

Uses of Client Side Paradata in Web Surveys

Dirk Heerwegh

Center for Survey Methodology

Katholieke Universiteit Leuven, Belgium

Overview

- Types of paradata
- Client side paradata (CSP)
- How CSP are recorded and analyzed: in paper
- Examples
 - Calibration of progress indicators
 - Response formats
 - Question wording
 - Attitude strength

Types of paradata

- Paradata are
 - Data that describe how respondent filled in survey questionnaire
- Collect web server events (server-side paradata): page visits.
 - Information on level of web survey page
- Collect on respondent's computer: client-side paradata (CSP)
 - Information on level of survey item

Client Side Paradata

- CSP
 - Describe actions of respondent within a web page
 - Clicking response alternatives
 - Changing answers
 - Clicking hyperlinks
 - Scrolling the page
 - Moving the mouse pointer
 - Interrupting the task
 - ... all with high-precision timestamps (ms)

Client Side Paradata

- CSP
 - Are collected using a JavaScript on the web page
 - » <http://perswww.kuleuven.ac.be/~u0034437/public/csp.htm>
 - Data are stored in hidden field on web page
 - Content of hidden field stored in data set on web server when respondent submits the page
 - More info on collecting and analyzing these data: see paper

Client Side Paradata

- CSP
 - Can be used to answer research questions that require information on answering process
 - But some limitations
 - CSP are lost if respondent abandons web survey
 - CSP response times include time needed to read Q
 - No direct insight into internal cognitive process of R, need to deduce it from manifest behavior

Examples

- Examples are in paper
 - With information on population, sample size, references to papers in which original analyses can be found or the analyses themselves
 - Most of the time: student samples (of around 1,000 units) or a web panel sample (of around 700 units). These samples are not representative for Belgian or Flemish population

Examples

- A. Use CSP for purely technical matters
 - Calibration of progress indicator
 - See paper

Examples

B. More methodological questions:

- The effect of response formats on data quality
 - Combination of technical question (response format) and methodological question (its effect on web survey data quality)
 - Radio buttons and drop boxes serve same purpose (select one response alternative among set)
 - But drop-boxes require two mouse clicks and hide response alternatives: more difficult to use than radio buttons

Examples

- Higher task difficulty: higher probability of satisficing
- Satisficing = producing appearingly reasonable answers, but without carefully performing all cognitive steps
- Manifestations of satisficing: a.o. "don't know" (DK) responses
- Outcome data from experiment: drop-boxes do not yield more DK answers than radio-buttons: so no effect of response formats?
- Look at CSP

Examples

- CSP showed that
 - Drop boxes slow respondents down (longer response times)
 - R change answer to drop-boxes less often (changing response requires too much effort?)
 - Using drop boxes in battery of items: respondents stick more to prescribed sequence
 - » perhaps more difficult question format drains so much cognitive resources from respondent that they cannot find more optimal perspective on the battery of items
- Thus, subtle indications that drop-boxes are indeed more difficult to use than radio-buttons
- These differences would not have emerged without CSP

Examples

C. Purely methodological questions

C1. Question wording

- 6 question versions on granting municipal suffrage to non-EU immigrants in Belgium
- 2 x 3 experiment:

	-	5 years or longer	Sufficiently long
-	1	2	3
Legal	4	5	6

Examples

- RAS model (Zaller 1992): reported opinions under the influence of activated beliefs (cf. Ajzen 2001)
- Since no "one true opinion", reported opinions will change according to which considerations are embedded into question
- Significantly different response distributions across conditions were found: support for RAS model (from 30% in favor to 60% in favor)

