gesis Leibniz Institute
for the Social Sciences

Web Paradata in Survey Research

Tanja Kunz and Patricia Hadler

December 2020, Version 1.0

Abstract

Paradata are information about the primary survey data collection process. This guide is intended for survey practitioners who want to collect and use paradata in web surveys. The guideline focuses on a typology and possible applications of web paradata and practical implications regarding the collection, post-processing, and documentation of web paradata.

Citation

Kunz, T. & Hadler, P. (2020). Web Paradata in Survey Research. Mannheim, GESIS - Leibniz Institute for the Social Sciences (GESIS - Survey Guidelines).

DOI: 10.15465/gesis-sg_037

1 Introduction

In general, paradata refer to information about the process of collecting survey data. In web surveys, paradata are typically collected as a by-product of computer-assisted data collection (Couper, 1998). Web paradata describe the respondent's survey-taking behavior, including information on the contact process, device type and questionnaire navigation. As they are collected unobtrusively and in the natural environment, web paradata are characterized by low reactivity. Paradata must be distinguished from several other data that accompany survey data and/or can be merged with survey data (see Figure 1).

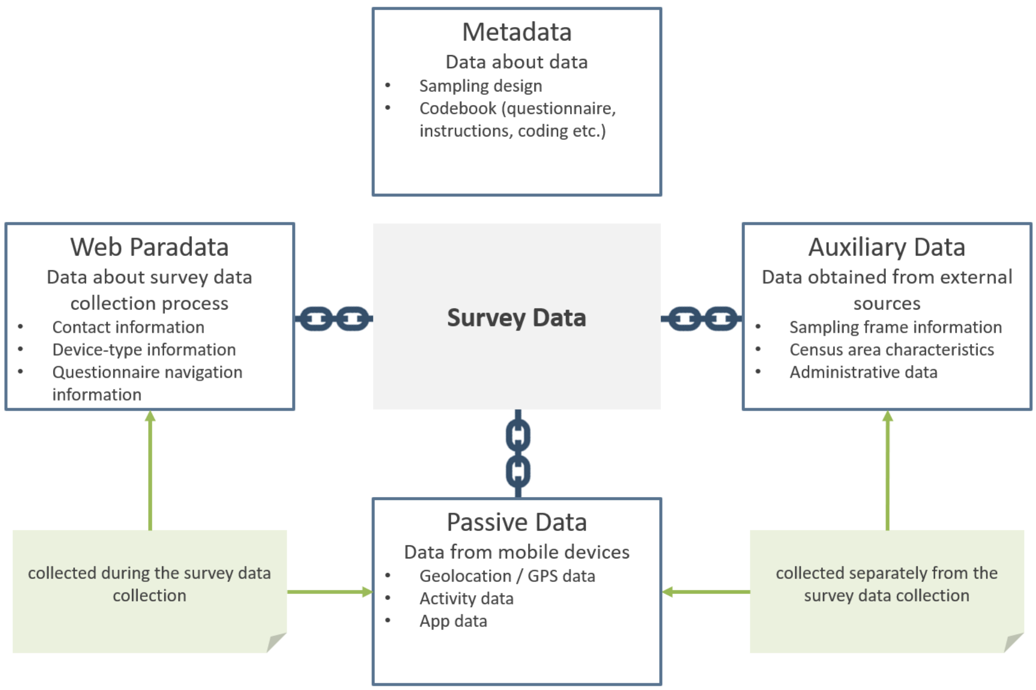

Figure 1: The distinction between web paradata and other types of data with some examples (own illustration).

Metadata are macro-level information about survey data, such as the sampling design or the codebook; thus, metadata are necessary to interpret survey data but are not linked to these on the micro-level of individual respondents. Auxiliary data can be obtained from external sources, such as census data or administrative data, and can, in some cases, be merged with the survey data (Sakshaug & Antoni, 2017). Passive data and paradata have in common that they are collected without (active) involvement of the respondent. Unlike paradata, passive data are usually used to describe actions or behaviors that exceed survey-taking behavior, such as motion data or online browsing behavior in other tabs, windows, or mobile apps. Passive data collection may be initialized during survey data collection but usually extends beyond the time taken to fill out the survey. In survey research, passive data is mostly collected via smartphone devices (Keusch, Struminskaya, Antoun, Couper, & Kreuter, 2019). Still, "the distinction between paradata, passive data, and auxiliary data is likely to be an ongoing discussion as technology and data collection efforts evolve" (McClain et al., 2019, p. 207).

2 Typologies and possible applications of web paradata

2.1 Categorizing web paradata

Types and purposes of web paradata are manifold. Paralleling the technological development of web paradata collection, a variety of categorizations have evolved.

The implementation of one of the first freely accessible paradata scripts in web surveys (Heerwegh, 2003) brought along the *technical distinction* between server-side and client-side paradata. *Server-side paradata* are collected at the server on which the web survey resides; they are captured by default by most online survey software. *Client-side paradata* collection is carried out on the respondent's (client's) device. This requires programming by the researcher, usually the implementation of JavaScript code in the survey software. Client-side data is richer than the server-side by offering detailed information on respondents' actions within survey pages (see section 3.3.1 for details).

Another categorization of paradata is based on the *object of description* (Callegaro, 2013; Callegaro, Lozar Manfreda, & Vehovar, 2015). *Contact-info paradata* include contact time, mode, and outcome during the recruitment of survey participants. *Device-type paradata* include information on screen size, resolution and orientation, browser, or internet connection. *Questionnaire navigation paradata* describe how the respondent fills out the survey. This may include page revisits, answer changes, or the use of non-question links.

In a similar approach, the typology of McClain et al. (2019) distinguishes in which of the four *phases of the data collection process* paradata are collected. These phases are *prior survey, recruitment, access,* and *response*. This categorization highlights the analytical use of paradata from previous waves in panel studies, such as the device used in previous waves to predict device use for the current wave of a study. Also, the typology presents which types of paradata from which phase can be used to analyze different errors from the Total Survey Error Framework (Groves et al., 2009). For instance, paradata gained during the recruitment phase are particularly useful to analyze coverage and nonresponse error, while paradata from the response phase mostly focus on measurement error.

In the following, we distinguish *contact information paradata* that are collected during the recruitment and access phase, *device-type paradata* gathered during the access and response phase, and *questionnaire navigation paradata* that are collected during the response phase (Callegaro, 2013; Heerwegh, 2011; McClain et al., 2019). Contact information, device-type, and questionnaire navigation paradata can also be taken from a survey other than the current one, which means that they are collected at an earlier stage—in the prior survey phase—and used in the current survey. Examples are paradata from previous waves in longitudinal studies or earlier stages in multi-stage surveys. Table 1 gives an overview of the most common types of web paradata and possible analytical purposes. In the next section, we provide more details on potential applications of the different kinds of web paradata.

Table 1: Typology and possible applications of web paradata

Type	Data collection phase	Examples	Derived measures	Relation to the Total Survey Error (TSE) framework and common analytical purposes
Contact information paradata	Recruitment and access	• Contact timing: time of day/day of week • Contact mode • Contact strategy • Outcomes of each contact attempt • Reasons for noncontact • Date and time of (first/last) access to questionnaire	• Date and time of completion • Time from open contact to access • Number of contact attempts • Number of login attempts • Session counts • Disposition codes/outcome • Survey outcome rates	Focus on nonresponse error • data collection efficiency – monitoring, evaluating and optimizing contact processes • survey estimates – assessing and adjusting for nonresponse bias
Device-type paradata	Access and response	• IP address • User agent string • Active scripting enabled (e.g., JavaScript, Flash, Cookies) • Screen size/resolution • Screen orientation • Browser window size • GPS coordinates	• Unique respondents • Device type • Device type switching • Operating system • Browser type/version • Screen orientation change • Browser window size change • Technical ability to access survey features • Location of filling out the survey (i.e., country, state, on company premises)	Focus on nonresponse and measurement error • data collection efficiency – monitoring and controlling contact processes • data quality – adapting questionnaire design – assessing response behavior
Questionnaire navigation paradata	Response	• Time stamps • Keystrokes • Mouse clicks/finger taps • Mouse movements • Mouse coordinates • Scrolling/swiping • Zooming/pinching • Detection of the current window	• Movements across the questionnaire (forward/backward) • Last page before dropout • Response times • Change of answers • Order of answering • Item nonresponse • Mouse movements within a page • Clicks on non-questions links (e.g., hyperlinks, FAQ, and help) • Appearances of prompts/error messages • Window switching	Focus on measurement error • data quality – determining the length of interview/questions – pretesting and evaluating questionnaire design, question format, order, and wording – guiding real-time interventions – assessing response behavior

2.2 Applications of web paradata

Web paradata can be used before, during, and after the collection of survey data for various purposes. Thus far, web paradata have mainly been used for methodological purposes, in particular, to address various sources of survey errors (see to address various sources of survey errors (see Table 1)). In this regard, possible applications can be assigned to three areas: improving *data collection efficiency, data quality,* and *survey estimates* (Kreuter, 2015). The use of web paradata for *substantive purposes* is still rather rare. For example, response times can be used to assess attitude stability (Heerwegh, 2011).

