FORS
explore. understand. share.
Translation, wording, question design, survey design: an empirical exploration of different challenges to the comparability of international survey data

3MC conference, 26th July 2016, Chicago

by Michèle Ernst Stähli, International Surveys, FORS
Research question

• What makes concepts measurements comparable?
• How far the specific wording matters?
• How far the scale matters?
• How far survey design features matter? (sampling, contact procedures, interview mode, surrounding questions, etc.)
• What is the impact of the different sources of error?
Research strategy

1) Comparisons of same or very similar questions on a national level
   - across surveys (different modes, topics, question designs)
   - across question designs (different designs implemented in same survey)

2) Comparisons of same question in same survey on an international level
   - across countries (with relative survey designs)
   - across institutional and cultural country characteristics

3) Comparisons of very similar question in different surveys on an international level
   - across surveys (with relative question and survey designs)
   - across countries (with relative survey designs)
First step: national comparisons

Same questions compared in different national surveys (in Switzerland):

• Concepts:
  Political interest & Satisfaction with democracy in own country

• Surveys considered:
  - ESS (European Social Survey)
  - MOSAiCH/ ISSP (International Social Survey Program)
  - SELECTS (Swiss Electoral Study)
  - SHP (Swiss Household Panel)
First step: national comparisons

Same questions compared in different national surveys (in Switzerland):

• Concepts:
  Political interest & Satisfaction with democracy in own country

• Surveys considered:
  - ESS (European Social Survey)
  - MOSAiCH/ ISSP (International Social Survey Program)
  - SELECTS (Swiss Electoral Study)
  - SHP (Swiss Household Panel)

  All with simple random sampling, same s. frame
  All fielded in Switzerland in 2010/2011
  All produced by FORS
First step: national comparisons

Same questions compared in different national surveys (in Switzerland):

• Concepts:
  Political interest & Satisfaction with democracy in own country

• Surveys considered:
  - ESS (European Social Survey)
  - MOSAiCH/ ISSP (International Social Survey Program)
  - SELECTS (Swiss Electoral Study)
  - SHP (Swiss Household Panel)

Different populations
Different field procedures
Different modes
Different survey topics

Differences in questions:
- Wording
- Scale
First step: national comparisons

Same questions compared in different national surveys (in Switzerland):

• Concepts:
  Political interest & Satisfaction with democracy in own country

• Surveys considered:
  - ESS (European Social Survey)
  - MOSAiCH/ ISSP (International Social Survey Program)
  - SELECTS (Swiss Electoral Study)
  - SHP (Swiss Household Panel)

Different question designs implemented in one survey (ESS 2014)
# First step: national comparisons

## Overview of central features of the surveys and items

<table>
<thead>
<tr>
<th>Survey</th>
<th>ESS (cross-sectional)</th>
<th>MOSAiCH-ISSP (cross-s)</th>
<th>SELECTS (cross-sect.)</th>
<th>SHP (longitudinal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>General</td>
<td>General</td>
<td>Politics, postelectoral</td>
<td>General</td>
</tr>
<tr>
<td>Mode</td>
<td>FtF</td>
<td>FtF</td>
<td>CATI, WEB</td>
<td>CATI</td>
</tr>
<tr>
<td>Target pop</td>
<td>15+</td>
<td>18+</td>
<td>18+ Swiss citizens</td>
<td>Private HH, 14+ for individuals</td>
</tr>
<tr>
<td>Periodicity</td>
<td>every 2 years (since 2002)</td>
<td>every 2 years (since 97/05)</td>
<td>every 4 years (since 1995)</td>
<td>every year (since 1999)</td>
</tr>
<tr>
<td>Sample</td>
<td>SRS 1500</td>
<td>SRS 1200</td>
<td>SRS 4000 -5000</td>
<td>SRS 4000 -5000 HH</td>
</tr>
<tr>
<td>Average response rate</td>
<td>50%</td>
<td>50%</td>
<td>35% (including coverage error)</td>
<td>65% of HH in W1, 85% of individuals within HH, 10% attrition</td>
</tr>
<tr>
<td>Edition(s) considered</td>
<td>2010 &amp; 2014</td>
<td>2011</td>
<td>2011</td>
<td>2011</td>
</tr>
</tbody>
</table>

