CENTRE
FOR SOCIAL
INFORMATICS

University of Ljubljana
Faculty of Social Sciences

Centre for Social Informatics

Faculty of Social Sciences

University of Ljubljana

Direct Paradata Usage for Analysis of Response Quality, Respondent Characteristics, and Survey Estimates

State-of-the-Art Review and Typology of Paradata

Vasja Vehovar, Gregor Čehovin

Ljubljana, June 2023

Direct Paradata Usage for Analysis of Response Quality, Respondent Characteristics, and Survey Estimates: State-of-the-Art Review and Typology of Paradata

Vasja Vehovar [a], Gregor Čehovin [a]

[a] Centre for Social Informatics, Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia. Email: vasja.vehovar@fdv.uni-lj.si, gregor.cehovin@fdv.uni-lj.si.

Funding

This work was supported by the Slovenian Research Agency [grant numbers P5-0399, J5-9334, J5-8233 and NI-0004].

Abstract

With the proliferation of web surveys, researchers have increasingly used paradata, such as response times or mouse movements, which are collected automatically during the response process or with external software. While paradata have allowed researchers to identify important sources of survey error, they remain underutilized as a comprehensive resource for analyzing a wide range of methodological and substantive indicators. To highlight this problem, this paper provides a state-of-the-art review and summarizes empirical studies on three ways paradata are used: (i) to predict differences in response quality, (ii) to predict differences in respondent characteristics (e.g., sociodemographics, personality traits, and emotions), and (iii) to predict differences in survey estimates (e.g., values and attitudes). The results point to certain blind spots in survey research: The majority of studies used paradata to examine only (i) response quality, and many studies used only a single paradata indicator (i.e., response times). The main opportunity for future research lies in using multiple paradata indicators to simultaneously examine response quality, respondent characteristics, and estimates. This paper also provides a typology of available paradata in web surveys that could be used as an operational framework in survey research.

Keywords: web surveys, paradata, response quality, data quality, respondent characteristics, survey estimates, state-of-the-art review

1 Introduction

Paradata are a relatively new and expanding source of information about the survey response process and provide data about the computer-human interaction in surveys, such as mouse movements, keystrokes and response times. Paradata are particularly valuable for investigating web survey response quality and also enable insights that range from errors and costs to web survey usability (Callegaro et al., 2015; Kreuter, 2013a). The expansion and importance of paradata has been amplified due to the proliferation of web surveys (ESOMAR, 2020) and the continued application of the Total Survey Error (TSE) paradigm in survey methodology (e.g., Biemer et al., 2017; Groves & Lyberg, 2010). On the other side, the complexity of paradata processing requires considerable resources (Kunz & Hadler, 2020), which is also an important barrier to their usage.

This paper is concerned with *direct paradata* (hereafter referred to as *paradata*), which are automatically collected using the web survey software or external software (see Callegaro et al., 2015) and can be divided into three categories. First, *contact info paradata* provide information about respondents' contact attempts, such as the outcome of email invitations and whether respondents accessed the survey introduction page. Next, *device type paradata* comprise information about the respondent's device, such as its type, operating system, and IP-based location. Third, *questionnaire navigation paradata* describe respondents' progress through the questionnaire with information about response times (e.g., per item and per page), whether the respondent clicked outside the questionnaire (i.e., focus-out events), keystrokes, mouse clicks, and touch input (Callegaro et al., 2015).

The purpose of this paper is a state-of-the-art review of research in survey methodology, in which paradata have been used that are particularly promising and fruitful for investigating the causes of survey errors. In addition, the aim is to present a typology of available paradata in web surveys. Three major areas of survey methodology in which paradata are used are examined: (i) response quality (e.g., speeding), (ii) respondent characteristics (e.g., personality traits), and (iii) survey estimates (e.g., specific target variables).

While various indicators of (i) *response quality* have been used in the past to identify respondents with low response quality (Callegaro et al., 2015), paradata are also a valuable source for uncovering response behaviors that would otherwise have gone unnoticed, such as speeding (e.g., Greszki et al., 2015). Studies have used paradata to predict (ii) *respondent characteristics*, such as emotions (e.g., Yamauchi & Xiao, 2018), health (e.g., Seelye et al., 2015) and personality traits (e.g., Cheng et al., 2020). Paradata have also been used to investigate response quality in relation to some respondent characteristics, such as personality (e.g., Bowling et al., 2016) and sociodemographics (e.g., Höhne, Schlosser, et al., 2020). Studies have used paradata to analyze differences in (iii) *estimates*, such as attitudes (e.g., Schneider et al., 2015), behavioral attributes (e.g., Tzafilkou & Nicolaos, 2018) and behaviors (e.g., Greszki et al., 2015).

This paper does not address *indirect paradata* that require additional external devices (e.g., eye-tracking, brain wave monitoring) or external observation (e.g., behavior coding). In addition, paradata related to mixed-mode surveys are not addressed here. These include multimode contact paradata, interviewer characteristics, interviewer observations, and audio and video recordings (Kreuter, 2013b). While longitudinal or modular survey designs may rely on paradata defined in a previous wave, data that exist prior to data collection for the web survey (e.g., sampling frame information) are excluded from the paradata definition. In addition, the paradata definition excludes auxiliary data and digital traces about decisions under the researcher's control regarding the details of the survey design before the survey (e.g., changes to the questionnaire) and after the survey (e.g., changes to responses), as well as auxiliary paradata generated outside of the data collection process (e.g., passive or commercial sources) (Callegaro et al., 2015; McClain et al., 2019).

Following a conceptual background on paradata, a typology of available paradata in web surveys is provided in the next section, which can serve as an operational framework for future studies. The empirical part of this paper examines whether multiple paradata areas are covered by research and whether the use of paradata has proven useful for survey research. The discussion and conclusion sections address insights that paradata can realistically provide so that researchers can avoid unattainable expectations in survey practice.

2 Background

In survey data collection, researchers dispose with (i) *questionnaire data*, e.g., substantive responses and records of missing data, (ii) *auxiliary data*, e.g., characteristics of the respondents obtained from a sampling frame or another external source, (iii) *metadata* that provide macro-level information about questionnaire data, e.g., information about the period of data collection and question measurement scales, (iv) *passive data* or non-reactive data that describe behavior beyond the survey data collection, e.g., motion data and web tracking data, and (v) *paradata* that analyze respondent's interaction with the web survey and provide insights about the survey data collection process and the respondents' behavior, ranging from errors and costs to web survey usability (Callegaro et al., 2015; Kunz & Hadler, 2020). Paradata are very rich, can be captured at a low cost, and can be recorded at the server of the web questionnaire, e.g., to describe the respondent's navigation through the questionnaire, or at the respondent's device, e.g., to record keystrokes and mouse clicks (Callegaro et al., 2015; McClain et al., 2019)**.**

This paper builds on typologies of paradata provided by Callegaro et al. (2015) and McClain et al. (2019). Under these frameworks, paradata can be structured into three main conceptual types of paradata in web surveys. First, *contact info paradata* describe the behavior surrounding survey contact attempts and the respondent's contact with the survey team (e.g., the timing and outcome of email invitations). For nonprobability convenience methods, paradata may describe behavior at the recruitment stream (e.g., a recruitment website or a list of contacts). Second, *device type paradata* are aggregated to the session level and describe the behavior surrounding one or more attempts of the recruited units to access the web survey (e.g., time elapsed from opening the invitation to accessing the survey and the number of access attempts). In mixed-mode surveys, the survey may be accessed through multiple channels.

Third, *questionnaire navigation* paradata are collected at the action level and aggregated to the level of the survey page, survey session, and respondent. These paradata can range from the most granular (e.g., keystrokes, clicks, and mouse movements), through the more processed ones (e.g., response times), to those specific to the survey instrument design (e.g., validation triggers, focus-out events). Besides the above-described three types of paradata, McClain et al. (2019) also define *prior survey paradata* that comprise various types of paradata aggregated to the respondent level and describe behaviors specific to online panels (e.g., device use, missing data, and response speed in previous survey waves) with the aim to inform survey design and help predict response quality in multi-wave surveys (e.g. Roßmann & Gummer, 2016).

