Exploring Web Survey Paradata to Improve Survey Design

Réjean Doiron

Collection Planning and Research Division

Statistics Canada

Fifth International Workshop on Business Data Collection

Lisbon, Portugal

September 2018

Canada

Acknowledgements

Support provided by Business Survey Methodology Division, Statistics Canada

- Susan Demedash (susan.demedash@canada.ca)

- Jessica Andrews (jessica.andrews@canada.ca)

Purpose

- Highlight uses of paradata at Statistics Canada.

- Outline new research initiatives.

What is paradata?

- Microsoft Internet Information Services (IIS) logs;

- Show actions taken in the web questionnaire by both respondents and interviewers;

  - E.g.: what pages were visited, in what order, time spent on a page, browser used.

- Do not contain response data or other websites visited by respondent.

4

Statistics Canada   Statistique Canada

Canada

Monthly Survey of Manufacturing
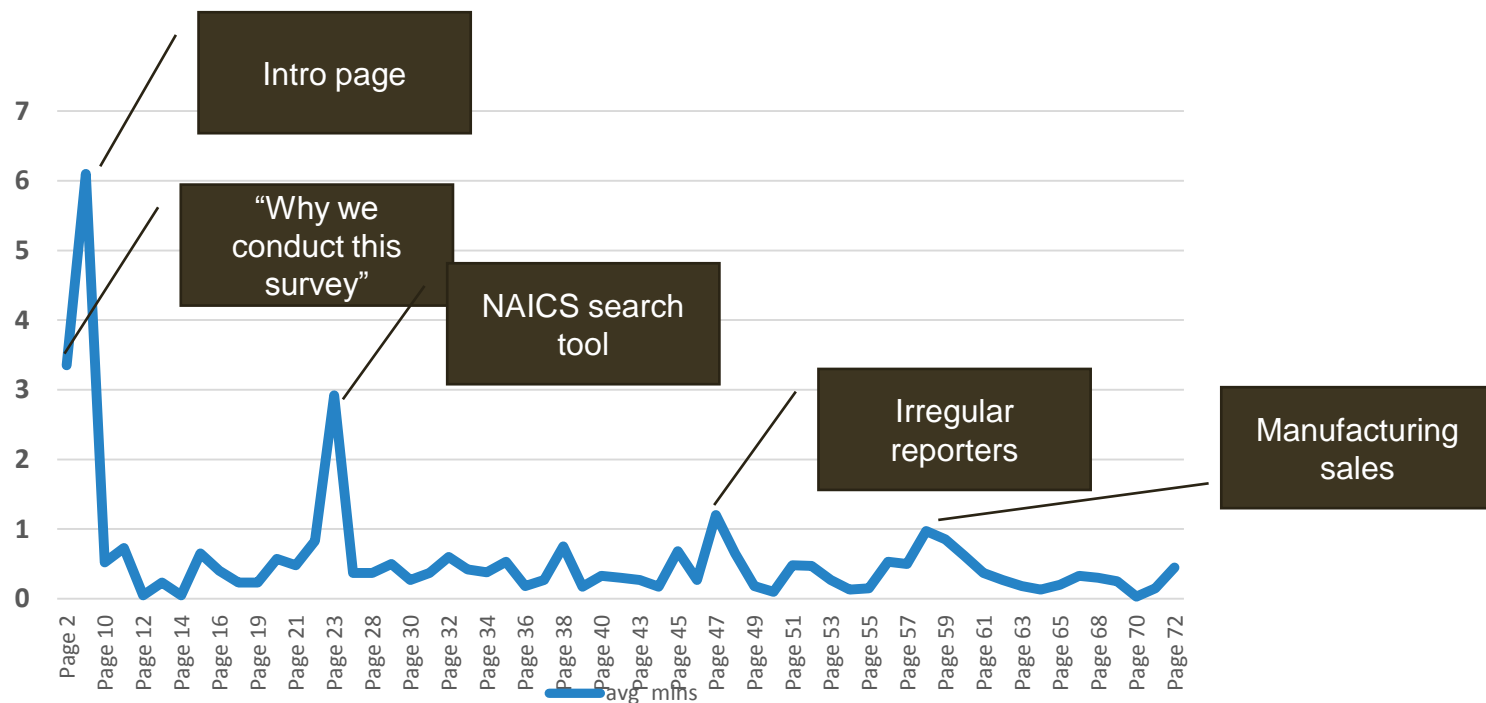
Monthly Survey of Manufacturing

- "Mission Critical" Survey
- Collects information on the sale of manufactured goods, inventories, unfilled orders and production capacity.
- Monthly sample = 5,800
- Target response rate = 95%

Time per page (mins)

- Overall EQ average = 14.26 mins.

