### Questionnaire development documentation in Generations and Gender Programme







Andrej Kveder GGP project manager kveder@nidi.nl Generations and Gender Programme



## Outline

### About GGP

Questionnaire development

- Documentation
  - Implementation
  - Automated routing scripts
  - para-data



## What is GGP?

### Generations and Gender Survey (GGS)

- strong theoretical base
- broad age range
  - 18-79
- large sample
- comparative pan-European
- retrospective life-histories
- prospective
  - panel sample design
- GGP contextual database
  - Theory based conceptual framework based on welfare state research
  - Data that influence timing & sequencing of events (Legal norms and regulations, Statistical norms)



### **Implementation parameters**

- centrally developed questionnaire
- sampling
  - random
  - ranging from register based to random route
- sample size
  - cca. 10,000 respondents wave 1 per country
    - 7,682 Australia 25,000 Hungary
  - aimed sample size 8,000 after wave 3
- Face-to-Face interviews
  - PAPI as well as CAPI
  - mixed mode in Norway
    - CATI drop-off pick-up register data







SEVENTH FRAMEWOR

N i D/

- 💠 Wave 1
  - 19 countries

Wave 2

- Bulgaria, France, Georgia, Germany, Netherlands
- Hungary, Italy, Russian Federation, Australia



## **Questionnaire elements**

### question types

- (auto) biographic
- attitudinal
- time scope
  - retrospective histories
  - prospective intentions
  - present state



- social networks



## **Questionnaire content**

- socio-demography
- HH composition
- demographic events
  - childbearing
  - partnership
  - leaving parental home
- economic activity
- social support
- TPB
- attitudes



## **Questionnaire development history**

- first theoretical concepts in 2000
- Wave 1 in 2005
  - pre-tested in UK and Russian Federation
- Wave 2 in 2007
  - new section on activity history
  - some "stable" measures temporarily removed
- Wave 3 in 2009
- \*\*\*\*
- SEVENTH FRAMEWORK PROGRAMME N 1 D 1

- "stable" measured re-introduced
- new section on grandchildren
- \*GGS 2015 in 2012

### **Recent GGP developments**

### methodological evaluation

- substantive
- measurement equivalence
- item nonresponse
- new questionnaire
  - CAI
  - implemented changes based on evaluation
  - mode equivalent
  - additions
    - big 5 15 items

### 💠 pilot

• 3 modes: CAPI, CATI, WEB



NiD

### **Implementation documentation**

### compliance

- question by question
- V1.0 Access database
  - not flexible enough
  - questionnaire based
- V2.0 Excel sheet
  - dataset (variable) based

### information

- availability
- deviations
  - country specific variables
  - country specific values



NiD

GG,





#### GGS Wave 1 Variables' availability and Country Specificities

Generations & Gender Programme

Caption								
NA	Variable Not Available in the dataset							
x additional values	Values from core questionnaire (see in codebook) + x additional values specific to the country. See list of specific values in the comment.							
x specific values	Only country specific values (values from core questionnaire not asked in this country). See list of values in the comment.							
opt.	Question from optional submodules							
cons.	Consolidated variable derived from original variables							
axxx_1xxx (ex: a866_1601)	Country specific variable: guestion slighty differed from the core questionnaire and only asked in this country. See details and justifications in the comment.							

Question number	Variable Name	Variable Label	Australia v4.1	Austria v4.1	Belgium v4.1
			Country code = 24	Country code = 21	Country code = 23
Question number	amonth	Mnth of interview			
	ayear	Yr of interview			i
1202	adwell	Type of dwelling where R lives	6 additional values	NA	) i
1203	afloor	Floor number on which R lives	NA	NA	8
3	aregion	Region or administrative unit of residence (country spec.)	8 specific values		3 specific values
	atype	Type of settlement (rural/ urban/ capital)	6 specific values		1 additional value
	aplace	Code of place of residence (City / town / village code)	NA	-	NA

### P Implementation documentation conclusion

### information collected

- at the time of harmonization
- PRO
  - very accurate information
  - compliance based on variables
    - constructs, and not on exact questions

### CON



- not timely documentation
  - sometime 2 or more years after data collection
- compliance based on variables
  - differences in questionnaires not documented

## **Questionnaire routing check script**

developed to check routing compliance between questionnaire and data

### elements

- standardized description of the questionaire
  - routing conditions
  - validity checks
- scripts
  - questionnaire reading
  - comparison with the data
  - automated corrections