After the paradata are captured, a considerable amount of processing is usually needed to aggregate the paradata to the appropriate level and create indicators, which can inform analyses on response quality and survey errors (McClain et al., 2019). The aggregation of these paradata can be structured into four levels (see Kaczmirek, 2009). The first level of aggregation includes data on individual actions of respondents (e.g., time stamps, mouse clicks) and survey invitation data (e.g., time stamps, invitation content). Typically, these data are recorded sequentially in a non-rectangular dataset, as the number of individual actions (e.g., on a screen) cannot be predetermined. The second level of aggregation comprises first level data across individual actions of each respondent (e.g., the number of mouse clicks on a page), which can be structured into a rectangular dataset. The third level includes second level data aggregated across respondents or variables (e.g., the mean number of answer changes per respondent or per question). The fourth level of aggregation is done across respondents and variables, providing a single value per survey (e.g., response time, item nonresponse rate).

The TSE framework has underlined paradata as a method of investigating the causes of survey errors and the quality of survey estimates (McClain et al., 2019). The prior survey paradata can be used to analyze nonresponse error; however, in panel surveys these paradata can be used to investigate measurement error over time. The contact info paradata enable the analysis of coverage error, while in combination with the prior survey paradata researchers can explore additional ways of reducing nonresponse error. The device type paradata provide insight into nonresponse error (e.g., due to breakoffs after unsuccessful access attempts) and measurement error (e.g., due to device type) (ibid.). Questionnaire navigation paradata enable the investigation of nonresponse and measurement error through different indicators of respondent behavior and response quality (e.g., Höhne et al., 2017; Sendelbah et al., 2016), while some studies have aimed to also predict breakoffs with these paradata (e.g., Mittereder, 2019).

3 Method

3.1 Research questions

In providing a typology of available paradata, this paper addresses two research questions:

RQ.1: Which types of paradata do researchers use in web surveys?

RQ.2: What types of indicators can researchers create from direct paradata?

The following research question is addressed in this paper's state-of-the-art review:
RQ.3: In which of the three major paradata areas (i.e., respondent quality, respondent characteristics, and estimates) were paradata used by previous studies in web survey methodology?

3.2 Literature search strategy for state-of-the-art review

The literature search was performed in between April and January 2022 based on three eligibility criteria:

1. an empirical study used paradata in some relation to web surveys;
2. the paradata were used to analyze response quality and/or respondent characteristics and/or survey estimates;
3. the basic web survey mode (Callegaro et al., 2015)—which uses a computerized self-administered questionnaire to collect responses on a server which the respondents access with a web browser—was used in relation to paradata about the respondent's interaction with the computerized questionnaire.

Four combinations of keywords were used: paradata in web surveys, paradata and response quality, paradata and respondent characteristics, and paradata and survey estimates. For each of the four keyword combinations, the approximately the first 75 results were retrieved, totaling to 301 results for the four iterations. In addition, reference lists of the eligible manuscripts were screened, and 589 results were identified on this basis in total. After removing duplicate results, 570 results remained. The literature selection process (shown in Figure 1) was performed on the basis of the eligibility criteria. The titles and abstracts of 570 results were checked for eligibility. On this basis, 467 results were excluded, and 103 results were used for further evaluation. Full-text manuscripts of the remaining 103 results were obtained and evaluated according to eligibility criteria. A total of 57 manuscripts were found eligible and included in the review, while 46 manuscripts were excluded.

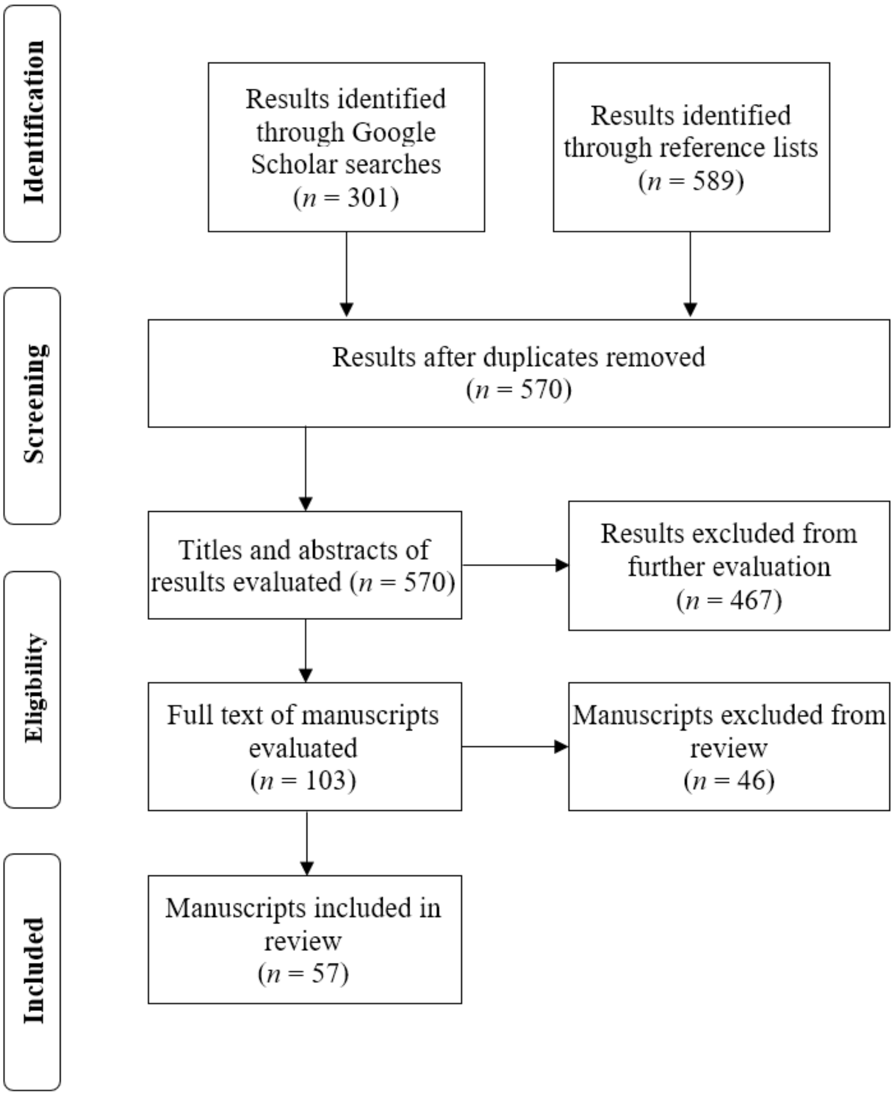

Figure 1: PRISMA diagram of the literature selection process

4 Paradata Typology

Using paradata generated by the respondent's interaction with the survey instrument, this paper identifies a set of key paradata about the respondent's device, the patterns of respondent's inputs and the patterns of respondent's navigation through the questionnaire.

Table 1: Types of paradata

Paradata type	Examples	TSE areas
Contact info paradata (recruitment phase)	- Outcomes of an email invitation - Access to the questionnaire introduction page - Last question answered before breakoff - Indicator and timing of e-mail opening - Times/dates of contacts - Recruitment stream (river/intercept samples)	- nonresponse error - measurement error in panels over time
Device type paradata (access phase)	- User agent string - Recruitment stream (river/intercept samples) - Cookie and IP recordings - Time from contact to access - Device characteristics (type, screen dimensions, resolution and orientation, etc.) - Login attempts (total count, count of successes, etc.)	- nonresponse error - measurement error
Questionnaire navigation paradata (response phase)	- Time spent per screen or response latency - Keystrokes and mouse-clicks - Change of answers - Real-time validation messages - Response times - Item missing data - Navigation (e.g., backups) and response changes - Loss of focus/out of browser window - Mouse, touch and keyboard actions	- nonresponse error - measurement error

5 Findings of the state-of-the-art review

Table 2 lists the 57 eligible manuscripts according to the paradata that were used and a summary of the findings. In addition,

Table 2 shows whether paradata were used to predict differences in response quality, respondent characteristics, and survey estimates.

