

Data Harmonization in Practice:

The Comparative Study of Electoral Systems

*Katharina Blinzler,
GESIS Data Archive*

*CESSDA Training Day
27-28 November 2019, Cologne*

 [cessda.eu](https://www.cessda.eu)  @CESSDA_Data



1

Project Overview



The Origins of the CSES

- The project was founded in 1994.
- Objectives:
 - To promote international collaboration among national election studies.
 - Micro-macro design, to study variations in electoral systems (and other political institutions).
 - Foremost a comparative project, but serves other purposes also.

The Rationale for the CSES...

...Make electoral research global.

- **There is wide variation across countries on...**
 - Electoral rules
 - Systems of governance (Presidential/parliamentary/mixed)
 - Federalism vs. unitary governments
 - Lines of political conflict ...and more

- *How do these variations impact individual attitudes and behaviours, especially voting and turnout?*

The CSES Project in Brief

- A CSES Module is a 10-15 minute questionnaire with a specific substantive theme that asks the same questions in different countries.
- The CSES Module is included in high quality national post-election surveys around the world.
- The data from all countries are merged into a single dataset along with administrative, demographic, district, and macro variables.
- A new theme & questionnaire every 5 years

Common questions/variables available in all CSES releases

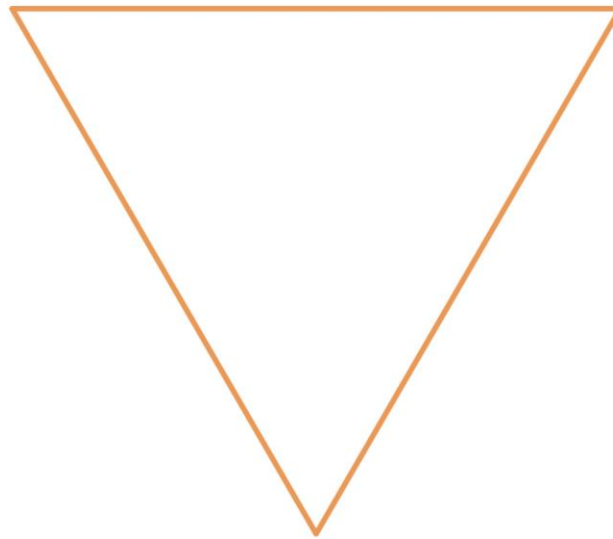
- Voter turnout (current and past election)
- Vote choice
- Satisfaction with democracy
- Political efficacy
- Party attachment
- Evaluations of parties (like-dislike scale)
- Ideological assessments of parties (left-right scale)
- Demographics (e.g.: age, gender, education etc.)

