

How to apply personality as a worldwide common concept?

Big Five, Big More, or should we?

Anke M. Weekers

ITC conference
Hong Kong, July 2010



Presentations

1. Neal Schmitt

Impact of Measurement Invariance on Construct correlations, Mean Differences, and Relationships with Extrenal Correlates: Big Five and RIASEC Measures

2. Filip de Fruyt

The Big Five in selection and development assessment: "blessing and yoke"



Presentations

3. Anke Weekers
Response processes in personality measurement
4. Alec Serlie
Are there Big Cultural differences in Personality?
5. Fons van de Vijver
Discussion



Cross-Cultural Measurement and Partial Invariance

Neal Schmitt
Michigan State University

Talk presented at the 7th Conference of the
International Test Commission
July 2010

Problem

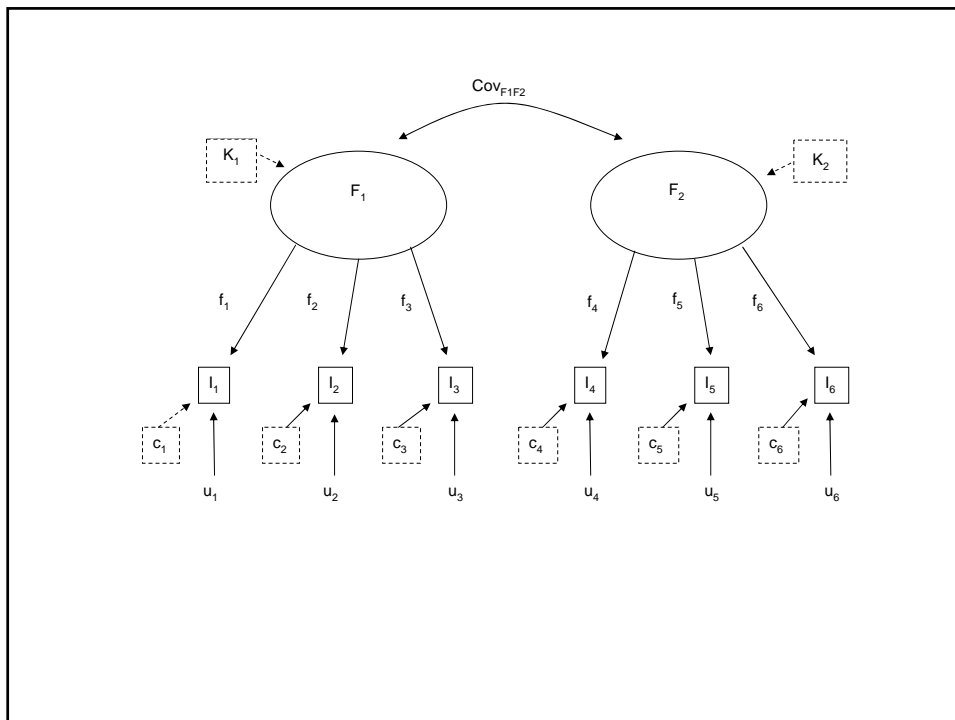
- Do measures constructed with and for respondents of one culture work the same way when used with members of another culture or subgroup?
- Do persons of different subgroups with the same observed score on some measure have the same standing on the construct that is the target of measurement

Solutions

- Measurement equivalence: Item response theory and confirmatory factor analysis (CFA)
- Relational equivalence usually analyzed by examining subgroup differences in regression analyses

Focus in this paper will be on the use of CFA to examine subgroup differences

- Meredith (1993) is generally credited with the formal specification of how various forms of invariance are established though others obviously provided earlier versions of the approach
- Test of configural invariance: Estimates of the same model across groups, but there are separate estimates of the factor loadings, intercepts, uniquenesses and factor variances and covariances. This overall model for one group is depicted on the next slide.



- In this figure each observed variable or indicator (I) is represented by this regression equation:
- $I = c + fF + u$
- (1)
- where I represents an observed measure or indicator, c is a constant or the intercept in the equation, F is a latent or unobserved variable, f is a regression coefficient or factor loading in CFA terms, and u is the residual or uniqueness associated with the regression.