Table 2: Studies using paradata to analyze response quality, respondent characteristics and survey estimates

#	Reference	Published	Paradata	RQ[(a)]	RC[(b)]	E[(c)]	Findings with paradata
1	(Fernández-Fontelo et al., 2022)	2022	- mouse actions	x			Mouse movement curves as a whole contain more information for prediction of question difficulty than mouse movement measures alone.
2	(Cepeda et al., 2021)	2021	- mouse actions - response time	x	x		Interactive pointer movements provide a promising approach of better understanding of the relationship between the dynamics of pointer movements and underlying perceptual, cognitive and affective psychological mechanisms.

#	Reference	Published	Paradata	RQ[a]	RC[b]	E[c]	Findings with paradata
3	(Fernández-Fontelo et al., 2021)	2021	- response time - mouse actions	x			Mouse movement paradata improves prediction of question difficulty compared to response-time-only models.
4	(Gummer et al., 2021)	2021	- response time	x			Instructed response items identify respondents who show an elevated use of straightlining, speeding, item nonresponse, inconsistent answers, and implausible statements throughout a survey
5	(Matjašič et al., 2021)	2021	- response time - device characteristics	x			Moderate and statistically significant correlation between RTs and share of low RQ units, number of RQ indicators failed and each of the eight RQ indicators. Quality of responses to open-ended question is the most suitable and sufficient response quality indicator for identifying speeders.
6	(Schroeders et al., 2021)	2021	- response time	x			Comparison between the results of the simulation and the online study showed that responses in real-world settings seem to be much more erratic than can be expected from the simulation studies.
7	(Cheng et al., 2020)	2020	- outcomes of reminders	x	x		Conscientiousness and openness to experience predict the incidence of unit nonresponse in subsequent survey waves, even after controlling for cognitive ability and demographic characteristics
8	(Höhne, Revilla, et al., 2020)	2020	- respondent motion - response time	x			Compared respondents' acceptance to comply with standing at a fixed point or walking around. The quality of responses to the five single questions did not differ significantly between the motion conditions
9	(Höhne, Schlosser, et al., 2020)	2020	- focus-out	x	x		On-device media multitasking is more common on PCs than on smartphones. Engaging in on-device media multitasking affects response quality.

#	Reference	Published	Paradata	RQ[(a)]	RC[(b)]	E[(c)]	Findings with paradata
10	(Horwitz et al., 2020)	2020	- mouse actions	x			Unordered response options, yes/no response format, and complex response options resulted in higher mouse movement values.
11	(Höhne & Schlosser, 2019)	2019	- respondent motion - response time	x			Respondents' motion level affects response times and the quality of responses given
12	(Sturgis et al., 2019)	2019	- response time	x	x		Lower satisficing propensity for higher conscientiousness, higher agreeableness, and higher response time
13	(Kühne & Kroh, 2018)	2018	- mouse actions - answer changes	x			The provision of feedback did not lead to respondents changing their answers. Only 0.24% of respondents used the back button to revise personality answers.
14	(Paas & Morren, 2018)	2018	- response time	x			The study 1 results show that IMC-failers often dedicate less time to the survey. Study 2 results show that attention-enhancing tools do not reduce IMC failure.
15[(d)]	(Tzafilkou & Nicolaos, 2018)	2018	- keyboard actions - mouse actions			x	Mouse patterns can be associated with perceived ease use, perceived usefulness, self-efficacy, willingness to learn or risk-perception. Some keystroke dynamics like can be associated with perceived ease of use or self-efficacy.
16	(Yamauchi & Xiao, 2018)	2018	- mouse actions - response time		x		Cursor movement patterns such as the area under curve and direction change help infer emotions of computer users.
17	(Andersen & Mayerl, 2017)	2017	- response time	x		x	Faster responses are associated with the reporting of desirable attitudes and behavior while slower responses are linked with those that are undesirable.
18	(Conrad et al., 2017)	2017	- response time	x	x		In addition to reducing speeding, interactive prompts increased response accuracy. Prompting also reduced later straightlining in one experiment.
19	(Hibbeln et al., 2017)	2017	- mouse actions		x		Mouse cursor distance and speed can be used to infer the level of negative emotion.

#	Reference	Published	Paradata	RQ[(a)]	RC[(b)]	E[(c)]	Findings with paradata
20	(Horwitz et al., 2017)	2017	- mouse actions - response time	x			Certain mouse movements are highly predictive of difficulty and add considerable value when used in conjunction with response times.
21	(Bowling et al., 2016)	2016	- response time	x	x		Insufficient Effort Responding is in part a manifestation of enduring individual differences.
22	(Roßmann & Gummer, 2016)	2016	- response time - device characteristics - previous survey participation	x			Some paradata (e.g., response times and participation history) improved the prediction of panel attrition, whereas others did not.
23	(Sendelbah et al., 2016)	2016	- response time - focus-out	x			Item nonresponse is associated with respondent multitasking, while non-differentiation is not. Overall, 60% of respondents have multitasked at least once.
24	(Greszki et al., 2015)	2015	- response time	x		x	Speeding affects substantive results only in some exceptional instances, if at all. Marginal distributions were not altered substantively.
25	(Gummer & Roßmann, 2015)	2015	- response time	x	x		Factors on both levels, the individual and the survey level, had effects on interview duration.
26	(Huang et al., 2015)	2015	- response time	x			The shortened infrequency scale was significantly correlated with total survey time and letters typed while the two objective outcomes of response effort were positively associated.
27	(Jenkins et al., 2015)	2015	- response time - mouse actions	x			Response uncertainty significantly influences mouse movement (area-under-the-curve). Low uncertainty and high uncertainty had significantly lower response times, suggesting that these respondent groups are fast because they know the answer or are just randomly guessing.
28	(Revilla & Ochoa, 2015)	2015	- response time	x			Significant relationship between response times and RQ. Longer response times associated with higher self-reported effort. RQ not linked to effort.

#	Reference	Published	Paradata	RQ[a]	RC[b]	E[c]	Findings with paradata
29	(Schneider et al., 2015)	2015	- response time - mouse actions			x	Mouse trajectory pull of the non-chosen evaluative option was mostly positively correlated with self-reports of subjective ambivalence.
30[d]	(Seelye et al., 2015)	2015	- mouse actions		x		Mild cognitive impairment was associated with making significantly fewer total mouse moves. Mouse movement measures were significantly associated with several cognitive domains.
31	(Maniaci & Rogge, 2014)	2014	- response time	x			Inattentive respondents provided self-report data of markedly poorer quality, sufficient to obscure meaningful regression results as well as the effects of experimental manipulations.
32	(Zhang & Conrad, 2014)	2014	- response time	x	x		Tendency to speed is related to several respondent characteristics, particularly age (younger respondents are more likely to speed). More speeding seems to be universally related to more straightlining, and this relationship is particularly strong among the less educated respondents.
33	(Horwitz et al., 2013)	2013	- response time - number of sessions - answer changes - validation prompts - device characteristics	x			Respondents who changed their answers took longer to respond.
34	(Huang et al., 2012)	2012	- response time	x			Effectiveness of four indices for detecting IE responses: response time, long string, psychometric antonyms, and individual reliability coefficients. Detection indices measured the same underlying construct and showed the improvement of psychometric properties after removing IE respondents.

