	14.1%
Have lived in all 50 states in the U.S.A.	16.7%
Love to chat	16.7%
Look at the facts	17.9%
Think a lot before complaining in a restaurant if my meal is not well-prepared	19.2%
Want to know the reasons why	19.2%

Some ask about specific behaviors or life experiences

(the population might not vary much between and/or within cultures on these items)

Results

Results by item | Results by nation pair | DTS contributions

Items with the fewest significant tests (< 20%)

Believe that there are many sides to most issues	7.7%
Want to understand things	9.0%
Enjoy the beauty of nature	10.3%
Know how to fly a helicopter	10.3%
Have little patience for slackers	11.5%
Am committed to principles of justice and equality	12.8%
Am able to see how other peoples' emotions can affect their behaviors	14.1%
Don't have a soft side	14.1%
Have lived in all 50 states in the U.S.A.	16.7%
Love to chat	16.7%
Look at the facts	17.9%
Think a lot before complaining in a restaurant if my meal is not well-prepared	19.2%
Want to know the reasons why	19.2%

Others were openness-related

Results

Results by item | Results by nation pair | DTS contributions

More openness-related items with fewer (< 30%) significant tests

Enjoy games of strategy
Try to identify the reasons for my actions
Go my own way
Love to learn new things
Don't like to travel
Get excited when I learn new things
Like to visit new places
Love to read challenging material
Love to travel to places that I have never been before
Understand things quickly
Want to increase my knowledge
Come up with alternatives
Am good at thinking of solutions that no one else thinks of

Prefer to analyze problems rather than just describe them
See beauty in things others might not notice
Have new ideas for improving things
Look at myself through others' eyes
Love to hear about other countries and cultures
Tend to think about something for hours
Am a true life-long learner
Can see different points of view
Can sense how things will turn out
Care about justice
Notice contradictions
Try to do difficult things

Results

Results by item | Results by nation pair | DTS contributions

Themes among openness items…

With the most significant tests:

With the fewest significant tests:

Results

Results by item | Results by nation pair | DTS contributions

Themes among openness items…

With the most significant tests:

- **Coming up with ideas/imagination**

Am able to come up with new and different ideas
Have a vivid imagination
Am full of ideas
Love to think up new ways of doing things

With the fewest significant tests:

- **Coming up with solutions, strategies**

Have new ideas for improving things
Come up with alternatives
Am good at thinking of solutions that no one else thinks of
Enjoy games of strategy

Results

Results by item | Results by nation pair | DTS contributions

Themes among openness items…

With the most significant tests:

- Coming up with ideas/imagination
- **"Quickness" of understanding**

Think quickly
Learn things slowly
Am quick to understand things
Catch on to things quickly

With the fewest significant tests:

- Coming up with solutions, strategies
- **Enjoyment of learning, challenges**

Am a true life-long learner
Love to learn new things
Love to read challenging material
Try to do difficult things
Want to increase my knowledge
Want to understand things

Results

Results by item | Results by nation pair | DTS contributions

Themes among openness items…

With the most significant tests:

- Coming up with ideas/imagination
- "Quickness" of understanding
- **Being different from others**

Like to be thought of as a normal kind of person
Would hate to be considered odd or strange
Am an original thinker

With the fewest significant tests:

- Coming up with solutions, strategies
- Enjoyment of learning, challenges
- **Perspective-taking**

Can see different points of view
Look at myself through others' eyes
Believe there are many sides to most issues

Results

Results by item | Results by nation pair | DTS contributions

Themes among openness items…

With the most significant tests:

- Coming up with ideas/imagination
- "Quickness" of understanding
- Being different from others
- **Art, politics, or religion**

Believe in the importance of art
Do not enjoy going to art museums
Believe in one true religion
Tend to vote for liberal (conservative) political candidates

With the fewest significant tests:

- Coming up with solutions, strategies
- Enjoyment of learning, challenges
- Perspective-taking
- **Cultural exploration**

Love to hear about other countries and cultures
Like to visit new places
Don't like to travel

Results

Results by item | Results by nation pair | DTS contributions

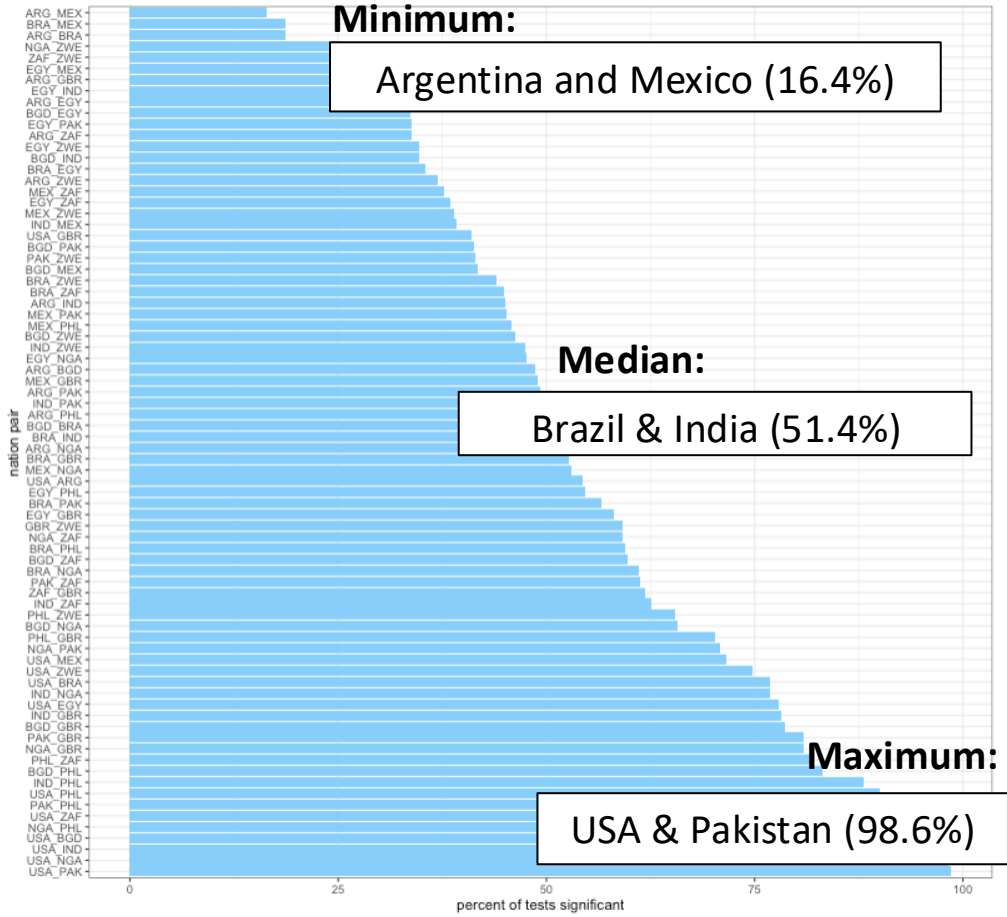
Results

Results by item | **Results by nation pair** | DTS contributions

The proportion of significant tests per nation pair ranged from **16.4%** to **98.6%**

- Mean = 55.8 %
- Median = 51.4 %
- SD = 21.2%

% of significant tests across nation pairs



Results

Results by item | **Results by nation pair** | DTS contributions

Nation pairs with the <u>fewest</u> significant tests

- Nations in close geographical proximity



<u>South/Central American nations</u>
Argentina and Mexico (16.4%), Brazil and Mexico (18.6%)
Argentina and Brazil (18.7%)