Invariance Defined

- Metric invariance: Specify that the factor loadings in the previous figure and equation are equal across groups
- Scalar invariance: Specify that the intercepts in the equation are equal across groups. Both metric and scalar invariance are usually deemed necessary to address differences in latent factor means.
- Strict invariance: Also specify that the uniquenesses are equal. This is considered necessary to make comparisons of observed factor means
- Then tests of structural equivalence follow: Tests of the covariance of latent factors, mean differences in factor means and relations with other constructs.

This approach to invariance has been used in hundreds of studies summarized in at least three reviews.

- Vandenberg and Lance (2000) and Schmitt and Kuljanin (2008) focused on a review of the manner in which these tests were conducted.
- Chen (2008) estimated the frequency and magnitude of invariance reported in 97 studies
 - In 75% of the studies, some portion of the factor loadings lacked invariance; the difference on average was .13 in factor loadings
 - In most cases, the factor loading was higher in the “reference” group than in the “focal” group
 - This usually results in a measure that is less reliable for the minority focal group than is true for the majority reference group

Common Solution When a Lack of Invariance is Found

- Allow for a partially invariant model in which separate subgroup parameters are estimated.
- The question addressed in this paper is what impact does employing a partially invariant model have on estimates of the structural parameters in models
 - Variance-Covariance matrices of the latent factors
 - Estimates of latent mean differences
 - Estimates of relations with outside variables

Method

315 African American and 338 Caucasian college students responded to measures of the Big Five and RIASEC dimensions

These measures were combined into three parcels for each of the 11 constructs, so there were 33 measures. In addition, we secured high school and college grade point averages for each student. The latter two variables represented external variables.

Analyses

Multiple groups CFA using LISREL software specifying a measurement model with 11 covarying latent factors each with three indicators. Configural model estimated separate subgroup parameters – factor loadings, intercepts, uniquenesses, factor variances and covariances. Metric, scalar, and strict invariance models were each estimated in turn with no consideration of statistically significant differences in parameters.

Results

Model	Chi-square	df	NNFI	CFI	RMSEA	Change in chi-square
Configural	1900.83	880	.9287	.9406	.05891	
Metric	1952.70	902	.9284	.9389	.05949	51.86(22)
Scalar	2079.27	924	.9232	.9328	.06136	126.58(22)
Uniqueness	2186.04	957	.9211	.9285	.06188	106.77(33)
Fac. Var.	2201.86	968	.9217	.9282	.06194	15.82(11)
Fac. Cov.	2314.71	1023	.9224	.9248	.06269	112.86(55)
Means	2449.90	1034	.9159	.9176	.06569	185.19(11)

Verbal summary of Table Results

Measurement Model

Configural fit is reasonable

Slight evidence of lack of metric invariance; chi-square and CFI change statistics

Fairly substantial evidence of lack of scalar and uniqueness invariance

Structural Model

Slight and nonsignificant difference in variances

Some evidence of differences in covariances

Fairly large difference in factor means

Evaluation of Partial Invariance Models

Model	Chi-square	df	NNFI	CFI	RMSEA	Change in Chi-square
Configural	1900.83	880	.9287	.9406	.059	
Partial metric	1911.83	895	.9302	.9408	.058	11.00(15)
Partial scalar	1936.83	911	.9308	.9403	.058	25.00(16)
Partial Uniq.	1975.41	936	.9318	.9395	.057	38.58(25)
Factor Var.	1986.63	947	.9325	.9395	.057	11.22(11)
Factor Cov.	2061.70	994	.9340	.9379	.057	75.07(47)
Equal Means	2166.77	1005	.9290	.9324	.059	105.17(11)

Verbal Summary of Partial Invariance Models

- Seven factor loadings, six intercepts, eight uniquenesses and eight factor covariances were separately estimated across groups in these analyses. All were significantly different by group.
- As you can see, adding these parameters produced models across levels of invariance that were reasonably similar.
- The factor means as a whole were still significantly different – see more detail on the next slide

Means with and without provision for lack of invariance

	Fully Invariant model		Partially Invariant Model	
Variable	Mean Diff.	Std. Difference	Mean Diff.	Std. Difference
Agreeableness	.01	.03	.04	.09
Conscientious	-.00	-.01	-.00	.00
Extraversion	-.01	-.02	.02	.03
Openness	.07	.14	.07	.14
Stability	-.08	-.11	.01	.01
Realistic	.33	.61*	.18	.31*
Investigative	.23	.28*	.23	.28*
Artistic	-.30	-.43*	-.30	-.42*
Social	-.28	-.45*	-.31	-.51*
Enterprising	-.32	-.45*	-.32	_.45*
Conventional	-.07	-.11	-.11	-.17

Verbal Summary of Table

- No differences in Big Five measures
- Significant differences on all RIASEC measures with the exception of Conventional interests
- Mean differences across models in which no provision for lack of invariance was made and those in which partial invariance was allowed were almost exactly the same with the exception of the means for Realistic interests.

