gesis
Leibniz Institute
for the Social Sciences

The longer, the better?
Selection of appropriate cross-national indicators of response quality in open-ended questions

Katharina Meitinger (GESIS & University of Mannheim)

Michael Braun & Dorothée Behr (GESIS)

Member of the
Leibniz Association

CSDI, Mannheim. March 17, 2017

OPEN-ENDED QUESTIONS

Open-ended questions are an important addition to closed items in the toolkit of social scientists:

- Exploration of unknown topics
- Verbal probing for validity and equivalence assessment

Recent increase in open-ended questions due to advances in technology that facilitate data collection and analysis (Poncheri et al. 2008)

But: Open-ended questions are cognitively more demanding than closed questions → Issues of response quality (e.g., high levels of nonresponse [Barrios et al. 2010; Denscombe 2008; Meitinger & Behr 2016])

→ Need for methodological experiments to improve the design of open-ended questions to reduce problems with response qualities

ASSESSING RESPONSE QUALITY IN OPEN-ENDED QUESTIONS

Number of methodological studies on open-ended questions:

- Survey stimuli:
  - Size of the answer boxes (Christian & Dillman 2004; Smyth et al. 2009; Israel 2010; Emde & Fuchs 2012; Behr et al. 2014)
  - Number of answer boxes (Fuchs 2009; Keusch 2014; Hofelich Mohr, Sell, & Lindsay 2015)
  - Use of motivational sentences (Oudejans & Christian 2011; Smyth et al. 2009; Kaczmirek, Meitinger, & Behr 2017)
  - Clarification features (Kunz & Fuchs 2012; Metzler, Kunz, & Fuchs 2015)
  - Position of the open-ended question in the web survey (Miller & Dumford 2014)
  - Counters indicating the number of characters written (Emde & Fuchs 2012)
  - Mode (Denscombe 2008)
- Respondents' characteristics (Andrews 2005; Barrios et al. 2010; Denscombe 2008; Miller & Dumford 2014; Smyth et al. 2009; Züll et al. 2015)

Most common indicators:　　Number of themes
　　　　　　　　　　　　　　　　Response length
　　　　　　　　　　　　　　　　Percentage of nonresponse
　　　　　　　　　　　　　　　　Response time

But: Most of these studies are based on a national sample.
Do these indicators also work in a cross-national context?

What could possibly go wrong?
Response Length

Are longer responses always better than short responses?

Cross-national context:

- Cross-national differences in response length (Meitinger, Braun, & Behr, forthcoming)

- **Linguistic reasons:** Languages might differ in the number of words necessary to express the same opinion

  ▸ Text expansion: Text length in-/decreases across languages

  ▸ Text boundaries: No space between words (Asian languages), compound words (German) → **Donaudampfschifffahrtsgesellschaftskapitänsmütze**

  ▸ Information provided: No use of gender, little personal pronouns in Asian languages (Usunier & Roulin 2010)

  ▸ Information density: Linguistic information per syllable (Pellegrino et al. 2011)

What could possibly go wrong?
Response Length

Information density: Linguistic information per syllable (Pellegrino et al. 2011)

LANGUAGE	INFORMATION DENSITY ID_L
English	0.91 (± 0.04)
French	0.74 (± 0.04)
German	0.79 (± 0.03)
Italian	0.72 (± 0.04)
Japanese	0.49 (± 0.02)
Mandarin	0.94 (± 0.04)
Spanish	0.63 (± 0.02)
Vietnamese	1 (reference)

What could possibly go wrong? Response Length II

Are longer responses always better than short responses?