2.2.1 Contact information paradata

Contact information paradata[12] usually include the time, date, and outcome of each contact attempt. In web surveys, email paradata provide information on how many invitations were sent, when they were sent, and with what outcome (e.g., whether the email was (not) delivered, whether it was opened, the time between delivery and email opening). This information can be used to improve data collection efficiency by understanding potential delivery problems and optimizing the timing and frequency of email communications to increase the probability that sample units will open the email (Hupp, Schroeder, & Piskorowski, 2017). The paradata are considerably less extensive if the respondents are invited to the web survey by mail and not by email (e.g., in a probability-based web survey, there is usually only a postal address but no email address for the first contact). In case of an invitation by mail, the only information available would be the date on which an invitation is sent. Moreover, contact mode (e.g., email vs. SMS invitation) (de Bruijne & Wijnant, 2014; Mavletova & Couper, 2014; McGeeney & Yan, 2016), contact strategy (e.g., timing and order of providing access to mixed-mode survey) (Holmberg, Lorenc, & Werner, 2010; Millar & Dillman, 2011), and reasons for noncontact (e.g., bounced emails) can be gathered. This information can be used to implement adaptive and responsive survey designs[3] before or during fieldwork to improve contact and cooperation (Lewis & Hess, 2017). Further important contact information paradata relate to the date and time of the [first/last] access to the questionnaire. Derived variables in this context include the date and time of completion, disposition codes (e.g., complete, partial, breakoff, refusal), and survey outcome rates (e.g., response rates, breakoff rates). For example, the number of breakoffs (i.e., those who drop out a survey), the number of suspends (i.e., those who resume the survey after a break), and the session counts indicate possible problems with survey length, respondent fatigue, or other issues. Early and late respondents can be compared for nonresponse bias analyses (Kypri, Samaranayaka, Connor, Langley, & Maclennan, 2011).

2.2.2 Device-type paradata

Device-type paradata are typically collected server-side on the first survey page during the access phase. A user agent string is transmitted every time a web browser connects to a website that contains, among others, information about the device type, browser name, and operating system. Device switching may occur during the completion of a single survey and between waves of a longitudinal study (Zijlstra, Wijgergangs, & Hoogendoorn-Lanser, 2018). Additional device-specific information relates to screen resolution,

[1] In interviewer-administered surveys, contact information paradata is often referred to as “call record data”, “contact form data”, or “level-of-effort paradata”.

[2] The distinction between contact information paradata and metadata is not always clear-cut, as some of the paradata listed here are also referred to as metadata (e.g., contact mode and strategy, outcome rates).

[3] Adaptive and responsive survey designs have in common that the characteristics of the survey design may differ for different (groups of) sample units depending on their characteristics. For instance, the effort expended in establishing contact and cooperation vary from one sample unit to another or in the course of data collection. In adaptive survey designs, tailor-made strategies are defined before the survey starts; in responsive survey designs, paradata that are collected in early phases of a survey are used for design decisions in later phases (Tourangeau,Brick,Lohr, & Li, 2017).

screen orientation (portrait or landscape), screen orientation change, and the browser window size. This information is often collected client-side during the response phase on several or even all survey pages. Device-type paradata can be used before and during the collection of survey data to implement adaptive and responsive survey designs with the aim of improving the efficiency of data collection, for example, by predicting device use based on longitudinal data to promote the use of a particular device (Haan, Lugtig, & Toepoel, 2019), or to reduce nonresponse and measurement error by assigning respondents to their preferred device (Metzler, 2020). Device-type paradata are also used during survey data collection to optimally adapt the questionnaire design to the characteristics of different devices (i.e., desktop and mobile devices), browsers, and operating systems (Antoun, Katz, Argueta, & Wang, 2018; Beuthner, Daikeler, & Silber, 2019; Callegaro, 2010), and after the fieldwork process to evaluate possible device-dependent differences in survey data quality (Antoun, Couper, & Conrad, 2017; Keusch & Yan, 2017; Lugtig & Toepoel, 2016; Mavletova, 2013; Sommer, Diedenhofen, & Musch, 2016; Toninelli & Revilla, 2016; Tourangeau et al., 2018; Verbree, Toepoel, & Perada, 2019).

2.2.3 Questionnaire navigation paradata

Questionnaire navigation paradata are often collected during the entire questionnaire completion process and include keystrokes, mouse actions, touch events, and timestamps stored along with the respondents' actions (e.g., (de-)selecting a radio button, entering a text, scrolling the browser window). Standard derived measures are response times spent per question/survey page, change of answers, the order of answering, and mouse movements within a survey page. In general, questionnaire navigation paradata allow conclusions about the interaction of respondents with the questionnaire and their response behavior. They can be used before survey data is collected, i.e., during pretesting to optimize questionnaire design, or post-survey to better understand the "black box" of respondent behavior and assess survey data quality. Long response times may, for example, indicate poor question design and respondent difficulties with particular questions (Lenzner, Kaczmirek, & Lenzner, 2010). Similarly, excessive mouse movements can be used to identify confusing questions (Horwitz, Kreuter, & Conrad, 2017), and the number of answer changes may indicate respondent problems in mapping the response on the provided response alternatives (Stern, 2008). The last page before dropout (i.e., the last answered question) allows conclusions about which questions were particularly problematic in terms of, among others, the topic, layout, or format of the questions (Peytchev, 2009). Extremely short response times (i.e., speeding) are often associated with satisficing behaviors such as straightlining (i.e., selecting the same response option to answer several rating scale items) and primacy effects (i.e., selecting the first response option that seems appropriate) and, thus considered an indicator of low respondent effort and low survey data quality (Malhotra, 2008; Revilla & Ochoa, 2015; Zhang & Conrad, 2013). Moreover, browser window/tab switching indicating that respondents are temporarily leaving the survey allows conclusions about multitasking during questionnaire completion (Sendelbah, Vehovar, Slavec, & Petrovcic, 2016) and over-optimizing response behavior in knowledge questions (Gummer & Kunz, 2019). In these cases, questionnaire navigation paradata are used to make inferences about the respondents, for instance, to classify them as multitaskers, satisficers versus optimizers, etc. Questionnaire navigation paradata can also be used to identify problematic response behavior and to guide appropriate interventions in real-time, for example by providing additional clarification of the question meaning due to respondents' inactivity (Conrad, Schober, & Coiner, 2007), by prompting respondents to select the desired number of responses in check-all-that-apply questions (Kunz & Fuchs, 2019a), or by asking speeders and straightliners to take more time for their answers and differentiate more in grid questions (Conrad, Tourangeau, Couper, & Zhang, 2017; Kunz & Fuchs, 2019b). However, the interpretation of questionnaire navigation paradata in particular is seldom straightforward and should be guided by a relevant theory (see section 3.2.4).

2.2.4 Prior survey paradata

Contact information, device-type, and questionnaire navigation paradata can stem from prior waves in panel surveys or screener interviews in multiple-stage surveys (household rosters). Among others, prior survey paradata (e.g., item nonresponse, response speed, participation history, the device used in previous waves) can be used to predict and correct for panel attrition (Kern, Weiss, & Kolb, 2019; Roßmann & Gummer, 2016; Tienda & Koffman, 2020). The effects of switching devices over successive waves of a panel on data quality and developments over time can be studied (Lugtig & Toepoel, 2016; Revilla, Toninelli, Ochoa, & Loewe, 2016). Moreover, invitation mode can be varied according to device preferences in previous waves, which may have effects on the likelihood of participation and data quality (Metzler, 2020).