Relationship with College GPA: Results of Regression of CGPA on Constructs

	Partial Invariance Model--AA	Partial Invariance Model--CA	Full Invariance Model-- AA	Full Invariance Model -- CA
Agreeableness	.47*	.44	.50*	.45*
Conscientious	.08	.12	.04	.10
Extraversion	-.24*	-.27*	-.23*	-.26*
Openness	.19	-.24	.16	-.20
Stability	-.18	.07	-.15	.07
Realistic	.23	-.02	.17	-.02
Investigative	-.13	.24*	-.10	.23*
Artistic	-.01	.17	.04	.14
Social	-.14	-.30	-.16	-.29
Enterprising	.21	.45*	.19	.42*
Conventional	.03	.16	-.10	-.14

Verbal Summary of Regressions

- There is almost no difference in the regression coefficients for fully and partially invariant models
- Agreeableness is positively related to CGPA for both groups of respondents
- Extraversion is negatively related to CGPA for both groups of respondents
- Enterprising interests are positively and significantly related to CGPA for Caucasian students, but nonsignificantly so for African Americans

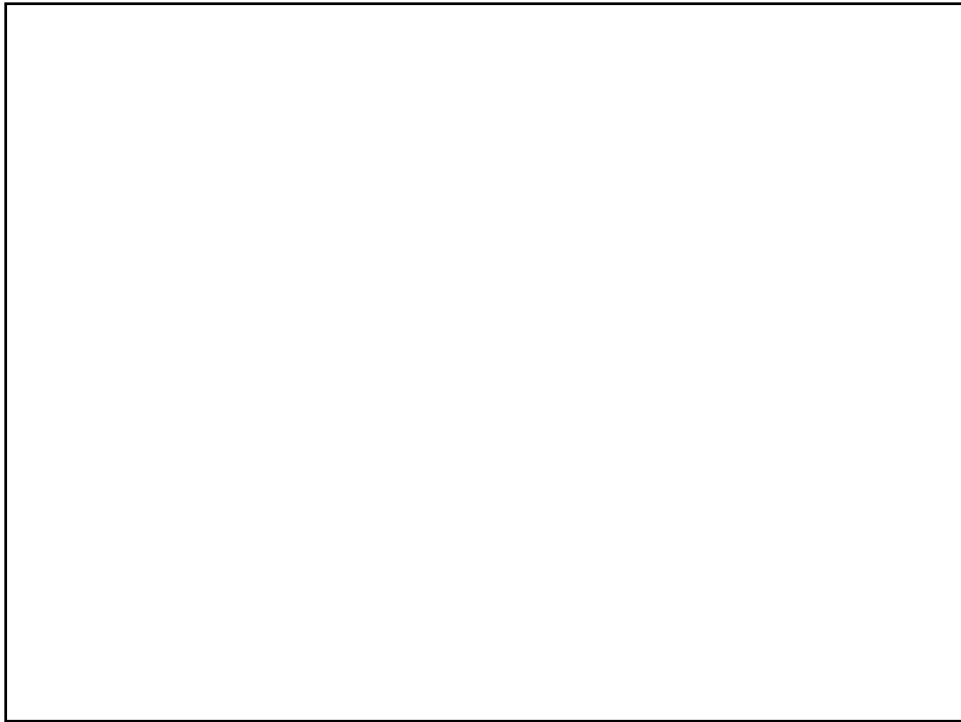
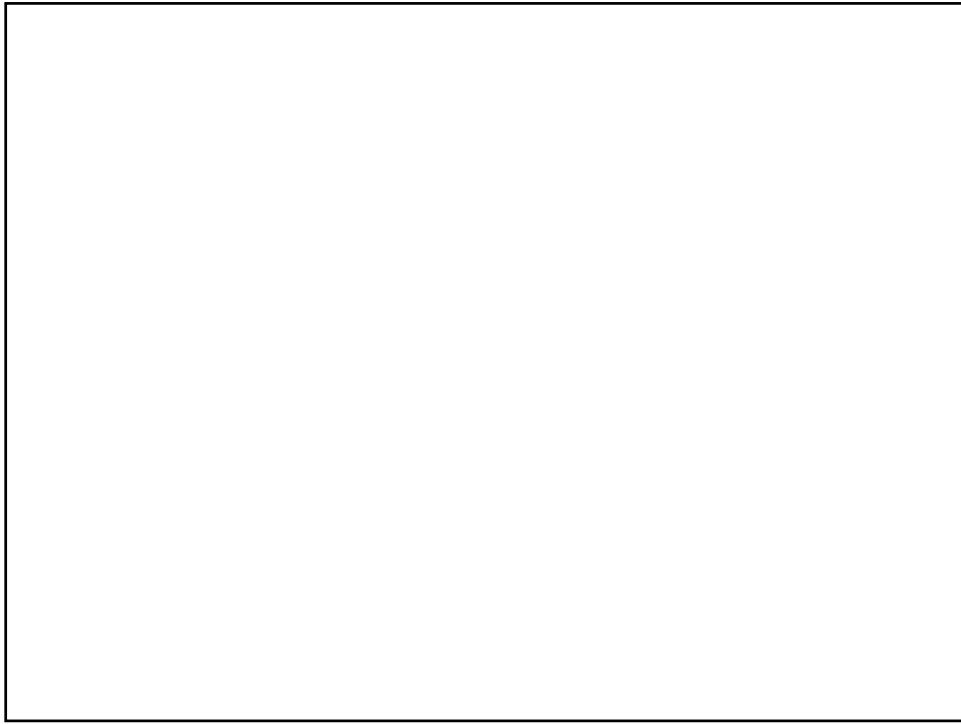
Summary and Conclusions