Cross-national context:

- **Culture-specific communication styles**
  - ▸ **"Overtness of message"** (Hall 1976)
    - Low-context cultures: Explicit communication: simple, linear, clear
    - High-context cultures: Implicit communication: reading between the lines, metaphors
  - ▸ **Succinct, exacting, and elaborate communication styles** (Gudykunst, Ting-Toomey, & Chua 1988)
    - Succinct: Statements that go to the point, frequent pauses (e.g., Japan)
    - Exacting: Provide the exact information that is necessary ("neither more nor less")
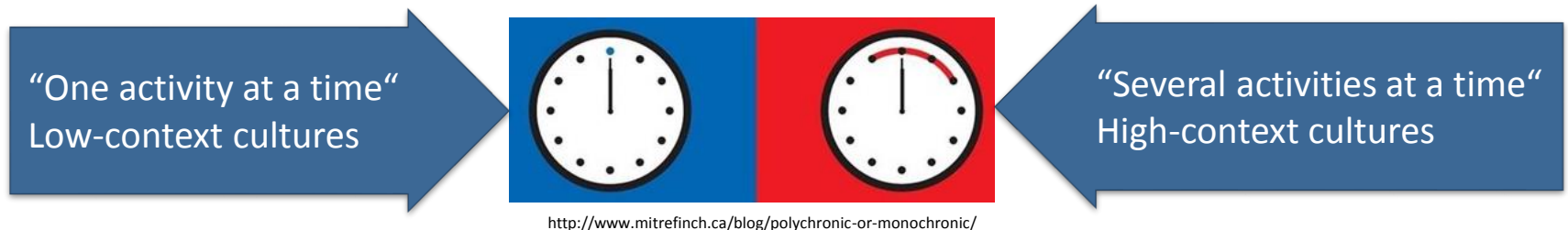   - Elaborate: Richness in language, repetitions, and expressiveness (e.g., Middle East)

U.S. context:

- Long open-ended answers might reduce intercoder reliability (Conrad et al. 2016)

What could possibly go wrong?
Response Time

What does it mean if respondents take longer?

- **"Flexibility of time":** Monochronic vs. polychronic time (Hall 1983)

"One activity at a time"
Low-context cultures

"Several activities at a time"
High-context cultures

http://www.mitrefinch.ca/blog/polychronic-or-monochronic/

- ➢ Closed items: Small group differences in response latencies between White Americans and Mexican Americans, African Americans as well as Korean Americans →Cultural differences in time perceptions and time utilization (Johnson et al. 2015)

- Response time in open-ended questions is related to response length. The longer the written response, the longer is the response time.

Cultural values potentially affecting nonresponse

Latin America:

Simpatía: Respondents aims to be polite, agreeable, likeable, and respectful in conversations (Triandis, Marin, Lisansky, & Betancourt 1984)

→ Might have an impact on nonresponse behavior.

Online Probing

Approach: Application of probing techniques from cognitive interviewing in web surveys

Goals of online probing:

▶ Reveal respondents' cognitive processes when answering a survey question

▶ Uncover equivalence problems in cross-national surveys

Three types of probes (Prüfer and Rexroth 2005; Willis 2005)**:**

- **Category-selection**: Inquires about the reasons why a certain answer category has been chosen

- **Specific**: Asks for additional information on a particular detail in the question

- **Comprehension**: Requests a definition of a specific term

Research Questions

1) Are there cross-national differences with regard to the different indicators of response quality in open-ended questions?

2) Do these differences hold with different topics?

3) Do the indicators arrive at similar or contradictory conclusions with regard to response quality?

METHOD & DATA

Data:

- 2 Web surveys
  - ▸ Survey 1: May 2014, 2,685 respondents: ISSP Module on National Identity
  - ▸ Survey 2: June 2014, 2,689 respondents: ISSP Modules on Citizenship and Family and Gender Roles
- Respondents from non-probability online-access panels
- Quota: Age, gender, and education
- Countries: Germany, Great Britain, the U.S., Spain, and Mexico
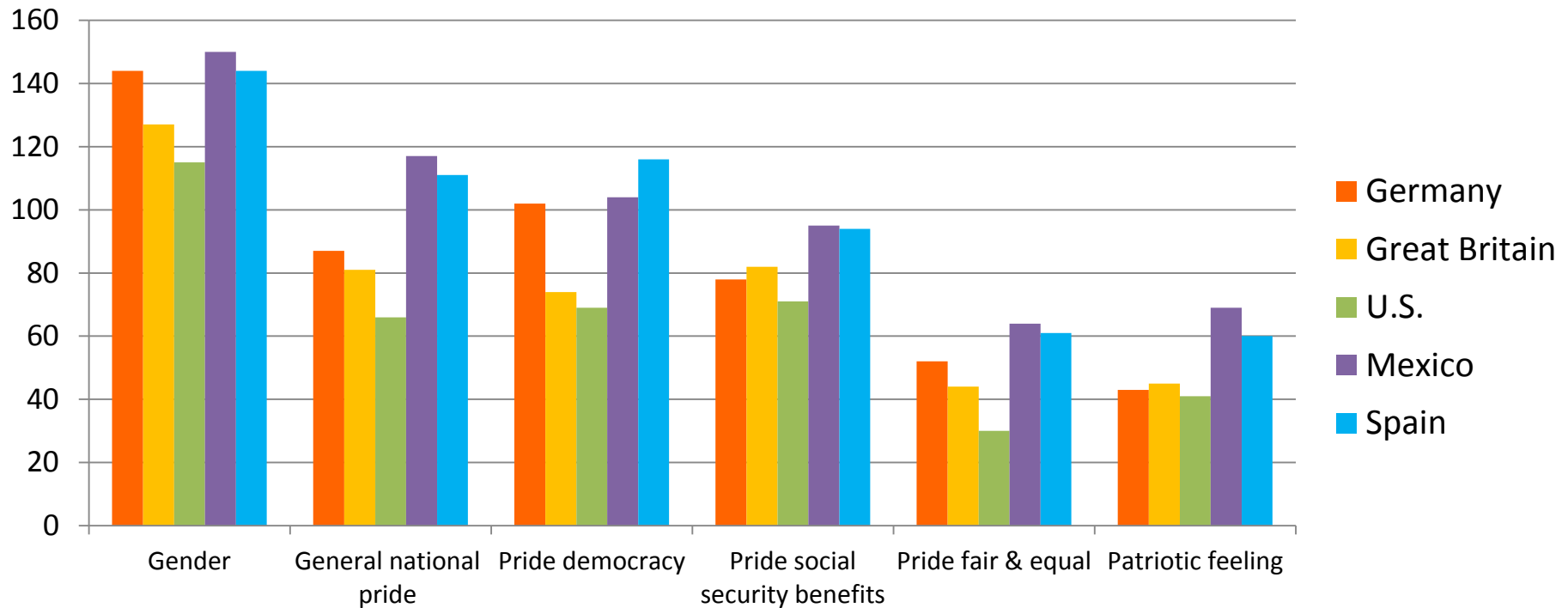- Languages: English, German, & Spanish

Used items:

- Gender item
- General national pride
- Specific pride: Democracy
- Specific pride: Social security benefits
- Specific pride: Fair & equal treat of all groups in society
- Patriotic feelings