Help Pages

- 50 of 5400 (0.9%) cases accessed at least one help page.

- Pages that generated the most help page requests:

Help Page	Number of Hits	% of Total
Page 62 - Production capacity	26	34.2
Page 58 - Report manufacturing sales	14	18.4
Page 59- Report book value of inventories	9	11.8
Page 19 - Confirm main business activity	7	9.2
Page 60 - Report value of unfilled orders	7	9.2

Edits

- Irregular reporters change answer >50% of the time.

- Edits on sales/inventory changed less frequently.

Page Triggering Edit	Responses Received	Respondent - EQ Edit Did Not Trigger	Respondent - EQ Edit Triggered, Corrected	Respondent - EQ Edit Triggered, Not Corrected	Total Edits Triggered	Help Page?
Page 45 - Reporting date wrong for irregular reporters	4,519	2,605	204	246	450	No
Page 58 - Sales value changed significantly	4,519	2,628	33	394	427	Yes
Page 58 - Total sales value changed significantly	4,519	2,648	45	362	407	Yes
Page 59 - Total value inventory changed significantly	4,519	2,116	53	223	276	Yes

Respondents switching between collection modes

		Initial Collection Mode Sent					
		Web		**Mail/Fax**		**Telephone**	
		Units	**%**	**Units**	**%**	**Units**	**%**
Mode Received	**Total**	4,213	100.0%	17	100.0%	644	100.0%
	Web	2,950	70.0%	0	0.0%	6	0.9%
	Mail/Fax	14	0.3%	3	17.6%	2	0.3%
	Telephone	1,222	29.0%	13	76.5%	635	98.6%
	Non-response	27	0.6%	1	5.9%	1	0.2%

Conclusions

- Time per page:
  - Production capacity questions not significantly more burdensome than other questions.

- Help pages:
  - Limited use of help pages: monthly respondents more used to the survey?
  - Consider adding a help page for irregular reporting period cases.

- Edits
  - Consider loosening the edit boundaries for key questions.
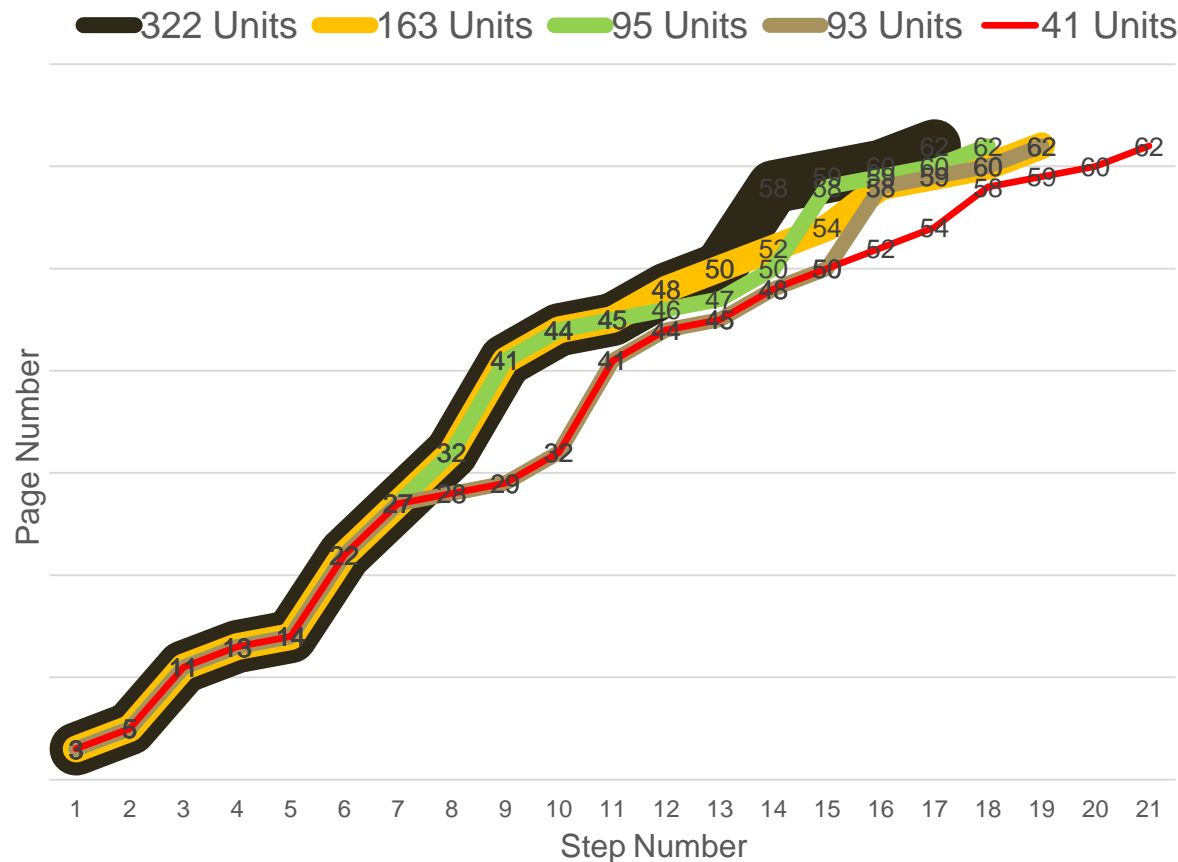
Other uses of paradata at Statistics Canada