1. There was evidence of a reasonably large amount of difference in subgroup parameters
2. Allowing for separate subgroup parameter estimation produced acceptable model fit
3. Allowing for partial invariance had minimal impact on estimates of factor mean differences with one exception – Realistic Interests
4. Allowing for partial invariance had minimal impact on any of the regression coefficients for either group
5. Overall, concerns about lack of invariance may be overstated.

Limitations

1. We had only American subgroups although there was a fair degree of lack of invariance and only one set of data
2. Allowing for partial invariance represents the type of specification searches decried by MacCallum (1986) in the modification and evaluation of a priori models.
3. We investigated the impact of partial invariance only on the predictor side of the equation. One could also examine similar problems on the outcome (CGPA) side.

Thank you for your attention!



Response processes in personality measurement

Anke M. Weekers

ITC conference
Hong Kong, July 2010



Overview

- Introduction
 - General
 - Response processes
- Study
 - Questionnaire
 - Participants, analysis
 - Results
 - Differences between USA and Dutch questionnaire
- Conclusions



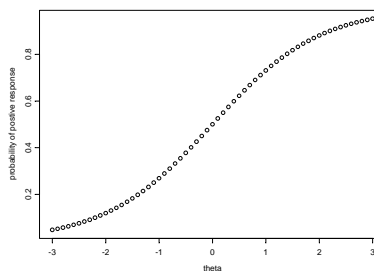
Introduction: general

- Self-report questionnaire
 - Being neat is not exactly my strength
 - Half of the time I do not put things in their proper place
 - I become annoyed when things around me are disorganized
- Construction and analysis
 - CTT and factor analysis
 - dominance IRT models



Dominance response processes

- I become annoyed when things around me are disorganized
- Maximum performance, but typical performance?
- Being neat is not exactly my strength



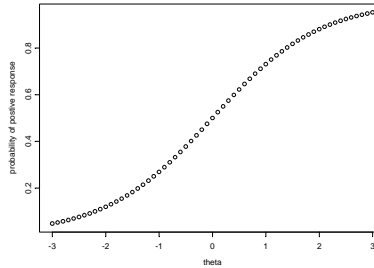
$$p_i(\theta) = P(X_i = 1 | \theta) = \frac{\exp(a_i(\theta - \beta_i))}{1 + \exp(a_i(\theta - \beta_i))}$$



Dominance response processes

What about:

- Half of the time I do not put things in their proper place

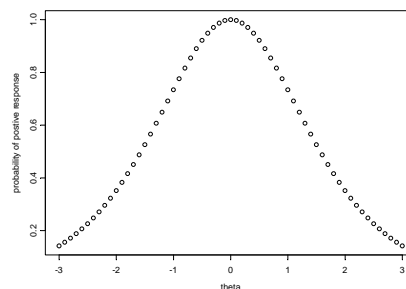


$$p_i(\theta) = P(X_i = 1 | \theta) = \frac{\exp(a_i(\theta - \beta_i))}{1 + \exp(a_i(\theta - \beta_i))}$$



Single-peaked response processes

- Half of the time I do not put things in their proper place
- responses have another meaning

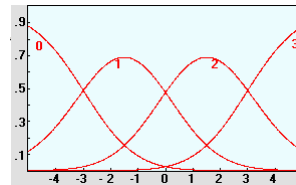


$$P_i(X=1|\theta) = \frac{\exp(\alpha_i[(\theta - \beta_i) - \tau_i]) + \exp(\alpha_i[2(\theta - \beta_i) - \tau_i])}{1 + \exp(\alpha_i[(\theta - \beta_i) - \tau_i]) + \exp(\alpha_i[2(\theta - \beta_i) - \tau_i]) + \exp(\alpha_i[3(\theta - \beta_i) - \tau_i])}$$



Single-peaked Response processes

Agree and disagree for two reasons



$$P_i(X=1|\theta) = \frac{\exp(\alpha_i[(\theta - \beta_i) - \tau_i]) + \exp(\alpha_i[2(\theta - \beta_i) - \tau_i])}{1 + \exp(\alpha_i[(\theta - \beta_i) - \tau_i]) + \exp(\alpha_i[2(\theta - \beta_i) - \tau_i]) + \exp(\alpha_i[3(\theta - \beta_i) - \tau_i])}$$

Study: research question

Which models describe personality data best?
Does this differ over countries?

Study: Questionnaire

- Order questionnaire
 - By Chernyshenko, Stark, Drasgow, & Williams (2007)
 - 20 items (6,6,8)
 - Selection based on
 - Content
 - Psychometric properties
 - Analysis: results USA
 - Worse fit under dominance IRT model
 - Ordering conform content
 - 6 single-peaked items

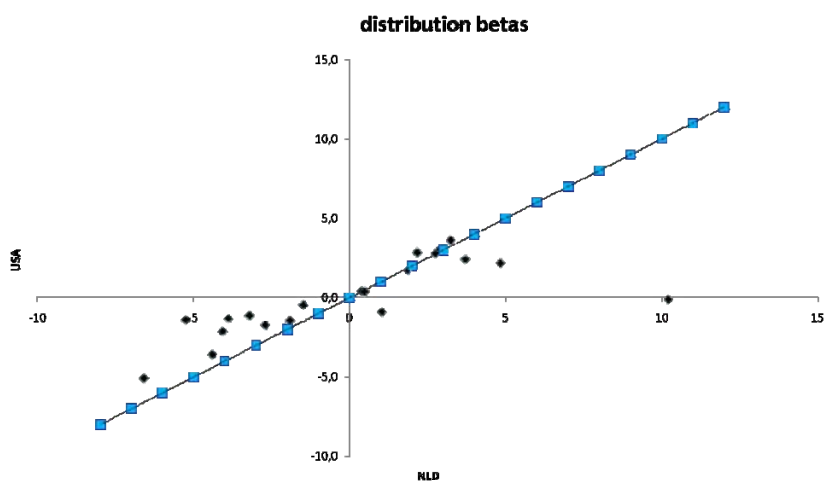
Study: participants, analysis

- Participants (N = 704)
 - Secondary school students
 - University psychology students
- Analysis
 - OPLM
 - GGUM