3 Practical implementation

Although web paradata are often described as a by-product of survey data collection, in most cases they are collected intentionally and involve some additional effort for the survey researcher. Given the extra effort associated with the collection and use of web paradata, survey researchers should explicitly decide before starting survey data collection which paradata are to be collected for which purpose and how they will be analyzed. This decision should preferably be based on a relevant theory. Also required is a certain standardization regarding the collection and post-processing of web paradata, quality assurance, and comprehensive documentation of the collection and preparation of the paradata (Couper, 1998; Kreuter, 2015; Smith, 2011).

3.1 Collection

3.1.1 Server-side and client-side web paradata

As mentioned above, an important technical distinction regarding web paradata concerns *server-side* (i.e., visits to a specific page) and *client-side* paradata (i.e., events within a page).

Practically all web survey data sets include at least some paradata, as most computer-assisted interviewing (CAI) software systems offer the possibility of collecting the most common server-side paradata by default (e.g., disposition code, date and time of (first/last) access, duration, device type). These measures are usually collected at the survey level (i.e., each time a respondent accesses the survey) or at the page level (e.g., elapsed time from loading a webpage to submitting it by clicking on the "Next" button, including request/response transmission times between server and client. If server-side paradata are collected at the page level, information content decreases with each additional question on a page.

Compared to server-side paradata, client-side paradata are "richer in detail, precision, and amount of information that can be collected" (Callegaro, 2013, p. 262). Client-side paradata can provide measurements for each respondent's action *within* a survey page (e.g., elapsed time between individual mouse clicks). Thus, client-side paradata contain more detailed information about specific respondent actions. Consequently, client-side paradata are especially advantageous when no strict paging design is applied with each question displayed on a separate survey page, but instead, multiple questions or—in the most extreme—all questions are presented on one survey page in a scrolling design (Mavletova & Couper, 2014; Peytchev, Couper, McCabe, & Crawford, 2006). One disadvantage of client-side paradata is that special scripts with usually JavaScript code are required to capture them. This means additional programming effort for the researcher and function control to ensure an error-free technical implementation. A pretest should be carried out to ensure proper functioning using different devices and browsers. Furthermore, it requires a certain flexibility of the survey software solution to enable the integration of such client-side paradata scripts (Callegaro, 2013; Heerwegh, 2002). Besides, client-side paradata are more susceptible to missing data (e.g., if JavaScript is disabled). Client-side paradata usually generate non-rectangular data in the form of strings or log files (also referred to as audit trails), which may require considerable effort to extract and prepare the data.

If client-side paradata are to be collected, it is advisable to use freely available client-side paradata scripts (e.g., scripts by Heerwegh, 2003; Kaczmirek & Neubarth, 2007; Schlosser & Höhne, 2020) to keep the additional effort for programming and function control low, to reduce the error-proneness of the technical implementation, and to enable a certain degree of standardization and comparability of the collected web paradata across studies. Depending on the amount and detailedness of information required, client-side paradata scripts can be implemented on every n^{th} to all survey pages.

3.1.2 Tailored set of web paradata

Following the data minimization principle (see Article 5 (1) lit c, EU-GDPR), it is strongly recommended that the collection of web paradata is "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed." Thus, researchers must determine before survey data collection, which set of paradata variables are best to be collected for which purposes. The general advice is to limit the collection of web paradata to the relevant ones, whereby the selection is best determined "by a research question and a relevant theory" (Yan & Olson, 2013, p. 89). Time stamps, keystrokes, and mouse clicks are certainly among the most frequently collected and used client-side web paradata (Olson & Parkhurst, 2013). If a freely available client-side paradata script is implemented, that makes it easy to obtain a whole range of variables, the selection of web paradata should nevertheless be tailored to the researchers' use to avoid hoarding "unnecessary" data that is never analyzed.

3.1.3 Informed consent

In the field of market, opinion and social research as well as in the scientific community, there is still a debate about the conditions under which the informed consent of respondents to the collection and use of their web paradata must be obtained. Different types of paradata can be associated with varying requirements of consent. Further information on when and how informed consent for the collection and use of web paradata should be obtained can be found in the GESIS Survey Guideline "Informing about Web Paradata Collection and Use" (Kunz, Beuthner, Hadler, Roßmann, & Schaurer, 2020).

3.2 Post-processing

3.2.1 Quality control

Quality control of collected paradata is crucial before data processing and analysis. How quality control is carried out is highly dependent on the type of paradata collected. We generally recommend paying close attention to missing and inconsistent data.

Missing data can be easily identified in the raw paradata files in case of predefined missing codes (e.g., -66, -99). However, there are usually other cases with inconsistent (e.g., fewer mouse clicks than substantial answers) or even incorrect (e.g., time stamp values less or equal to zero) values. During data processing, these cases are usually set to missing or—less invasively—marked by a flag variable (i.e., indicating that there are discrepancies in the data). Additional information in the form of survey data (i.e., substantive responses) or other paradata variables (e.g., device type, disposition code, JavaScript disabled) is often required to identify missing data and inconsistent data. This is especially the case if a differentiated missing value scheme is used, and different reasons for missing values shall be distinguished.

In general, missing data and inconsistent data for web paradata can be due to *technical problems* (e.g., temporary interruption of the script, an unsupported function of the script) or due to *respondent behavior* (e.g., skipping a question, dropping off the survey, backing up to earlier questions). In both cases, the occurrence of missing data and inconsistent data also depends on the kind of technical implementation of the paradata script functions. For example, a distinction is made between asynchronous and synchronous communication between server and client. In the latter case, paradata are only stored after they have been transmitted to the server, resulting in missing data if the "Next" button is not clicked (Schlosser & Höhne, 2020). Or, when respondents use the "Back" button to navigate to a previous page and change a previously given answer, for instance, paradata that were gathered during the first visit of the page will either be continued/added by new paradata (e.g., by storing additional time stamps in a new variable) or overwritten (e.g., by replacing previously stored time stamps in the same variable), depending on the technical implementation. In the latter case, paradata are only partially stored, which can lead to inconsistent data.

Missing data

Missing data may occur for individual or all respondents, single or all paradata variables, and on an individual or all survey pages. If *all respondents* are affected, this strongly points to a problem with the script (e.g., because the entire script or parts of it do not work (anymore) in current browsers/devices or the survey software used) or a mistake while implementing the script (e.g., mistakenly giving the same variable name to two paradata variables will lead to data being captured for only one of them).

However, even when the paradata script is fully functioning and correctly implemented, missing data for *individual respondents* is to be expected. In this case, missing data may occur:

- For *all* paradata variables across the *entire* survey (completely missing at the survey level): indicating that the respondent's set-up did not allow for the web paradata script to run (e.g., JavaScript disabled) or no paradata were collected/transmitted to the server due to unit nonresponse (e.g., respondent refused to cooperate or was screened out).
- For *all* paradata variables on *individual* survey pages (completely missing at the page level): indicating that the script was temporarily not working/interrupted or no paradata were collected/transmitted to the server, for example, due to partial nonresponse (i.e., respondent abandoned the survey on this page) or missing by design (i.e., respondent skipped the page due filter conditions).
- For a *single* paradata variable across the *entire* survey (completely missing at the variable level): indicating that a particular script function was not applicable for specific devices (e.g., desktop/mobile, responsive layout), or not supported by the operating system (e.g., Android, iOS, Windows) or browser type/version.
- For a *single* paradata variable on *individual* survey pages (partially missing at the variable level): indicating that a particular script function was temporarily not working/interrupted or not called, for example, due to item nonresponse (i.e., respondent skipped question(s) on this page) or (missing) behavior patterns (e.g., no scrolling, no window switching, no mouse clicking but only keyboard input).

The percentage of general Internet users who have JavaScript disabled can be considered low, as more and more websites require JavaScript to render any content at all, just as virtually all social media applications require some form of active scripting (Couper & Zhang, 2016). For instance, Höhne, Schlosser, & Krebs (2017) report a share of about 1 percent. Similarly, in our studies based on samples from online access panels, we found that about 1.5 percent of respondents had JavaScript disabled.[4]

[4] The data are available from the first author on request.