Currently - Module 5: 2016-2021

Theme: Democracy Divided? People, Politicians and the Politics of Populism

Attitudes towards
elites

Attitudes towards
majority rule and
representative democracy



Attitudes towards out-
groups

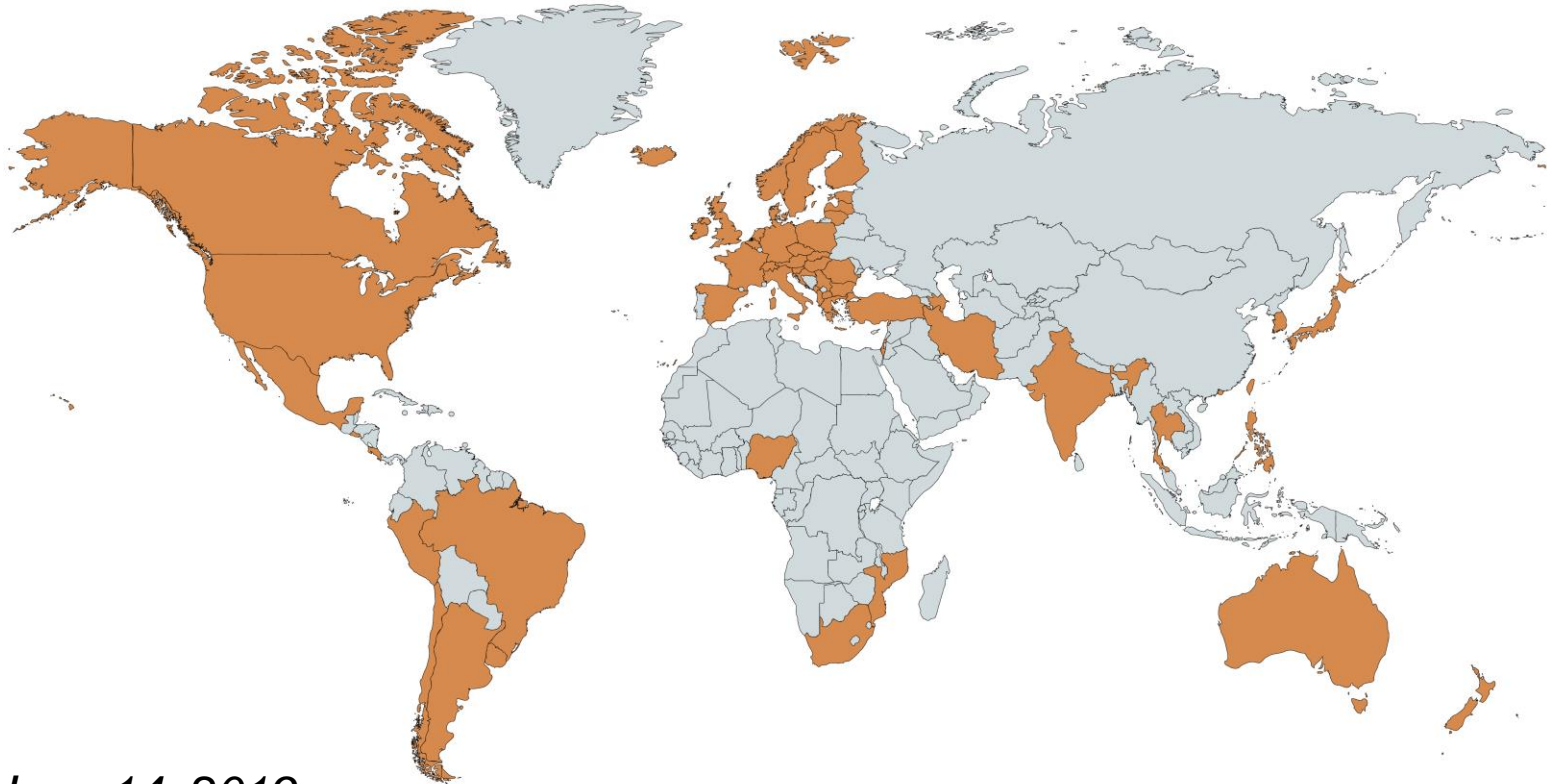
*CSES Planning
Committee
Module 5 Final
Report, p. 6*

Module 5: 2016-2021 – specific variables

- Political interest
- Agreement with attitudinal statements (for example):
 - Most politicians do not care about the people
 - The people, not politicians, should make most important decisions
 - Minorities should adapt to the customs and traditions of [country]
 - The will of the majority should always prevail, even over the rights of minorities
 - Immigrants are generally good for [country]’s economy
- How widespread is corruption?

Coverage:

Module 5 Expressions of Interest



As of June 14, 2019.

For comprehensive coverage overview of all CSES Modules see:

<https://cses.org/data-download/download-data-documentation/election-studies/>

Organization and Governance

- Collaborators
- Secretariat
- Planning Committee
- Users

- Planning Committee (PC), elected at Plenary, oversees study.
- Chair of PC oversees Secretariat.
- Secretariat consists of 6 staff members (2.5 FT).