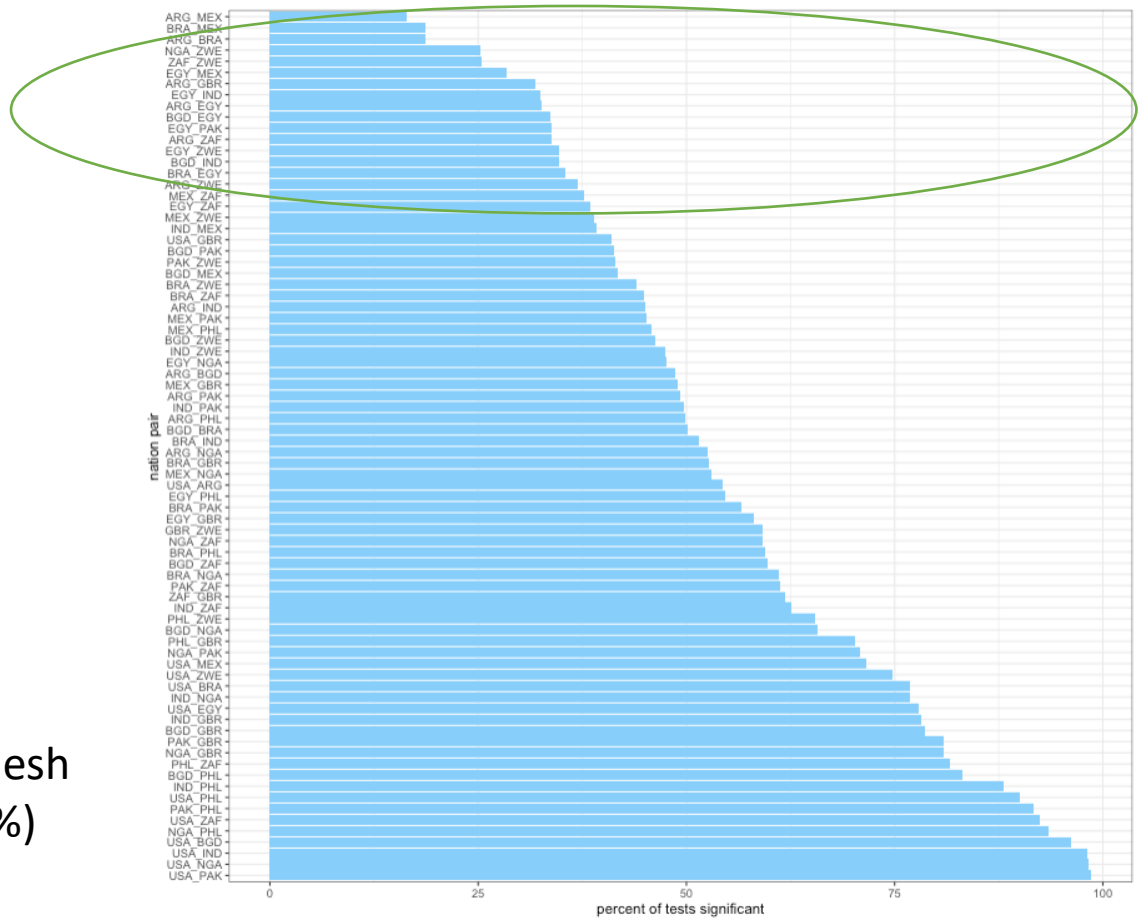
<u>African nations</u>
Nigeria and Zimbabwe (25.2%),
South Africa and Zimbabwe (25.4%)

- Tests between Egypt and other nations



Mexico (28.4%), India (32.5%), Argentina (32.6%), Bangladesh (33.7%), Pakistan (33.8%), Zimbabwe (34.7%), Brazil (35.5%)

<u>% of significant tests across nation pairs</u>



Results

Results by item | **Results by nation pair** | DTS contributions

Nation pairs with the most significant tests

- Tests between United States and other nations



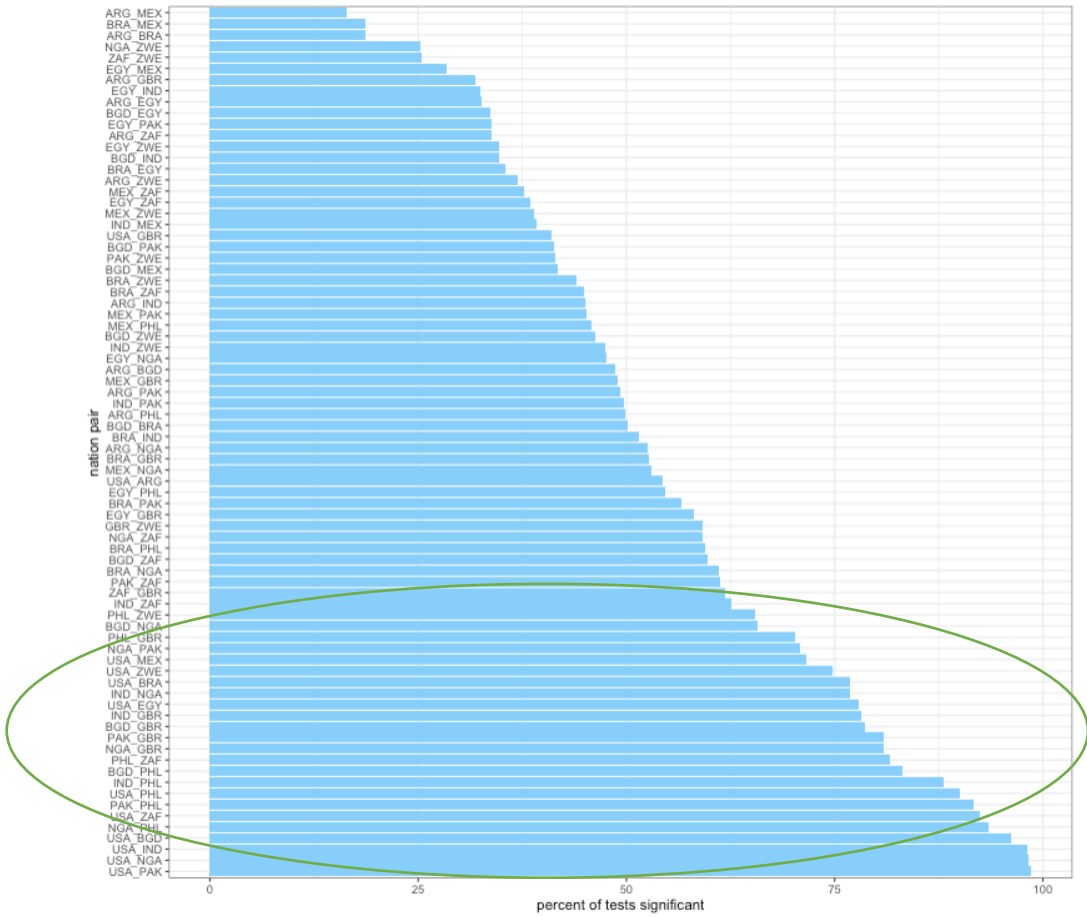
Pakistan (98.6%), Nigeria (98.2%), India (98.1%), Bangladesh (96.1%), South Africa (92.5%), Philippines (90.1%), Egypt (77.9%), Brazil (76.8%), Zimbabwe (74.8%), Mexico (71.6%)

- Tests between the Philippines and other nations



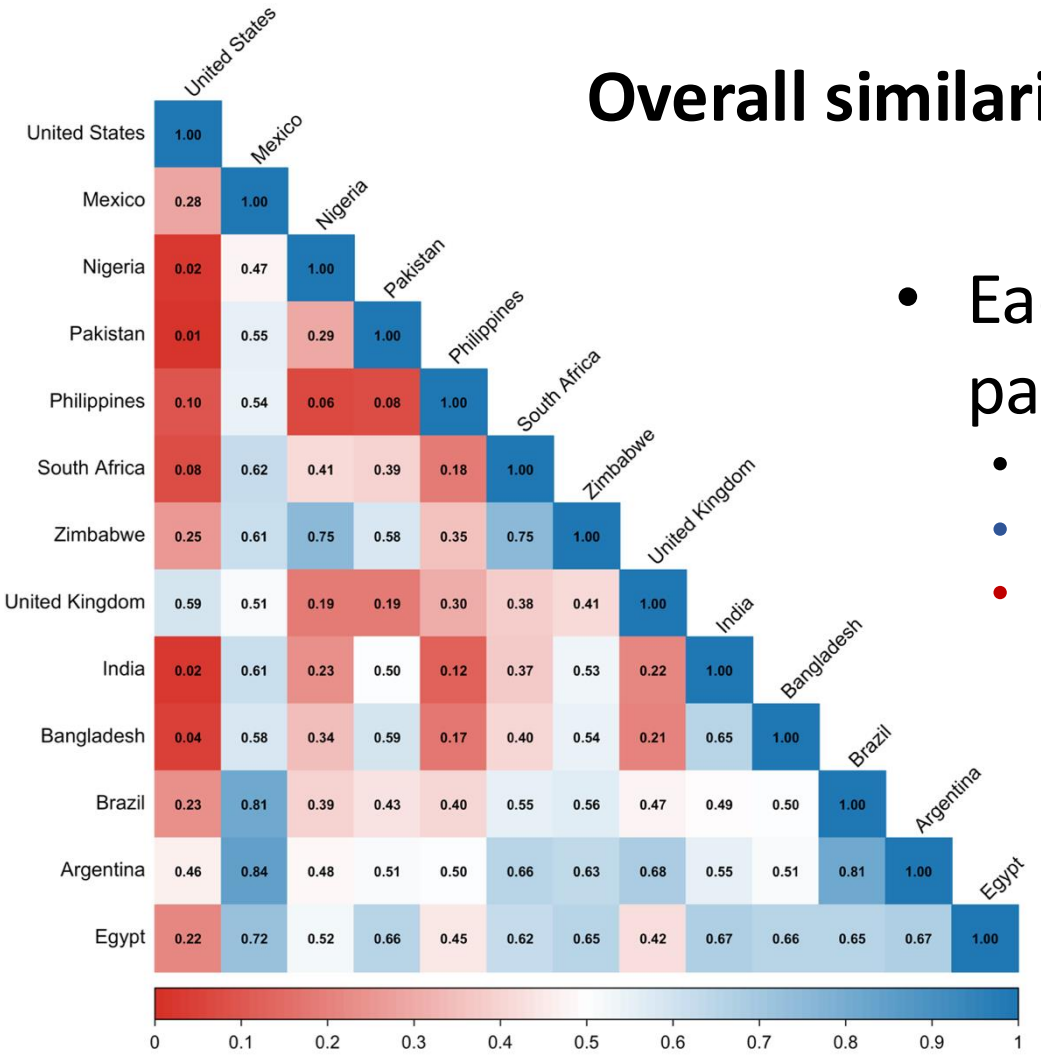
Nigeria (93.5%), Pakistan (91.7%), United States (90.1%), India (88.1%), Bangladesh (83.1%), South Africa (81.6%)