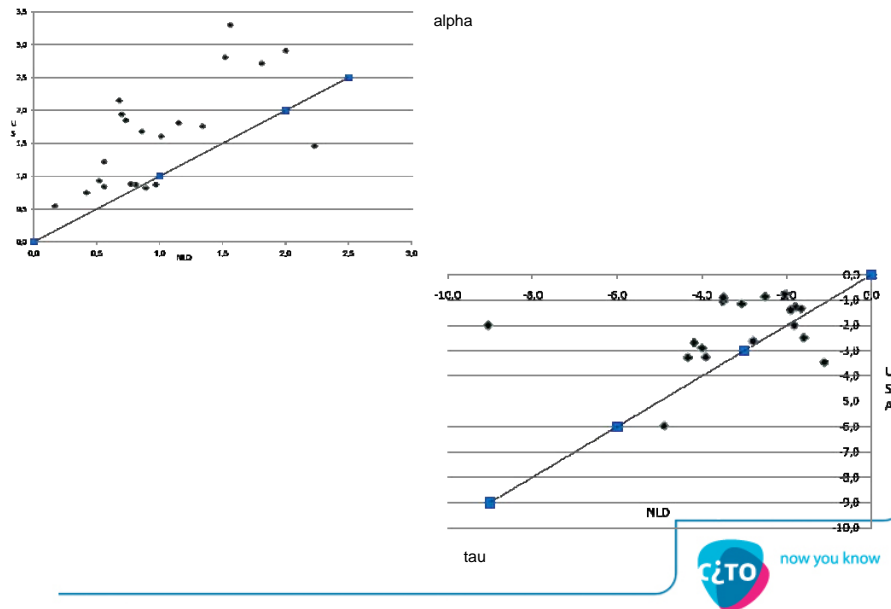
Study: Results Netherlands

- Cronbach's alpha = 0.74
- Average item-test correlation = 0.30
- OPLM
 - Not all items fit the model
 - 1 deleted: My room neatness is about average
 - Average discrimination: 4.2 2.5 2.8
- GGUM
 - 8 decreasing
 - 4 single-peaked
 - 8 increasing

Differences location parameters



Shape parameters



Conclusions



- Neutrally worded items can add to measurement precision
- Difficult to model neutrally worded items under dominance model
- Unfolding models and single-peaked response processes are useful
- Location results similar between USA and NLD
- Shapes less steep, more wide NLD

1 GITP Big Cultural differences? Alec Serlie 19-07-2010

Are there Big Cultural differences in Personality?


Annemarie M.F. Hiemstra*, Alec W. Serlie*
 Rob G.J. van Leeuwen ¥,
 Madelijn Bazen ‡

*Erasmus University Rotterdam / GITP
 ¥ GITP
 ‡ Leiden University





2 GITP Big Cultural differences? Alec Serlie 19-07-2010

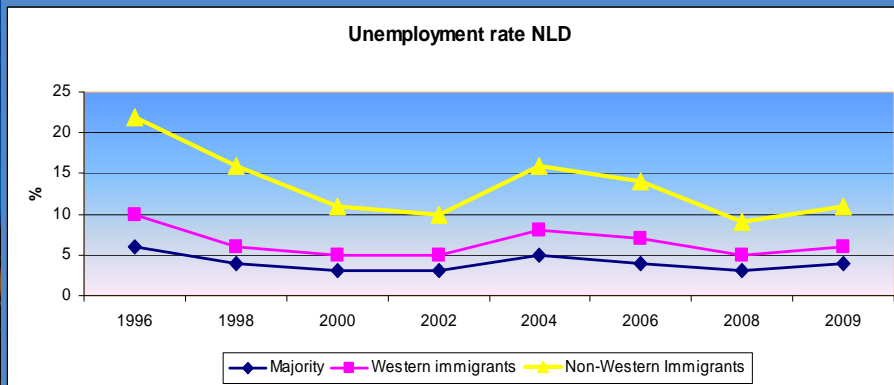
Introduction: Dutch population



	total population	Majority		Minority	
2000	15.863.950	13.088.648	83%	2.775.302	17%
2009	16.485.787	13.198.081	80%	3.287.706	20%
2010	16.577.612	13.215.458	80%	3.362.154	20%



Introduction: Unemployment rate



gitpl

Introduction

- > Need for Culture free testing
- > No irrelevant variables e.g.
 - Language
 - Testing experience
- > Similar interpretation results

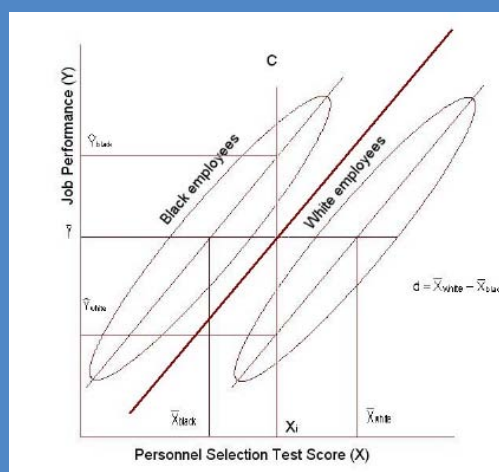
gitpl

Introduction

> Van Leest (2001):