#	Reference	Published	Paradata	RQ[a]	RC[b]	E[c]	Findings with paradata
35	(Meade & Craig, 2012)	2012	- response time	x			Response time failed to load onto a factor, potentially due to the hypothesized nonlinear relationship between time and response quality. Results indicated that there are two distinct patterns of careless response (random and non-random) and that different indices are needed to identify these different response patterns.
36	(Funke et al., 2011)	2011	- response time	x			Higher response times and a higher break-off rate with respondents with a low formal education. Slider scales lead to statistically significantly higher break-off rates and substantially higher response times.
37	(Gutierrez et al., 2011)	2011	- response time	x		x	10% of those completing the survey flagged as speeders. Speeders less likely to report owning a battery charger among 20 check boxes. Speeding did not affect survey estimates.
38	(Lenzner et al., 2010)	2010	- response time	x			Respondents answering the suboptimal questions had longer response times. The overall effect of text features on total response times was highly significant. Data quality was only partially found to be affected by the text features.
39	(Stieger & Reips, 2010)	2010	- mouse actions - keyboard actions - response time - device characteristics	x			10.5% of participants showed more than five single behaviors (mouse and keyboard) with highly negative influence on data quality in the whole online questionnaire (out of 132 possible single behavior judgments)
40	(Wells et al., 2010)	2010		x	x		Spanish-language panelists took from 75% to 150% longer to complete their online surveys. Education, age, and broadband internet connection also have large and statistically significant effects on response time.
41	(Callegaro et al., 2009)	2009	- response time	x			Respondents who engage in satisficing spend less time answering questions.

#	Reference	Published	Paradata	RQ[a]	RC[b]	E[c]	Findings with paradata
42	(Galesic & Bosnjak, 2009)	2009	- response time	x			The further away from the beginning a block of questions was asked, the shorter were response times and the lower was the variability of answers to questions in grids.
43	(Kaczmirek, 2009)	2009	- mouse actions - response time - device characteristics	x			- The number of respondents who clicked more than necessary was between 35% and 46%. - Answer time is better suited for the detection of differences in experimental conditions than page time and server time. - The least common denominator in available technology on client-side does not necessarily need to be pure HTML-code.
44	(Malhotra, 2008)	2008	- response time - response order	x	x		Low-education respondents who were the fastest were most prone to primacy effects for unipolar rating scales. Age and education are statistically significant predictors of response time.
45	(Stern, 2008)	2008	- response time - response order - answer changes	x			Visual layout of survey questions affects the number and also the types of response changes. For long-list questions, grouped listings have the longest response time, followed by an alphabetical list, while the fastest response times were for an open-ended input.
46	(Yan & Tourangeau, 2008)	2008	- response time	x	x		Response times are affected by question characteristics and respondent characteristics
47	(Conrad et al., 2007)	2007	- response time	x			Respondents answered particularly accurately with mixed-initiative clarification, where respondents could initiate clarification or the system could provide unsolicited clarification when respondents took too long to answer. Although clarification dialogue increased response times, respondents preferred being able to request clarification than not.

#	Reference	Published	Paradata	RQ[(a)]	RC[(b)]	E[(c)]	Findings with paradata
48	(Healey, 2007)	2007	- response time - mouse actions	x			Drop downs led to higher item nonresponse and longer response times. The 76% of respondents using scroll mice to complete the survey were prone to accidentally changing an answer if presented with drop-down questions.
49	(Conrad et al., 2006)	2006	- response time - mouse actions	x			The number of clarification requests is quite sensitive to the number of clicks involved. Response times were faster when definitions were obtained with one click than with multiple clicks. For the mouse-over design there were more requests for clarification (22.4%) than for the click design (13.8%).
50	(Smyth et al., 2006)	2006	- response time	x			Respondents endorse more options and take longer to answer in the forced-choice format than in the check-all format. The check-all format may encourage a weak satisficing response strategy.
51	(Haraldsen et al., 2005)	2005	- response time - response changes	x			Longer response times for open-ended questions. More answer changes for questions requiring judgement and multiple-choice questions.
52	(Wise & Kong, 2005)	2005	- response time	x			Psychometric characteristics of RTE scores were empirically investigated and supportive evidence for score reliability and validity was found. In addition, both RTE scores and self-reported effort showed very similar motivation filtering effects.
53	(Tourangeau et al., 2004)	2004	- response time	x			Inconsistent order of response options resulted in higher response times. Presenting grid items on a single screen rather than multiple screens resulted in lower response times.

#	Reference	Published	Paradata	RQ[(a)]	RC[(b)]	E[(c)]	Findings with paradata
54	(Heerwegh, 2003)	2003	- response time - response changes	x			Respondents who changed their answers took longer to respond. Respondents who did not know the answer to knowledge questions took longer to respond.
55	(Heerwegh & Loosveldt, 2002)	2002	- response time	x			Respondents took longer to answer with dropdowns that with radio buttons in one of the two experiments.
56	(Heerwegh, 2002)	2002	- response time - response changes - response order	x			About 30% of respondents changed their answer at least once in a grid of 8 items. About 10% of respondents did not respond to the 8 items sequentially. Responses to the first item in a grid take longer. Dropdowns produce less answer changes, less non-sequential responses, and longer response times.
57	(Crawford et al., 2001)	2001	- response time	x			There was no association between the burden indicator (8 to 10 minutes vs. 20 minutes) and time before breakoff.
	Total			**52**	**13**	**6**	

[(a)] Response quality column comprises studies that used paradata to examine specific response quality indicators (e.g., item nonresponse, satisficing) and/or characteristics of the responding process (e.g., response uncertainty, response difficulty).

[(b)] Respondent characteristics column comprises studies that investigated the relation of paradata to sociodemographic characteristics and other personality, psychological or cognitive measures (e.g., Big Five personality dimensions).

[(c)] Estimates column comprises that examined paradata in relation to scores of some attitudinal or opinion scores (e.g., risk perception, willingness to learn). Studies that analyzed self-reports related directly to response quality (e.g., self-reported multitasking) are not included in this column (they are already reported in the response quality column).

[(d)] The study initially collected paradata through computer use of the participants, followed by the collection of survey data through a computerized survey.

5.1 Paradata and response quality

Response time paradata were the most common set of paradata indicators, which the literature utilized to analyze response quality. Response time paradata refer to the time that the respondents requires to answer an item, a set of items or the entire questionnaire (Matjašič et al., 2018). In this context, the literature reported that unusually short (i.e., speeding) or unusually long response times (e.g., due to focus-out events) indicate worse response quality (e.g., Revilla & Ochoa, 2015; Stieger & Reips, 2010). It was found that response times are affected by question type and characteristics (e.g., Funke et al., 2011). Open-ended questions were identified as the most suitable and sufficient response quality indicator for identifying speeders (Matjašič et al., 2021).

Mouse actions are another notable set of paradata indicators. While mouse movement have been previously studies with eye tracking (e.g., Galesic et al., 2008), this requires special equipment and physical presence of the respondent, while the paradata approach is a low-cost and scalable alternative (Cepeda et al., 2021). Here, simple mouse action paradata (e.g., pointer movements and mouse clicks) have been used as indicators of response quality (e.g., Horwitz et al., 2020; Stieger & Reips, 2010), where excessive movements and clicking exhibited a negative influence of response quality. More complex paradata have in addition been used to extract specific mouse patterns (e.g., Tzafilkou & Nicolaos, 2018) and to predict respondent difficulty using machine learning (e.g., Fernández-Fontelo et al., 2021).

Studies have used focus-out event paradata—which indicate that the respondent switched from the web browser’s tab or window to another activity—to detect respondent multitasking. Focus-out events were found to be statistically significantly associated with item nonresponse (Sendelbah et al., 2016).

On smartphones, response quality has been analyzed in relation to respondent motion (e.g., walking), which was monitored using acceleration paradata. A study that included walking and climbing the stairs found that a high motion level resulted in longer completion times and a more pronounced primacy effect (Höhne & Schlosser, 2019). Conversely, a study where the motion condition only included walking found no statistically significant different in response quality between the different levels of respondent motion (Höhne, Revilla, et al., 2020).