### «Political interest» (backtranslated)
- “How interested would you say you are in politics?”
  - 4 point scale, fully labeled
  - “Very interested”
  - “Quite interested”
  - “Hardly interested”
  - “Not interested at all”
- With showcards

### «Satisfaction with democracy» (backtranslated)
- “On the whole, how satisfied are you with the way democracy works in Switzerland?”
  - 11 point scale, endpoint labels:
    - “Extremely dissatisfied”
    - “Extremely satisfied”
- With showcards

### «Satisfaction with democracy» (backtranslated)
- “On the whole, are you ... satisfied with the way democracy works in Switzerland?”
  - 4 point scale, fully labeled
  - “Very satisfied”
  - “Rather satisfied”
  - “Rather not satisfied”
  - “Not satisfied at all”
- With showcards

### «Satisfaction with democracy» (backtranslated)
- “In general, are you ... satisfied with the way democracy works in our country?”
  - 4 point scale, fully labeled
  - “Very satisfied”
  - “Rather satisfied”
  - “Rather not satisfied”
  - “Not satisfied at all”
- With showcards

### «Satisfaction with democracy» (backtranslated)
- “Generally, what is your level of satisfaction with the way democracy works in our country?”
  - 11 point scale, endpoint labels:
    - “Not at all satisfied”
    - “Completely satisfied”

### «Satisfaction with democracy» (backtranslated)
- “Generally, how interested are you in politics? Are you...”
  - 4 point scale, fully labeled
  - “Very interested”
  - “Rather interested”
  - “Rather not interested”
  - “Not interested at all”
- With showcards

### «Satisfaction with democracy» (backtranslated)
- “Generally, how interested are you in politics?”
  - 11 point scale, endpoint labels:
    - “Not at all interested”
    - “Very interested”
- With showcards
First step: national comparisons

Overview of central features of the surveys and items

<table>
<thead>
<tr>
<th>Survey</th>
<th>ESS (cross-sectional)</th>
<th>MOSAiCH-ISSP (cross-s)</th>
<th>SELECTS (cross-sect.)</th>
<th>SHP (longitudinal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>General</td>
<td>General</td>
<td>Politics, postelectoral</td>
<td>General</td>
</tr>
<tr>
<td>Mode</td>
<td>FtF</td>
<td>FtF</td>
<td>CATI, WEB</td>
<td>CATI</td>
</tr>
<tr>
<td>Target pop</td>
<td>15+</td>
<td>18+</td>
<td>18+ Swiss citizens</td>
<td>Private HH, 14+ for individuals</td>
</tr>
<tr>
<td>Periodicity</td>
<td>every 2 years (since 2002)</td>
<td>every 2 years (since 97/05)</td>
<td>every 4 years (since 1995)</td>
<td>every year (since 1999)</td>
</tr>
<tr>
<td>Sample</td>
<td>SRS 1500</td>
<td>SRS 1200</td>
<td>SRS 4000 -5000</td>
<td>SRS 4000 -5000 HH</td>
</tr>
<tr>
<td>Average response rate</td>
<td>50%</td>
<td>50%</td>
<td>35% (including coverage error)</td>
<td>65% of HH in W1, 85% of individuals within HH, 10% attrition</td>
</tr>
<tr>
<td>Edition(s) considered</td>
<td>2010 &amp; 2014</td>
<td>2011</td>
<td>2011</td>
<td>2011</td>
</tr>
</tbody>
</table>

«Political interest» (backtranslated)

- "How interested would you say you are in politics?"
  4 point scale, fully labeled
  "Very interested"
  "Quite interested"
  "Hardly interested"
  "Not interested at all"
  With showcards

«Satisfaction with democracy» (backtranslated)

- "On the whole, how satisfied are you with the way democracy works in Switzerland?"
  11 point scale, endpoint labels:
  "Extremely dissatisfied"
  "Extremely satisfied"
  With showcards
First step: national comparisons