- Differences in personality questionnaires due to
 - 'expressions' eg 'Is the life and soul of the party'
 - uncommon phrases
 - Social desirability

Introduction: Cleary model



Introduction: DIF

- > Are items answered in the same manner by different groups
- > DIF occurs when answer pattern depends on the group the respondent belongs to

Introduction: DIF Methods

- < Mantel – Haenszel
- < Logistic Regression
- < Item Respons Theory

Introduction: DIF LR

Effect size Zumbo & Thomas (1997)

Type A items – negligible DIF: $\Delta R^2 < .13$

Type B items – moderate DIF $.13 \leq \Delta R^2 \leq .26$

Type C items – large DIF $\Delta R^2 > 0.26$

Effect size Jodoin & Gierl (2001)

Type A items – negligible DIF: $\Delta R^2 < .035$

Type B items – moderate DIF $.035 \leq \Delta R^2 \leq .070$

Type C items – large DIF $\Delta R^2 > 0.070$

Hypothesis

- > There are items showing DIF for personality items, when comparing majority and minority groups

Method: Participants

		Majority	Minority	Total
Gender	F	99	80	179
	M	62	39	101
Total		161	119	280

$\chi^2 = .98$ (ns)

Age

Majority 22,1 years (SD: 2,4)

Minority 24,6 years (SD: 3,9)

F: 44,5 p <.01

gtp

Method: Participants

	Frequency	Valid Percent
Majority	161	57,5
1st generation Immigrant	60	21,4
2nd generation Non-Western Immigrant	52	18,6
2nd generation Western Immigrant	7	2,5
Total	280	100,0

gtp

Method: Participants GMA



	Mean	Std. Error of Mean	F	Sig.	Eta Squared
Majority	4,74	,178	5,037	,026	,019
Minority	4,13	,207			

gitpl

Method: G5R-reliability



	mean	SD	S	E	O	A	C	SP
Stability	85.5	10.7	(.90)					
Extraversion	98.8	13.1	.30**	(.92)				
Openness	183.6	15.0	.32**	.53**	(.90)			
Agreeableness	108.8	9.1	.22**	.26**	.44**	(.85)		
Conscientiousness	161.5	15.6	.31**	.23**	.34**	.52**	(.94)	
Selfpresentation	136.9	15.3	.37**	.62**	.61**	.24**	.51**	(.93)

N=52,764

gitpl

Method: G5R-test-retest reliability

		measurement 1 (t1)					
		Stability	Extraversion	Openness	Agreeabl	Conscien	SelfPres
Measurement 2 (t2)	Stability	.66	.22	.18	.19	.23	.20
	Extraversion	.19	.64	.19	.07	.09	.35
	Openness	.26	.28	.53	.25	.17	.33
	Agreeableness	.14	.09	.16	.48	.24	.05
	Conscientiousness	.23	.15	.16	.33	.61	.29
	Selfpresentation	.31	.43	.30	.15	.31	.61

N=386

t2-t1: mean: 390 days (SD 293)

gitpl

Method: G5R-validity



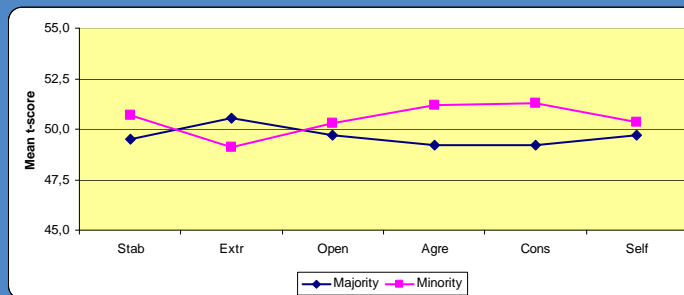
		G5R					
		Stability	Extraversion	Openness	Agree	Conscien	SelfPres
NEO-PI-R							
Neuroticism		-.65**	-.29**	-.27**	-.21**	-.42**	-.40**
Extraversion		.29**	.70**	.45**	.27**	.28**	.59**
Openness		.11	.37**	.56**	.27**	.08	.20**
Agreeableness		.09	-.06	.02	.49**	.12	-.21**
Conscientiousness		.38**	.24**	.28**	.27**	.62**	.47**