Among other less frequently used paradata indicators to study response quality is prior survey participation. A split panel study (Roßmann & Gummer, 2016) found that some paradata were robust predictors of attrition, namely the number of item nonresponses, the response speed index, and the respondents’ participation history. On the other hand, results were less clear for the usage of tablets and smartphones, the number of correctly answered control questions, the number of “don’t know” answers.

5.2 Paradata and respondent characteristics

Few studies used paradata to analyze only respondent characteristics, as analyzing both response quality and respondent characteristics was more prevalent in the literature (see Section 5.4). In a study addressing cognitively intact older adults and older adults with mild cognitive impairment, mouse movements were captured during one week of routine home computer use (Seelye et al., 2015). Participants' health and cognitive status was clinically assessed during annual in-home visits using standardized questionnaire scales as well as weekly online health surveys. It was found that mild cognitive impairment manifested in fewer total mouse movements with lower efficiency and longer pauses between movements during routine home computer use.

A study on inferring emotions with paradata (Yamauchi & Xiao, 2018) found that mouse movements were statistically significantly correlated to self-reported anxiety scores, to certain positive emotions in relation to music, to positive affect and attentiveness scores in females and self-assurance scores in males in relation to film clips, and to valence and arousal scores in relation to emotion-elicitation stimuli.

An observational study analyzed the efficacy of using the mouse cursor distance and speed paradata to infer the self-reported level of negative emotion (Hibbeln et al., 2017). Mouse movements were monitored while participants completed five goal-directed tasks, such as configuring a laptop or a car on a website. Participants' emotion was allowed to vary without direct manipulation and after completing each task, participants reported their level of negative emotion on a 9-point pictorial scale of emotion. The study found that greater cursor distance and slower cursor speed are both statistically significantly correlated with lower self-reported level of pleasure.

5.3 Paradata and estimates

In a set of experiments where participants rated attitude objects (Schneider et al., 2015) the pull of the participants' mouse trajectories was greater between two opposing evaluations. In two of the three studies, mouse trajectory pull of the non-chosen evaluative option and response time were positively correlated also with self-reported subjective ambivalence. A study that analyzed self-reported effort found that longer response time was statistically significantly related to higher effort scores (Revilla & Ochoa, 2015). However, worse response quality did not result in lower scores of self-reported effort.

A field test of 30 student participants (Tzafilkou & Nicolaos, 2018) captured mouse and keyboard actions where the participants had to create a web form. After the task, they were administered a questionnaire measuring self-reported behavioral attributes. The results showed that certain mouse movements were statistically significantly correlated with self-reported usefulness, ease of use, self-efficacy, willingness to learn and risk perception. Certain keyboard actions were significantly correlated with self-reported ease of use and self-efficacy.

5.4 Paradata, response quality and respondent characteristics

Several studies used paradata to investigate response quality in relation to Big Five personality traits. Among them is the study on insufficient effort (i.e., low quality of responses due to failure to read or comply with instructions), which found that differences in the level of effort reflected the respondent's self-reported personality traits (Bowling et al., 2016). The results showed that openness was unrelated to insufficient effort responding, while conscientiousness, agreeableness, extraversion, and emotional stability were negatively related to insufficient effort responding (ibid.).

The Understanding America Study found that conscientiousness and openness to experience predicted the incidence of unit nonresponse in subsequent survey waves (Cheng et al., 2020). In the study, both paradata and personality traits were found to capture independent information that explained unit nonresponse. Another study employing Big Five personality traits found that a lower level of nondifferentiation was correlated with a higher agreeableness score and with longer response times (Sturgis et al., 2019). In addition, fewer "don't know answers" signaled a link to higher conscientiousness score and shorter response times.

A study that monitored mouse actions with paradata also analyzed self-reported personality traits (Cepeda et al., 2021). It was found that mouse micro-behaviors predicted 78% of response uncertainty events. For example, it was found that respondents moved the pointer more frequently from one response option to another when uncertain and that respondents more frequently visited items associated with uncertainty.

Besides personality traits discussed above, studies also compared general sociodemographics on the basis of some paradata indicators. In an experiment comprising six web surveys, paradata were used to monitor interactive feedback prompts to speeders (Conrad et al., 2017). It was found that the prompt reduced speeding, improved response quality, and reduced straightlining. When comparing for sociodemographics, younger respondents (especially in the 18–34 age group) were found to have sped on significantly more questions than did older respondents in all of the surveys.

When analyzing respondent characteristics, the study found that focus-out events were more common for young and highly educated respondents. A study that monitored focus-out events with paradata found that switching away from the web survey was associated with selecting the middle response category, and self-reported multitasking (Höhne, Schlosser, et al., 2020).

A study on speeding found that very short response times were related to a higher level of straightlining regardless of the demographics, but the link was particularly strong among less educated respondents (Zhang & Conrad, 2014). Speeding was found to be related to several respondent characteristics, particularly age where younger respondents were more likely to speed. A multilevel analysis of 21 web surveys on political attitudes and behavior found that both the individual and the survey design characteristics level had effects on interview duration (Gummer & Roßmann, 2015). The larger share of the variation in interview duration was explained by respondent characteristics. In a study on response times (Yan & Tourangeau, 2008), they were found to be affected by question characteristics, complexity of the response options, and location of the question within the questionnaire, as well as age, education, experience with the internet and experience with completing web surveys.

5.5 Paradata, response quality and estimates

Three of the 57 eligible studies analyzed response quality and survey estimates with response time paradata. The first study found that removing speeders did not alter marginal distributions of estimates of attitudes and behaviors (Greszki et al., 2015). Their findings suggested that responding too fast primarily adds random noise to the data and weakens correlations. The study also found that with response times it is possible to detect some satisficing behavior that would have gone unnoticed on the basis of only traditional indicators, such as “don’t know” answers and item nonresponse (ibid.). A different study found that response times can be used as proxies to identify socially desirable response behavior (Andersen & Mayerl, 2017). Faster responses were found to be associated with the reporting of desirable attitudes and behavior. The third study (Gutierrez et al., 2011) identified speeders as 10% of respondents. Speeding was significantly related to straightlining; however, speeding did not have a statistically significant effect one stimates.

6 Discussion

6.1 Research Questions

Regarding the type of paradata used by researchers (**RQ.1**), questionnaire navigation paradata were the predominant type of paradata used among 57 eligible manuscripts (see Table 3). These paradata included response times, mouse action and keyboard action paradata (e.g., Stieger & Reips, 2010), focus-out events (e.g., Höhne, Schlosser, et al., 2020), and respondent motion paradata (e.g., Höhne, Revilla, et al., 2020). Device type paradata were used less frequently, for example when comparing PCs vs. smartphones (e.g., Höhne, Schlosser, et al., 2020) and different screen resolutions (e.g., Kaczmirek, 2009). Even fewer studies used contact info paradata, which included paradata indicators from previous survey participation (e.g., Roßmann & Gummer, 2016) and outcomes of reminders (e.g., Cheng et al., 2020).

Table 3: Types of paradata indicators derived from direct paradata in 57 eligible manuscripts

Paradata indicator type	Number of manuscripts
Response time	48
Mouse actions	16
Response changes	6
Device characteristics	5
Response order	3
Keyboard actions	2
Respondent motion	2
Focus-out	2
Outcomes of reminders	1
Previous survey participation	1
Validation prompts	1
Number of sessions	1

As for the specific indicators that the researchers derived from the paradata (**RQ.2**) in 57 eligible manuscripts (Table 3), several predominant types of indicators can be underscored. Among the 48 manuscripts, which derived paradata indicators of response time, these indicators were used for different purposes, including analyzing the relationship of response times to response quality (e.g., Revilla & Ochoa, 2015), to respondent characteristics (e.g., Yan & Tourangeau, 2008), and to survey estimates (e.g., Andersen & Mayerl, 2017), as well as with the aim of identifying speeders (e.g., Conrad et al., 2017). Mouse action paradata, which were derived in 16 manuscripts were used in contexts, such as mouse clicks (e.g., Conrad et al., 2006) and pointer distance and speed (e.g., Hibbeln et al., 2017). Only a few studies captured other paradata indicators, including on device characteristics, keyboard actions, and respondent motion (Table 3).