Aims: Detect and compare
- Coverage error
- Nonresponse bias
- Measurement errors due to:
  - presence of an interviewer
  - mode of data collection
  - question wording
  - number and labels of answer categories

Strategy:
- Compare sample composition (design weighted)
- Weight on socio-demographics
- Compare means
- Retest in same survey
### “Political interest” across CH surveys

<table>
<thead>
<tr>
<th>Scale</th>
<th>Scale wording</th>
<th>Quest word</th>
<th>ESS10 FtF (N~1200)</th>
<th>MOS11 FtF (N~1000)</th>
<th>SEL11 cati (N~2300)</th>
<th>SEL11 web (N~430)</th>
<th>SHP11 cati (N~6200)</th>
<th>ESS14 FtF (N~1140/570)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mean (SE)</td>
<td>mean (SE)</td>
<td>mean (SE)</td>
<td>mean (SE)</td>
<td>mean (SE)</td>
<td>mean (SE)</td>
</tr>
<tr>
<td>4</td>
<td>quite/hardly VS rather/rather not</td>
<td>'-' vs 'generally'</td>
<td>1.711 .024</td>
<td>1.746 .028</td>
<td>1.878 .020</td>
<td>1.871 .037</td>
<td></td>
<td>2.215 (2.234 retest)</td>
</tr>
<tr>
<td>11</td>
<td>not at all/very</td>
<td>generally</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.365 (~1.95)</td>
</tr>
</tbody>
</table>

### “Satisfaction with democracy” across CH surveys

<table>
<thead>
<tr>
<th>Scale</th>
<th>Scale wording</th>
<th>Quest word</th>
<th>ESS10 FtF (N~1200)</th>
<th>MOS11 FtF (N~1000)</th>
<th>SEL11 cati (N~2300)</th>
<th>SEL11 web (N~430)</th>
<th>SHP11 cati (N~6200)</th>
<th>ESS14 FtF (N~1140/570)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mean (SE)</td>
<td>mean (SE)</td>
<td>mean (SE)</td>
<td>mean (SE)</td>
<td>mean (SE)</td>
<td>mean (SE)</td>
</tr>
<tr>
<td>11</td>
<td>not at all/completely VS extr diss/extr sat</td>
<td>generally , level of sat VS 'on the whole'</td>
<td>7.022 (~2.55) .056</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.171 (~2.24) .024</td>
</tr>
<tr>
<td>4</td>
<td>very/rather/rather not/not at all</td>
<td>globally</td>
<td>2.019 .020</td>
<td>2.004 .015</td>
<td>1.820 .031</td>
<td></td>
<td></td>
<td>1.944 .022</td>
</tr>
</tbody>
</table>
“Political interest” 4pt scale (ESS) vs 11pt scale (SHP)

**Orig by Retest**

81% keep same answer (gamma = .93***)

9% say to be more interested
10% say to be less interested

81% keep «same» answer (gamma = -.85 ***)

8% say to be more interested
11% say to be less interested

First step: national comparisons
First step: national comparisons

Results:

- Compared and corrected for coverage error
- Compared and corrected for nonresponse bias

⇒ Coverage and NR errors are mostly along the same socio-demographic characteristics: we miss foreigners, single households, non-married

⇒ But poststrat. weights do not correct or alter the means of the considered two items

- Measurement errors due to:
  presence of an interviewer, mode of data collection, question wording, number and labels of answer categories.

⇒ The differences in means in a single country are bigger across surveys than across question versions
Second step: international comparisons in same survey

Same question compared in different countries of the same survey (ESS):

• Concepts:
  Political interest

• Surveys considered:
  - ESS (European Social Survey)
    Round 7 (2014)
    21 countries
  (- ISSP 2014)

Identical question
Identical context (place in questionnaire)
Very similar field procedures

Different sampling frames,
Different translations,
Same mode but CAPI/PAPI
Produced by different institutions
Second step: international comparisons in same survey
Second step: international comparisons in same survey
Second step: international comparisons in same survey
Second step: international comparisons in same survey
Second step: international comparisons in same survey

Observations pondérées par Design weight
Second step: international comparisons in same survey

Observations pondéréées par Post-stratification weight including design weight
Second step: international comparisons in same survey
Second step: international comparisons in same survey
Conclusions