N=261

gitpl

Results: mean differences

	Majority	Minority	F	sig
Stab	49,5	50,7	,95	ns
Extr	50,6	49,1	1,42	ns
Open	49,7	50,3	,22	ns
Agre	49,2	51,2	2,61	ns
Cons	49,2	51,3	2,92	ns
Self	49,7	50,3	,26	ns



gtp

Results: DIF items

	Total items	DIF items		Majority	Minority
Stability	23	4	17,4%	2	2
Extraversion	27	2	7,4%	2	
Openness	28	1	3,6%	1	
Agreeableness	47	0	0,0%		
Conscientiousness	40	0	0,0%		
Self Presentation	36	4	11,1%	3	1
Total	201	11	5,5%	8	3

gtp

		DIF χ^2	sig	DIF R^2	effect size
Stability	Is able to deal with tension	7,78	0,02	0,026	A
	Worries about possible setbacks	17,68	< ,01	0,048	B
	Is easily irritated by some people	28,5	< ,01	0,008	A
	Is often irritated	8,29	0,02	0,007	A
Extraversion	Likes telling stories	12,05	< ,01	0,035	B
	Keeps a low profile	12,56	< ,01	0,045	B
Agreeableness	Helps those who are worse off	17,06	< ,01	0,005	A
SelfPresentation	Has boundless energy	9,83	< ,01	0,033	A
	Is willing to go to great lengths to achieve success	11,67	< ,01	0,135	C
	Is extremely demanding of him/herself	11,52	< ,01	0,029	A
	Is career-oriented	16,72	< ,01	0,050	B

gtp

- ## Conclusion
- > When comparing Higher educated Minority / Majority groups on P-questionnaire
 - > Hypothesis accepted
 - > 5.5% of the items show DIF
 - > DIF spread over different P-dimensions
 - > Minority → higher score
 - > No DIF for 'expression' items eg *Is the life and soul of the party*
 - > Stop stereotyping?

gtp

Limitations & Recommendations

- > Large diversity in Minority group
 - Western / non western
 - 1st and 2nd generation immigrants
- > No ethnic registration in NLD

- > → Larger sample, not only students
- > → additional variables
 - Years in country
 - Indication of acculturation

Thank You

- > Alec.Serlie@gitp.nl
- > Serlie@fsw.eur.nl



Discussion

Fons van de Vijver

General

- Rich set of good papers

Neal Schmitt: Impact of Measurement Invariance on Construct Correlations, Mean Differences, and Relationships with External Correlates

- Deals with fundamental issue of bias: same score and still subgroup differences (partial invariance)
- Partial invariance in each stage of invariance testing
- Not much difference in using estimates based on partial invariance and estimates neglecting partial invariance
 - Comment: Findings generalizable?
 - Cf. studies of removal of DIF
- All sources of bias considered?
 - Comment: Method bias? Acquiescence,

Filip de Fruyt: The Big Five in selection and development assessment: “blessing and yoke”

- Review of research
 - Five-factor model is replicable
 - Do we need all five in IO? → Conscientiousness
 - Comment: Not necessarily comprehensive → CPAI-2
 - Less analysis of criterion variables
 - Comment: More susceptible to bias
 - Use of IPIP items
 - Comment: Psychometric analysis of these item banks not tried out
 - “We can import and borrow more than we ever thought”
 - Comment: If tested properly
 - All kinds of interesting ideas on how to go beyond standard five-factor model

Anke Weekers: Response processes in personality measurement

- Puzzle of positively and negatively formulated items
 - Comment: when do these items (not) split up across different factors with a slight negative correlation?
- Different processes for maximum and typical performance: dominance vs. single-peaked
 - Comment on findings: Location parameters show quite some cross-cultural variation, notably in the discrimination parameters
 - Reasons for differences not very clear to me
- Interesting way of dealing with neutral items

Alec Serlie: Are there big cultural differences in Personality?

- Need for culture-free testing
 - Comment: reduce bias as much as possible
- DIF in personality
 - Less frequently studied, though very important
 - Implicit culture in much personality assessment
- Not too much DIF found
 - Substantive reasons for the DIF not very clear (“life and soul of the party” not biased)
 - Acculturation items?