Inconsistent data

In addition to missing values, paradata variables that seem to be stored correctly at first glance may be inconsistent when compared with substantive data (e.g., fewer mouse clicks than responses) or when comparing different paradata measures of the same construct (e.g., server-side vs. client-side vs. self-reported device type information). Furthermore, it can also happen that paradata variables have incorrect values (e.g., timestamp values less than or equal to zero) and therefore lead to inconsistencies. In most cases, it is advisable to set inconsistent data to missing. Thus, even when paradata output appears complete on first sight, we strongly recommend quality checks of all paradata variables used for analysis.

Examples of inconsistencies between (different types of) paradata and survey data are:

- *Number of clicks/time stamps.* The number of mouse clicks/finger taps or time stamps of mouse/finger entries documented in the paradata output should be equal to or higher than the number of items answered on that page plus the click on the "Next" button. If the number is lower, this may indicate that the script was interrupted or, for instance, the respondent used the keyboard instead of the mouse cursor to enter the responses.
- *Keystroke files.* All keystrokes made by a respondent to enter or edit open-ended answers or navigate within or between survey pages (i.e., alpha-numeric characters, function keys such as tab or arrow keys) can be captured. For open-ended questions, the number of recorded keystrokes must be equal to or higher than the number of characters logged into an open-ended text field by the respondents.
- *Response time measures.* Typically, sever-side and client-side response time measures should be similar, with server-side response times being slightly longer because they include the transmission time between server and client systems (Yan & Tourangeau, 2008). In general, it is recommended to use more detailed client-side response time measures when available (Heerwegh, 2011). However, when server- and client-side response times have both been captured, the discrepancy between the two should be examined. Researchers must decide and document how they deal with cases in which the two measures are highly discrepant, or when server-side measures show a shorter response time than client-side measures.
- *Device type measures.* Server-side device-type information (i.e., PC/laptop, tablet, smartphone) as captured by many online survey software providers based on the user agent string can be compared to, for example, screen sizes and screen orientation changes captured by client-side paradata scripts, or to the respondents' self-reports (see Table 2). There are alternative approaches for the classification of device type. Although we cannot give a general recommendation for a preferred data basis to classify device types, we recommend comparing alternative classifications on different (para)data (if available) and applying the chosen classification of the device type consistently within (and between) studies.
- *Device type and survey focus.* Some JavaScript functions may differ depending on the device used. For example, survey focus events are captured by two separate client-side functions for respondents using desktop or mobile devices in some scripts (e.g., in the ECSP script by Schlosser & Höhne, 2020). The data is then also stored in two variables (i.e., "SurveyFocus" for desktop users and "MobileSurveyFocus" for mobile users). In this case, a respondent should have valid values for only one of the two variables at a time.

Table 2: Device type (in %) based on user agent string paradata and self-reports (n=4,302)

	Self-report					
Paradata	**No answer**	**Desktop**	**Smartphone**	**Tablet**	**Others**	**Total**
Unknown	0.0	0.1	3.1	0.9	0.0	4.1
Desktop	0.1	**67.4**	0.3	0.6	0.0	68.5
Smartphone	0.0	0.1	**20.6**	0.5	0.0	21.3
Tablet	0.0	0.0	0.1	**5.6**	0.1	5.8
Console	0.0	0.0	0.0	0.0	**0.0**	0.0
Phablet	0.0	0.0	0.3	0.0	**0.0**	0.3
	0.2	67.7	24.5	7.5	0.1	100.0

Note. Data comes from a web survey conducted in October 2018 among members of a large German opt-in online panel provided by respondi AG

Appendix A gives an example of how to detect missing and inconsistent data in non-rectangular paradata strings.

3.2.2 Unit of analysis

Web paradata can be analyzed in varying degrees of detail depending on the *level of aggregation*, usually distinguishing between the *action, question, page,* and *survey level* (see Figure 2). Ideally, the researcher determines in advance of the data collection which paradata variables are needed, at which level of detail they must be collected, and in which unit of analysis they are to be analyzed. In some cases, it may be necessary to collect very detailed paradata at the action level; in other cases, however, it may be sufficient to collect them already in the aggregated form at a higher level. Thus, web paradata are already *measured* at a higher level or *aggregated* to a higher level during data post-processing. The decision at which level of aggregation paradata are to be collected and analyzed is primarily guided by the goal of the analysis (i.e., research question). In addition to theoretical considerations, pragmatic research reasons can also be decisive (based on empirical results). For instance, especially in the case of rare events, it may be advisable to aggregate to a higher level.

While lower-level measurements can be aggregated to any higher level during data post-processing, the reverse process of decomposing higher-level paradata into lower-level measurements is only possible by accepting considerable inaccuracies (e.g., the overall time spent per survey can be divided by the number of questions, resulting in a very rough average time per page). Therefore, the level of aggregation should be chosen very carefully in advance of data collection and, in case of doubt, a more detailed measurement is preferable "in order to prevent being unable to answer interesting follow-up research questions that require first-level [action-level] paradata or an alternative organization of first-level paradata, which cannot be arrived at from available higher-level paradata" (Heerwegh, 2011, p. 327).

In the most granular form, client-side paradata are collected on the *action level* for each respondent. For instance, client-side time stamps can be captured for each respondent's action (e.g., keystroke entry, mouse click, mouse movement) within each survey question and page. Because respondents may carry out a different number of actions within a page, action-level paradata are typically non-rectangular, meaning that a different number of observational points are recorded for respondents, usually in the form of string variables. The researcher must therefore decide which unit of analysis is required and how the paradata variables must be aggregated (Kaczmirek, 2008; Yan & Olson, 2013).

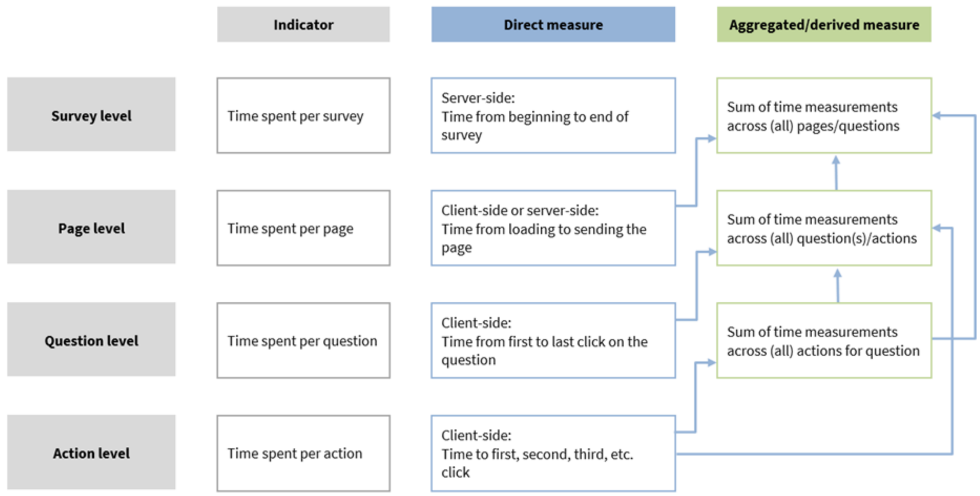

Figure 2: Different levels of aggregation of web paradata, using the example of response times (own illustration)

In general, action-level paradata can be aggregated to the question, page, or survey level. In a one-question-per-page design, question and page level aggregation is identical. For instance, time measures can be aggregated across all actions taken by a respondent to answer a question (i.e., action-level time stamps aggregated on the *question* or *page level*), indicating the "time spent per question/page." Similarly, time measurements can be aggregated across all questions/pages on the *survey level*, indicating the "time spent per survey."

Once the *level of aggregation* is determined (e.g., from action to question level), paradata measures are usually subject to further aggregation for analysis. The *direction of aggregation* can be horizontal at the respondent level or vertical across respondents. Figure 3 shows time measures that have been aggregated to the question level for a data set containing five respondents and three survey questions. This data can be aggregated *horizontally* by calculating the average *on respondent level*. This is done by adding the time spent on each survey question and dividing it by the number of survey questions. This results in one value for each respondent, which indicates how long a respondent took on average to complete a survey question. However, it is more common in such a setting to aggregate *vertically across respondents*, resulting in one value per indicator studied. In this case, the average completion time for a survey question is calculated by taking each respondent's time to answer the survey question and calculating the mean across respondents.