Question wording (translated from Dutch)	Response (translated from Dutch)					Total (<i>n</i>)
	Totally disagree	disagree	neither agree, nor disagree	Agree	Totally agree	
1. Immigrants residing in Belgium should be granted municipal suffrage	26,96	21,74	22,17	19,13	10,00	100 (230)
2. Immigrants residing in Belgium for a sufficiently long period of time should be granted municipal suffrage	18,18	18,61	18,61	26,84	17,75	100 (231)
3. Immigrants residing in Belgium for 5 years or longer should be granted municipal suffrage	25,96	21,70	20,43	19,57	12,34	100 (235)
4. Immigrants residing legally in Belgium should be granted municipal suffrage	17,52	23,50	22,22	25,64	11,11	100 (234)
5. Immigrants residing legally in Belgium for a sufficiently long period of time should be granted municipal suffrage	14,35	13,90	15,25	38,57	17,94	100 (223)
6. Immigrants residing legally in Belgium for 5 years or longer should be granted municipal suffrage	13,19	21,28	22,13	30,21	13,19	100 (235)

$\chi^2(20)=63.1034, p<0.0001$

Examples

- Supplemented with CSP data:
 - Different question versions take different amounts of time to respond to
 - Probably not only because of length of survey item
 - More embedded considerations lead to more processing time needed, especially when a vague consideration ("sufficiently long period of time") is used

Examples

- Thus the CSP data allow us to get a more fine-grained view on what happens when respondents are confronted with different versions of a question

Examples

C2. Attitude strength

- Weak attitudes or non-attitudes (Converse 1964)
- Detection of weak attitude
 - Ask question
 - Present counterargument in follow-up question
 - Observe respondents change their mind in follow-up question. Changing of mind is indication of weaker attitude

Examples

- Two hypotheses:
 1. R's with **weak** attitude **need more time** to answer initial question because they need to **fabricate** response
 2. R's with **strong** attitude **need more time** to answer initial question because they are **more involved** and want to review as many considerations as possible (cf. RAS model)
- Bassili & Fletcher (1991) tested which of these hypotheses hold in a telephone survey
- They found that H1 holds, not H2
- Is this also true for web survey respondents?

Examples

- Web survey used comparable quotas question as Bassili & Fletcher study and similar counter argument in follow-up question
- Found similar results: respondents who change answer (weak attitude) to follow-up question had a significantly longer response time to initial question
 - Respondents with weak attitude need to fabricate a response which takes time
 - Respondents with strong attitude appear not to bother reviewing as many considerations as possible (supports view of survey respondents as "cognitive misers") or they know the opinion from memory (but this is not so evident – see RAS model: reported opinions are constructed based on salient beliefs)

Examples

- Additional analysis in web survey
 - Changing answers to knowledge questions
 - » Correlates with fewer correct responses
 - » Would indicate uncertainty
- CSP allowed to test a hypothesis that could otherwise not have been tackled in a web survey environment

Other possible uses of CSP

- Effects of visual grouping of answer options (Smyth, Dillman, et al. 2004 AAPOR paper)
- Effects of alternative question versions (Stern, Dillman, et al. 2004 WAPOR paper)
- Effects of age, internet experience, education and other respondent characteristics on response times (indications of cognitive capacity, ability to respond to questions) (Yan & Tourangeau 2004 AAPOR paper)

Other possible uses of CSP

- Test navigation aids (words, symbols,...) for later use in mail surveys
- Pretesting survey questions (difficult questions: long response times, changing of answers)
- Collect information on setting in which survey was completed (task interruptions)

Conclusions

- CSP have some limitations but also some useful applications
 - Assist in purely technical matters
 - Answer research questions that are a mixture of technical and methodological issues
 - Answer purely methodological questions
 - Can be very useful, but could do with some more fine-tuning

Conclusions

- Ethical issues
 - Should we inform the respondents?
 - Is it a potential invasion of privacy?
 - Eg. Income question or other sensitive question to which respondent answers and then resets the answer (e.g. "don't know"). We should not use CSP to retrieve the initial answer
 - Use CSP only if absolutely necessary, only for genuine research questions that do not threaten a respondent's privacy