FORS
explore.understand.share.

Does adding a survey language reduce coverage bias?

Oliver Lipps
Michael Ochsner

Swiss Centre of Expertise in the Social Sciences (FORS)

ITSEW Bergamo, June 2019

Content

- Framework: Aspects when adding language(s) to reduce coverage bias of migrants in surveys
- Contextualization: Potential Study
- Results of contextualization
- Conclusion

FORS
explore.understand.share.

Which aspects play a role when adding language(s)?

FORS
explore.understand.share.

Aspects when adding language(s)

Coverage success of adding sample members depends (at least) on:

- **Additional potential** of new language(s)
- Language **mastery needed** to complete survey
- Survey **Topic /** main **person subgroups**

… and complex interplay

FORS
explore.understand.share.

Framework: languages

- **National language**(s)
  - Language(s) of most population surveys
- **Lingua franca**
  - used in international or inter-lingual exchanges
- **Migration language**
  - spoken by significant groups of migrants (proficient)

FORS
explore.understand.share.

Framework: language mastery

Survey-specific necessary **language mastery** depends on:

- Survey **mode** (presence interviewer, etc.)
- **Complexity** of survey (admin vs. scientific, etc.)
- Target survey **quality** (measurement quality, etc.)

FORS
explore.understand.share.

Framework: survey topic

Language added can reduce OR increase coverage bias

- **Reduce bias if underrepresented are added**
  (e.g., difficult survey / migration language)

- **Increase bias if overrepresented are added**
  (e.g., difficult survey / lingua franca)

FORS
explore.understand.share.

Contextualization: Potential Study

Pooled Swiss Structural Survey (census) 2010-2014; 1.5 Million observations, adults 18+

Context: languages added (example: relig. majority)

- **Portuguese**
  - Catholics
- **Serbo-Croatian**
  - Orthodox Christians
- **Albanian**
  - Muslims

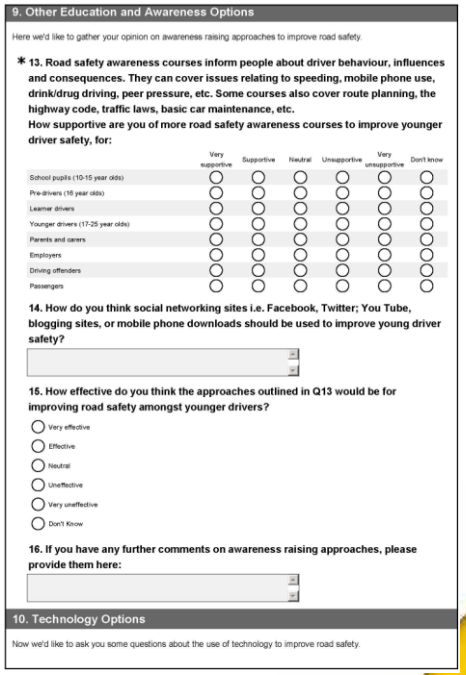
- **English**
  - “lingua franca”
  - spoken by many migrants (basic level)
  - convenient (translation, interviewers, etc.)

Context: language mastery

- **good language mastery:** survey language is main language

- **basic language mastery:** survey language is main language *or*

  language is spoken at home / at work / at education



9. Other Education and Awareness Options

Here we'd like to gather your opinion on awareness raising approaches to improve road safety.

* **13. Road safety awareness courses inform people about driver behaviour, influences and consequences. They can cover issues relating to speeding, mobile phone use, drink/drug driving, peer pressure, etc. Some courses also cover route planning, the highway code, traffic laws, basic car maintenance, etc.**
How supportive are you of more road safety awareness courses to improve younger driver safety, for:

	Very supportive	Supportive	Neutral	Unsupportive	Very unsupportive	Don't know
School pupils (10-15 year olds)	○	○	○	○	○	○
Pre-drivers (16 year olds)	○	○	○	○	○	○
Learner drivers	○	○	○	○	○	○
Younger drivers (17-25 year olds)	○	○	○	○	○	○
Parents and carers	○	○	○	○	○	○
Employers	○	○	○	○	○	○
Driving offenders	○	○	○	○	○	○
Passengers	○	○	○	○	○	○

14. How do you think social networking sites i.e. Facebook, Twitter; You Tube, blogging sites, or mobile phone downloads should be used to improve young driver safety?