- There are some differences in means across surveys in a single country, even if controlled for SD coverage and NR bias
- These differences are bigger across surveys than across question versions
- (Still not totally disentangled the effect of the different sources of error)
- Additional effects of the way to present the survey?
- In an international context, need to disentangle survey design effects from real country differences, but even apparently slight differences of design could play a role...
More information

- **Links and references:**

- **Contact:**
  Dr. Michèle Ernst Stähli
  Head of unit “International surveys”
  FORS, c/o University of Lausanne
  Tél. +41 21 692 37 36
  Michele.ErnstStaehli@fors.unil.ch
THANK YOU!

And now your questions …
## Overview of central features of the surveys and items

<table>
<thead>
<tr>
<th>Study</th>
<th>Topic</th>
<th>Periodicity</th>
<th>Target pop</th>
<th>Sample size</th>
<th>Mode</th>
<th>Average response rate</th>
<th>«Political interest» item (backtranslated)</th>
<th>«Satisfaction with democracy» item (backtranslated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESS (cross-sect)</td>
<td>General</td>
<td>every 2 years (since 2002)</td>
<td>15+ SRS 1500</td>
<td>FtF</td>
<td>50%</td>
<td>&quot;How interested would you say you are in politics?&quot; 4 point scale, fully labeled &quot;Very interested&quot; &quot;Quite interested&quot; &quot;Hardly interested&quot; &quot;Not interested at all&quot; With showcards</td>
<td>&quot;On the whole, how satisfied are you with the way democracy works in Switzerland?&quot; 11 point scale Endpoint labels: &quot;Extremely dissatisfied&quot; &quot;Extremely satisfied&quot; With showcards</td>
<td></td>
</tr>
<tr>
<td>MOSAiCH -ISSP (cross-sect)</td>
<td>General</td>
<td>every 2 years (since '97/2005)</td>
<td>18+ SRS 1200</td>
<td>FtF</td>
<td>50%</td>
<td>&quot;How interested would you say you are in politics?&quot; 4 point scale, fully labeled &quot;Very interested&quot; &quot;Quite interested&quot; &quot;Hardly interested&quot; &quot;Not interested at all&quot; With showcards</td>
<td>&quot;On the whole, are you ... satisfied with the way democracy works in Switzerland?&quot; 4 point scale, fully labeled &quot;Very satisfied&quot; &quot;Rather satisfied&quot; &quot;Rather not satisfied&quot; &quot;Not satisfied at all&quot; With showcards</td>
<td></td>
</tr>
<tr>
<td>SELECTS (cross-sect)</td>
<td>Electoral politics</td>
<td>every 4 years (since 1995)</td>
<td>18+ Swiss citizens</td>
<td>SRS 4000 - 5000</td>
<td>CATI  (web)</td>
<td>35%</td>
<td>&quot;Generally, how interested are you in politics? Are you...&quot; 4 point scale, fully labeled &quot;Very interested&quot; &quot;Rather interested&quot; &quot;Rather not interested&quot; &quot;Not interested at all&quot;</td>
<td>&quot;In general, are you ... satisfied with the way democracy works in our country?&quot; 4 point scale, fully labeled &quot;Very satisfied&quot; &quot;Rather satisfied&quot; &quot;Rather not satisfied&quot; &quot;Not satisfied at all&quot;</td>
</tr>
<tr>
<td>SHP (longitudinal)</td>
<td>General</td>
<td>every year (since 1999)</td>
<td>Private HH, 14+ for individuals</td>
<td>SRS 4000 - 5000 HH</td>
<td>CATI</td>
<td>65% of HH in W1, 85% of individuals within HH, 10% attrition</td>
<td>&quot;Generally, how interested are you in politics?&quot; 11 point scale Endpoint labels: &quot;Not at all interested&quot; &quot;Very interested&quot;</td>
<td>&quot;Generally, what is your level of satisfaction with the way democracy works in our country?&quot; 11 point scale Endpoint labels: &quot;Not at all satisfied&quot; &quot;Very satisfied&quot;</td>
</tr>
</tbody>
</table>