2

Data Harmonization Approaches



Data Harmonization in “3MC” Surveys

“...generic term for **procedures aimed at achieving or at least improving the comparability** of different surveys” (Granda et al. 2010, p.315)

“...refers to all efforts that **standardize inputs and outputs** in ... ,3MC‘ surveys.” (Granda & Blasczyk, 2016)

“We consider measures to be **comparable if similarities or differences in measurements ... across countries reflect similarities in the measured trait** and cannot be attributed to method” (Wolf et al. 2016, p. 503)

→ Overreaching Goal: (Functional) **Comparability!**
(Spatial / Temporal)

How to Think About Harmonization I

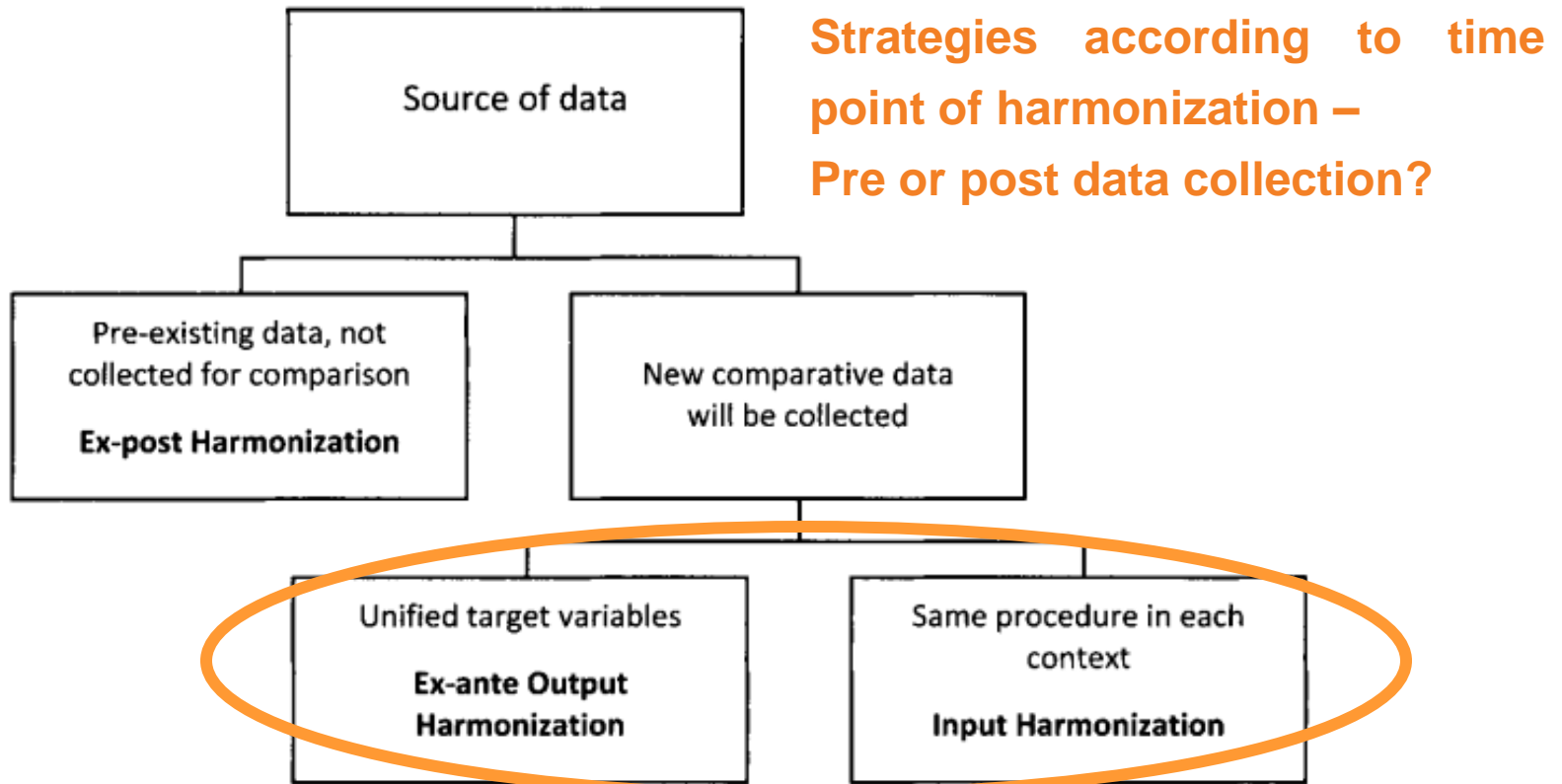
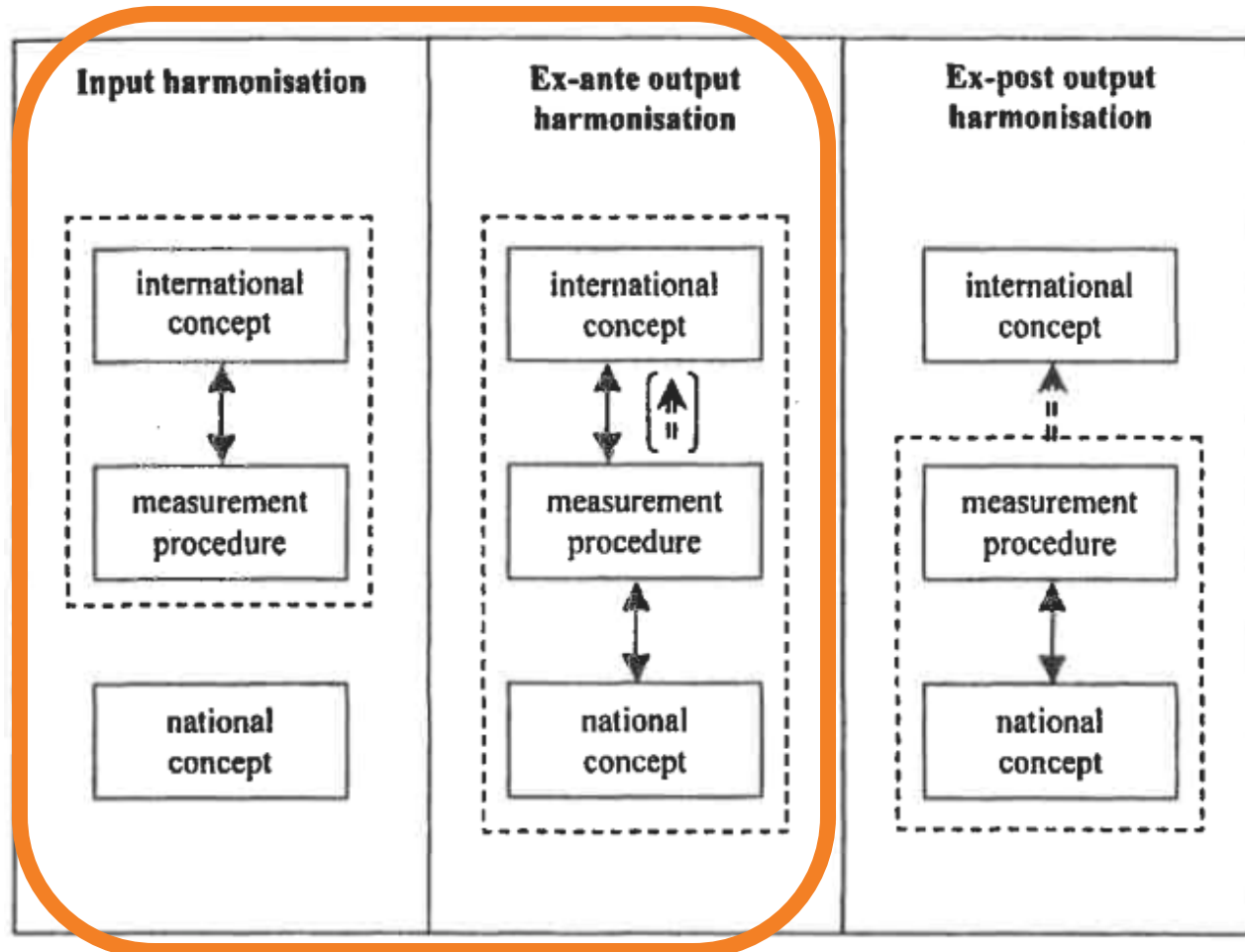