### **Questionnaire routing - example**



\*\*\*\*

GGP



### **GPStandardized questionnaire** description

// section 1	
gridstart ahg1_ ahg2_ ahg3_ ahg4_ ahg5_ ahg6m_ ahg6y_ ahg7m_ ahg7y_ ahg8_ ahg9_	, exempt(ahg1_)
check ahg1_ range 1,95	
check ahg2_ range 1,2	
check ahg3_ range 0,17	
check ahg4_ range 1,2	
check ahg5_ range 0,996	
check ahg6m_ range 1,12 range 21,25	
check ahg6y_ range 1900,2010	
cond c (inlist(ahg3_,4,5,6))	
check ahg7m_ range 1,12 range 21,25	
check ahg7y_ range 1900,2010	
endcond	
cond c (ahg5_>13) & !missing(ahg5_)	
check ahg8_ range 1,10	
endcond	
check ahg9_ range 1,2	
gridend	
check al05 range 1,2	
cond a105==1	
check a106a_1 range 1,996	
check a106a_2 range 1,996	
endcond	
cond alu5==2	
check al06D range 1,9996	
check al0/m range 1,12 range 21,25	
cneck alo/y range 1900,2010	
chock a100	
CHECK alos	

## Routing script – potential use

comparison of questionnaires

generation of para-data

- question sequencing
- any information based on collected data
- encoding question information to the variables



## **Questionnaire Para-data**

### question classification

- question typology
- calculation of quality indicators
  - item nonresponse
  - mode effects



## **Question typology**

#### Context type

- 1. standard
- 2. HH roster
- 3. event history
- 4. social network

#### Question type

- 1. biographic fact
- 2. biographic frequency
- 3. date / time
- 4. attitude / opinion
- 5. estimation
- 6. intention

#### Direction

- 1. respondent
- 2. proxy

#### Time scope

- 1. past
- 2. future
- 3. present
- Centrality (subsection)

#### Sensitivity

#### Response type

- 1. categories (choice)
- 2. Yes / No
- 3. Frequencies (count)
- 4. magnitudes
- 5. date
- 6. open

#### Scale type

- 1. nominal
- 2. ordinal
- 3. numeric
- Topic
  - 1. Socio-demo
  - 2. Family
  - 3. Sex
  - 4. Society
  - 5. Economics
  - 6. Values
  - 7. personal
  - 8. support
- Social desirability



N i D/i

GGP

## **GP** Item nonresponse notation

response to a survey question is defined as p

 $\rho = \begin{bmatrix} 1 & if \ valid \ response \\ 0 & if \ nonresponse \end{bmatrix}$ 

individual level response rate is then defined as:

$$RR_i^R = \frac{\sum_{k=1}^{n_i^Q} \rho_{i,k}}{n_i^Q}$$

where  $\rho_i^k$  is the outcome for respondent i to question k question level response rate is thus defined as:



$$RR_k^Q = \frac{\sum_{i=1}^{n_k^R} \rho_{i,k}}{n_k^R}$$

## Mode effect measures

- estimated for new GGS 2015
- Effect size measures
  - calculated for each variable (1,500)
  - mode pairwise calculation
  - sample limit
    - 15 respondents per variable per mode
- Numeric variables
  - Cohen's D



Ni

### Nominal Variables

- group Canonical correlation of indicator variables
- Fisher's Z transformation of rho

## **Future work**

### documentation

- encoding of country specific implementation
- transformation into database
- 💠 para-data
  - multiple coders
  - semantics (Saris & Gallhofer)
- automated scripts
  - encode documentation into datasets
  - generate structural para-data



Ni



## Thank you!

# http://www.ggp-i.org





### **Nonresponse coding**

#### System missing

- . (STATA) sysmis (SPSS)
- Empty (blank) data cells

#### User missing

- Missing information that should be provided
  - Due to errors in survey process
- Do not know
  - 7, 97, 997, ...
  - Recall difficulty
- Refusal
  - 8, 98, 998, ...
  - Unwillingness to disclose
- Other NR and not applicable
  - 9, 99, 999, ...
  - Other reasons for missing information
  - Not applicable in spite of routing directions



GG,



### Why are there country differences?

### Different implementation

- Omitted questions
- Different questionnaire structure
- Different survey implementation
  - Sample
  - Data collection methods

### Different context

SEVENTH FRAMEWORK N 1 D 1

- Context specific measurement
- Different quality

## **Country specificities**

#### Why included?

- Minor deviations in response categories
- Country specific lists
- Measuring same concept but with different approach
- Country specific response values
  - Question following model questionnaire
  - Answers not at all or partly compatible
  - Separate values preceded by country code
    - France = 15 → 1501, 1502
- Country specific variables
  - Questions different from model questionnaire
  - Measuring the same concept
  - Separate variable with country code in suffix
    - a105 → a105\_1501