Most Frequent Flow Patterns

Legend: 322 Units, 163 Units, 95 Units, 93 Units, 41 Units

Y-axis: Page Number
X-axis: Step Number

Source: Jessica Andrews, Statistics Canada
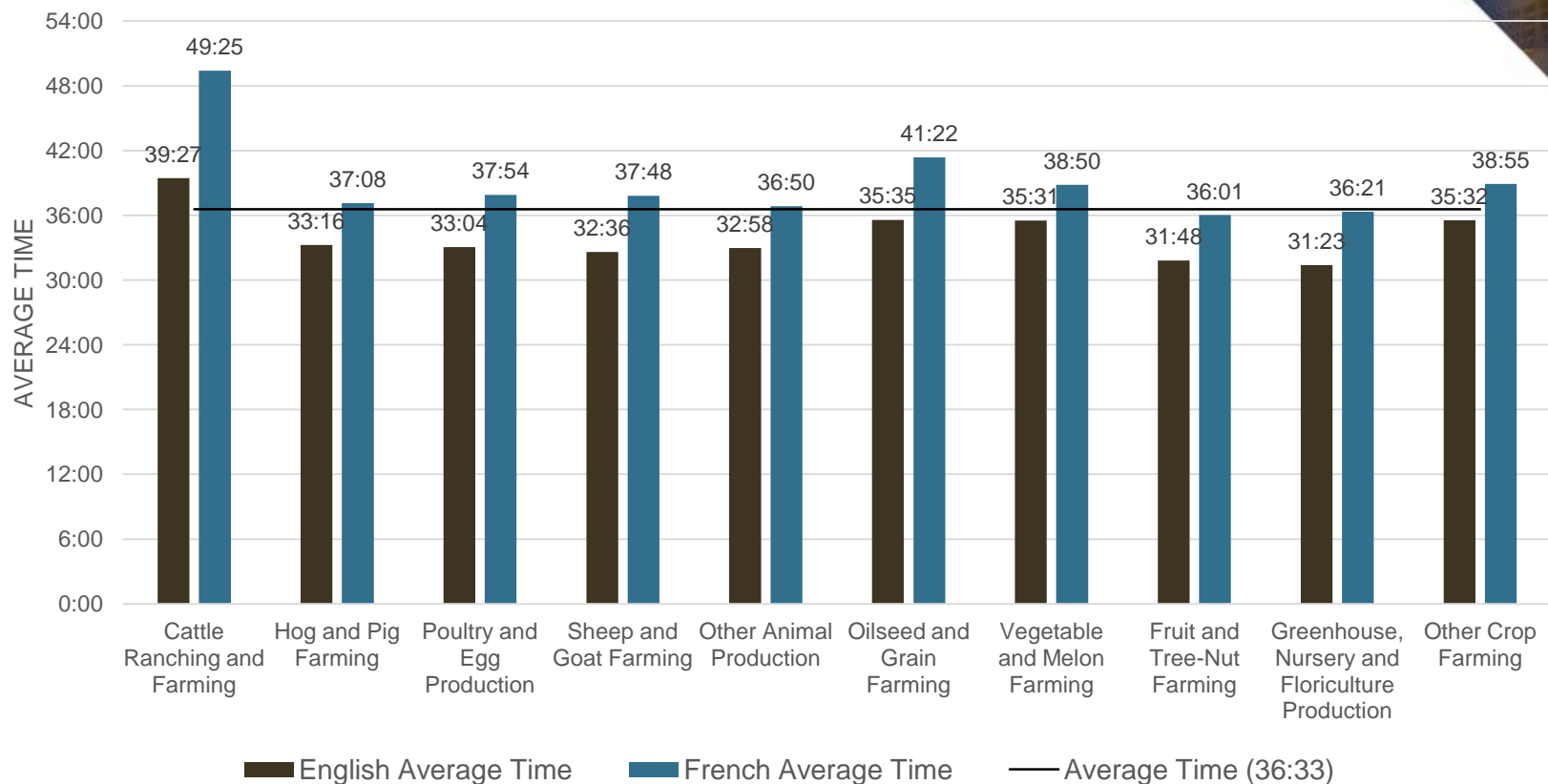
13