Figure 17.1. Three Approaches to Harmonization

Granda et al. 2010, p. 318.

How to Think About Harmonization II



Strategies according to study design – national or international?

Ehling 2003, p. 23.

↔ ... coordination of the survey design ⇒ ... subsequent conversion

3

Input Harmonization in the CSES



3 Steps to Input Harmonization (Wolf et al. 2016)

For every single variable:

1. Establish a common understanding of the **theoretical concept** to be measured (What to measure?)
2. Specify cross-culturally **comparable empirical indicators** of that concept (How to measure?)
3. Choose **questionnaire items** that capture the indicators of interest (How to measure?)

CSES Design Process: Module 5 I

Multiple years of effort go into a CSES Module before any data are collected.

Idea development

- Transparent process - Open call to user community via e-mail list and social media soliciting ideas (start: June 2013)
- Special Subcommittee of Planning Committee (PC) setup to evaluate proposals and present them to the wider PC.
- Proposals also presented and discussed in Plenary meeting (October 2014)
- Planning Committee selects themes at PC meeting (March 2015)

CSES Design Process: Module 5 II

Module development

- Having decided on a theme, PC breaks into several subcommittees (March 2015):
 - Theme and Module Specific Questions
 - Methods
 - Demographics Data
 - Macro and District Data
 - Political Knowledge
- Subcommittees present two reports to PC – one preliminary for initial feedback and one final report with recommendations incorporating feedback, which is eventually presented to the Plenary (after pre-testing).

CSES Design Process: Module 5 III

Pretesting:

- PC signs off on draft module questionnaire for pre-testing (October 2015).
- Nov 2015-Jan 2016: Draft questionnaire pretested in several polities: Greece, Ireland, South Korea, Sweden, Switzerland, and Taiwan.

Finalization:

- Complete Subcommittee reports and Pretest results presented to Plenary meeting (Summer 2016)
- Plenary adopts (or make suggested changes) to Module.
- Autumn 2016: Module finalized and available for field.

Why Collective International Questionnaire Design Matters

Not all concepts or indicators are equally relevant in or have the same meaning across all polities & cultures!

- Left-right ideological party placements
- Turnout as general indicator for political participation
- M5: Asking about National Identity in Montenegro collides with main cleavage (“pro Montenegrin” vs. “pro Serbian”)
- M5: “Ethnic Minorities” in Italy refers to linguistic minorities living close to Austrian, French, and Slovenian Borders
- M5: “Dominant Religion” as aspect of national identity: No such dominant religion in South Korea or Hong Kong

Beyond Questionnaire Design – Crucial Role of CSES Collaborators

Decentralized Structure of the CSES: Collaborators participating in the CSES project must raise their own local funding

→ in the end it is collaborators who collect high quality datasets suitable for comparative analysis

→ ***“...as much attention needs to be given to developing the collaborative network and making participation rewarding ... as is given to the development of the scientific instruments themselves.”***
(Howell 2010, 527)

→ No formal application process: recruited from among their country's foremost social scientists; often recommended by existing network, usually interested in academic research, should be capable of running a high quality national survey

Beyond Questionnaire Design – CSES Standards for Including Studies

- Only post-election studies accepted
 - Random sampling procedures at all stages with adequate coverage (voting eligible population)
 - No fewer than 1,000 interviews (required since M5)
 - Interviews conducted within 6 months after the election
 - Aiming for face-to-face interviews
 - Detailed documentation of sampling procedures
- **Detailed documentation of sampling procedures provided in Design Reports made public to users.**

3

(Ex-ante) Output
Harmonization in CSES



3 Steps to Ex-Ante Output Harmonization

(Wolf et al. 2016)

1. Specification of a **comparative target variable** to represent the identified concepts
2. Development of **country-specific questionnaire items**, response categories, and their mapping to the target variable
3. After data collection: **Execution of harmonization recodes**

(Ex-ante) Output Harmonization in CSES - Demographics

- There is great international variation in the ways that collaborators will go about soliciting this information from respondents.
- The CSES objective with demographic is not standardization of the way collaborators ask the questions. Instead **we go for standardization to a common, cross-national scheme**, which is achieved by the CSES Secretariat.
- It is up to the collaborators to what extent they ask specific demographic variables and its their choice how!

