



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

**Comparability of personality trait scores across adulthood
(contribution to an invited symposium: Developmental aspects of
personality)**

Citation for published version:

Mottus, R, Realo, A & Allik, J 2013, 'Comparability of personality trait scores across adulthood (contribution to an invited symposium: Developmental aspects of personality)' Paper presented at 1st World Conference on Personality, Stellenbosch, South Africa, 19/03/13 - 23/03/13, .

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Comparability of personality trait scores across adulthood

Beyond technical issues

René Mõttus^{ab}, Anu Realo^a, Jüri Allik^a

^aUniversity of Tartu ^bUniversity of Edinburgh

The Big Five and The Mighty Age

An increasing number of studies available

- NEO go down, AC go up
- The pattern does not always replicate but it often does

It's not all that simple though

The Big Five domains don't always stick together

- Facets of the same domains show quite different age-differences
- In an item-level study, the same was true for the items of the same domain

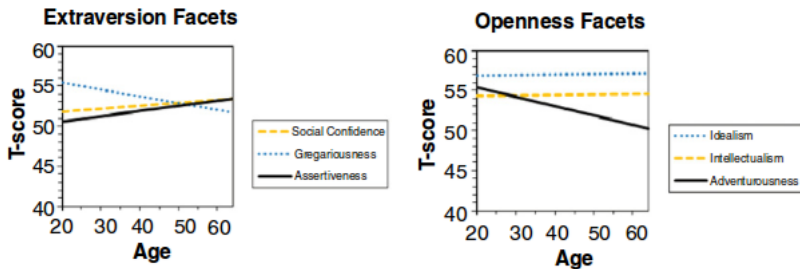


Figure: Soto and John, 2012, J Pers

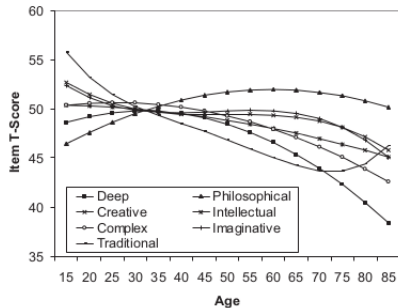
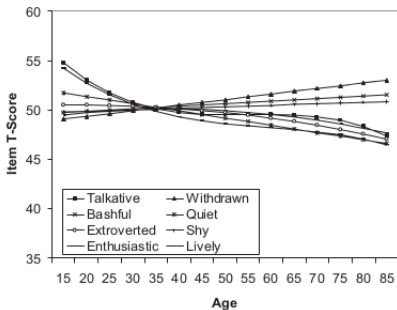


Figure: Lucas and Donnellan, 2009, Dev Psych

However, is the within-trait variability significant?

A lot of wobbling may happen by chance alone, and due to differential factor loadings

- It can be formally tested by testing the suitability of the same measurement models at different ages
- Known as measurement invariance (MI) testing

Configural, weak, strong and strict MI

Strong is our man

- Fit the same measurement model in different age-groups
- It fits without any parameter equality constraints – configural MI
- It still fits with loading equality constraints – weak MI
- **It still fits with intercept equality constraints – strong MI**
- It still fits with residual variance equality constraints – strict MI

Lack of strong MI

Is a bad thing

- Trait indicators wobble in **significantly** different directions

Some recent studies have shown MI across age-groups

But the question may not be **completely** settled yet

- Some studies have addressed short time-periods
- Others have used short personality measures (e.g., 3 items per trait)
- Some studies have aggregated items into parcels (3 to 4 parcels per trait), which *a priori* suppresses some of the item-specific variance
- Not many studies addressing **long time-periods** and using **comprehensive personality tests** (e.g., the NEO-PI-R) have sought to establish **strong MI**

Should we care though?