Contact information and device-type paradata are typically aggregated on the respondent level. For each respondent, one value is then noted for the number of contacts, the time of first and last access, the device they used, etc. If respondents have changed device or browser type, this data may still be aggregated at the respondent level to indicate the binary outcome of whether a change has occurred (or not). However, survey settings that require multiple devices are imaginable, in which case a more detailed unit of analysis may become necessary. The same applies to changes in screen orientation, which are often aggregated on the respondent level but may be of analytical interest to researchers at the question or page level in the course of pretesting and improving questionnaire design.

Figure 3: Direction of aggregation of web paradata for analysis, using the example of response times (own illustration)

Questionnaire navigation is often carried out on the level of the question or page. Back-up to previous survey pages are aggregated to the page level. Response times, answer changes, or prompts are typically aggregated to the question or page level. However, in some cases, a higher level of aggregation becomes necessary. For instance, in an experiment on answer changes depending on different response scale designs, Heerwegh (2011) found too few answer changes on question level to carry out meaningful analysis, and therefore aggregated to the page level, marking whether an answer change took place at all throughout several questions. Rare events such as answer changes, window switching, and scrolling are thus likely to require aggregation to a higher level for analysis.

When paradata from prior survey waves are part of the analysis (e.g., number of contacts in wave X, previous mobile device use, response times), these data are also usually aggregated on the respondent level.

3.2.3 Outliers, skewness and zero inflation

Paradata that constitute metric variables are subject to outliers (unusually high or low values) and potentially skewed distribution. Thus, outlier treatment and decisions about a possible transformation of these paradata are important analytic decisions (see, for example, Leys, Ley, Klein, Bernard, & Licata, 2013; Matjašic, Vehovar, & Lozar Manfreda, 2018; Ratcliff, 1993). Response time measures are a typical case in which these decisions must be made; however, potentially all metric web paradata may be subject to outliers and skewed distribution, such as the number of mouse clicks and answer changes, or the number of contacts.

Several outlier definitions exist. They can be based on the mean value and standard deviations (e.g., $\bar{x} \pm 2SD$), which is, in most cases, the method of choice (Yan & Olson, 2013). As both the mean value and standard deviation are, however, affected by outliers, using the median and the interquartile range has been suggested (e.g., $Q_2 \pm 1.5[Q_3 - Q_1]$) (Höhne & Schlosser, 2018; Leys et al., 2013). Other studies use cut-off values at the top/bottom one (or five) percentiles to define outliers (e.g., Q_1 and Q_{99}). We recommend examining the descriptives of the raw data before choosing an outlier definition. Furthermore, we recommend testing the stability of analyses by performing them with more than one outlier definition, as many current studies do (e.g., Revilla & Couper, 2018).

Moreover, there are several ways of dealing with outliers. They can be excluded from analysis (i.e., set to missing), or replaced by alternative values, such as the cut-off point (e.g., the top/bottom percentile), mean or median value.

For analyses requiring data with normal distribution, response times, and other skewed measures generally need to be transformed. Standard procedures to transform skewed data are logarithmic, square root, and reciprocal transformation. The effect of different transformation methods on reducing skewness should be carefully examined (Stocké, 2004).

In cases where paradata measures are count variables with low arithmetic mean (typically <10), appropriate regression models must be used (e.g., Poisson regression, binomial regression) (Coxe, West, & Aiken, 2009; Heerwegh, 2011). To account for zero-inflation and possible under- or overdispersion with count data, a marginalized zero-inflated (generalized) Poisson regression can be performed (Cummings & Hardin, 2019).

3.2.4 Interpretation

Depending on the type of web paradata, interpretation of the paradata-based measures may be more or less straightforward.

In the case of contact information and device-type paradata, interpretation is generally unambiguous. Often, measures derived from these paradata are used as explanatory variables when interpreting (non-) response behavior and data quality. For instance, a researcher may hypothesize that the length of response to open-ended questions is higher among respondents using desktop devices such as PCs or laptops than among respondents using handheld devices such as smartphones. In this case, the researcher's primary concern will not be how to interpret the device type. Instead, the main effort lies in processing the device-type paradata to assign respondents to larger and smaller devices correctly.

In contrast, questionnaire navigation paradata and the measures derived from them are generally used as indicators for an underlying construct of interest, which are mainly aspects of cognitive processing. The research design and survey setting strongly determine the interpretation of these measures. For example, long response times can be a good, bad, or no indication of response quality. In some settings, long response times point to poor question design, consequent respondent confusion, and respondent difficulties answering. In contrast, in other settings, they act as a sign of increased cognitive effort and respondent motivation. Table 3 shows common questionnaire navigation paradata indicators and their possible interpretations (represents only an extract).

Table 3: Examples of questionnaire navigation paradata indicators and their exemplary interpretation (example studies in parentheses)

Indicating:	Response times	Answer changes	Window switching
Poor question design	Longer response times indicate poor or difficult question design (Healey, 2007; Lenzner et al., 2010; Revilla, Toninelli, & Ochoa, 2017; Selkälä & Couper, 2018; Smyth & Olson, 2018; Stern, 2008)	Answer changes indicate poor or difficult question design (Healey, 2007; Höhne et al., 2017; Revilla & Couper, 2018; Stern, 2008; Stieger & Reips, 2010)	-
Cognitive effort and optimizing	Longer response times indicate optimizing behavior, as active and intensive cognitive response processing requires more effort and time (Callegaro, Yang, Bhola, Dillman, & Chin, 2009; Höhne et al., 2017; Kaczmirek, 2009; Lenzner et al., 2010; Toepoel, Das, & van Soest, 2008)	Answer changes indicate optimizing behavior, as respondents try to select the most appropriate answer option (Heerwegh, 2003; Kunz, Landesvatter, & Gummer, 2020; Selkälä & Couper, 2018; Stern, 2008)	Frequent window switching indicates potential boredom or frustration (Revilla & Couper, 2018) *or* low response engagement (Kunz, Landesvatter, et al., 2020)
Multitasking and distraction	Longer response times indicate that respondents spend time on something other than the survey (Antoun & Cernat, 2020; Höhne & Schlosser, 2018; Revilla & Ochoa, 2015; Sendelbah et al., 2016)	-	Window switching indicates respondent distraction or multitasking (Höhne, Schlosser, Couper, & Blom, 2020; Revilla & Couper, 2018)
Confusion	Longer response times indicate difficulties with processing and understanding the question (Christian, Parsons, & Dillman, 2009; Couper, Tourangeau, Conrad, & Singer, 2006; Funke, Reips, & Thomas, 2011)	Reciprocal answer changes indicate respondent confusion (e.g., regarding response scale or correct data entry) (Giroux, Tharp, & Wietelman, 2019; Heerwegh, 2011; Stern, 2008)	-
Practice effects	Shorter response times reflect practice due to repetition of the same task (Couper et al., 2006; Couper & Zhang, 2016; Revilla et al., 2017)	-	-
For knowledge questions: Lack of knowledge	Longer response times indicate a lack of knowledge (Heerwegh, 2003)	Answer changes indicate guessing/lack of knowledge (Heerwegh, 2003)	-
For knowledge questions: Looking up answers	Longer response times indicate looking up an answer (Clifford & Jerit, 2016; Munzert & Selb, 2015) *vs* shorter response times indicate looking up an answer (Jensen & Frolund Thomsen, 2014)	-	Window switching indicates looking up answers and thus cheating (Diedenhofen & Musch, 2017) *vs* over-optimizing (Gummer & Kunz, 2019)
For attitude questions: Attitude stability	Longer response times indicate unstable attitudes (Heerwegh, 2003)	Answer changes indicate unstable attitudes (Heerwegh, 2003)	-

Researchers are generally advised to follow two main premises regarding the interpretation of web paradata:

1. The interpretation of paradata measures must be guided by a relevant theory (Yan & Olson, 2013).
2. Researchers should not rely on one paradata variable or paradata-based indicator, but validate their results by examining multiple paradata-based measures (Revilla & Couper, 2018) or examining paradata-based measures in conjunction with other indicators (Antoun & Cernat, 2020; Revilla & Ochoa, 2015; Zhang & Conrad, 2013).

