SERISS Experiment for Testing Close versus More Adaptive Questionnaire Translation Approaches:
First Results from the Translation Sessions and from the Field *(Work in progress)*

Brita Dorer, GESIS (Mannheim – Germany)
Lydia Repke, GESIS (Mannheim – Germany)
Ana Villar (London)

CSDI, Limerick, 26 March 2018

1

Overview

- Background: close versus adaptive translation
- Background: Research goals / research design
- Translations & first findings (qualitative)
- Measurement approach (quantitative)

1. Background

Dilemma: close vs. adaptive translation

- Most surveys: "Ask-the-same-question (ASQ)"

- No definition of what 'same question' means

- Researchers make different interpretations
  - 'As close as possible' at the word level
  - Others focus on making stimulus functionally equivalent and believe this requires adaptation

- No empirical evidence yet on which interpretation leads to more comparable data

Deciding level of closeness: example

How much of the electricity used in [country] should be generated from each energy source?

Energy sources: Coal, Natural gas, Hydroelectric power, Nuclear power, Solar power, Wind power, Biomass

❑ A very large amount

❑ A large amount

❑ A medium amount

❑ A small amount

❑ None at all

Deciding level of closeness: example (cont'd)

- Advance translation and translation verification suggested in most languages 'amount'

    $\rightarrow$ share / portion / part

- BUT portion posed limitations on the answers: you can't have a very large portion of all energy sources

- TEP: Natural language use vs. safe translation?

- Which version yields most comparable data?

2. Research goals / Research design

Research goals / questions

- What to do when "dilemmas" come up?

- Which approach is better: 'as close as possible' vs. 'encouraging adaptation'
  - Better = yielding more comparable data

- Under which conditions is that approach better? Does the effect differ by…
  - Language: are some languages more resilient to close translations than others?
  - Type of adaptation: factual, linguistic…
  - Degree of adaptation: are some adaptations "too much"?

Overall model

Independent variable:
Experimental manipulation
2 translation approaches

Dependent variable:
Measurement quality
Measurement comparability

Moderating variables:
1 Distance between
1a. Language (linguistic structure)
1b. Country/culture (different realities)