RESULTS

RESPONSE LENGTH

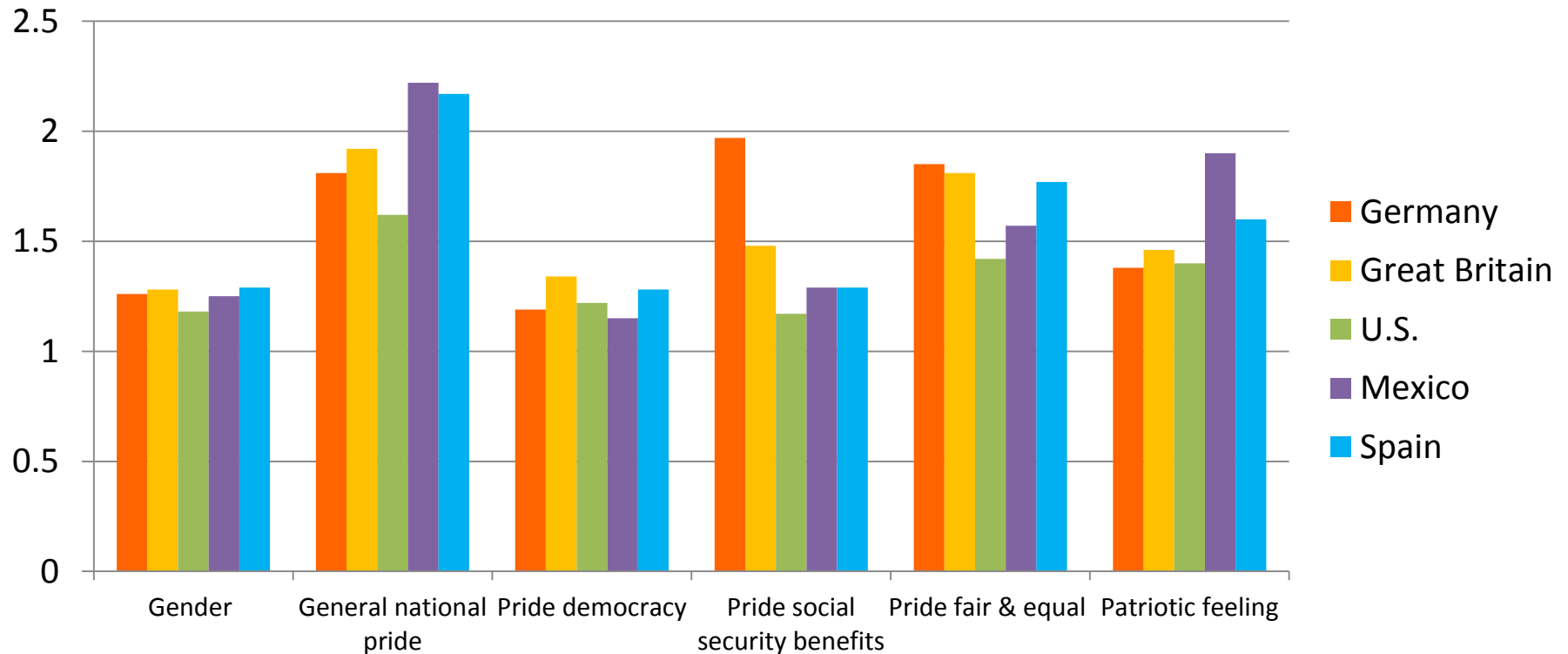**Response length:** Average number of characters

U.S.: Shortest responses for each probe
Great Britain & Germany: Depending on topic: long or middle-long responses
Mexico & Spain: Longest responses

Number of Themes

**Number of themes:** Average number of themes mentioned by respondents

U.S.: Lowest number of themes (exception: "Pride Democracy")
Other countries: Dependent on topic

Nonresponse

Nonresponse: Percentage of respondents that refuse answering, leave empty answer boxes, or provide don't know or unintelligible responses