Regarding the three major paradata areas (i.e., respondent quality, respondent characteristics, and estimates) of survey methodology studies (**RQ.3**), previous studies addressed a single paradata area or at most two paradata areas, suggesting that the scope of paradata use is still somewhat limited. The most prevalent use of paradata was related to response quality (52 manuscripts), while the other two areas (i.e., respondent characteristics and survey estimates) were addressed less frequently (Table 4).

Table 4: Major paradata areas among 57 eligible manuscripts

Major paradata area	Number of units	Number of mentions
Response quality	38	52
Respondent characteristics	3	14
Survey estimates	2	5
Response quality & respondent characteristics	*11*	*11*
Response quality & survey estimates	*3*	*3*
Respondent characteristics & survey estimates	*0*	*0*
Response quality, respondent characteristics & survey estimates	*0*	*0*
Total	57	85

Few studies examined two major paradata areas simultaneously (Table 4) and none of the studies addressed both respondent characteristics and survey estimates paradata areas. Furthermore, none of the papers used the available paradata to simultaneously examine response quality, respondent characteristics, and survey estimates, indicating a blind spot in research.

6.2 Broader context

This state-of-the-art review shows that the usage of the notion of paradata in increasing. Namely, there were 20 eligible manuscripts between 2017 and 2022 in this paper, which is roughly comparable to the period between 2001 and 2011 (i.e., 22). Of course, this increase refers only to experimental research with direct paradata in web surveys (see Limitations and future search), which is much rarer and is growing slower. For example, paradata have also been discussed in relation to other computerized survey modes (e.g., Bassili & Krosnick,

2000) and other contexts, which are not limited to direct paradata and the basic web survey mode (see Callegaro et al., 2015).

We explored the trend of paradata in the literature, starting with the year when a paper by Couper (2000) explicitly mentioned paradata. We used the keyword combination of ‘paradata’ and ‘survey’ in Google Scholar. While the insight that is provided in this way is mostly illustrative and lacks experimental insight, it demonstrates a sharp increase in the usage of paradata in the context of surveys after 2010 (right-hand side Y axis in Figure 2). Web of Science (WoS) and Scopus, two of the most recognized citation indexes, revealed a similar trend (left-hand side Y axis in Figure 2). Scopus was search with ‘TITLE-ABS-KEY (paradata AND survey)’ and WoS with ‘AB=(paradata AND survey) OR TI=(paradata AND survey) OR KP=(paradata AND survey)’.

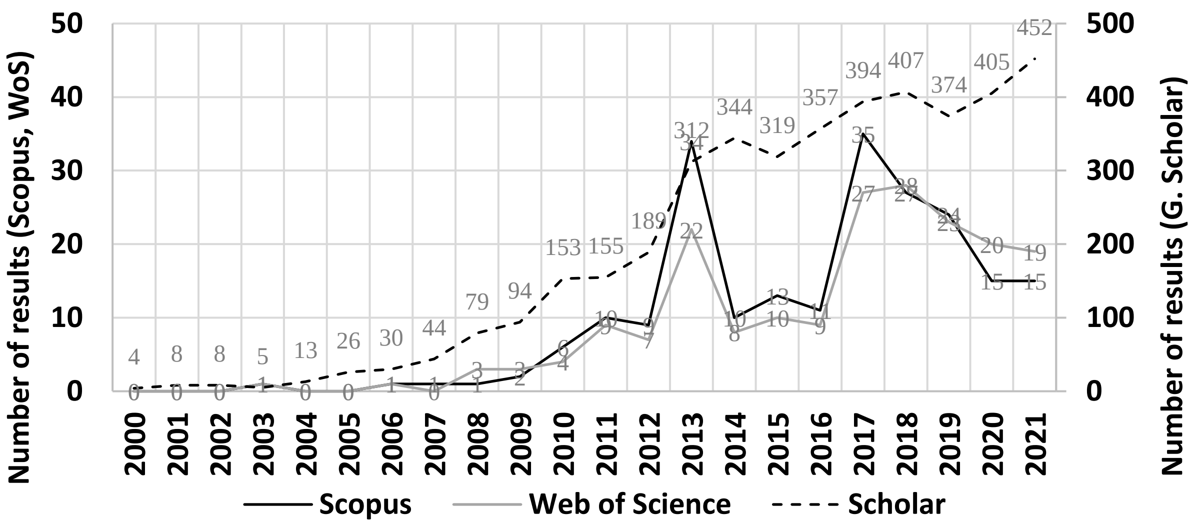

Figure 2: The number of search results for the sarch terms 'paradata' and 'survey' on Google Scholar (right Y-axis), Scopus, and Web of Science-WoS (left Y-axis) by year from 2000 to 2021

6.3 Limitations and future research

Several limitations and caveats in relation to this paper’s findings are discussed below. The first limitation concerns the paradata typology, which was used as the basis for identifying and classifying manuscripts in the state-of-the-art review. The typology is exhaustive in terms of direct paradata; however, we did not include indirect paradata, metadata, auxiliary data, passive data, and digital traces, as these are beyond the scope of this paper. These types of data and paradata could be explored by future studies for their potential relevance to the analysis of response quality, respondent characteristics, and survey estimates. In addition, the eligibility criteria of this paper required that direct paradata were used together with a web survey. Direct paradata collected with other computerized modes of data collection (see Callegaro et al., 2015) are thus beyond the scope of this paper, for example paradata in Computer Assisted Personal Interviewing (CAPI, e.g., Biemer et al., 2013), Computer Assisted Telephone Interviewing (CATI, e.g., Bassili & Krosnick, 2000), and Computer Assisted Self-Interviewing (CASI, e.g., Couper & Kreuter, 2013).

The second limitation refers to the method of literature search and selection, which was based on the exhaustive typology of direct paradata and yielded 570 initial results excluding duplicates. The purpose of the criteria, which resulted in 57 eligible manuscripts was to ensure that the manuscripts were of high quality and that the scope of the results remained manageable. Consequently, the results were limited to published scientific manuscripts in English, and gray literature was not examined because paradata in unpublished studies might be less comprehensive and less reliable. While focusing on published and preferably peer-reviewed manuscripts ensured a certain level of quality, there is a risk that some references may have been overlooked. In the future, a review could consider limiting the scope of the results to only a narrow range of paradata (e.g., only response quality), which may allow them to also examine unpublished studies and assess their quality.

The manuscripts included in the state-of-the-art review used different sets of paradata to analyze specific survey methodology issues, and the sets of paradata were often chosen arbitrarily, under unspecified circumstances, or following the practices of other researchers. It thus remains unclear whether these paradata sets are exhaustive and what possible combinations of paradata indicators might all be useful, for example, to examine a particular aspect of response quality. Future studies should therefore aim to identify and harmonize paradata indicators that are particularly meaningful, rich, and cost-effective in informing researchers about survey errors. In this context, systematic identification of the indicators with the best added value would be useful, as the complexity of documenting, cleaning, processing, aggregating, and evaluating paradata requires a significant amount of time and effort on the part of researchers. Subsequent studies should also identify the paradata indicators that would be meaningful and relevant for simultaneous analysis of all three major paradata areas, as previous studies addressed one or at most two major paradata areas.

7 Conclusion

This paper presented a state-of-the-art review and identified paradata that have been previously used by studies in web survey methodology. A total of 57 manuscripts were identified in which the causes of survey errors were investigated with paradata from at least one of the three major paradata areas: (i) response quality, (ii) respondent characteristics, and (iii) survey estimates. The findings showed that in most of the manuscripts (i.e., 91%), paradata have been used in the paradata area of (i) response quality. Conversely, fewer studies have used paradata to analyze (ii) respondent characteristics (i.e., 23%), and (iii) survey estimates (i.e., 11%). Moreover, no study to date has used paradata to examine all three major paradata areas simultaneously; previous studies addressed at most two major paradata areas.