% of significant tests across nation pairs



Results

Results by item | **Results by nation pair** | DTS contributions



Overall similarity in responding across nations

- Each cell represents the degree of similarity for a pair of nations
  - Similarity = 1 - proportion of significant comparisons
  - Nations with more similarity (values above .5) are shown in blue
  - Nations with less similarity (values below .5) are shown in red

Results

Results by item | Results by nation pair | DTS contributions

Does DTS offer advantages over the *t*-test and *F*-test?

Results

Results by item | Results by nation pair | DTS contributions

Does DTS offer advantages over the *t*-test and *F*-test? ☒

Regarding whether it detects a higher proportion of differences across groups, not particularly…

- **DTS** detected 56% of comparisons as statistically significant
- ***t*-test** detected 50% of comparisons as statistically significant
- ***F*-test** detected 30.7% of comparisons as statistically significant

---

Only a small number of comparisons—2,691 (1.3%) of the 204,594—were uniquely identified as statistically significant by DTS

Results

Results by item | Results by nation pair | DTS contributions

Does DTS offer advantages over the *t*-test and *F*-test? ☑

Regarding the *kind* of difference it can detect, yes.

- DTS detects differences in overall shape of the distribution (e.g., skew, multimodality).
- Items more consistently identified only by DTS assess various interpersonal norms.

Items more consistently detected by DTS only….

I apologize a lot
I believe that most people would lie to get ahead
I disclose my intimate thoughts
I get to the point quickly
I have a quiet laugh
I have no secrets
I talk even when I know I shouldn't
I work too much

Results

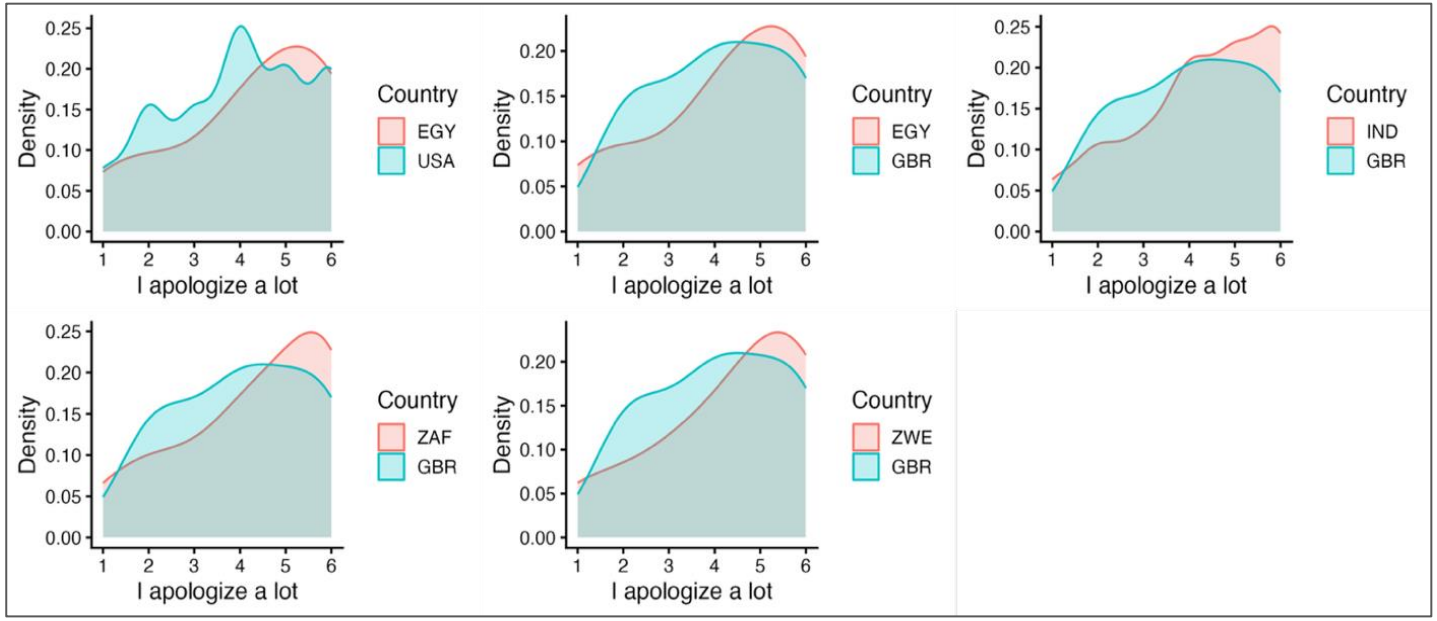
Results by item | Results by nation pair | DTS contributions

Does DTS offer advantages over the *t*-test and *F*-test? ☑

Regarding the *kind* of difference it can detect, yes.

- DTS detects differences in overall shape of the distribution (e.g., skew, multimodality).
- Items more consistently identified only by DTS assess various interpersonal norms.

Comparisons for the item ***I apologize a lot***



Results

Results by item | Results by nation pair | DTS contributions

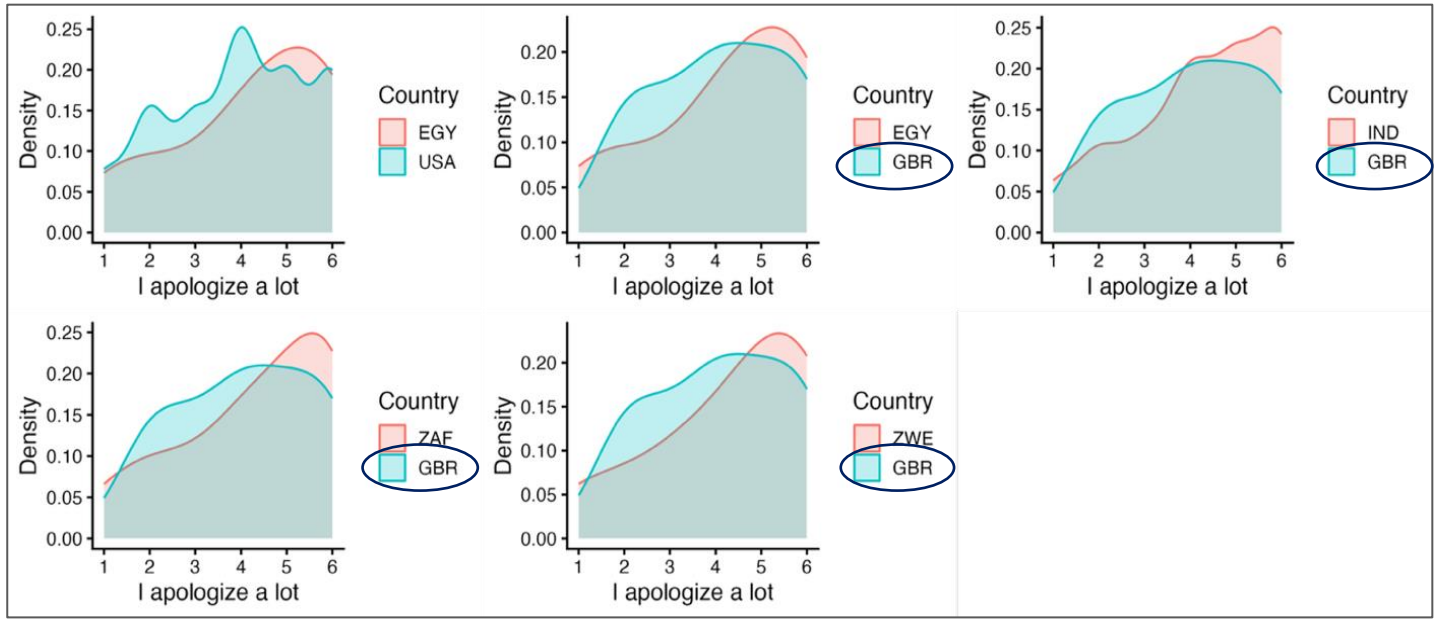
Does DTS offer advantages over the *t*-test and *F*-test? ☑

Regarding the *kind* of difference it can detect, yes.