Hard nonresponse: Respondents leave empty answer boxes
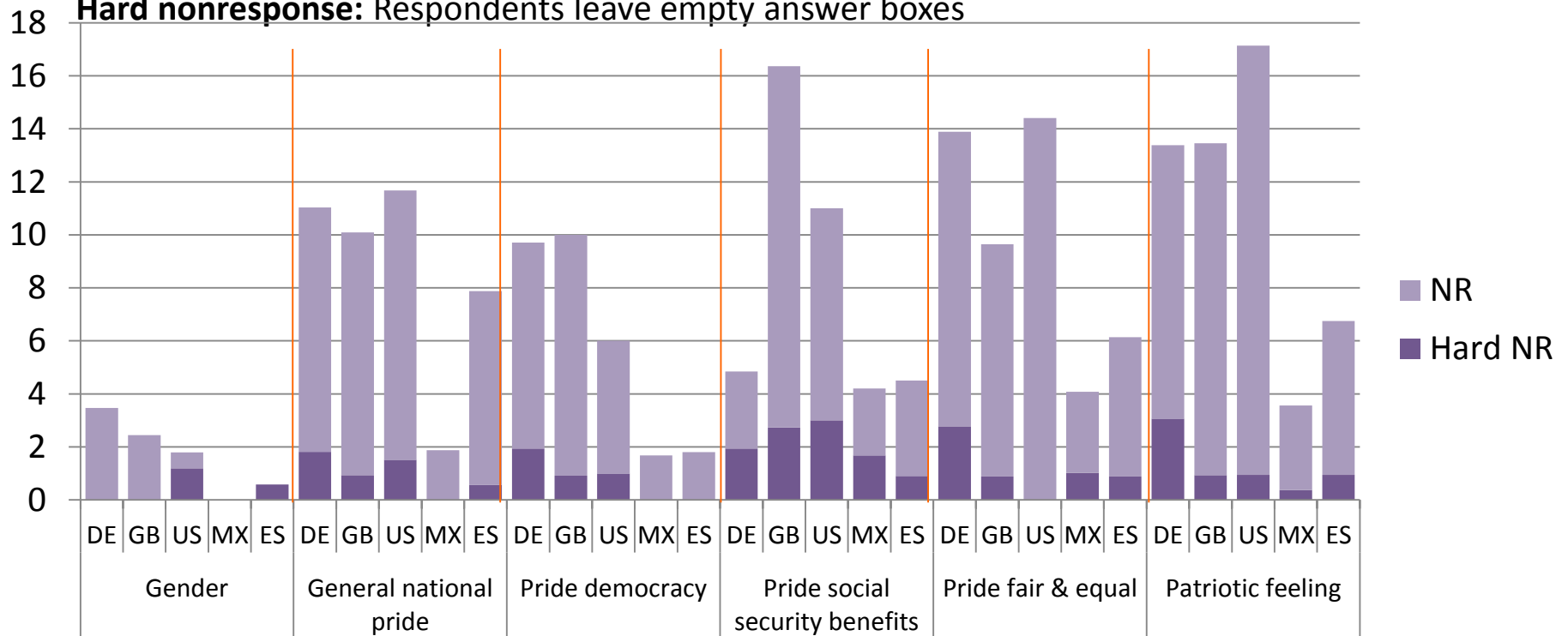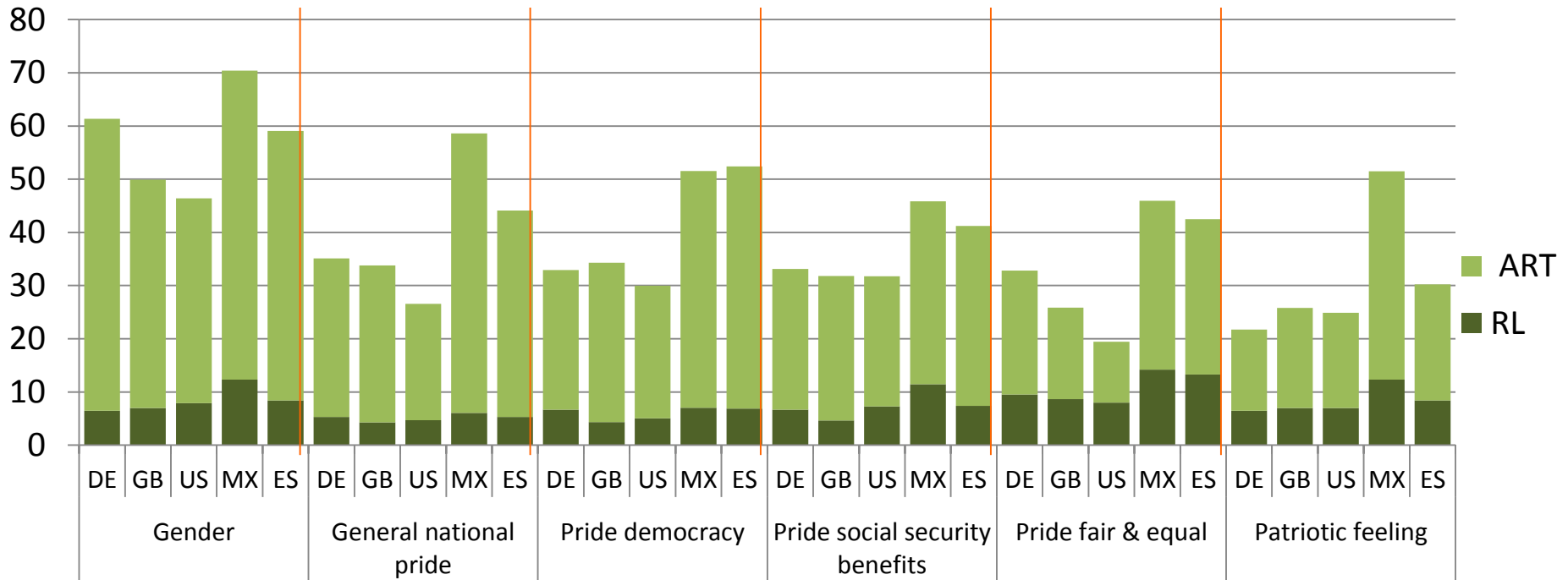