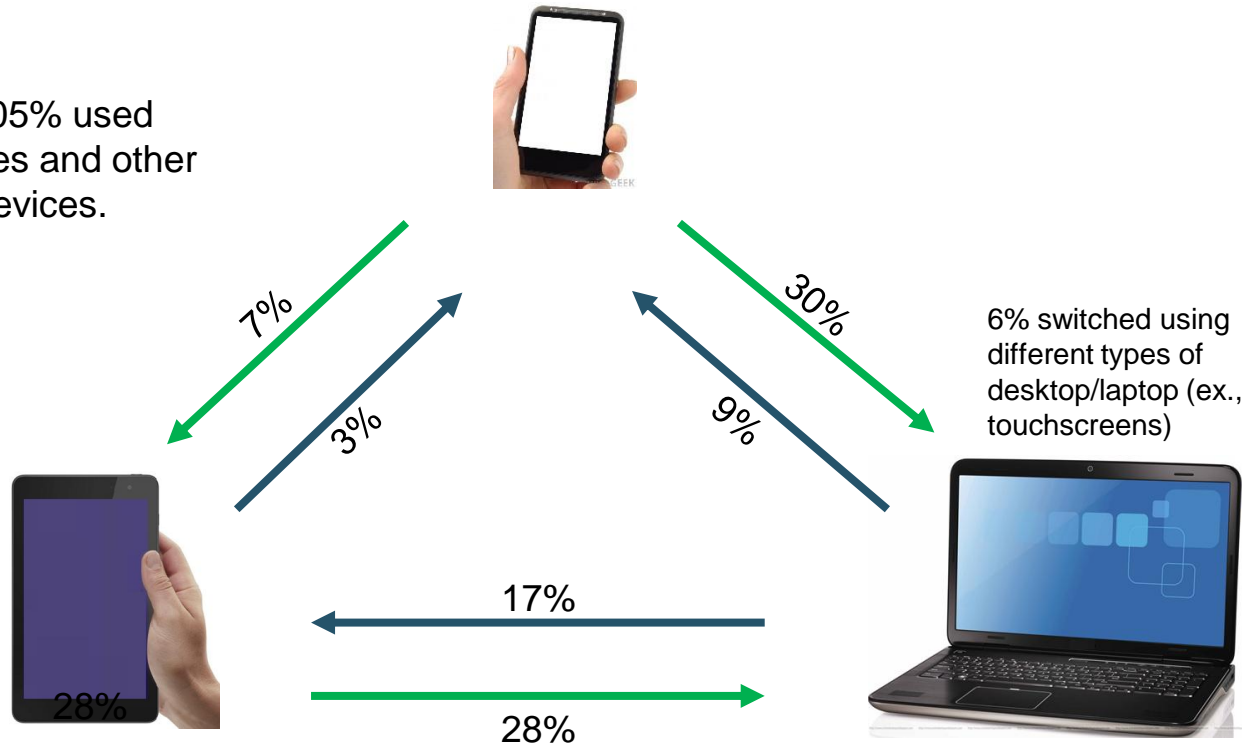# Time per Questionnaire (REQ) by Farm Type and Response Language

Source: Jessica Andrews, Statistics Canada

14

Switching Devices: Long Form Census = 1.51%

- People were more likely to switch from a smaller device to a larger device.