3.3 Documentation

Given the open science efforts in scientific research, survey researchers should strive for transparency and reproducibility of their web paradata collection and use. This implies comprehensive documentation of the measurement and processing of web paradata.

Documentation should include:

- Technical implementation (i.e., server- and/or client-side, software and/or script)
- Overview of collected paradata (e.g., user agent string, time stamps)
- Consent procedure (i.e., implicit or explicit informed consent)
- Data quality of the paradata (i.e., how data quality was assessed, how imperfect data were handled)
- Data processing (e.g., outlier definitions, the number of outliers removed and their value, transformation)
- Unit of analysis (i.e., survey, page, question or respondent level)

For general information on documentation of online surveys, we refer interested readers to Schaurer, Kunz, & Heycke (2020).

4 References

Antoun, C., & Cernat, A. (2020). Factors Affecting Completion Times: A Comparative Analysis of Smartphone and Pc Web Surveys. *Social Science Computer Review*, *38*(4), 477–489. https://doi.org/10.1177/0894439318823703

Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of Mobile Versus Pc Web on Survey Response Quality. A Crossover Experiment in a Probability Web Panel. *Public Opinion Quarterly*, *81*(S1), 280–306. https://doi.org/10.1093/poq/nfw088

Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018). Design Heuristics for Effective Smartphone Questionnaires. *Social Science Computer Review*, *36*(5), 557–574. https://doi.org/10.1177/0894439317727072

Beuthner, C., Daikeler, J., & Silber, H. (2019). *Mixed-Device and Mobile Web Surveys.* Mannheim, GESIS – Leibniz Institute for the Social Sciences (GESIS – Survey Guidelines). https://doi.org/10.15465/gesis-sg_en_028

Callegaro, M. (2010). Do You Know Which Device Your Respondent Has Used to Take Your Online Survey? *Survey Practice*, *3*(2). https://doi.org/10.29115/SP-2010-0028

Callegaro, M. (2013). Paradata in Web Surveys. In F. Kreuter (Ed.), *Improving Surveys with Paradata. Analytic Uses of Process Information* (pp. 261–279). Hoboken, NJ: Wiley & Sons.

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015). *Web Survey Methodology*. Los Angeles: 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

Christian, L. M., Parsons, N. L., & Dillman, D. A. (2009). Designing Scalar Questions for Web Surveys. *Sociological Methods and Research*, *37*(3), 393–425. https://doi.org/10.1177/0049124108330004

Clifford, S., & Jerit, J. (2016). Cheating on Political Knowledge Questions in Online Surveys. An Assessment of the Problem and Solutions. *Public Opinion Quarterly*, *80*(4), 858–887. https://doi.org/10.1093/poq/nfw030

Conrad, F. G., Schober, M. F., & Coiner, T. (2007). Bringing Features of Human Dialogue to Web Surveys. *Applied Cognitive Psychology*, *21*, 165–187. https://doi.org/10.1002/acp.1335

Conrad, F. G., Tourangeau, R., Couper, M., & Zhang, C. (2017). Reducing Speeding in Web Surveys by Providing Immediate Feedback. *Survey Research Methods*, *11*(1), 45–61. https://doi.org/10.18148/srm/2017.v11i1.6304

Couper, M. P. (1998). *Measuring Survey Quality in a Casic Environment*. Proceedings of American Statistical Association: Survey Research Methods Section, 41-49. Retrieved from http://ww2.amstat.org/sections/srms/Proceedings/.

Couper, M. P., Tourangeau, R., Conrad, F. G., & Singer, E. (2006). Evaluating the Effectiveness of Visual Analog Scales. A Web Experiment. *Social Science Computer Review*, *24*(2), 227–245. https://doi.org/10.1177/0894439305281503

Couper, M. P., & Zhang, C. (2016). Helping Respondents Provide Good Answers in Web Surveys. *Survey Research Methods*, *10*(2), 49–64. https://doi.org/10.18148/srm/2016.v10i1.6273

Coxe, S., West, S. G., & Aiken, L. S. (2009). The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives. *Journal of Personality Assessment*, *91*(2), 121–136. https://doi.org/10.1080/00223890802634175

Cummings, T. H., & Hardin, J. W. (2019). Modeling Count Data with Marginalized Zero-Inflated Distributions. *The Stata Journal*, *19*(3), 499–509. https://doi.org/10.1177/1536867X19874209

de Bruijne, M., & Wijnant, A. (2014). Improving Response Rates and Questionnaire Design for Mobile Web Surveys. *Public Opinion Quarterly*, *78*(4), 951–962. https://doi.org/10.1093/poq/nfu046

Diedenhofen, B., & Musch, J. (2017). Pagefocus: Using Paradata to Detect and Prevent Cheating on Online Achievement Tests. *Behavior Research Methods*, *49*(4), 1444–1459. https://doi.org/10.3758/s13428-

016-0800-7
Funke, F., Reips, U.-D., & Thomas, R. K. (2011). Sliders for the Smart: Type and Rating Scale on the Web Interacts with Educational Level. *Social Science Computer Review*, *29*(2), 221–231. https://doi.org/10.1177/0894439310376896
Giroux, S., Tharp, K., & Wietelman, D. (2019). Impacts of Implementing an Automatic Advancement Feature in Mobile and Web Surveys. *Survey Practice*, *12*(1). https://doi.org/10.29115/SP-2018-0034
Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Hoboken, NJ: Wiley.
Gummer, T., & Kunz, T. (2019). Relying on External Information Sources When Answering Knowledge Questions in Web Surveys. *Sociological Methods & Research*, *online first*. https://doi.org/10.1177/0049124119882470
Haan, M., Lugtig, P., & Toepoel, V. (2019). Can We Predict Device Use? An Investigation into Mobile Device Use in Surveys. *International Journal of Social Research Methodology*, *22*(5), 517–531. https://doi.org/10.1080/13645579.2019.1593340
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. (2002). *Describing Response Behavior in Web Surveys Using Client Side Paradata*. Paper presented at the International Workshop on Web Surveys, Mannheim, Germany.
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. (2011). Internet Survey Paradata. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet* (pp. 325–348). New York: Routledge.
Höhne, J. K., & Schlosser, S. (2018). Investigating the Adequacy of Response Time Outlier Definitions in Computer-Based Web Surveys Using Paradata Surveyfocus. *Social Science Computer Review*, *36*(3), 369–378. https://doi.org/10.1177/0894439317710450
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*, *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
Holmberg, A., Lorenc, B., & Werner, P. (2010). Contact Strategies to Improve Participation Via the Web in a Mixed-Mode Mail and Web Survey. *Journal of Official Statistics*, *26*(3), 465–480.
Horwitz, R., Kreuter, F., & Conrad, F. (2017). Using Mouse Movements to Predict Web Survey Response Difficulty. *Social Science Computer Review*, *35*(3), 388–405. https://doi.org/10.1177/0894439315626360
Hupp, A., Schroeder, H., & Piskorowski, A. (2017). Use of Email Paradata in a Survey of Sustainability Culture. In W. Leal Filho, R. W. Marans, & J. Callewaert (Eds.), *Handbook of Sustainability and Social Science Research* (pp. 387–399). Springer International Publishing.
Jensen, C., & Frolund Thomsen, J. P. (2014). Self-Reported Cheating in Web Surveys on Political Knowledge. *Quality & Quantity*, *48*, 3343–3354. https://doi.org/10.1007/s11135-013-9960-z
Kaczmirek, L. (2008). *Human-Survey Interaction. Usability and Nonresponse in Online Surveys.* Mannheimer sozialwissenschaftliche Abschlussarbeiten, 08/005.
Kaczmirek, L. (2009). *Human-Survey Interaction. Usability and Nonresponse in Online Surveys*. Köln: Halem.
Kaczmirek, L., & Neubarth, W. (2007). Nicht-Reaktive Datenerhebung: Teilnahmeverhalten bei Befragun-

gen mit Paradaten evaluieren. In M. Welker & O. Wenzel (Eds.), *Online-Forschung 2007. Grundlagen und Fallstudien* (pp. 293–311). Köln: Herbert von Halem Verlag.