15. How effective do you think the approaches outlined in Q13 would be for improving road safety amongst younger drivers?

- ○ Very effective
- ○ Effective
- ○ Neutral
- ○ Uneffective
- ○ Very uneffective
- ○ Don't Know

16. If you have any further comments on awareness raising approaches, please provide them here:

10. Technology Options

Now we'd like to ask you some questions about the use of technology to improve road safety.

FORS
explore.understand.share.

Results of contextualization

FORS
explore.understand.share.

Measuring coverage bias:

1. Std. deviation of the *percent coverages* of person groups (no account of frequencies)
   -> coverage

2. Cramér's V between *numbers of people* mastering or not the language across categories
   -> representativeness

Example: coverage by mode of travel

proficient ⟷ basic

Transp. mode	Nat l.	+E	+P	+A	+S	Nat l.	+E	+P	+A	+S
Non-motorized	.893	.906	.920	.902	.907	.927	.951	.948	.931	.934
Car/motorbike	.920	.928	.944	.931	.933	.952	.964	.969	.955	.956
Public Transp.	.896	.915	.915	.902	.905	.923	.958	.939	.926	.928
Std Dev.	.015	.011	.015	.017	.015	.015	.006	.015	.016	.015
Cramér's V	.045	.032	.053	.053	.050	.057	.024	.068	.059	.058

- Bias small
- Robust across measures / mastery
- adding English would reduce bias
- adding Portuguese or Albanian would increase bias

Example: coverage by religious affiliation

proficient ⟷ basic

Relig affiliate.	Nat l.	+E	+P	+A	+S	Nat l.	+E	+P	+A	+S
Catholic	.925	.930	.962	.928	.932	.944	.952	.973	.945	.947
Protestant	.990	.994	.991	.990	.990	.993	.997	.994	.993	.993
Other Christ.	.775	.800	.788	.775	.880	.843	.883	.853	.844	.902
Jew	.811	.885	.820	.814	.825	.849	.948	.857	.851	.857
Muslim	.564	.574	.567	.764	.619	.714	.729	.716	.833	.749
Other Relig.	.523	.598	.530	.532	.528	.632	.753	.637	.638	.636
Atheist	.899	.927	.916	.905	.904	.916	.959	.931	.919	.919
Std Dev.	.180	.168	.185	.151	.172	.126	.104	.131	.113	.123
Cramér's V	.353	.356	.401	.277	.324	.276	.274	.317	.228	.255

- Bias high
- Robust across measures / mastery
- adding Albanian would decrease / Portuguese increase bias

FORS
explore.understand.share.

Summary and conclusion

FORS
explore.understand.share.

Summary: our framework

(at least) three dimensions to evaluate potential of additional language to reduce coverage bias

- **Specific language** (people added by this language)
- **Language mastery** necessary
- Topics to define **subgroups**

Plus: different **statistics** to measure bias (distribution vs. impact of person group size)

Summary: our contextualization

- **Most important:**
  Interaction between
  Variable used to define **subgroups** of survey and **language** added

  - Coverage difference: +/-0% (other Christ., prof., adding A)
    …+20% (Muslim, prof., adding A)
  - Cramér's V difference: -.08 (relig, prof., adding A)
    …+.05 (relig, prof., adding P)
  - Std.dev. difference: -.29 (relig, prof., adding A)
    …+.05 (relig, adding P)

- **Mastery** needed / **coverage measure** less important

FORS
explore.understand.share.

Discussion points

Generalisation: adding survey languages

- visual versus aural **modes**
- **Multi-topic** surveys
- **Countries** with little knowledge about language competence

- Effects on nonresponse / measurement

Language coverage by mastery

good ⟷ basic

[% covered]	Nat l.	+E	+P	+A	+S	Nat l.	+E	+P	+A	+S
All individuals	**.91**	**.92**	**.93**	**.92**	**.92**	**.93**	**.95**	**.95**	**.94**	**.94**

Language combination robust against socio-demographic variables (pred. probability):

- Sex, age, marital status
- Survey year
- Region
- Length of stay in municipality