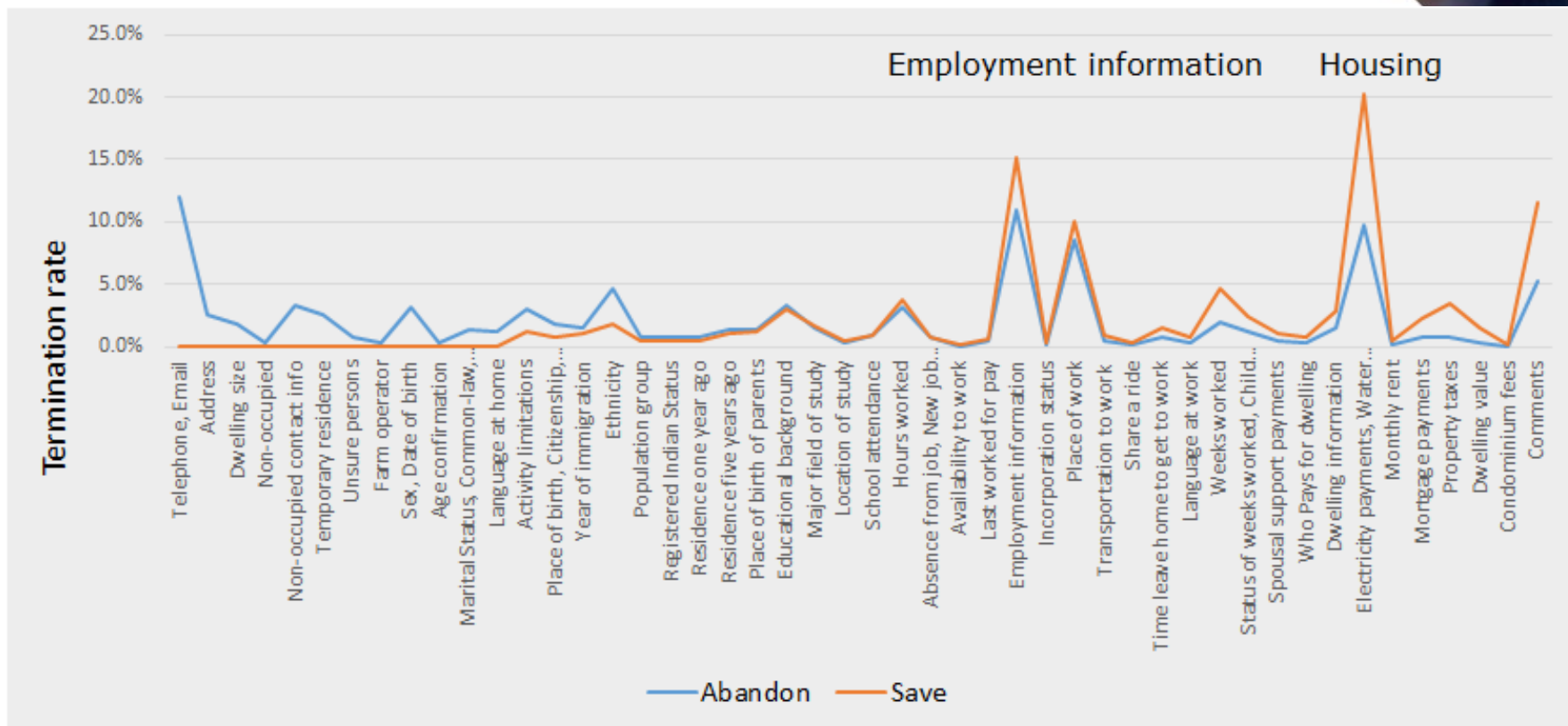*Less than 0.05% used gaming devices and other unidentified devices.

7%

3%

30%

9%

6% switched using different types of desktop/laptop (ex., touchscreens)

17%

28%

28%

Source: Anthony Bremner, Statistics Canada

Pages where respondents abandoned or saved (Long-form)

- People most often abandoned or saved the questionnaire when asked about Employment or Housing (bills).

Source: Anthony Bremner, Statistics Canada

New directions for paradata use at Statistics Canada

Improving Collection Strategies

Optimizing reminders:

- Does day of week or time of day matter for reminder?

- What is optimal number of reminders?

- What is optimal time between reminders?

Predicting response to surveys using web logs and call history

- Aim is to build a model that can be used to predict if a respondent will reply by a certain point in collection

- Plan would be to move survey units who were predicted to respond lower on the list of prioritization for follow up calls

Question

- What has been your experience using web questionnaire paradata?

- What new directions are you exploring that we can learn from?

THANK YOU!

Réjean Doiron
Assistant Director, Collection Planning and Research Division
Statistics Canada
rejean.doiron@canada.ca

For more information, please visit
www.statcan.gc.ca

#StatCan100