Mexico: Lowest nonresponse
Spain: Low nonresponse
Other countries: Dependent on topic

Response Time (in seconds/median)

Absolute response time (ART): Response time till last click

Response latency (RL): Response time till first key stroke

Mexico: Longest response time (exception: "Pride democracy")
U.S.: Shortest response time (exception: "Patriotic feeling")

Do we arrive at similar conclusions with all indicators?

Correlation of number of themes and response length

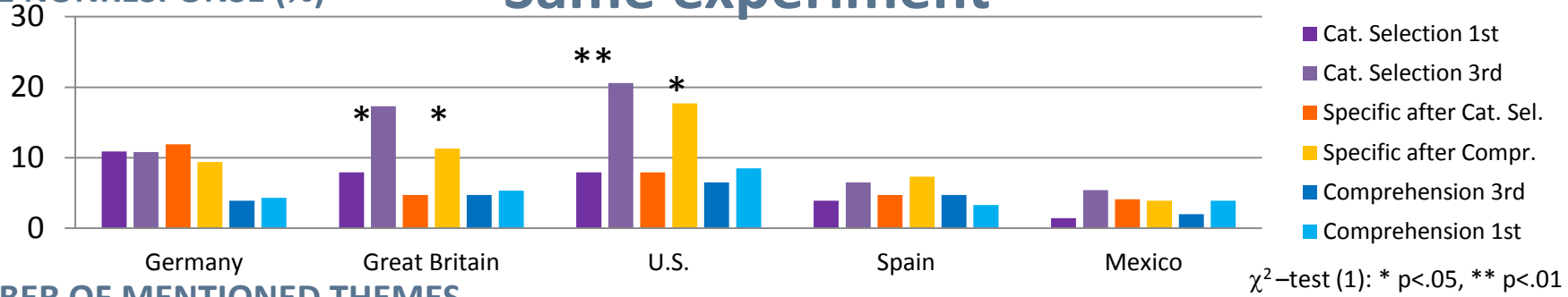|	Germany	Great Britain	U.S.	Mexico	Spain
Gender	.40	.58	.44	.45	.36
General nat. pride	.61	.68	.57	.11	.66
Pride democracy	.42	.41	.49	.40	.69
Pride social security benefits	.08	.31	-.03	.22	.48
Pride fair & equal	.09	.55	.52	.48	.42
Patriotic feeling	.47	.53	.50	.58	.57