The above suggests that paradata are underutilized by studies in survey methodology. It should be noted that the scope and granularity of the paradata to be collected depend on the circumstances and preferences of survey research projects. Therefore, the challenge remains for future studies to identify paradata indicators, which could be implemented routinely and simultaneously in survey datasets, as well as serve as meaningful predictors of methodological and substantive issues in survey research. Here, the challenge for web survey software is to facilitate the use of paradata indicators by survey researchers. Support for automated collection of paradata worthy of being routinely collected for general use in survey datasets would reduce the difficulty of processing paradata in the research community.

References

Andersen, H., & Mayerl, J. (2017). Social Desirability and Undesirability Effects on Survey Response latencies. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *135*(1), 68–89. https://doi.org/10.1177/0759106317710858

Bassili, J. N., & Krosnick, J. A. (2000). Do Strength-Related Attitude Properties Determine Susceptibility to Response Effects? New Evidence From Response Latency, Attitude Extremity, and Aggregate Indices. *Political Psychology*, *21*(1), 107–132. https://doi.org/10.1111/0162-895X.00179

Biemer, P. P., Chen, P., & Wang, K. (2013). Using level-of-effort paradata in non-response adjustments with application to field surveys. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *176*(1), 147–168. https://www.jstor.org/stable/23355181

Biemer, P. P., Leeuw, E. de, Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, N. C., & West, B. T. (2017). *Total Survey Error in Practice*. Wiley.

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, *111*(2), 218–229. https://doi.org/10.1037/pspp0000085

Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web survey methodology*. Sage.

Callegaro, M., Yang, Y., Bhola, D. S., Dillman, D. A., & Chin, T.-Y. (2009). Response Latency as an Indicator of Optimizing in Online Questionnaires. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *103*(1), 5–25. https://doi.org/10.1177/075910630910300103

Cepeda, C., Dias, M. C., Rindlisbacher, D., Gamboa, H., & Cheetham, M. (2021). Knowledge extraction from pointer movements and its application to detect uncertainty. *Heliyon*, *7*(1). https://doi.org/10.1016/j.heliyon.2020.e05873

Cheng, A., Zamarro, G., & Orriens, B. (2020). Personality as a Predictor of Unit Nonresponse in an Internet Panel. *Sociological Methods & Research*, *49*(3), 672–698. https://doi.org/10.1177/0049124117747305

Conrad, F. G., Couper, M. P., Tourangeau, R., & Peytchev, A. (2006). Use and Non-use of Clarification Features in Web Surveys. *Journal of Official Statistics*, *22*(2), 245–269. http://www.websm.org/db/12/919/

Conrad, F. G., Schober, M. F., & Coiner, T. (2007). Bringing features of human dialogue to web surveys. *Applied Cognitive Psychology*, *21*(2), 165–187. https://doi.org/10.1002/acp.1335

Conrad, F. G., Tourangeau, R., Couper, M. P., & Zhang, C. (2017). Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods*, *11*(1), Article 1. https://doi.org/10.18148/srm/2017.v11i1.6304

Couper, M. P. (2000). Usability evaluation of computer-assisted survey instruments. *Social Science Computer Review*, *18*(4), 384–396.

Couper, M. P., & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 271–286. https://doi.org/10.1111/j.1467-985X.2012.01041.x

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web Surveys: Perceptions of Burden. *Social Science Computer Review*, *19*(2), 146–162. https://doi.org/10.1177/089443930101900202

ESOMAR. (2020). *Global Market Research 2020*. ESOMAR. https://ana.esomar.org/documents/global-market-research-2020

Fernández-Fontelo, A., Henninger, F., Kieslich, P. J., Kreuter, F., & Greven, S. (2022). *Classification ensembles for multivariate functional data with application to mouse movements in web surveys* (arXiv:2205.13380). arXiv. https://doi.org/10.48550/arXiv.2205.13380

Fernández-Fontelo, A., Kieslich, P. J., Henninger, F., Kreuter, F., & Greven, S. (2021). Predicting Question Difficulty in Web Surveys: A Machine Learning Approach Based on Mouse Movement Features. *Social Science Computer Review*, 08944393211032950. https://doi.org/10.1177/08944393211032950

Funke, F., Reips, U.-D., & Thomas, R. K. (2011). Sliders for the Smart: Type of Rating Scale on the Web Interacts With Educational Level. *Social Science Computer Review*, *29*(2), 221–231. https://doi.org/10.1177/0894439310376896

Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, *73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-Tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding. *Public Opinion Quarterly*, *72*(5), 892–913. https://doi.org/10.1093/poq/nfn059

Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the Effects of Removing "Too Fast" Responses and Respondents from Web Surveys. *Public Opinion Quarterly*, *79*(2), 471–503. https://doi.org/10.1093/poq/nfu058

Groves, R. M., & Lyberg, L. E. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, *74*(5), 849–879. https://doi.org/10.1093/poq/nfq065

Gummer, T., & Roßmann, J. (2015). Explaining Interview Duration in Web Surveys: A Multilevel Approach. *Social Science Computer Review*, *33*(2), 217–234. https://doi.org/10.1177/0894439314533479

Gummer, T., Roßmann, J., & Silber, H. (2021). Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementation. *Sociological Methods & Research*, *50*(1), 238–264. https://doi.org/10.1177/0049124118769083

Gutierrez, C., Wells, T., Rao, K., & Kurzynski, D. (2011). *Catch Them When You Can: Speeders and Their Role in Online Data Quality*. Midwest Association for Public Opinion Research (MAPOR) 2011. http://www.websm.org/db/12/16145/

Haraldsen, G., Kleven, Ø., & Sundvoll, A. (2005). *Big scale observations gathered with the help of client side paradata*. Quest Workshop. http://www.websm.org/db/12/15969/

Healey, B. (2007). Drop Downs and Scroll Mice: The Effect of Response Option Format and Input Mechanism Employed on Data Quality in Web Surveys. *Social Science Computer Review*, *25*(1), 111–128. https://doi.org/10.1177/0894439306293888

Heerwegh, D. (2003). Explaining Response Latencies and Changing Answers Using Client-Side Paradata from a Web Survey. *Social Science Computer Review*, *21*(3), 360–373. https://doi.org/10.1177/0894439303253985

Heerwegh, D. (2002). *Describing response behavior in websurveys using client side paradata*. Web survey Workshop and Symposium. http://www.websm.org/db/12/345/

Heerwegh, D., & Loosveldt, G. (2002). An Evaluation of the Effect of Response Formats on Data Quality in Web Surveys. *Social Science Computer Review*, *20*(4), 471–484. https://doi.org/10.1177/089443902237323

Hibbeln, M. T., Jenkins, J. L., Schneider, C., Valacich, J., & Weinmann, M. (2017). *How Is Your User Feeling? Inferring Emotion Through Human-Computer Interaction Devices* (SSRN Scholarly Paper ID 2708108). Social Science Research Network. https://papers.ssrn.com/abstract=2708108

Höhne, J. K., Revilla, M., & Schlosser, S. (2020). Motion instructions in surveys: Compliance, acceleration, and response quality. *International Journal of Market Research*, *62*(1), 43–57. https://doi.org/10.1177/1470785319858587

Höhne, J. K., & Schlosser, S. (2019). SurveyMotion: What can we learn from sensor data about respondents' completion and response behavior in mobile web surveys? *International Journal of Social Research Methodology*, *22*(4), 379–391. https://doi.org/10.1080/13645579.2018.1550279

Höhne, J. K., Schlosser, S., Couper, M. P., & Blom, A. G. (2020). Switching away: Exploring on-device media multitasking in web surveys. *Computers in Human Behavior, 111*, 106417. https://doi.org/10.1016/j.chb.2020.106417