United Kingdom participants report less apologizing: potential reference group effect?



Comparisons for the item ***I apologize a lot***



Results

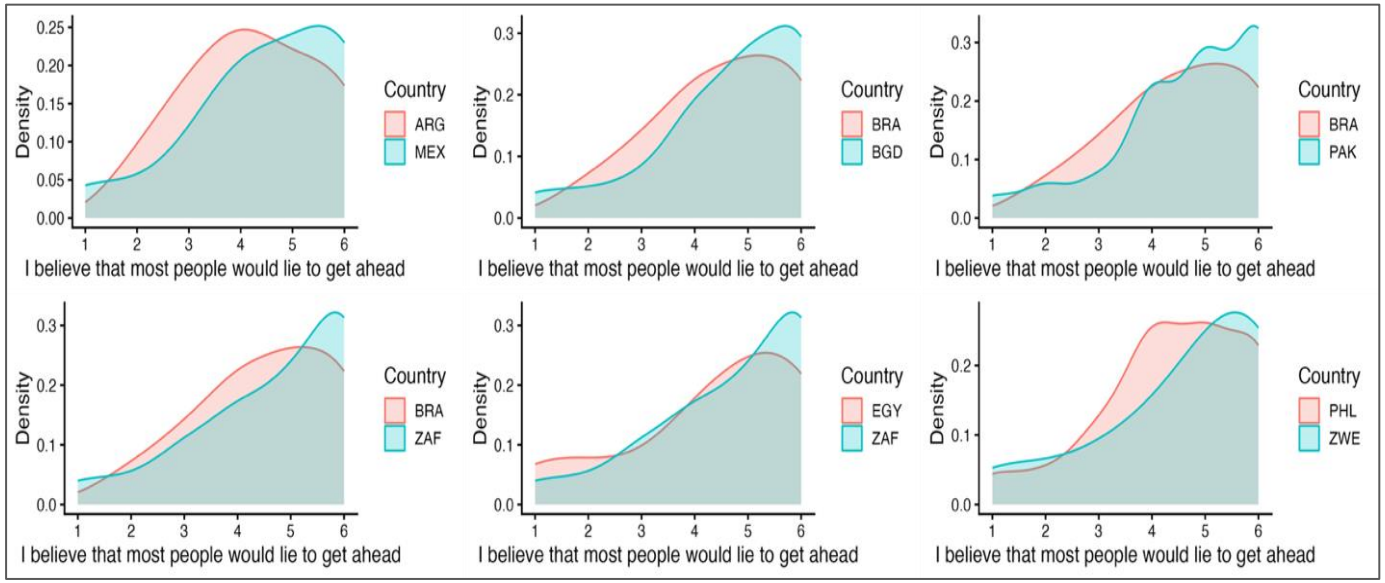
Results by item | Results by nation pair | DTS contributions

Does DTS offer advantages over the *t*-test and *F*-test? ☑

Regarding the *kind* of difference it can detect, yes.

- DTS detects differences in overall shape of the distribution (e.g., skew, multimodality).
- Items more consistently identified only by DTS assess various interpersonal norms.

Comparisons for the item ***I believe that most people would lie to get ahead***



Results

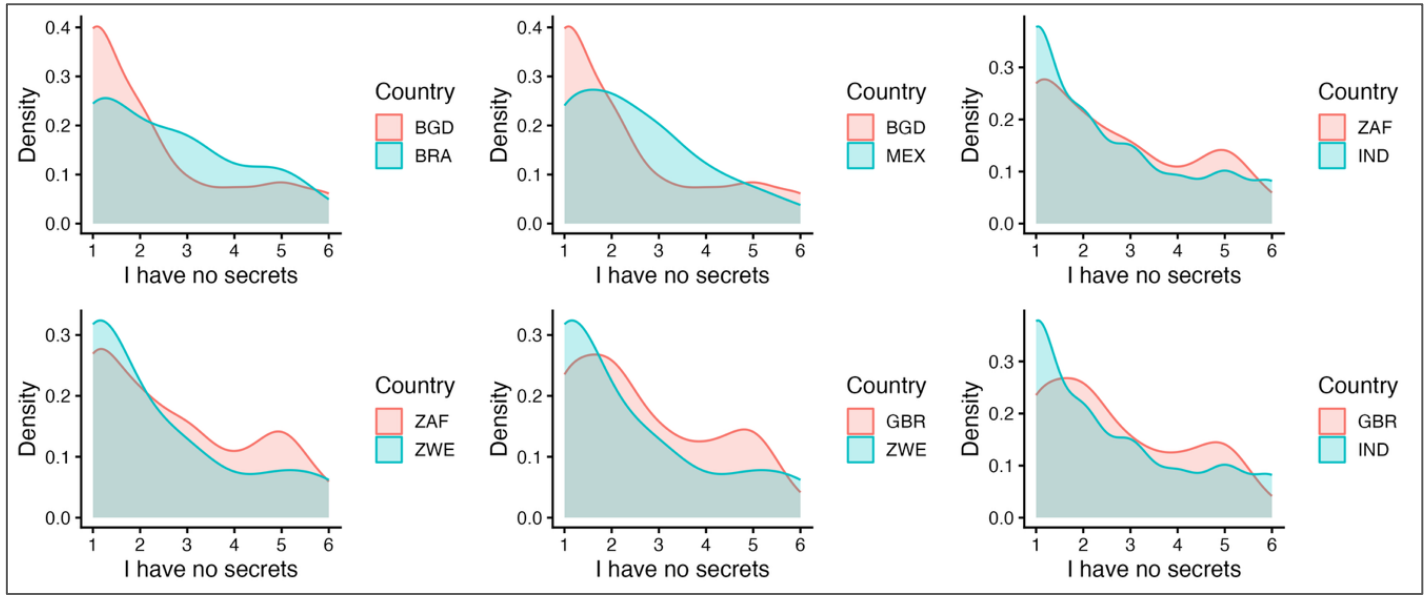
Results by item | Results by nation pair | DTS contributions

Does DTS offer advantages over the *t*-test and *F*-test? ☑

Regarding the *kind* of difference it can detect, yes.

- DTS detects differences in overall shape of the distribution (e.g., skew, multimodality).
- Items more consistently identified only by DTS assess various interpersonal norms.

Comparisons for the item ***I have no secrets***



Results

Results by item | Results by nation pair | DTS contributions

Does DTS offer advantages over the *t*-test and *F*-test? ☑

Regarding the *kind* of difference it can detect, yes.

- DTS detects differences in overall shape of the distribution (e.g., skew, multimodality).
- Items more consistently identified only by DTS assess various interpersonal norms.