Examples of Harmonized Variables in CSES Survey Data

- **Retaining original idiosyncrasies:** Ethnicity, Race, Region of Residence, Primary Electoral District...
- Harmonized variables including (limited) number of **election study specific codes:** Employment Status
- Applications of **international standards:** Education (ISCED 2011) and Occupation (ISCO08)
- **Hierarchical Coding:** Occupation (ISCO08) and Religious Denomination

Strategy for Complex Variables: Hierarchical Coding

- Useful for complex variables with a great number of codes
- Idea: Most studies will provide data on the first and most general level – more diverse studies may be collapsed to this level by users to allow for an extensive comparative analysis
- More specific codes not relevant / employed in all countries are retained at the lower levels - allowing an in-depth analysis
- **Example in CSES: Religious Denomination**

Hierarchical Coding – Example

Religious Denomination

E2013

>>> RELIGIOUS DENOMINATION

D10. Religious denomination.

.....

CHRISTIAN

1000. CHRISTIAN (NO DENOMINATION GIVEN)

CATHOLIC

1101. ROMAN CATHOLIC

1102. EASTERN (GREEK RITE) CATHOLIC CHURCHES

1199. CATHOLIC, OTHER [SEE ELECTION STUDY NOTES]

PROTESTANT

1200. PROTESTANT, NO DENOMINATION GIVEN

1201. ADVENTIST

1203. BAPTIST

1204. CONGREGATIONAL

1205. EUROPEAN FREE CHURCH (ANABAPTISTS,
MENNONITES)

1206. HOLINESS

1207. FUNDAMENTALIST

1208. LUTHERAN

1209. METHODIST

1210. PENTECOSTAL

1211. PRESBYTERIAN

Applying International Standards ... Isn't it Straightforward?

- Applying international standards is **not** straightforward –
All CSES Secretariat staff receive a special training on how to do this!

Example: Harmonizing Education: International **S**tandard **C**lassification of **E**ducation (ISCED) by the UNESCO Institute for Statistics (UIS)

Its structure is derived from patterns of education found to exist in many countries but does not reflect the conditions in any one country.

Measuring Education in M4-M5 (ISCED 2011)

“This item (education) should report the respondent's highest level of education that has been completed.”

- 01. Level 0 - Early childhood education
- 02. Level 1 - Primary education
- 03. Level 2 - Lower secondary education
- 04. Level 3 - Upper secondary education
- 05. Level 4 - Post-secondary non-tertiary education
- 06. Level 5 - Short-cycle tertiary education
- 07. Level 6 - Bachelor or equivalent
- 08. Level 7 - Master or equivalent
- 09. Level 8 - Doctoral or equivalent

Why Application Isn't Always Straightforward - Example I

- In CSES Module 4, the 2013 Australian election study presented the biggest challenge.
- The collaborators used 3 items to collect educational background information from respondents via mail-back questionnaire.
 - *G1. How old were you when you left secondary school?*
 - AGE given
 - 1. No formal schooling
 - 2. Went to primary school only
 - 3. Still at school
 - *G2. In all, how many years of tertiary study have you completed since you left secondary school? If your tertiary study was part-time, give the number of years of equivalent full-time study.*

Why Application Isn't Always Straightforward - Example II

- *G3. Have you obtained a trade qualification, a degree or a diploma, or any other qualification since leaving school? What is your highest qualification?*
 - 1. No qualification since leaving school
 - 2. Postgraduate Degree or Postgraduate Diploma
 - 3. Bachelor Degree (including Honours)
 - 4. Undergraduate Diploma
 - 5. Associate Diploma
 - 6. Trade qualification
 - 7. Non-trade qualification
- The collaborators did not provide us (at the time) with a recoding schema.

Why Application Isn't Always Straightforward - Example III

- The solution was to combine the information about the number of years of completed education with that about particular qualifications obtained by respondents.
- For this, we used the ISCED 2011 specifications about the duration of each education level to approximate placement of respondents into the different ISCED categories (Stata-code on next slides).

