Do national coding indexes code the same occupations similarly?

ESRA 2023 Conference 20 July 2023

Kea Tijdens, WageIndicator Foundation and University of Amsterdam, Netherlands

The issue – two examples from coding indexes

- **How is the plumber coded across countries? – 2 out of 4**
  - **no_NO**   **Blikkenslager**   **7213**
  - **sk_SK**   **Klampiar**   **7213**
  - **es_ES**   **Plomero**   **7312**
  - **en_SA**   **Plumber**   **7126**

- **How is the historian coded across countries? – 2 out of 4**
  - **ro_RO**   **Istoric**   **2443**
  - **az_AZ**   **Tarixçi**   **2633**
  - **bg_BG**   **Историк**   **2633**
  - **es_ES**   **Historiador**   **2822**

Coding indexes

- **International occupational classification: ISCO**
  - Latest version ISCO-08 from 2008, update scheduled for 2028

- **European Union**
  - National Statistical Offices (NSO) used to have their own classification
  - But in 2008 European Union has adopted ISCO as the standard classification
  - Some countries adopted ISCO, other maintain their own with cross-over tables
  - Eurostat urges NSOs to deliver ISCO coded labour force data
  - Candidate countries must develop an ISCO coding index

- **Coding indexes**
  - ISCO has an English coding index with approx. 2,000 titles
  - NSOs developed national coding indexes, downloadable from their websites

Research objective

- **Occupational coding in multi-country surveys is mostly a black box: do national survey agencies classify the same occupational titles into the same category across countries?**

- **Survey agencies typically use coding index of NSO in their country**

- **We aim to validate the coding indexes across countries: are the same occupations coded similarly?**

- **Using coding indexes from NSO, provided 5-digit ISCO-08 coding**

Translating occupations

The translation challenge

- **Occupations are country-specific**
- **Translation is not as straightforward as in other areas**
- **Translations (English > natlang) cause large numbers of complaints**
- **Checks needed from local labour market experts**

Are occupations coded similarly across countries?

- **The only way to validate is translation in English (natlang > English)**

Translating occupations – countries vary ..

with respect to their labour force composion

- **If higher educated, the division of labour in high-skilled occupations is larger >> more occupational titles, similarly for lower educated**

with respect to their certification/licensing policies

- **Certification is not harmonized across Europe or the world**
- **Except when insurance companies require so, f.e. aircraft pilots**

with respect to their educational systems

- **Labour markets in countries with general education systems are more likely to adapt career ladders in labour markets**

Classifying occupations – the challenge

- **National stocks of job titles are …**
  - large   >> 10,000's of job titles in any national labour force
  - unstructured   >> vague boundaries between job titles, except licensed occ's
  - unlimited   >> no fixed list
  - distribution   >> extremely skewed: many nurses, few C+ programmers
  - dynamic   >> changes within every decade, many entries and exits over time

- **The challenge**
  - To classify this messy list into a classification scheme
  - To do so consistently across countries and over years
  - No international board exists, overseeing the classification

Two studies to test across-country classification

Study 1 – 2018

- **merged 20 coding indexes, translated the occupations using Google translate**

- Tijdens K.G. & Kaandorp C.S. (2018), Validating occupational coding indexes for use in multi-country surveys. Survey Insights: Methods from the Field. Retrieved from https://surveyinsights.org/?p=10422 DOI:10.13094/SMIF-2018-00007

Study 2 – 2023

- **merged 25 coding indexes (most overlapping with study 1)**

- **translated the occupations using DEEPL translation package and Google translate**

- **no publication yet**

Classification strategy

Strategy Study 1

- **We identified similarity of translated titles (= English) based on text similarity, regardless the initial ISCO code**
- **One translator: Google T**
- **Typical similarity problems related to words as Chief, Supervisor, Manager**

Strategy Study 2

- **We identified similarity of translated titles (= English), searching for the best matching title in the WageIndicator Occupation database with the same ISCO-08 4-digit code as the initial code in the coding index**
- **Two translators: GoogleT and DEEPL**
- **WageIndicator Occupation database for 55 languages with more than 4,000 titles**

occupations_ISCO08_5dgt_55languages_4000titles_with_mapping_surveycodings_20230425.xlsx

Study 1 – 2018

Findings

- **Aim: to validate classification of 5-digit titles into 4-digit ISCO-08 by comparing the English translations of 20 national ISCO-08 indexes**

- **merged database of 70,489 records of which
4.2% could not be translated & 10.3% had non-existent codes
of remaining 60,559 records only 31% has at least one duplicate title**

- **when aggregating the duplicate occupational titles (5,350 records),
64% have the same 4-digit code across the indexes,
70% has so when limiting to 3-digit codes, 74% has for 2-digit codes,
and 80% has same codes when comparing the 1-digit code**

Study 2 – 2023

Coding indexes from 25 countries

- **25 countries with 83,279 records (=5-digit occupational titles)**
- **Ranging from Austria with 13,314 titles to Finland with 206**