Comparisons for the item ***I disclose my intimate thoughts***



Chapter 3 takeaways

The present study used a large and varied item-bank to explore the generalizability of items in a set of culturally and geographically diverse nations

---

Several findings offer further support for the value of item-focused analyses

- Many items assessing emotional expression and openness were among those showing the most differences across nations
- Other openness-related items were among those showing the fewest differences across nations, aligning with more recent work proposing a “universal core” of Openness
- Findings for conceptually similar items (e.g., *I disclose my intimate thoughts, I have no secrets*) point to possible psychometric issues

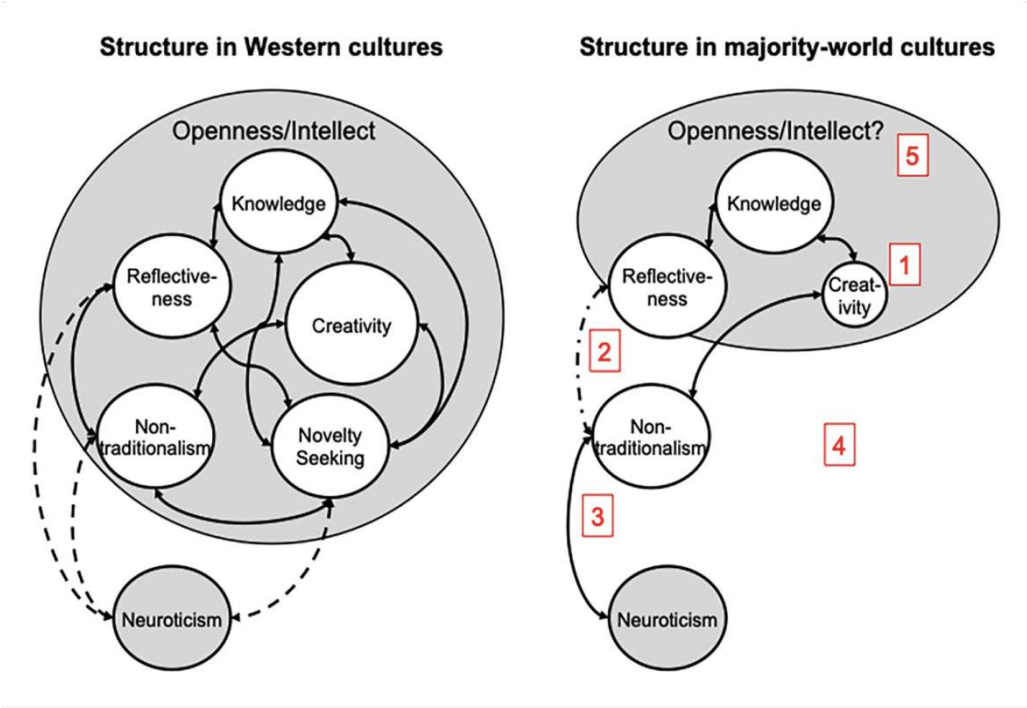


Image from Thalmayer & Schwaba (2025)

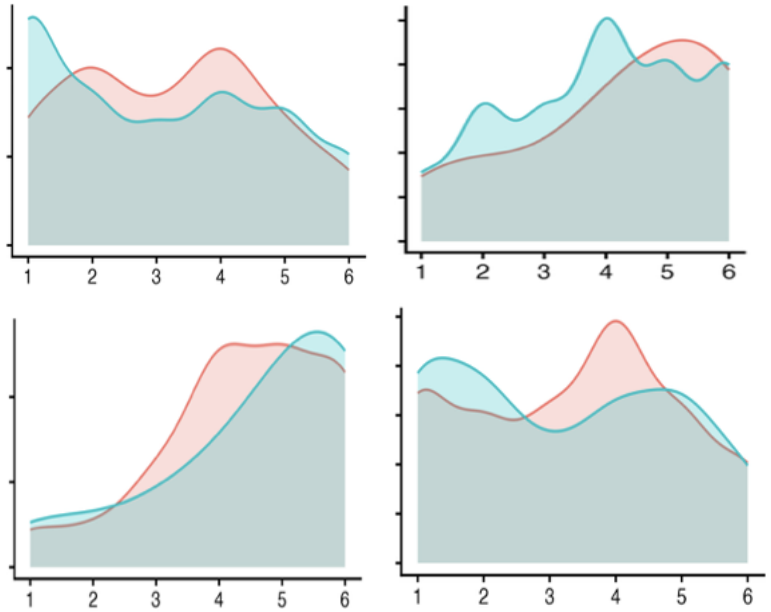
Chapter 3 takeaways

The present study used a large and varied item-bank to explore the generalizability of items in a set of culturally and geographically diverse nations

---

DTS detects differences in distribution shape—using a single, omnibus test—that more familiar approaches miss

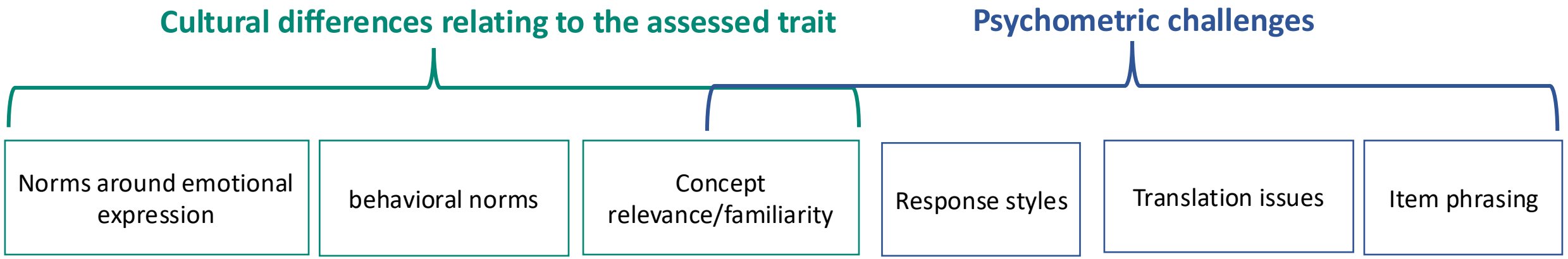
- Tests focused on means (t-tests) and variances (F-tests) do not detect cross-national differences in other distributional features such as skew, clustering, and multimodality
- DTS uniquely detected differences in various interpersonal norms, such as frequency of apologizing, sharing secrets, and beliefs regarding others' tendency to be honest



Study limitations and future directions

Findings are descriptive, and differences can arise for many reasons…

- E.g., True cultural differences, response styles, interpretation issues?
- Cross-national differences do not necessarily reflect cross-*cultural* differences
- Can't explain differences in responding here—need qualitative work to evaluate the items further



Study limitations and future directions

Need to adjust for multiple comparisons—particularly when interpreting individual tests

- Over 200,000 tests were run, many of which will no longer be significant after applying the necessary correction.

Set of items was developed in Western context

- Future work should include items developed in non-western contexts

Limitations of DTS

- Lack of benchmarks for interpreting effect sizes
- Future work should explore whether adjusting for item variance influences results

Chapter 4
Conclusions & Takeaways

Challenges and future directions for item-focused research

Domain level measures are popular, in part, because they offer a more tractable starting point for interpretation

- When administering a large number of items, the most central or important findings may become less clear
- Findings that are not explored as thoroughly could still be useful to others in the field

Developing a central repository for descriptive findings in personality research (as proposed by Mõttus et al., 2020) might be a useful strategy

- Similar resources have been developed in fields like neuroscience and genetics

OSF

HARVARD DATAVERSE

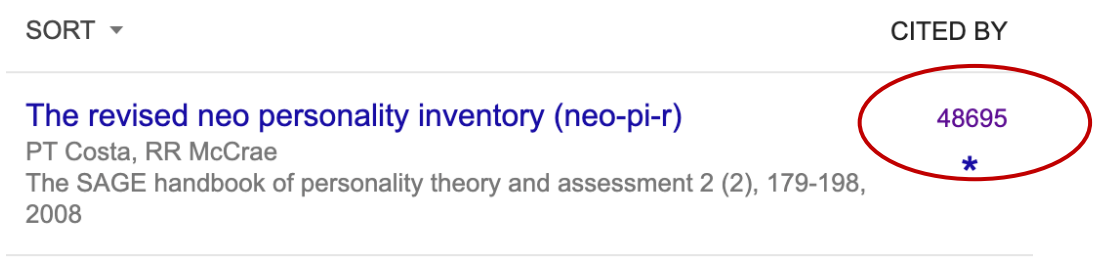
NCBI National Center for Biotechnology Information

HUMAN Connectome PROJECT

Challenges and future directions for item-focused research

Situating item-level findings within the broader literature will likely involve some challenges….