Why Application Isn't Always Straightforward - Example IV

- First, we assumed that the theoretical starting age of all children, based on Australian ISCED data, is 4.
 - by subtracting '4 years' from the age of the Australian respondent at which they left school, we obtained the years of secondary education they had.

```
gen years = G1age - 4 if G1age > 0
```

- Second, we coded respondents who mentioned to have no education as '96. NONE (NO EDUCATION)' and those who said they went to 'primary school' as '02. ISCED LEVEL 1 – PRIMARY'. We also separated the 'don't know' and 'missing' answers into the appropriate categories.

```
gen D2003 = 96 if G1 == 2
recode D2003 . = 2 if G1 == 3
recode D2003 . = 98 if G1 == 4
recode D2003 . = 99 if G1 == -1
```

Why Application Isn't Always Straightforward - Example V

- Third, we sorted people into the different ISCED levels by virtue of how many years they spent in school.

```
gen D2003s = 96 if years < 4
recode D2003s .= 1 if years <= 7
recode D2003s .= 2 if years <= 9
recode D2003s .= 3 if years <= 13
recode D2003s .= 4 if years <= 15
recode D2003s .= 5 if years <= 15
recode D2003s .= 6 if years <= 22
recode D2003s .= 7 if years > 22
```

- Fourth, we took into account the information about years of tertiary studies.

```
replace D2003s = 5 if G2 == 0
replace D2003s = 6 if G2 > 0 & G2 <= 2
replace D2003s = 7 if G2 > 2
```

Why Application Isn't Always Straightforward – Example VI

- Finally, we took into account the qualifications obtained by respondents.

```
replace D2003s = 7 if G3==6 | G3==7 | G3==3  
replace D2003s = 8 if G3==4 | G3==5  
replace D2003s = 9 if G3==2
```

- The final distribution of education among respondents is:

D2003	Freq.	Percent	Cum.
1	1	0.03	0.03
2	104	2.63	2.66
3	237	6.00	8.65
4	22	0.56	9.21
5	563	14.24	23.45
6	178	4.50	27.95
7	1,726	43.66	71.62
8	474	11.99	83.61
9	505	12.78	96.38
96	35	0.89	97.27
98	7	0.18	97.44
99	101	2.56	100.00
Total	3,953	100.00	

CSES Philosophy of Documentation I

→ **Thorough documentation is key!**

- **The imperfections of a study should not be hidden, but highlighted:**
 - Improves the quality of resulting analyses
 - Allows proper comparisons using the data
- Codebook notes anything we know of that has a possible impact on quality, comparability, or analytical outcomes → large codebooks

D27. This variable reports the respondent's region of residence. Regions are usually (but not always) based upon the social, cultural, or historical differences (though some correspond to administrative regions) that manifest themselves in political cleavages

.....

01.-80. REGION CODES [SEE ELECTION STUDY NOTES]

99. MISSING

| VARIABLES NOTES: D2028

| Data are unavailable for HONG KONG (2012), IRELAND (2011),
| ISRAEL (2013), KENYA (2013), MEXICO (2012).

| ELECTION STUDY NOTES - AUSTRALIA (2013): D2028

CSES Code	Election Study Code/Category
01.	New South Wales
02.	Victoria
03.	Queensland
04.	South Australia
05.	Western Australia
06.	Tasmania
07.	Northern Territory
08.	Australian Capital Territory

| ELECTION STUDY NOTES - AUSTRIA (2013): D2028

| The Austrian Bundesland Vienna is not included in
| D2028, although a number of persons report a ballot cast in one
| of Vienna's electoral districts, according to D2032.

CSES Code	Election Study Code/Category
01.	Burgenland
02.	Carinthia
03.	Lower Austria
04.	Upper Austria
05.	Salzburg
06.	Styria
07.	Tyrol
08.	Vorarlberg

| ELECTION STUDY NOTES - BULGARIA (2014): D2028

CSES Philosophy of Documentation II

- **More is better – let the practitioner(s) decide.**
- Original collaborator documents are made available for public download:
 - Original language questionnaire
 - English language questionnaire translations
 - Macro report
 - Sample design and data collection (methodology) report

4

Also Part of Harmonization – Data Quality Controls



Data Quality Controls I

- Quality doesn't end after with data collection guidelines.
- Collaborators clean data to their national standard and then pass it on to the Secretariat, who reviews, cleans, and harmonizes the data to a cross-national standard.
- Each election study is gone through meticulously by a data processing specialist (mean 30hrs per election study).
- Secretariat usually asks 15-35 questions of each national collaborator prior to finalizing a single election study.
- Processed election study returned to collaborator for a last review as a final check on data quality and documentation.
- Additional quality controls at the cross-national level.