2 Type of adaptation
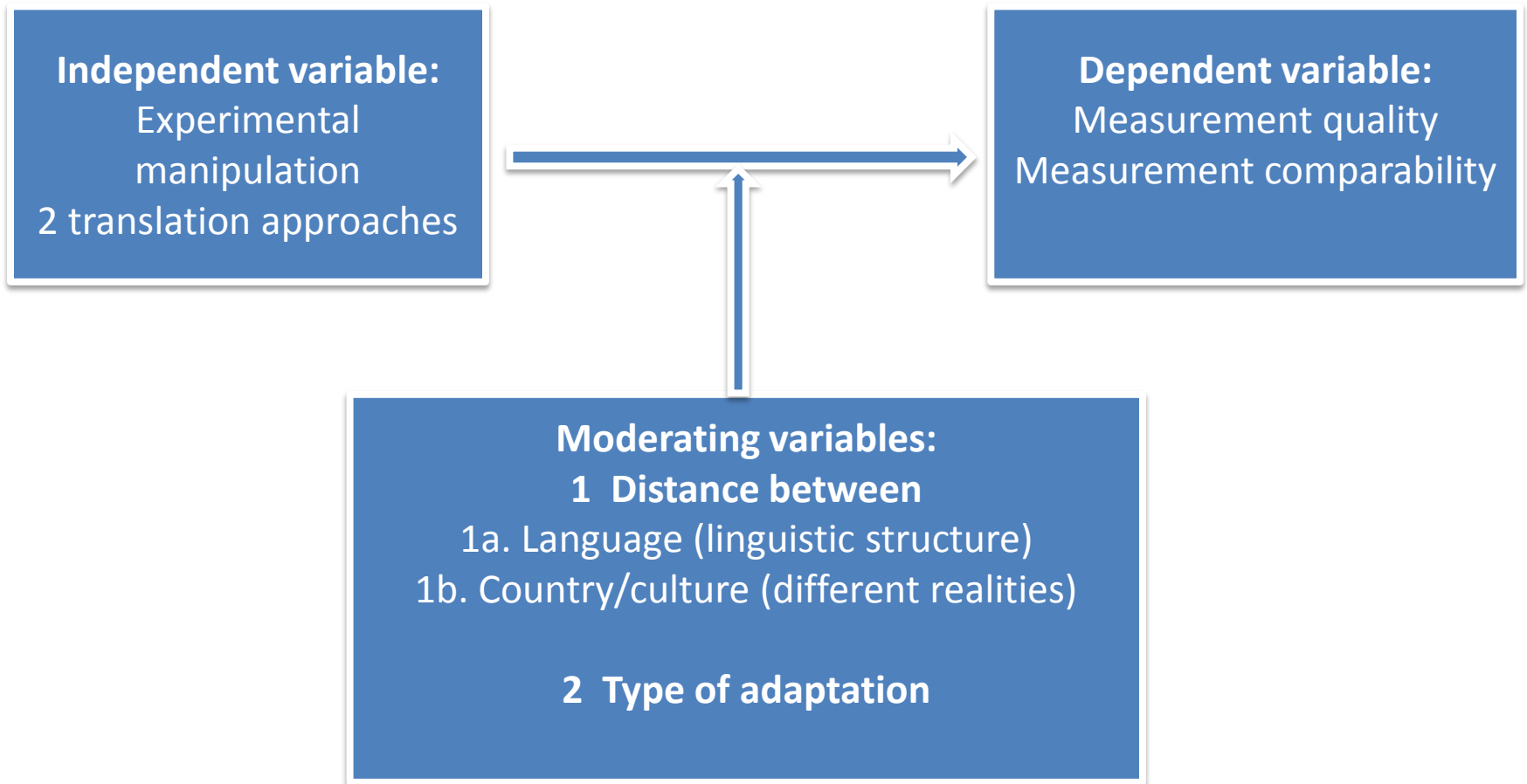
Research design

- 2 sets of instructions produced: close vs adapted
- 2 languages: Slovene (SL) & Estonian (ET)
- For each language 3 teams and 3 sets of items:
  - 2 parallel Translations following experimental instructions
  - Review / Adjudication meeting with instructions present
- Manipulation checks
  - Audio taping
  - Follow up interviews with translators
- Data collection: Cross-National Online Survey (CRONOS) panel
  - 60 items

Close = 'As close as possible'
Adapt = 'Encouraging adaptation'

	Team 1	Team 2	Team 3
Items 1-20	Close	Adapt	Adapt
Items 21-40	Adapt	Close	Adapt
Items 41-60	Close	Adapt	Close

3. Insights from the translations

Translation sessions: Estonian (ET) and Slovene (SL)

- Took place in summer 2017
- Team approach (2 translators + 1 reviewer/adjudicator)
- All reviewers experienced in questionnaire translation
- Almost all translators experienced in questionnaire translation
- None of them from ESS, ISSP, SHARE translation teams
- Each team 3 batches
- At least 1 week between the 3 batches
- 2 instructions documents (close – adaptive)
- All Review sessions recorded (audio): Manipulation check

Findings from translation sessions

- Follow-up interviews with all 6 reviewers (3 ET and 3 SL)

- Follow-up interviews with translators of 3 last teams (2 SL teams + 1 ET team)

➢ Not everybody realised the difference between the approaches! (1 ET, 1 SL team)

➢ Most common approach: usually all experienced reviewers translate as close as possible to source (comparability) but adapt where necessary

Ongoing: qualitative aspects

- Native speakers in ET and SL

- Analyse
  a) translation templates / final translations
  b) review sessions (audio recordings)