- Most previous studies focus on the domain-level
- Discussing personality mostly in terms of item-level constructs would likely be too challenging



Reporting findings at multiple levels of analysis would offer the strongest approach for connecting findings to existing work

- Lower-dimensional structures (e.g., domain-level findings) can be derived from higher-dimensional structures (e.g., item- or facet-level findings), but the reverse is not possible
- Maintains connection between personality research and some of its broader purposes
  - accessibility to a more general audience and in more practical settings such as workplaces, classrooms, or clinical settings—where personality is often discussed using broad, imprecise terms

Conclusions

- Item-level constructs offer many advantages….
  - As standalone constructs (e.g., increased predictive power, sources of cross-cultural differences)
  - For broader scale development efforts—for developing a more generalizable measure of Openness to Experience

- The item-level is overlooked partially due to strong norms regarding test construction and measurement
  - Most statistical approaches upheld as more rigorous by the field (e.g. latent variable modeling, multigroup CFA), are not suited for addressing research questions at the item-level

- Broad trait models (e.g., the Big five, six) will remain useful in personality research
  - Measuring constructs at broader levels is the strongest option in some studies
  - Broad trait models offer a useful common language for personality researchers

Personality psychology would likely benefit most from viewing different levels of the trait hierarchy as complementary approaches to personality assessment, and identifying statistical techniques that are well-suited for addressing research questions at each level

Thank you!







PIE lab

Advisors

- David Condon
- Sara Weston

Lab mates

- Andrew Castillo
- Rachel Jacobson
- Sarah McDougald
- Ian Shryock
- Amala Someshwar
- Dillon Welindt

Dissertation Committee members

- Nash Unsworth
- Jake Searcy

Collaborators

- Dr. Gerard Saucier
- Dr. Amber Gayle Thalmayer

Questions?

Chapter 2: Descriptives

Age range, mean, and SD for each nation

Nation	mean_age	sd_age	min_age	max_age
ARM	49.04	17.10	18	91
CZE	49.11	17.11	18	91
GBR	51.82	18.42	18	99
IND	35.80	16.33	16	90
KEN	30.74	10.07	18	84
LBY	40.22	12.71	18	85
MAR	37.22	13.53	18	82
MDV	37.65	14.10	18	86
NLD	53.36	16.38	19	94
SVK	52.66	16.14	18	90
URY	49.83	18.07	18	90
UZB	35.90	12.89	18	83
VEN	38.31	15.25	18	85

Average DTS effect size per item

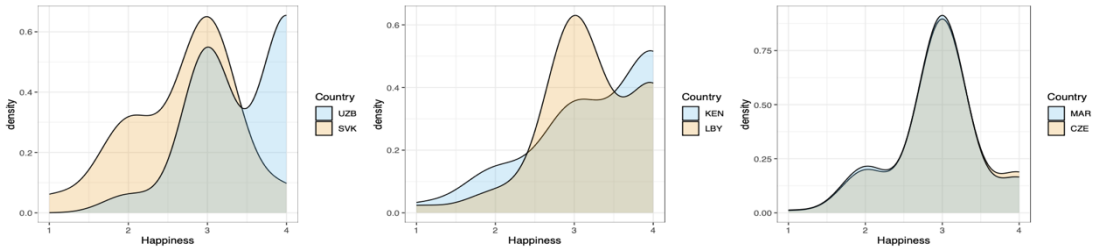
item	item_content	DTS_mean	DTS_sd	proportion_05	proportion_01
Q50	Financial Situation	90.61	55.21	0.99	0.97
Q49	Life Satisfaction	82.26	41.49	0.99	0.97
Q48	Freedom	67.70	29.84	0.99	0.96
Q46	Happiness	26.15	13.31	0.99	0.97
Q47	Health	31.90	18.30	0.96	0.92

Item descriptives

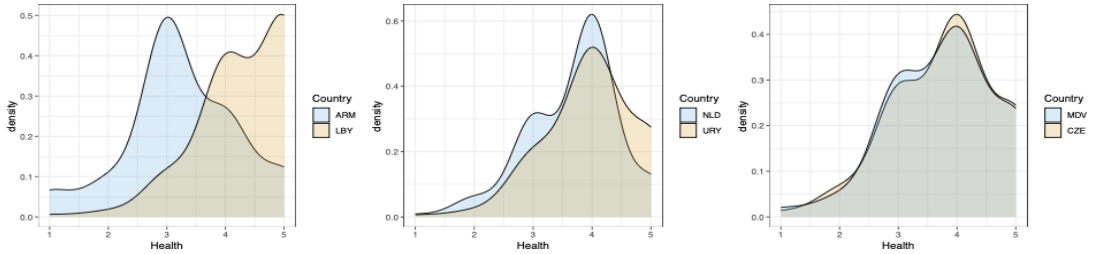
variable	item	mean	sd	min	max
Q46	Happiness	3.17	0.70	1	4
Q47	Health	3.83	0.90	1	5
Q48	Freedom	7.35	2.23	1	10
Q49	Life Satisfaction	7.11	2.24	1	10
Q50	Financial Situation	6.32	2.45	1	10

Chapter 2:
examples of smaller and larger cross-national differences for each item