Kern, C., Weiss, B., & Kolb, J.-P. (2019). *A Longitudinal Framework for Predicting Nonresponse in Panel Surveys.* Working Paper. Retrieved from https://arxiv.org/abs/1909.13361

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to Participate in Passive Mobile Data Collection. *Public Opinion Quarterly*, *83*(S1), 210–235. https://doi.org/10.1093/poq/nfz007

Keusch, F., & Yan, T. (2017). Web Versus Mobile Web: An Experimental Study of Device Effects and Self-Selection Effects. *Social Science Computer Review*, *35*(6), 751–769. https://doi.org/10.1177/0894439316675566

Kreuter, F. (2015). The Use of Paradata. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Eds.), *Improving Survey Methods. Lessons from Recent Research* (pp. 303–315). New York: Routledge.

Kunz, T., Beuthner, C., Hadler, P., Roßmann, J., & Schaurer, I. (2020). *Informing about Web Paradata Collection and Use.* Mannheim, GESIS – Leibniz Institute for the Social Sciences (GESIS – Survey Guidelines). https://doi.org/10.15465/gesis-sg_036

Kunz, T., & Fuchs, M. (2019a). Dynamic Instructions in Check-All-That-Apply Questions. *Social Science Computer Review*, *37*(1), 104–118. https://doi.org/10.1177/0894439317748890

Kunz, T., & Fuchs, M. (2019b). Using Experiments to Assess Interactive Feedback That Improves Response Quality in Web Surveys. In P. J. Lavrakas, M. W. Traugott, C. Kennedy, A. L. Holbrook, E. D. de Leeuw, & B. T. West (Eds.), *Experimental Methods in Survey Research: Techniques That Combine Random Sampling with Random Assignment* (pp. 247–274). Hoboken, NJ: John Wiley & Sons.

Kunz, T., Landesvatter, C., & Gummer, T. (2020). Informed Consent for Paradata Use in Web Surveys. *International Journal of Market Research*, *62*(4), 396–408. https://doi.org/10.1177/1470785320931669

Kypri, K., Samaranayaka, A., Connor, J., Langley, J. D., & Maclennan, B. (2011). Non-Response Bias in a Web-Based Health Behaviour Survey of New Zealand Tertiary Students. *Preventive Medicine*, *53*(4), 274–277. https://doi.org/10.1016/j.ypmed.2011.07.017

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

Lewis, T., & Hess, K. (2017). The Effect of Alternative E-Mail Contact Timing Strategies on Response Rates in a Self-Administered Web Survey. *Field Methods*, *29*(4), 351–364. https://doi.org/10.1177/1525822X17715865

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting Outliers: Do Not Use Standard Deviation around the Mean, Use Absolute Deviation around the Median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lugtig, P., & Toepoel, V. (2016). The Use of Pcs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error. *Social Science Computer Review*, *34*(1), 78–94. https://doi.org/10.1177/0894439315574248

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

Matjašic, M., Vehovar, V., & Lozar Manfreda, K. (2018). Web Survey Paradata on Response Time Outliers: A Systematic Literature Review. *Metodoloski Zvezki*, *15*(1), 23–41.

Mavletova, A. (2013). Data Quality in Pc and Mobile Web Surveys. *Social Science Computer Review*, *31*(6), 725–743. https://doi.org/10.1177/0894439313485201

Mavletova, A., & Couper, M. P. (2014). Mobile Web Survey Design: Scrolling Versus Paging, Sms Versus E-Mail Invitations. *Journal of Survey and Methodology*, *2*(4), 498–518. https://doi.org/10.1093/jssam/smu015

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

McGeeney, K., & Yan, H. Y. (2016). *Text Message Notification for Web Surveys. Sending Texts to Survey Panel Members Shortens Response Time.* Numbers, Facts and Trends shaping the World, 1-17. Retrieved from https://www.pewresearch.org/methods/2016/09/07/text-message-notification-for-web-surveys/

Metzler, A. (2020). *The Effect of Assigning Sample Members to Their Preferred Device on Nonresponse and Measurement in Web Surveys.* Doctoral Dissertation (PhD thesis, Darmstadt University of Technology). Retrieved from http://tuprints.ulb.tu-darmstadt.de/8788/

Millar, M. M., & Dillman, D. A. (2011). Improving Response to Web and Mixed-Mode Surveys. *Public Opinion Quarterly*, *75*(2), 249–269. https://doi.org/10.1093/poq/nfr003

Munzert, S., & Selb, P. (2015). Measuring Political Knowledge in Web-Based Surveys: An Experimental Validation of Visual Versus Verbal Instruments. *Social Science Computer Review*, *35*(2), 167–183. https://doi.org/10.1177/0894439315616325

Olson, K., & Parkhurst, B. (2013). Collecting Paradata for Measurement Error Evaluations. In F. Kreuter (Ed.), *Improving Surveys with Paradata. Analytic Uses of Process Information* (pp. 43–72). Hoboken: Wiley.

Peytchev, A. (2009). Survey Breakoff. *Public Opinion Quarterly*, *73*(1), 74–97. https://doi.org/10.1093/poq/nfp014

Peytchev, A., Couper, M. P., McCabe, S. E., & Crawford, S. D. (2006). Web Survey Design. Paging Versus Scrolling. *Public Opinion Quarterly*, *70*(4), 596–607. https://doi.org/10.1093/poq/nfl028

Ratcliff, R. (1993). Methods for Dealing with Reaction Time Outliers. *Psychological Bulletin*, *114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Revilla, M. A., & Couper, M. P. (2018). Comparing Grids with Vertical and Horizontal Item-by-Item Formats for Pcs and Smartphones. *Social Science Computer Review*, *36*(3), 349–368. https://doi.org/10.1177/0894439317715626

Revilla, M. A., & Ochoa, C. (2015). What Are the Links in 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

Revilla, M. A., Toninelli, D., & Ochoa, C. (2017). An Experiment Comparing Grids and Item-by-Item Formats in Web Surveys Completed through Pcs and Smartphones. *Telematics and Informatics*, *34*(1), 30–42. https://doi.org/10.1016/j.tele.2016.04.002

Revilla, M. A., Toninelli, D., Ochoa, C., & Loewe, G. (2016). Do Online Access Panels Need to Adapt Surveys for Mobile Devices? *Internet Research*, *26*(5), 1209–1227. https://doi.org/10.1108/IntR-02-2015-0032

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

Sakshaug, J. W., & Antoni, M. (2017). Errors in Linking Survey and Administrative Data. In P. P. Biemer, E. De Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, … B. T. West (Eds.), *Total Survey Error in Practice* (pp. 557–573). https://doi.org/10.1002/9781119041702.ch25

Schaurer, I., Kunz, T., & Heycke, T. (2020). *Documentation of Online Surveys.* Mannheim, GESIS – Leibniz Institute for the Social Sciences (GESIS – Survey Guidelines). https://doi.org/10.15465/gesis-sg_en_031

Schlosser, S., & Höhne, J. K. (2020). *Ecsp - Embedded Client Side Paradata*. Retrieved from https://zenodo.org/record/3782592#.XuxrAmgzY2w

Selkälä, A., & Couper, M. P. (2018). Automatic Versus Manual Forwarding in Web Surveys. *Social Science Computer Review*, *36*(6), 669–689. https://doi.org/10.1177/0894439317736831

Sendelbah, A., Vehovar, V., Slavec, A., & Petrovcic, A. (2016). Investigating Respondent Multitasking in Web Surveys Using Paradata. *Computers in Human Behavior*, *55*, 777–787. https://doi.org/10.1016/j.