## What do the routing checks do?

- uniform data pattern
- 3 possible outcomes
  - valid response (as defined by the codebook)
  - nonresponse code (DK, refusal, other NR)
  - system missing or empty cell
- system missing
  - indicating skip and only skip
- NR code
  - error due to
    - respondent
    - interviewer
    - routing mistake
    - routing differentiation across datasets







SEVENTH FRAMEWORK

N i D/i

# NESSTAR

		DESCRIPTION TABULATION ANALYSIS	x@				
	GGP data files First wave	Dataset: Generations and Gender Survey France Wave 1 Variable a405a_a : Who makes decisions about hh: Routine hh purchases					
	Bulgaria     France     Generations and Gender						
	Survey France Wave 1	PreQuestion Text					
	Metadata     We have already talked about the various tasks that have to be done in a household. Now I wo						
	<ul> <li>Administrative variables</li> <li>Household</li> </ul>	ask you some questions about decisions. Who makes d household? Card 401-2	lecisions about the following issues in your				
	Children     Partnersbins	Literal Question					
	<ul> <li>Household Organisation</li> </ul>	Routine purchases for the household	Question wording				
	and Partnership Quality	Descriptive Text					
Click on a variable to	<ul> <li>Decision-making</li> </ul>	Q 405 a a core questionnaire	•				
display its description	Who makes decisions	variable's orig	gin Countries that have				
aispidy its acscription	about hh: Routine hh	Available: BUL FRA GEO GER ROM RUS	implemented the question				
	Who makes		implemented the question				
	decisions about hh:	Values Categories	N				
	for the hh	1 alwaysR	1351 22.2%				
	Who makes	2 usually R	703 11.6%				
	decisions about hh: Way of shild raising	3 R and partner equally	2657 43.7%				
	<ul> <li>Way of clinic faising</li> <li>Who makes</li> </ul>	4 usually partner	568 9.3%				
	decisions about hh:	5 always partner	797 13.1%				
	Social activities	6 always or usually other persons in the househ	iold 7   0.1%				
	organized	7 always or usually someone not living in the ho	usehold 2 U.U%				
***	🔄 🗉 Partnership quality	97 does not know	U A				
	Home	90 reiusai 99 pot opplicable/po regeneres					
****	🗉 Fertility	Svemice	3991 Distribution				
	Health and Well-Being	Gysiniss	3331				
	Income	Summary Statistics					
	🗄 Partner's Activity and	Valid cases 6085					
SEVENTH FRAMEWORK	Income Household Possessions	Missing cases 3994					
PROGRAMME	Income and Transfers	This variable is numeric					
- / - /	Value Orientations and Attitudes	The maine	rea (this field doos not annoar when th				
1 D/1	<ul> <li>User defined variables</li> </ul>	Universe ine univer	se tuns jield does not appear when th				
	🗄 Bookmarks	If R lives together with a partner <b>question v</b>	vas administered to all respondents)				
Lawrence	H Georgia						





- SEVENTH FRAMEWORK N I D/1
- 15 wave 1 countries (16 datasets)
  - Australia (Hilda), Austria (18-55), Belgium, Bulgaria, Estonia (FFS II), France, Georgia, Germany, Hungary (pre-GGP), Italy (ISMS), Lithuania, Netherlands (NKPS), Norway, Romania, Russian Federation
  - Turkish supplemental sample Germany
- soon to come: Japan
- 5 wave 2 countries
  - Hungary, Italy, Russian Federation, Australia



SEVENTH FRAMEWORK

N i D/i

# **G**<sup>G</sup><sup>P</sup>**Data flagging**

						a105	a106a	a106b	a107m	a107y	a108
						ОК	ОК	ОК	ОК	ОК	ОК
						ОК	-Err	ОК	ОК	ОК	ОК
	1	1	1	1		ОК	ОК	ОК	ОК	ОК	ОК
a105	a106a	a106b	a107m	a10/y	a108	ОК	ОК	ОК	<del>,</del> Err	_Err	ОК
Vee	()					ОК	OK	ОК	OK	ОК	ОК
Yes	xy				4	FTT	ОК	ОК	ОК	ОК	ОК
yes No			•			Err	Err	ОК	ОК	ОК	ОК
NO		ХУ <i>"</i> ху"	4			ОК					ОК
NO Vac	ND	ХУ				Err	ОК	ОК	ОК	ОК	ОК
765					1	T.					1
	· · · · · · · · · · · · · · · · · · ·										
No	^y "vv"	¦							_		
NR		)		i • i	1						
				t t							
$\searrow$											





# Intentions

#### Theory of Planned Behaviour - TPB (Ajzen)

- intentions to commit behaviour
- attitudes
- subjective norms
- perceived behavioural control

 leaving parental home, childbearing, partnership formation (dissolution), retiring