- WHAT was actually asked, what were the differences

- Combine with qualitative results

4. Measurement approaches / insights from the field

Analytical strategy

- Dependent on the items

- Possible methods
  - Multigroup Confirmatory Factor Analysis (MGCFA)
  - Item response theory
  - Web probing
  - Questions from benchmark surveys
  - Content analysis of the review meetings: also learn about how instructions worked
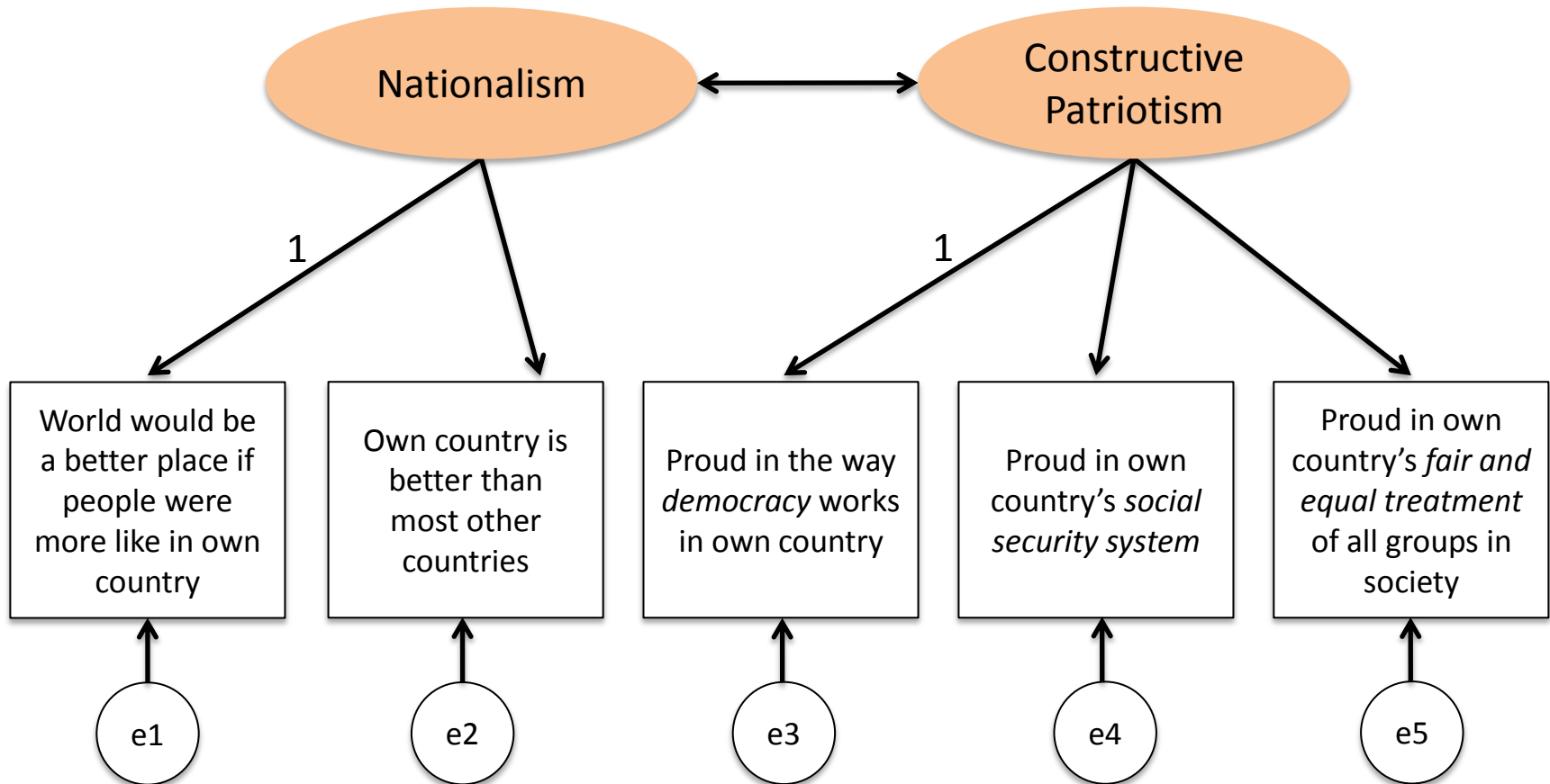
Sample

Country/ exp. group	EE	GB	SI	Total
1	173	209	194	576
2	153	223	217	593
3	173	201	204	578
Total	499	633	615	1.747

Note. In GB the three experimental groups are the same with respect to the questionnaire.

Example – National Identity

- 2013 ISSP Module on National Identity

- Dimensions measured by 5 items
  - Nationalism (2 items, 5-point scale)
  - Constructive patriotism (3 items, 4-point scale)

- Method: MGCFA (Davidov, 2009; Meitinger, 2017)

- 7 groups to test for measurement equivalence
  - EE (3 groups)
  - GB (3 "groups" ≙ 1 group)
  - SI (3 groups)

Example – National Identity

- Scalar invariance not supported by previous findings (Davidov, 2009)

- Possible problems with terms
  - E.g.: pride, social security system, democracy, all groups of society
  - Pride with how social security system works has been shown to load negatively on nationalism in GB (Davidov, 2009)

Next steps - Discussion

- Carry out / complete quantitative and qualitative analyses

- Discussion:
  How to test for measurement equivalence for single items?

THANKS !

Brita, Lydia and Ana