Correlations change in size across topics and countries
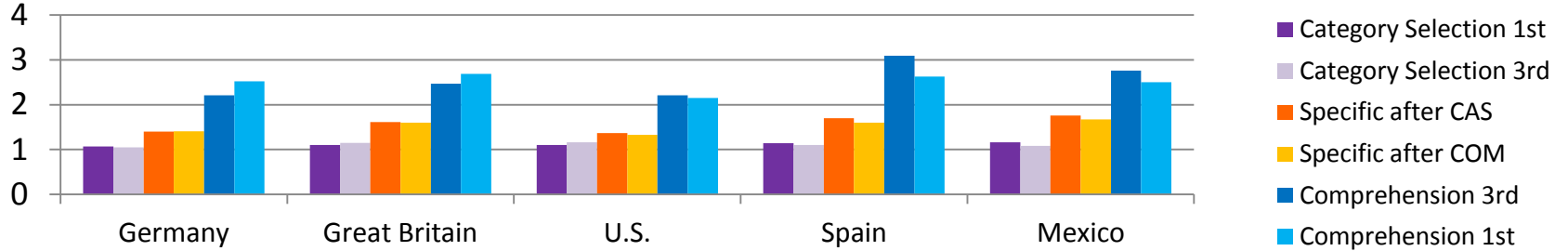→ Response length might not substitute for number of themes as an indicator
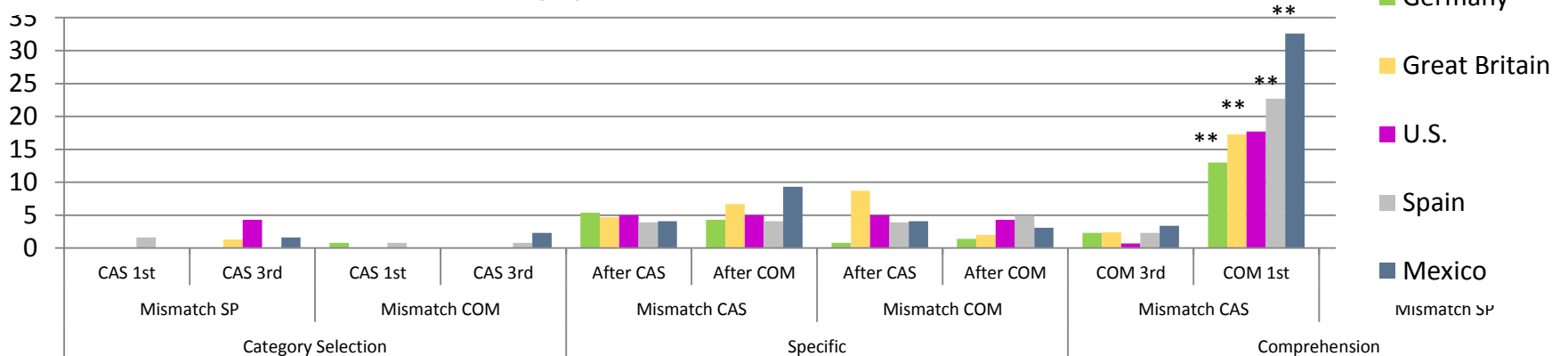
Why does this matter?

Same experiment

PROBE NONRESPONSE (%)

Legend (top chart):
- Cat. Selection 1st
- Cat. Selection 3rd
- Specific after Cat. Sel.
- Specific after Compr.
- Comprehension 3rd
- Comprehension 1st

χ^2–test (1): * p<.05, ** p<.01

NUMBER OF MENTIONED THEMES

Legend (middle chart):
- Category Selection 1st
- Category Selection 3rd
- Specific after CAS
- Specific after COM
- Comprehension 3rd
- Comprehension 1st

MISMATCHING PROBE RESPONSES (%)

