

The Role of Paradata in Measuring and Reducing Measurement Error in Surveys

Mick P. Couper, PhD
(Survey Research Center, University of Michigan)

NCRM Network for Methodological Innovation 2009:
The Use of Paradata in UK Social Surveys

[Paradata is ... or paradata are?](#)

UNIVERSITY OF MICHIGAN

Introduction

- An early interest in paradata focused on the keystroke files, trace files, or audit trails generated by many CAI systems
- Automatically generated as a byproduct of the system, used for technical purposes (error diagnosis, recovery from failure, etc.)
- Early analyses suggest paradata might provide insight into the question asking and answering process, and how interviewers or respondents interact with the CAI systems
- From these humble beginnings, a cottage industry has grown up around enhancing the capture, processing, and analysis of paradata focusing on measurement error

UNIVERSITY OF MICHIGAN

2

Uses of Paradata I

- Paradata are being used to evaluate many different parts of the process
 - Interviewers
 - Respondents
 - Questions
 - Instruments/systems
- Paradata are being used in different modes of data collection
 - Interviewer-administered (e.g., CATI, CAPI)
 - Self-administered (e.g., CASI, audio-CASI, Web, IVR)

Uses of Paradata II

- Analysis of paradata supplements, rather than replaces, other methods of process or data quality evaluation
- More useful at or near the end of the process than the beginning
 - Post-survey evaluation rather than development
 - Fits within a continuous quality improvement framework
- Paradata typically identify where problems occur but not why
 - Indirect measures of what happened
- Useful in identifying areas to focus attention using more expensive and time-consuming methods
 - Real-time monitoring
 - Digital recording and behavior coding
 - Cognitive testing, usability evaluations
 - Laboratory or field experiments
 - Eye-tracking studies
 - Etc.

Advantages and Disadvantages of Paradata

■ Advantages

- Cheap – automatically generated by software, or added with small modifications
- Wide – available on all cases at varying levels of granularity

■ Disadvantages

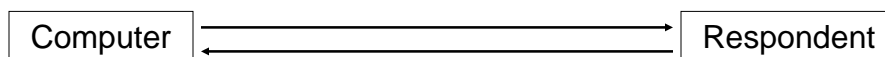
- Difficult to analyze – data not intended for analysis; non-rectangular; messy
- Thin – record of actions and events using the software, not thoughts, intentions, words, etc.

Interviewer- and Self-Administration

■ Interviewer-administered surveys (CAI)



■ Self-administered surveys (CSAQ)



- Paradata may thus be more useful in self-administered modes

Early Examples of Paradata Analysis in CAI

- Couper, Hansen, and Sadosky (1997)
 - Evaluation of interviewer performance in a CAPI survey
- Couper, Horm, and Schlegel (1997) and Couper and Schlegel (1998)
 - Evaluation of NHIS CAPI instrument
- Caspar and Couper (1997)
 - Evaluation of respondent use of audio-CASI

Recent Examples of Paradata Use in CAI

- Couper, Tourangeau, and Marvin (2009)
 - NSFG audio-CASI
- Couper and Kreuter (work in progress)
 - NSFG Cycle 6

Paradata Use in Web Surveys

- Because the respondent interacts directly with the computer—without the interviewer as intermediary—the paradata give us more insight into what they are doing
- Paradata analysis has been widely used in Web surveys to explore a number of design issues

Types of Paradata in Web Surveys

- User metrics or browser characteristics
- Server-side paradata
 - Captures submissions to server
 - Page-level times and breakoffs
- Client-side paradata
 - Embedded JavaScript code captures user actions
 - Item-level times, response latencies, changed answers, etc.
- Mouse movement
 - Embedded code captures mouse position every n-th millisecond

Examples of Paradata Use in Web Surveys

- Breakoffs
- Browser dimensions
- Response order effects
- Evaluation of grid design
- Navigation in long lists
- Use of visual analog scales

Summary on Web Paradata

- Both server-side and client-side paradata are widely used in Web surveys
- Technical challenges associated with client-side paradata
- As with all types of paradata, analysis is messy
- Ethical issues related to use of client-side paradata to evaluate respondent behavior
 - Should we inform respondents of the capture and use of these paradata?

General Summary

- The use of paradata for the evaluation and reduction of measurement error lags behind that of paradata for nonresponse error
 - Some may think of the instrument as more “fixed” and less subject to “responsive” or “adaptive” design views
- Similarly, paradata analysis for Web surveys – where respondents interact directly with the software – is further ahead than that for interviewer-administered surveys
- Paradata already being used in a number of ways in practice
- Work remains to demonstrate the utility of analyzing paradata for measurement error in real time
 - Management of ongoing projects
 - Continuous quality improvement perspective

Thank You

ESRC National Centre for
Research
Methods