Höhne, J. K., Schlosser, S., & Krebs, D. (2017). Investigating cognitive effort and response quality of question formats in web surveys using paradata. *Field Methods*, *29*(4), 365–382. https://doi.org/10.1177/1525822X17710640

Horwitz, R., Brockhaus, S., Henninger, F., Kieslich, P. J., Schierholz, M., Keusch, F., & Kreuter, F. (2020). Learning from Mouse Movements: Improving Questionnaires and Respondents' User Experience Through Passive Data Collection. In *Advances in Questionnaire Design, Development, Evaluation and Testing* (pp. 403–425). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119263685.ch16

Horwitz, R., Kreuter, F., & Conrad, F. G. (2017). Using Mouse Movements to Predict Web Survey Response Difficulty. *Social Science Computer Review*, *35*(3), 388–405. https://doi.org/10.1177/0894439315626360

Horwitz, R., Tancreto, J. G., Zelenak, M. F., & Davis, M. (2013). *Use of Paradata to Assess the Quality and Functionality of the American Community Survey Internet Instrument*. 79. https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Horwitz_01.pdf

Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting Insufficient Effort Responding with an Infrequency Scale: Evaluating Validity and Participant Reactions. *Journal of Business and Psychology*, *30*(2), 299–311. https://doi.org/10.1007/s10869-014-9357-6

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and Deterring Insufficient Effort Responding to Surveys. *Journal of Business and Psychology*, *27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Jenkins, J. L., Larsen, R., Bodily, R., Sandberg, D., Williams, P., Stokes, S., Harris, S., & Valacich, J. S. (2015). A multi-experimental examination of analyzing mouse cursor trajectories to gauge subject uncertainty. *2015 Americas Conference on Information Systems, AMCIS 2015*. http://www.scopus.com/inward/record.url?scp=84963625990&partnerID=8YFLogxK

Kaczmirek, L. (2009). *Human survey-interaction: Usability and nonresponse in online surveys*. Herbert Von Halem Verlag. https://ub-madoc.bib.uni-mannheim.de/2150/1/kaczmirek2008.pdf

Kreuter, F. (2013a). *Improving Surveys with Paradata: Analytic Uses of Process Information*. Wiley & Sons.

Kreuter, F. (2013b). Improving surveys with paradata: Introduction. In F. Kreuter (Ed.), *Improving surveys with paradata: Analytic use of process information* (pp. 1–11). Wiley.

Kühne, S., & Kroh, M. (2018). Personalized Feedback in Web Surveys: Does It Affect Respondents' Motivation and Data Quality? *Social Science Computer Review*, *36*(6), 744–755. https://doi.org/10.1177/0894439316673604

Kunz, T., & Hadler, P. (2020). *Web Paradata in Survey Research*. GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_037

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, *24*(7), 1003–1020. https://doi.org/10.1002/acp.1602

Malhotra, N. (2008). Completion Time and Response Order Effects in Web Surveys. *Public Opinion Quarterly*, *72*(5), 914–934. https://doi.org/10.1093/poq/nfn050

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Matjašič, M., Petrovčič, A., & Vehovar, V. (2021). *Combining response times and response quality indicators to identify speeders with low response quality in web surveys*.

Matjašič, M., Vehovar, V., & Lozar Manfreda, K. (2018). Web Survey Paradata on Response Time Outliers: A Systematic Literature Review. *Metodološki Zvezki*, *15*(1), 23–41. http://ibmi.mf.uni-lj.si/mz/2018/no-1/Matjasic2018.pdf

McClain, C. A., Couper, M. P., Hupp, A. L., Keusch, F., Peterson, G., Piskorowski, A. D., & West, B. T. (2019). A Typology of Web Survey Paradata for Assessing Total Survey Error. *Social Science Computer Review*, *37*(2), 196–213. https://doi.org/10.1177/0894439318759670

Meade, A., & Craig, B. (2012). Identifying Careless Responses in Survey Data. *Psychological Methods*, *17*, 437–455. https://doi.org/10.1037/a0028085

Mittereder, F. K. (2019). *Predicting and preventing breakoff in web surveys*. Doctoral dissertation. Ann Arbor, MI. https://deepblue.lib.umich.edu/handle/2027.42/149963

Paas, L. J., & Morren, M. (2018). PLease do not answer if you are reading this: Respondent attention in online panels. *Marketing Letters*, *29*(1), 13–21. https://doi.org/10.1007/s11002-018-9448-7

Revilla, M., & Ochoa, C. (2015). What are the links in a web survey among response time, quality, and auto-evaluation of the efforts done? *Social Science Computer Review*, *33*(1), 97–114. https://doi.org/10.1177/0894439314531214

Roßmann, J., & Gummer, T. (2016). Using Paradata to Predict and Correct for Panel Attrition. *Social Science Computer Review*, *34*(3), 312–332. https://doi.org/10.1177/0894439315587258

Schneider, I. K., van Harreveld, F., Rotteveel, M., Topolinski, S., van der Pligt, J., Schwarz, N., & Koole, S. L. (2015). The path of ambivalence: Tracing the pull of opposing evaluations using mouse trajectories. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00996

Schroeders, U., Schmidt, C., & Gnambs, T. (2021). Detecting Careless Responding in Survey Data Using Stochastic Gradient Boosting. *Educational and Psychological Measurement*, 00131644211004708. https://doi.org/10.1177/00131644211004708

Seelye, A., Hagler, S., Mattek, N., Howieson, D. B., Wild, K., Dodge, H. H., & Kaye, J. A. (2015). Computer mouse movement patterns: A potential marker of mild cognitive impairment. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, *1*(4), 472–480. https://doi.org/10.1016/j.dadm.2015.09.006

Sendelbah, A., Vehovar, V., Slavec, A., & Petrovčič, A. (2016). Investigating respondent multitasking in web surveys using paradata. *Computers in Human Behavior*, *55, Part B*, 777–787. https://doi.org/10.1016/j.chb.2015.10.028

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing Check-All and Forced-Choice Question Formats in Web Surveys. *Public Opinion Quarterly*, *70*(1), 66–77. https://doi.org/10.1093/poq/nfj007

Stern, M. J. (2008). The Use of Client-side Paradata in Analyzing the Effects of Visual Layout on Changing Responses in Web Surveys. *Field Methods*, *20*(4), 377–398. https://doi.org/10.1177/1525822X08320421

Stieger, S., & Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, *26*(6), 1488–1495. https://doi.org/10.1016/j.chb.2010.05.013

Sturgis, P., Schober, M. F., & Brunton-Smith, I. (2019). *Am I being neurotic? Personality as a predictor of survey response styles*. 8th European Survey Research Association

Conference (ESRA 2019). https://www.europeansurveyresearch.org/conferences/programme2019?day=1

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2004). Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly*, *68*, 368–393. https://doi.org/10.1093/poq/nfh035

Tzafilkou, K., & Nicolaos, P. (2018). Mouse behavioral patterns and keystroke dynamics in End-User Development: What can they tell us about users' behavioral attributes? *Computers in Human Behavior*, *83*, 288–305. https://doi.org/10.1016/j.chb.2018.02.012

Wells, T., Vidalon, M., & DiSogra, C. (2010). *Differences in Length of Survey Administration Between Spanish-Language and English-Language Survey Respondents*. 6. http://www.asasrms.org/Proceedings/y2010/Files/400137.pdf

Wise, S. L., & Kong, X. (2005). Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests. *Applied Measurement in Education*, *18*(2), 163–183. https://doi.org/10.1207/s15324818ame1802_2

Yamauchi, T., & Xiao, K. (2018). Reading Emotion From Mouse Cursor Motions: Affective Computing Approach. *Cognitive Science*, *42*(3), 771–819. https://doi.org/10.1111/cogs.12557

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, *22*(1), 51–68. https://doi.org/10.1002/acp.1331

Zhang, C., & Conrad, F. G. (2014). Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, *8*(2), Article 2. https://doi.org/10.18148/srm/2014.v8i2.5453