Legend (bottom chart):
- Germany
- Great Britain
- U.S.
- Spain
- Mexico

→ Depending on indicator different conclusions

CONCLUSION
&
FUTURE RESEARCH

CONCLUSION

- Clear cross-cultural differences with regard to the different indicators of response quality
- Extreme cases are Mexico and the U.S.
    - Mexico: Long responses, many themes, reduced nonresponse, longest response time
    - U.S.: Shortest responses, relatively few themes, shortest response time, inconsistent nonresponse

- This outcome is in line with:
    - Linguistic differences (response length & response time: Spanish language with low information density)
    - But: Differences within languages (Mexico vs. Spain; U.S. vs. GB) → Cultural factor
    - Culture-specific communication styles (response length: explicit communication in the U.S.)
    - Differences in time perception (response time: monochronic time use in the U.S.)
    - Cultural values (nonresponse: Simpatía in Mexico)

- Also: Correlations between indicators are not necessarily stable across questions and countries

→ Recommendation to use multiple indicators

FUTURE RESEARCH

- More fine-grained statistical analysis that controls for gender, age, education, etc.

- Check different types of open-ended answers

- Include more countries and languages

- Find more theoretical explanations for cross-cultural differences. Any suggestions?

Thank you!

Contact: katharina.meitinger@gesis.org

Publications about Online Probing

- Behr, Dorothée, Wolfgang Bandilla, Lars Kaczmirek, and Michael Braun. 2014. "Cognitive probes in web surveys: on the effect of different text box size and probing exposure on response quality." *Social Science Computer Review* 32 (4): 524-533. doi: http://dx.doi.org/10.1177/0894439313485203.

- Behr, Dorothée, Michael Braun, Lars Kaczmirek, and Wolfgang Bandilla. 2014. "Item comparability in cross-national surveys: results from asking probing questions in cross-national web surveys about attitudes towards civil disobedience." *Quality & Quantity* 48 (1): 127-148. doi: http://dx.doi.org/10.1007/s11135-012-9754-8.

- Braun, Michael, Dorothée Behr, and Lars Kaczmirek. 2013. "Assessing cross-national equivalence of measures of xenophobia: evidence from probing in web surveys." *International Journal of Public Opinion Research* 25 (3): 383-395. doi:http://dx.doi.org/10.1093/ijpor/eds034.

- Behr, Dorothée, Michael Braun, Lars Kaczmirek, and Wolfgang Bandilla. 2013. "Testing the validity of gender ideology items by implementing probing questions in web surveys." *Field Methods* 25 (2): 124-141. doi: http://dx.doi.org/10.1177/1525822X12462525.http://fmx.sagepub.com/content/25/2/124.

- Behr, Dorothée, Lars Kaczmirek, Wolfgang Bandilla, and Michael Braun. 2012. "Asking probing questions in web surveys: which factors have an impact on the quality of responses?." *Social Science Computer Review* 30 (4): 487-498. doi:http://dx.doi.org/10.1177/0894439311435305.

- Behr, Dorothée, and Michael Braun. 2015. "Satisfaction with the way democracy works: How respondents across countries understand the question." In *Hopes and anxieties. Six waves of the European Social Survey* , edited by P. B. Sztabinski, H. Domanski, and F. Sztabinski, 121-138. Frankfurt am Main: Lang.

- Braun, Michael, Dorothée Behr, Lars Kaczmirek, and Wolfgang Bandilla. 2014. "Evaluating cross-national item equivalence with probing questions in web surveys." In *Improving survey methods : lessons from recent research*, edited by Uwe Engel, Ben Jann, Peter Lynn, Annette Scherpenzeel, and Patrick Sturgis, 184-200. New York: Routledge

- Meitinger, Katharina, and Dorothée Behr. 2016 (Forthcoming). "Comparing Cognitive Interviewing and Online Probing: Do They Find Similar Results?." *Field Methods*. doi:10.1177/1525822X15625866