Data Quality Controls II

- **Cross-National Checks:** conducted by CSES Secretariat. Takes about 20-30hrs.
- **Duplicates Check:**
Classification of respondents with corresponding answers to all questions.
- **Inconsistency Checks:**
 - Identifies sets of variables which are inconsistent, or could be perceived as inconsistent (e.g., strange skip pattern, incompatible answers to related questions).
 - The CSES convention is not to change data. Instead, inconsistencies are noted in the CSES codebook, allowing users to make the final determination.

Data Quality Controls III

- Irregular code checks:

Sometimes these irregular codes are legitimate in the sense that they may be accounted for by a polity deviation on a particular variable.

- Variable and value label checks.

- Theoretical Checks:

- Explore theoretically expected relationships between variables (e.g., correlation between Political Efficacy and Satisfaction with Democracy; turnout and age; religiosity and religious service attendance).
- Examine distributions, correlations, and undertake regression analysis.
- Relatively unique to CSES to our knowledge.

5

Key Takeaway Points



Key Takeaway Points

- Harmonization is about making data **comparable**
- Harmonization strategies vary according to time point and study design and are not always straightforward to apply
- Harmonization matters throughout the complete survey lifecycle and comprises more than designing / harmonizing questionnaire items – **the earlier in the process you engage with it, the better**
- In “3MC” Surveys, engage international experts already in study design phase to ensure cross-cultural applicability
- **Thorough documentation of both the harmonization process & data imperfections** empowers other researchers to make best use of your data

Data Availability

- Public access, **free of charge**
- Download from CSES website: www.cses.org or [via the GESIS data catalogue](#)
- Archived at GESIS & ICPSR
- Full release every five years but advance releases of data every year

To learn more about CSES, or to
download data:

www.cses.org

...or email your questions to:

cses@umich.edu

References

- Ehling, Manfred. 2003. “Harmonising Data in Official Statistics. Development, Procedures, and Data Quality.” In: Jürgen H. P. Hoffmeyer-Zlotnik / Christof Wolf (Ed.). *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*. New York: Springer Science + Business Media: 17 – 32.
- Granda, Peter/Christof Wolf/Reto Hadorn. 2010. “Harmonizing Survey Data”. In: Janet A. Harkness/Michael Braun/Brad Edwards/Timothy P. Johnson/Lars Lyberg/Peter Ph. Mohler/Beth- Ellen Pennell/Tom W. Smith (Ed.). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Hoboken: Wiley: 315-332.
- Granda, Peter/Emlly Blasczyk. 2016. Data Harmonization. In: *Cross-Cultural Survey Guidelines*: 617 – 635. (Online Ressource: <https://ccsg.isr.umich.edu/index.php/chapters/data-harmonization-chapter>; Last accessed: November 26, 2019)
- Hobolt, Sara/Eva Anduiza/Ali Carkoglu/Georg Lutz/Nicolas Sauger. 2016. *CSES Planning Committee Module 5 Final Report. CSES Module 5: Democracy Divided? People, Politicians, and the Politics of Populism*. (Online Ressource: https://cses.org/wp-content/uploads/2019/03/CSES5_ContentSubcommittee_FinalReport.pdf; Last accessed: November 22, 2019)

References

- Howell, David A. 2010. “Enhancing Quality and Comparability in the Comparative Study of Electoral Systems (CSES).” In: Janet A. Harkness/Michael Braun/Brad Edwards/Timothy P. Johnson/Lars Lyberg/Peter Ph. Mohler/Beth-Ellen Pennell/Tom W. Smith (Ed.). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Hoboken: Wiley: 525-534.
- Wolf, Christof/Silke L. Schneider/Dorothee Behr/Dominique Joye. 2016. “Harmonizing Survey Questions Between Cultures and Over Time.” In: Christof Wolf/Dominique Joye/Tom W. Smith/Yang-chih Fu (Ed.). *The SAGE Handbook of Survey Methodology*. SAGE: 502-524.

This presentation is property of the CSES Secretariat. It should be cited as follows:

CSES (2019). Harmonization in Practice: The Comparative Study of Electoral Systems. Presentation at CESSDA Training Days, 28 November 2019.