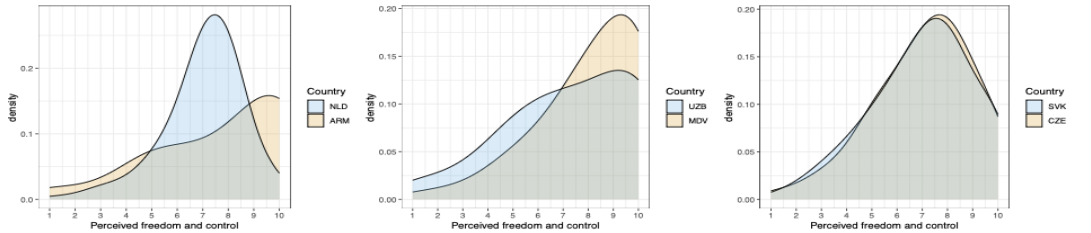
Happiness



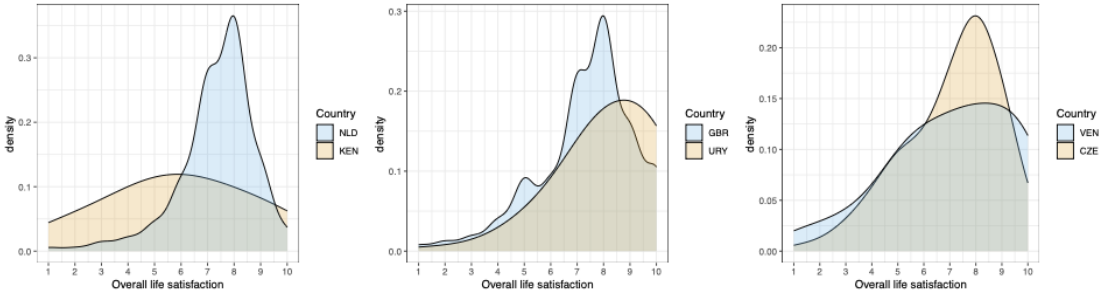
Health



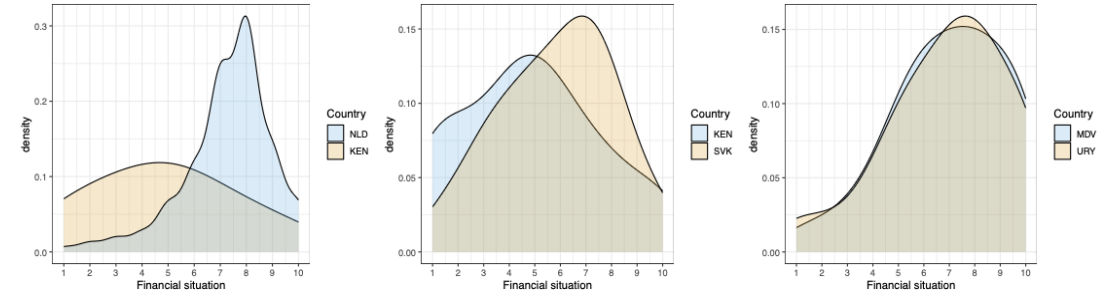
Perceived freedom



Overall life satisfaction



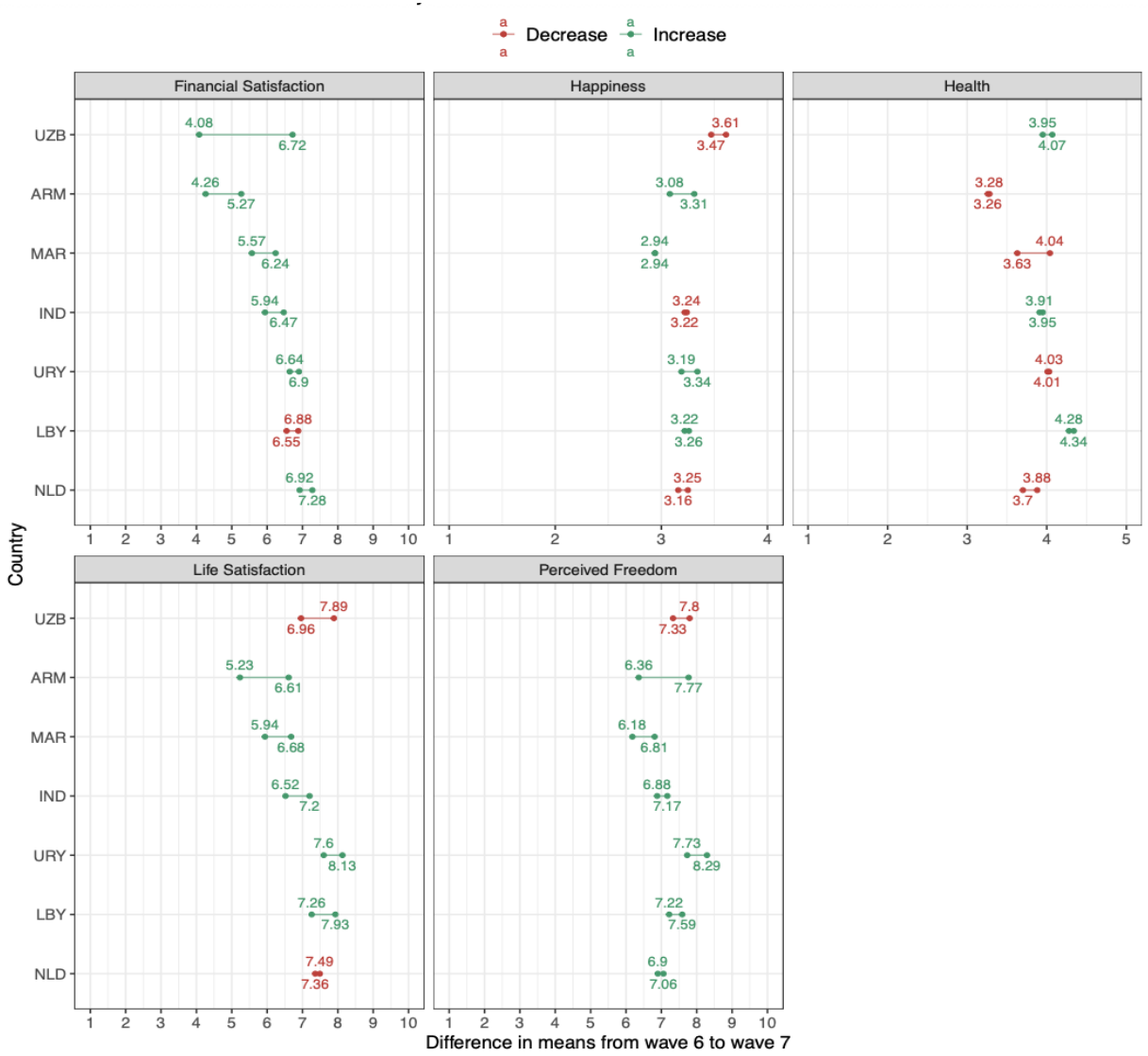
Satisfaction with financial situation



Chapter 2: Descriptives for nation-level variables

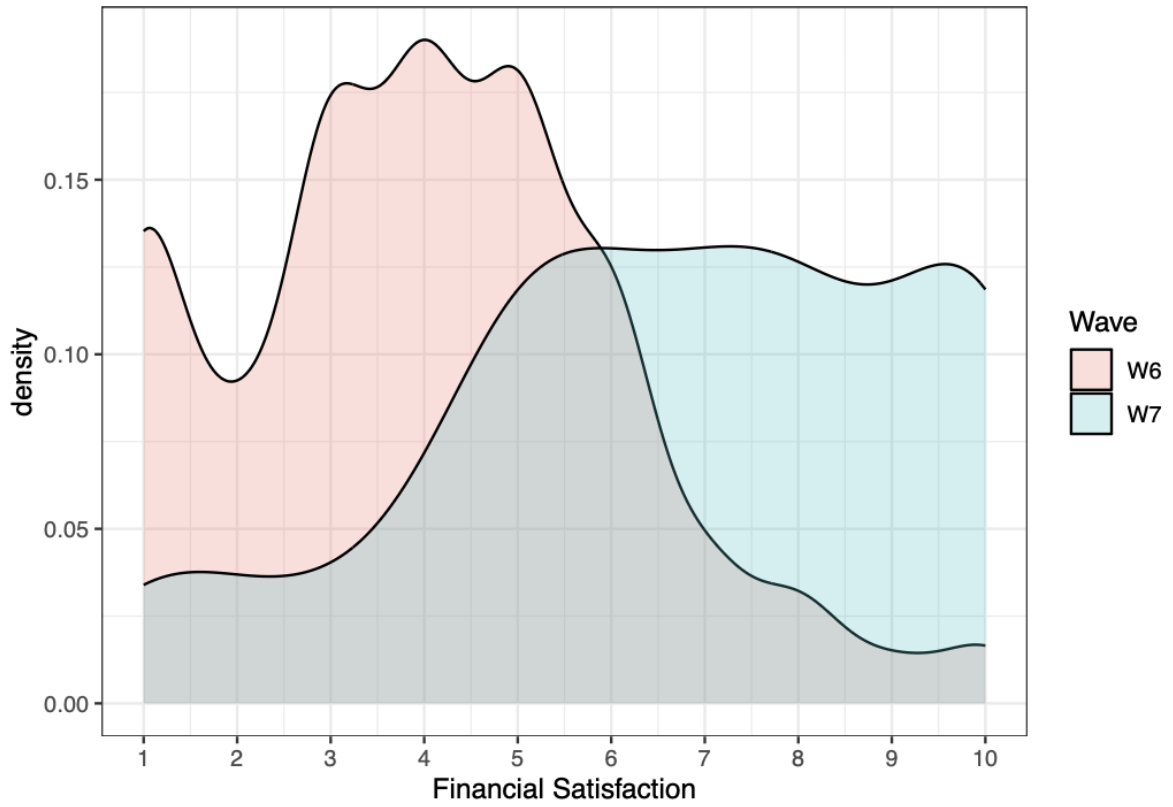
Nation	Global Freedom status	Political rights	Civil liberties	Gender Inequality
LBY	9	7	6	0.172
UZB	10	7	6	0.303
VEN	16	7	6	0.458
MAR	37	5	5	0.492
MDV	40	4	5	0.367
KEN	48	4	4	0.545
ARM	53	4	4	0.259
IND	71	2	3	0.501
SVK	88	1	2	0.191
CZE	91	1	1	0.137
GBR	94	1	1	0.119
URY	98	1	1	0.275
NLD	99	1	1	0.041
	(from 0 to 100)	(1=high, 7= low)	(1=high, 7= low)	(0=most equality, 1= least equality)