chb.2015.10.028
Smith, T. W. (2011). The Report of the International Workshop on Using Multi-Level Data from Sample Frames, Auxiliary Databases, Paradata and Related Sources to Detect and Adjust for Nonresponse Bias in Surveys. *International Journal of Public Opinion Research*, *23*(3), 389–402. https://doi.org/10.1093/ijpor/edr035
Smyth, J. D., & Olson, K. (2018). The Effects of Mismatches between Survey Question Stems and Response Options on Data Quality and Responses. *Journal of Survey Statistics and Methodology*, *7*(1), 34–65. https://doi.org/10.1093/jssam/smy005
Sommer, J., Diedenhofen, B., & Musch, J. (2016). Not to Be Considered Harmful: Mobile-Device Users Do Not Spoil Data Quality in Web Surveys. *Social Science Computer Review*, *Article first published online*. https://doi.org/10.1177/0894439316633452
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*, 1488–1495. https://doi.org/10.1016/j.chb.2010.05.013
Stocké, V. (2004). Measuring Information Accessibility and Predicting Response-Effects: The Validity of Response-Certainties and Response-Latencies. *Metodoloski zvezki*, *1*(1), 33–55.
Tienda, M., & Koffman, D. (2020). Using Paradata to Evaluate Youth Participation in a Digital Diary Study. *Social Science Computer Review*, *online first*. https://doi.org/10.1177/0894439320929272
Toepoel, V., Das, M., & van Soest, A. (2008). Effects of Design in Web Surveys. Comparing Trained and Fresh Respondents. *Public Opinion Quarterly*, *72*(5), 985–1007. https://doi.org/10.1093/poq/nfn060
Toninelli, D., & Revilla, M. A. (2016). Smartphones Vs Pcs: Does the Device Affect the Web Survey Experience and the Measurement Error for Sensitive Topics? A Replication of the Mavletova & Couper's 2013 Experiment. *Survey Research Methods*, *10*(2), 153–169. https://doi.org/10.18148/srm/2016.v10i2.6274
Tourangeau, R., Brick, J. M., Lohr, S., & Li, J. (2017). Adaptive and Responsive Survey Designs: A Review and Assessment. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *180*(1), 203–223. https://doi.org/10.1111/rssa.12186
Tourangeau, R., Sun, H., Yan, T., Maitland, A., Rivero, G., & Williams, D. (2018). Web Surveys by Smartphones and Tablets: Effects on Data Quality. *Social Science Computer Review*, *36*(5), 542–556. https://doi.org/10.1177/0894439317719438
Verbree, A.-R., Toepoel, V., & Perada, D. (2019). The Effect of Seriousness and Device Use on Data Quality. *Social Science Computer Review*, *online first*. https://doi.org/10.1177/0894439319841027
Yan, T., & Olson, K. (2013). Analyzing Paradata to Investigate Measurement Error. In F. Kreuter (Ed.), *Improving Surveys with Paradata. Analytic Uses of Process Information* (pp. 73–95). Hoboken: Wiley.
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*, 51–68. https://doi.org/10.1002/acp.1331
Zhang, C., & Conrad, F. G. (2013). Speeding in Web Surveys: The Tendency to Answer Very Fast and Its Association with Straightlining. *Survey Research Methods*, *8*(2), 127–135. https://doi.org/10.18148/srm/2014.v8i2.5453
Zijlstra, T., Wijgergangs, K., & Hoogendoorn-Lanser, S. (2018). Traditional and Mobile Devices in Computer Assisted Web-Interviews. *Transportation Research Procedia*, *32*, 184–194. https://doi.org/10.1016/j.trpro.2018.10.033

5 Appendix A - Missing and inconsistent data: An example

To visualize how to identify missing and inconsistent paradata, we use the first three paradata strings from the client-side script UCSP (Kaczmirek & Neubarth, 2007).

The first paradata string of the UCSP (String A) might look like this:

#6#0;0;18452;20371;21363;4;0;1536;750;0;0;0;0;1;315;0

Position	String	Content
1	#6#0;	version of script
2	0;	do answer check carried out [1=yes]
3	18452;	time to first click [ms]
4	20371;	time to second to last click [ms]
5	21363;	time to last click [ms]
6	4;	number of clicks on page
7	0;	number of double clicks on page
8	1536;	window width [pixel]
9	750;	window height [pixel]
10	0;	maximum width of scrolling [pixel]
11	0;	maximum height of scrolling [pixel]
12	0;	time to first keystroke [ms]
13	0;	time to last keystroke [ms]
14	1;	survey focus [window switching=1]
15	315;	duration of loss of survey focus [ms]
16	0	string from previous page visit

Figure 4: Example of non-rectangular string paradata

Figure 4 provides an overview of the variables captured in this string. String A summarizes client-side questionnaire navigation paradata that were collected on the survey page level, such as the total number of clicks (position 6 of the string) or the total time spent on the survey page (position 5), but also device-type paradata such as window size (positions 8 and 9).

Figure 5 shows the same paradata string for five respondents. For the third respondent, all paradata variables have the value “0” or “undefined” for this survey page. While “0” is a plausible entry for some positions of the string, it is undoubtedly not plausible for response times or window size. It seems that

ID	UCSP String A
respondent01	#6#0;0;8624;41711;42502;9;0;1680;914;0;0;0;0;0;0;0
respondent02	#6#0;0;7951;35918;37184;9;0;1680;802;0;0;0;0;0;0;0
respondent03	#6#0;0;0;0;0;0;0;undefined;undefined;0;0;0;0;0;0;0
respondent04	#6#0;0;3623;0;3623;1;0;958;927;0;0;7506;11750;1;1359;0
respondent05	#6#0;0;1337;2191;8753;3;0;1536;750;0;0;4926;6333;0;0;0

Figure 5: Example of missing paradata

all paradata variables are missing in this string. The researcher should check the paradata collection of the other paradata strings from this survey page and also paradata collection on other survey pages to determine whether the script did not function in general for this respondent. The paradata strings of the other four respondents appear plausible at first sight.

ID	V1	V2	V3	V4	V5	V6	V7	V8	UCSP String B	UCSP String C
respondent01	1	4	8	2	2	1	1	1	v_1x1;v_2x4;v_3x8;v_4x2;v_5x2;v_6x1;v_7x1;v_8x1;os;	8624;5246;6937;7487;3720;3017;3343;3337;791;
respondent02	2	4	5	4	2	2	2	2	v_1x2;v_2x4;v_3x5;v_4x4;v_5x2;v_6x2;v_7x2;v_8x2;os;	7951;4918;4142;3837;4964;2626;3236;4244;1266;
respondent03	1	5	5	5	1	1	1	1	-99	-99
respondent04	4	2	4	2	2	2	4	2	v_1x2;	3623;
respondent05	2	4	4	2	2	5	5	4	TD;v_1x2;os;	1337;854;6562;

Figure 6: Example of inconsistent paradata

Figure 6 shows the survey answers (V1 to V8) and the second and third paradata string (UCSP String B and C) for the same five respondents. String B depicts each mouse click. String C shows the time stamps for each click. The missing values in strings B and C for respondent 3 confirm that no paradata were collected on this survey page. If the other survey pages also contain no paradata for this respondent, the script possibly did not function on the respondent's device.

Examining the survey data shows that paradata collection is at least partially incorrect or incomplete for respondents 4 and 5. Both respondents answered all eight survey items, as can be seen by the valid entries for V1 to V8. However, according to paradata string A, respondent 4 only clicked once, and respondent 5 only three times. The number of clicks on the survey page and the survey data are inconsistent. A closer look at strings B and C show that in both cases, the paradata script captured the answer to the first item only. For respondent 5, it also captured a click beside the radio button and the click on the "Next" button. Based on this, the researcher must assume that the response times for these respondents are incorrect as they are not based on all entries made on the page. Indeed, while respondents 1 and 2 spent 42 and 37 seconds, respectively, on the survey page, respondents 4 and 5 only spent 3 and 8 seconds on the survey page. In such a case, it is recommendable to set the time measures and the number of clicks to missing for analysis.

Other paradata variables of respondents 4 and 5, such as window size, do not appear to be affected. String A indicates that respondent 4 switched windows on this survey page. Possibly, multitasking leads to an interruption of the script for this respondent. However, this is merely speculation, and it is recommended to examine the paradata strings of these respondents on other survey pages to test whether the script functions were only interrupted on this individual survey page, or whether certain functions did not work for these respondents throughout the survey.

In summary, all paradata strings must be examined in detail for missing and inconsistent data. Documentation should describe in which ways the data was checked. Data handling of such cases, such as setting specific values to missing, must be done systematically. Documentation should include how many cases were declared missing or inconsistent.