I guess we should

- Comparing aggregate scores is a little questionable when its constituents wobble in different directions:
 - Potentially valuable information gets masked
 - Technically incorrect (apples and oranges)
- Substantively speaking, development may be driven by specific, narrowly-operating mechanisms rather than by general mechanism

Here's another study

Given the importance of the problem

- We explicitly investigated the extent to which:
 - Facets of the same Big Five domains vary in age-differences
 - Items of the same facets vary in age-differences
- MI framework: measurement models compared across age-groups

Data

- A largish population-sample of Estonians from ages 18 to 91
- NEO-PI-3 ratings (240 items, 30 facets, five domains)
 - Self-ratings (N = 2,711)
 - Informant-ratings (partners, parents, children, friends etc; N = 2,658)

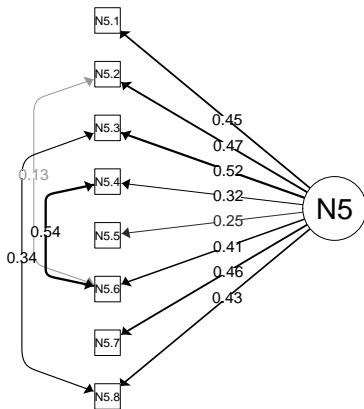
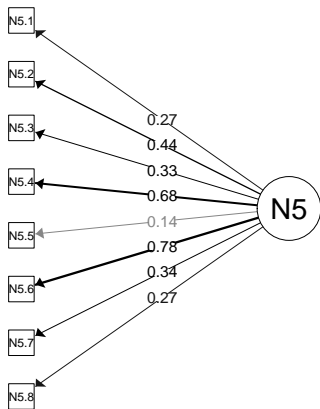
A measurement model for each facet

It wasn't as simple as that

- Unidimensional models with 8 items loading on a single trait
- MLR estimator, models fitted with 'lavaan' (Rosseel, 2012)
- Poor fit in 55 of the 60 facets:
 - Comparative Fit Index (CFI) < .95 and/or
 - Root Mean Square Error of Approximation (RMSEA) > .08
 - From 1 to 5 pairs of residual correlations were needed (typically 2 or 3)

N5: Impulsiveness

CFI = .56, RMSEA = .152 [.144, .159] → CFI = .96, RMSEA = .049 [.042, .057]



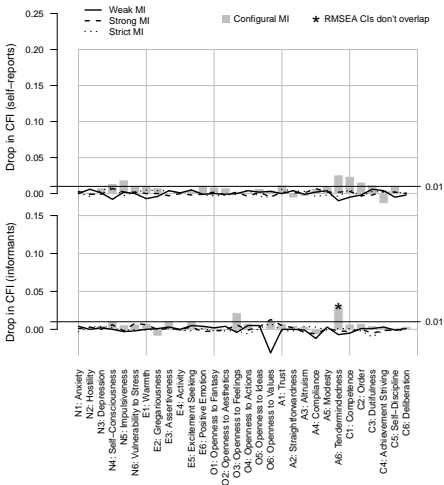
All models were then fitted as multi-group models

12 age-groups; N = 149 to 310

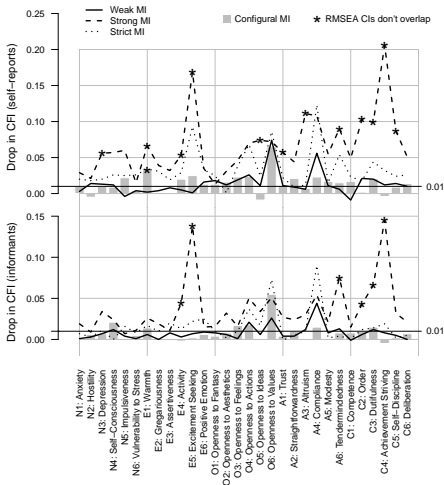
- Configural MI: means 0 in all groups
- Weak MI: means 0, loadings equal in all groups
- Strong MI: loadings, intercepts equal in all groups
- Strict MI: loadings, intercepts, residuals equal in all groups