Chapter 2: Within nation changes from wave 6 to wave 7

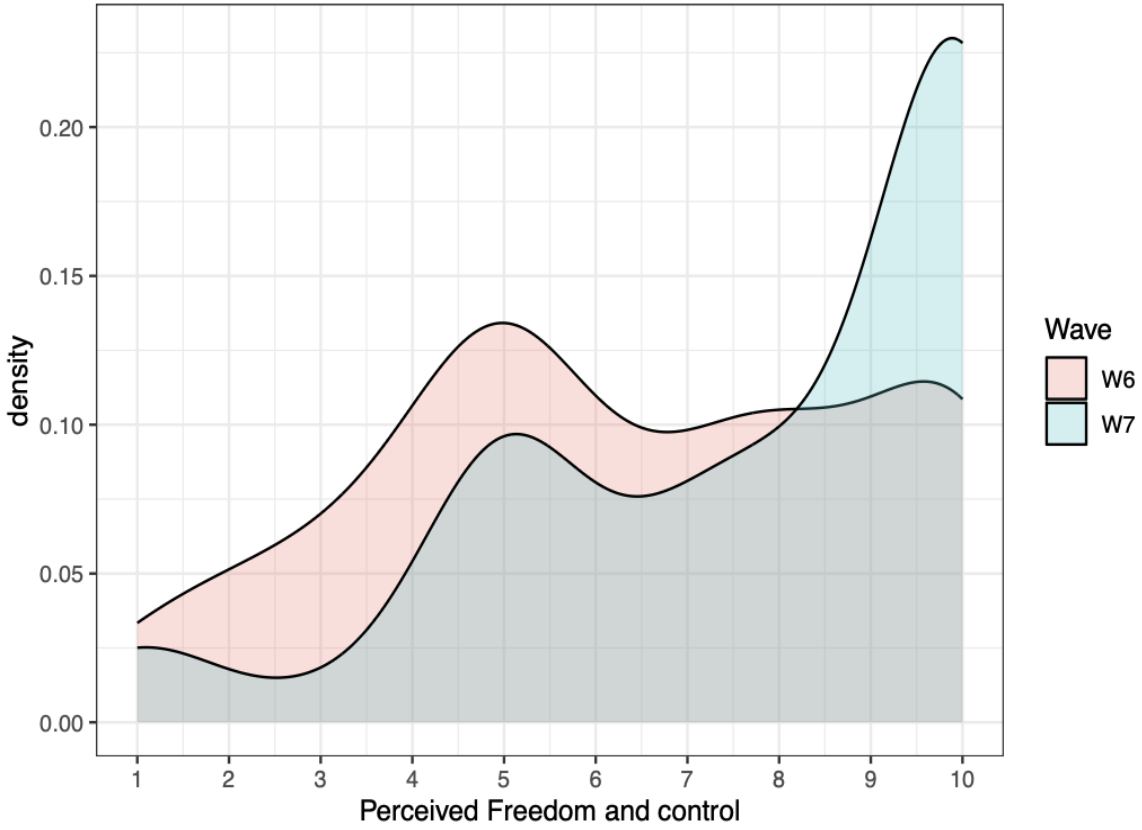


Chapter 2: Outliers among within nation comparisons

Uzbekistan—Satisfaction with financial situation



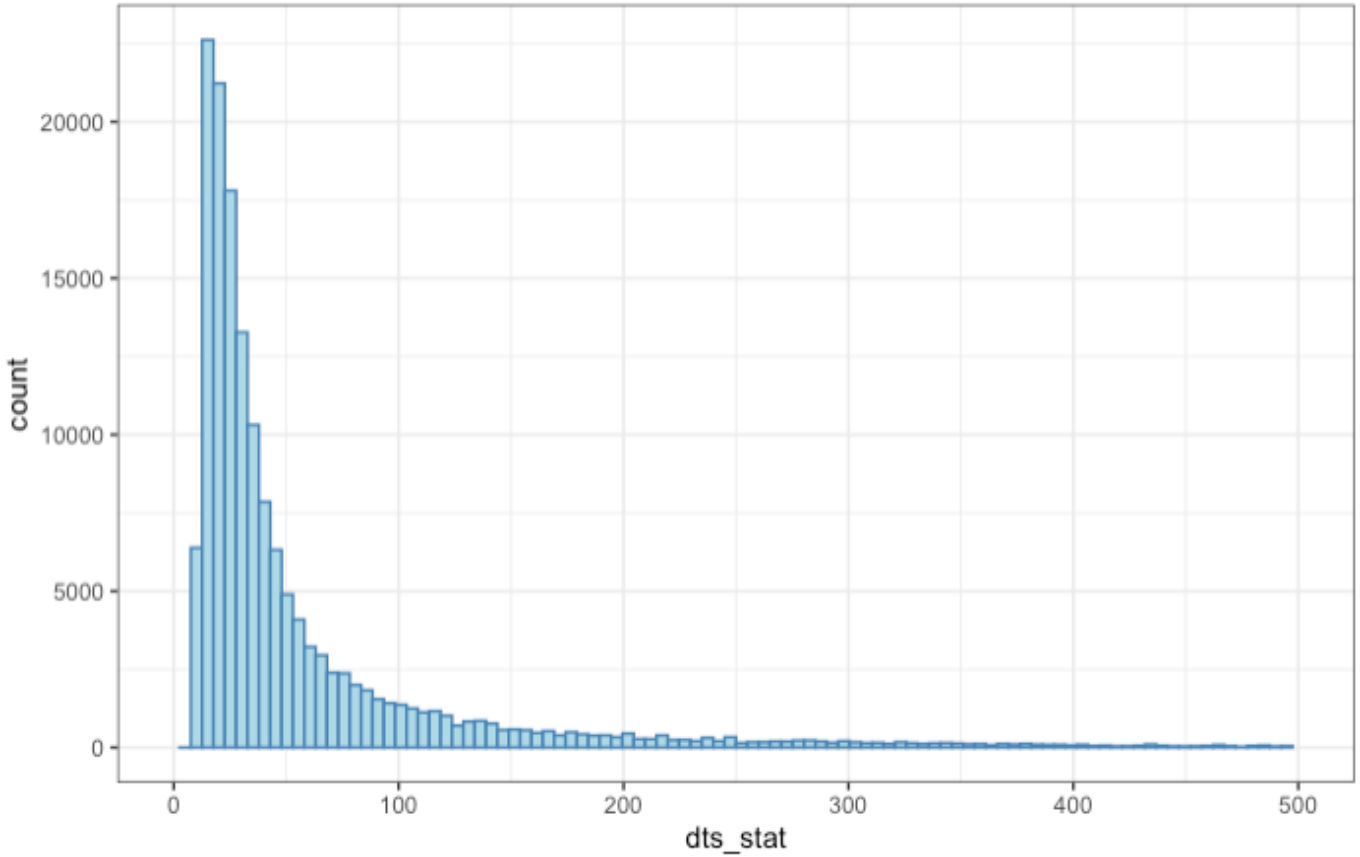
Armenia —Perceived Freedom and Control



Chapter 3: descriptives for DTS effect sizes

DTS effect sizes ranged from 7.43 to 2,286.41 (*M* = 65.02; *SD* = 103.29)

Distribution of significant DTS effect sizes



DTS effect size descriptives across nations

Nation	Mean DTS	SD DTS	Min DTS	Max DTS
ARG	31.36	59.47	0.39	1698.89
BGD	39.32	81.16	0.99	1907.64
BRA	34.79	68.26	0.92	1537.60
EGY	32.54	69.77	0.83	1985.92
GBR	39.52	68.16	0.55	1486.23
IND	52.07	96.60	1.10	1376.51
MEX	32.21	66.87	0.39	1559.06
NGA	47.62	79.90	0.76	1802.68
PAK	46.20	91.52	0.93	2286.41
PHL	47.83	80.37	0.63	1105.24
USA	99.50	158.22	2.55	2286.41
ZAF	34.88	62.13	0.76	1094.88
ZWE	33.55	61.95	0.92	1535.38

Chapter 3

DTS effect size descriptives for NEO facets (arranged by mean DTS)

Construct	Mean	SD	Min	Max
Liberalism	410.78	490.64	11.96	2286.41
Immoderation	122.03	129.78	11.60	849.67
Anxiety	119.43	193.92	10.75	985.25
Cheerfulness	117.22	178.53	10.38	1306.20
Self-discipline	116.12	162.77	10.54	1097.53
Anger	109.29	168.95	11.28	1194.16
Adventurousness	100.32	100.56	10.63	606.17
Cautiousness	99.27	149.95	10.49	1032.20
Activity-Level	97.71	93.19	11.22	497.72
Modesty/Humility	94.78	127.53	11.09	917.12
Emotionality	94.32	101.86	10.74	644.75
Vulnerability	91.35	160.80	10.56	1165.59
Trust	90.20	93.82	11.28	521.69
Excitement-seeking	82.96	110.98	10.53	847.05

Construct	Mean	SD	Min	Max
Imagination	75.39	88.35	11.15	488.01
Gregariousness	72.72	103.19	11.41	666.83
Depression	69.23	85.54	11.09	550.83
Morality	67.09	94.64	11.17	775.47
Achievement-striving	54.28	66.72	10.89	478.64
Self-consciousness	51.51	78.50	10.77	709.02
Dutifulness	51.29	74.75	10.15	601.60
Altruism	42.30	47.60	10.46	391.58
Aesthetic Appreciation/Artistic Interests	38.71	26.60	14.91	143.58
Self-efficacy	34.50	23.70	9.86	149.36
Orderliness	32.92	15.31	10.75	82.24
Sympathy	32.45	20.35	10.62	123.36
Cooperation	31.96	20.38	10.42	123.53
Assertiveness	24.80	9.41	10.62	58.56

Chapter 3

DTS effect size descriptives for NEO domains (arranged by mean DTS)

Construct	Mean	SD	Min	Max
Neuroticism	101.13059	146.60402	11.257350	985.24529
Openness To Experience	84.34072	93.01179	10.231937	573.41217
Conscientiousness	64.82145	121.50763	10.534919	1097.53144
Agreeableness	60.76026	75.81647	9.958428	508.66665
Extraversion	28.79035	17.24723	10.490953	93.86611