Coding indexes from 25 countries

- **1 index excluded because only ISCO08 3dgt: Italy**
- **2 indexes in English: Singapore, South Africa**
- **4 languages not in DEEPL: Albania, Azerbaijan, Iceland, Montenegro**
- **1 index: Google Translate not used: Austria, too many records**
- **1 country provided already English translations: Finland**
- **For 16 countries the translations could be compared**

Translators compared

Examples

locale	index	Google T	DEEPL
bg_BG	Главен възпитател	Chief educator	Senior Educator
et_EE	Käsitsi nõudepesija	Hand-washing dishes	Manual dishwasher

Translation evaluation

- DEEPL provided better quality translations
- Though for fewer countries (4 less)
- 56,442 records could be translated with both GoogleT and DEEPL
- 15,396 records had equal translations (27%)
- 40,886 records: DEEPL was preferred over GoogleT, using manual check (99%)

Classification strategy, example code 2432

Country	Nat ISCO	Natlang	Best translation	Best classification	Final code
Netherl	2432	Communicatieadviseur	Communication advisor	Communications consultant	2432000200018
Netherl	2432	<mark>Communicatiemedewerker</mark>	Communication officer	Communications consultant	2432000200018
Sweden	2432	Kommunikationsstrateg	Communication strategist	Communications consultant	2432000200018
Sweden	2432	Kommunikatör, informatör	Communicator, information officer	Communications consultant	2432000200018
Austria	2432	Berater, Kommunikation	Consultant, Communication	Communications consultant	2432000200018
Austria	2432	Planer in Betriebskommunikation	Planner in corporate communications	Communications consultant	2432000200018
Sweden	2432	Katalogredaktör	Catalogue editor	Copy editor	<mark>2642</mark>000300018
S Africa	2432	Public relations copy writer	Public relations copy writer	Copywriter public relations	2432000400018
Turkey	2432	Halkla ilişkiler metni yazarı	Public relations copywriter	Copywriter public relations	2432000400018

Classification results

Non-existing codes

- **ISCO08 has only 436 4-dgt codes, but merged db 1,300 different codes, excl Italy with only 3-dgt codes**

- **>> 864 non-existing codes (66% of codes) for 7,539 of records (9.7%)**

- **>> Countries assign new codes at will, but luckily this applies only to small minority of occupational titles**

Similar codes

- **Coded: 47%   (37,336 of 77,557 records, excl. Italy)**

- **Coded similarly as in original index: 71.6% (26,757 of 37,336 titles)**

- **Coding work not yet finished, to be continued**

SURVEY CODINGS and FURTHER READING

- **SURVEY CODINGS**
  - WageIndicator Databases free downloadable on https://www.surveycodings.org/articles/home
- **Further reading**

  - Tijdens KG (2014) Drop-out rates during completion of an occupation search tree in web-surveys, *Journal of Official Statistics*, doi.org/10.2478/jos-2014-0002
  - Tijdens KG, De Ruijter E, De Ruijter J (2014) Comparing work tasks of 160 occupations across eight European countries, *Employee Relations*, doi.org/10.1108/ER-05-2013-0046
  - Tijdens KG (2015) Self-identification of occupation in web surveys: requirements for search trees and look-up tables, *Survey Methods: Insights from the Field*, doi.org/ 10.13094/SMIF-2015-00008
  - Belloni M, Brugiavini A, Meschi E, Tijdens K (2016) Measurement error in occupational coding: an analysis on SHARE data, *Journal of Official Statistics*, doi.org/10.1515/JOS-2016-0049
  - Tijdens KG, Kaandorp CS (2018) Validating occupational coding indexes for use in multi-country surveys, *Survey Insights: Methods from the Field*, doi.org/ 10.13094/SMIF-2018-00007
  - Tijdens KG, Beblavý M, Thum-Thysen A (2018) Skill mismatch comparing educational requirements vs attainments by occupation, *International Journal of Manpower*, doi/full/10.1108/IJM-10-2018-0328
  - Belloni M., Tijdens KG (2017) Occupation > industry predictions for measuring industry in surveys, SERISS Deliverable 8.11, 10.13140/RG.2.2.31328.02566
  - Tijdens, K.G. (2020). Managing surveys: ten lessons learned from web-surveys
  - Tijdens KG (2022) The importance of occupation coding quality: lessons for EU-SILC from SHARE and other international surveys, Ch 29 (pp 413-426) in European Commission, Eurostat, Improving the measurement of poverty and social exclusion in Europe : reducing non-sampling errors : 2022 edition, Lyberg, L.(editor), Lynn, P.(editor), Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2785/18803
  - Tijdens KG, Brugiavini A, Buia RE (2022) A 20th century version of the occupation multilingual ontology, SSHOC Deliverable 3.10, 10.5281/zenodo.6490676

Thank you for listening

- **Questions? k.g.tijdens@uva.nl**