- Drop in CFI ($\Delta\text{CFI} \leq .01$) compared to a one step less constrained model = evidence for MI
- Overlap in RMSEA 90 percent confidence intervals = evidence for MI

MI based on random age-groupings



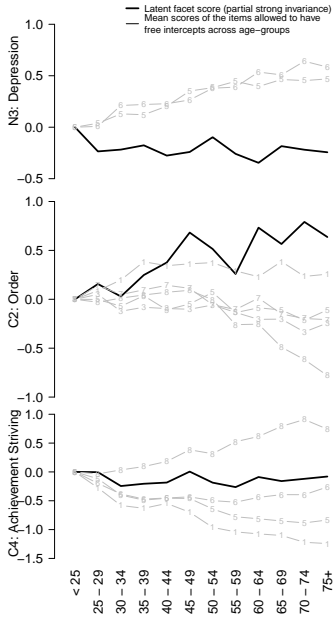
MI based on true age-groupings



Several loadings/intercepts needed to wobble freely

Known as partial MI

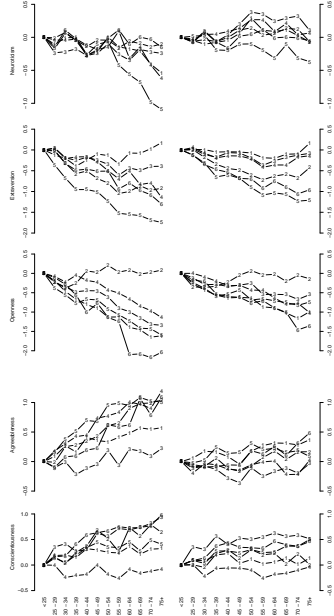
- If there was no weak MI in the first place, loading constraints were relaxed as appropriate
 - In self-ratings 17 facets 'healed' with 1 to 5 loadings being freed
 - In informant-ratings 7 facets 'healed' with 1 to 3 loadings being freed
- If there was no strong MI in the first place, intercept constraints were relaxed as appropriate
 - All self-report facets 'healed' with 1 to 6 intercepts being freed
 - 28 informant-report facets 'healed' with 1 to 4 intercepts being freed
- Overall, 101 of 240 intercepts had to be freed in self-ratings (70 in informant-ratings)



Item-specific residual variance and age

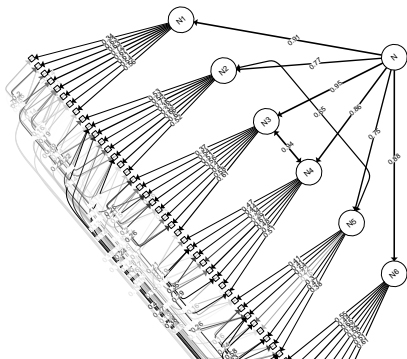
Each item residualized for its facet score

- In self-reports, unique variances of 46 percent of items had significant correlations with age ($p < .0002$)
- It was 34 percent in informant-ratings
- 27 percent of items had similarly significant residual age-correlations in both types of ratings
- Lost information when comparing scale-scores only



Did facets define the Big Five domains in the same way at different ages?

First attempt: hierarchical models ... well, they fell over



Did facets define the Big Five domains in the same way at different ages?

Second attempt: facets as principal component scores

- Weak MI held except for self-reported Agreeableness
 - Loadings of A1: Trust and A4: Compliance were freed
- Strong MI never held
 - Intercepts of 1 to 4 facets per domain were freed to get well-fitting models

The same measurement models rarely apply at different age levels

As far as Estonians and NEO-PI-3 are concerned

- This may sound like a technical glitch but it is not (only that)
- Instead, this reflects systematic properties of personality traits
 - Age-differences are specific

Implications for personality development

To the extent that these age-differences reflect development

- Personality development may be more complex than age-differences in domains or even facets
- Driven by specific and narrowly-operating mechanisms rather than by broad mechanisms whose impact is aligned according to the Big Five domains
- Looking at the domain or facet scores alone may imply losing valuable information

