National Research University
Higher School of Economics

As a manuscript

Daniil Lebedev

Paradata opportunities to measure and improve data quality in web surveys

Thesis for the purpose of obtaining academic degree PhD in Sociology

Academic Supervisor:
PhD in Sociology
Aigul Klimova

Moscow, 2023

Development of the research question

Data collection in the social sciences relies heavily on computer-assisted data survey modes. These include web surveys, computer-assisted personal interviewing (CAPI), computer-assisted self- interviewing (CASI), and computer-assisted telephone interviewing (CATI). It is worth noting that CASI interviews are conducted in the presence of an interviewer, with a respondent filling out the questionnaire themselves, while online surveys imply self-completion of a web questionnaire by a respondent without the presence of an interviewer. Web survey mode has become the most prominent method of data collection in public opinion research (ESOMAR, 2018). However, data quality in these survey modes raises a lot of concerns, including sampling bias (Couper, 2000), data equivalence with other modes (Hox, De Leeuw and Zijlmans, 2009) or between different devices within one mode (Mavletova, 2013; Mavletova, Couper and Lebedev, 2018), respondents' multitasking when completing a web survey (Sendelbach et al., 2016; Höhne et al., 2020), noncompliance with instructions (Gummer and Kunz, 2019), fabrication and falsifications by the interviewers (Murphy et al., 2016), etc. All of these may lead to lower data quality and biased results that can produce wrong economic, social, and policy decisions. In this regard, survey methodologists have done a lot to improve data quality in computer-assisted survey modes.

Recent advances in addressing this issue include augmenting survey data with additional data, such as administrative (Kreuter, Müller and Trappmann, 2010), mobile apps (Strumiskaya et al., 2020), or social media (Kühne and Zindel, 2020). Another promising and widely used approach to measure and improve data quality in web surveys is ***paradata collection***. That is the focus of my thesis.

Paradata can be described as additional information collected in the process of interview completion (Couper, 1998; Durrant and Kreuter, 2013; West, 2011; Matjasic, Vehovar and Manfreda, 2018). This is data about the process itself, including the behavioral characteristics of the interviewer, respondent, and description of the survey environment. It also includes data measured using supplementary devices (Kreuter and Casas-Cordero, 2010). This data may include

information on survey completion time, interview location, interviewer reports, interview audio recordings, survey navigation, browser focus change, mouse movements, pupil fixations and diameter dynamics (McClain et al., 2019), etc.

Paradata are not available prior to fieldwork, but are created and modified duringthe process, providing an opportunity to reveal some patterns of respondent andinterviewer behavior during the interview (Kaczmirek, 2008; McClain et al., 2019).

Survey research focuses on bias and variance at every survey design stage, as well as on validity and reliability as key data quality indicators (Groves, 1987; Groves and Lyberg, 2010). Total Survey Error (TSE) is the most inclusive theoretical framework for data quality control and assessment that addresses all survey errors from a construct and inferential population (representation) to a survey statistic in both measurement and representation survey quality perspectives (Groves and Lyberg, 2010). These errors include coverage error, sampling error, nonresponse error, specification error, data processing error, measurement error, and others. Overall, the methodological goal is to minimize the sum of all possible survey errors. In this research, I focus on measurement error which is "one of the most damaging sources of error" (Biemer, 2010: 823).

Measurement error is the difference between the true value and provided value. Measurement error may occur when respondents provide incorrect information, interviewers fabricate or falsify the survey data, or unintentionally influence respondents' answers in personal interviews. That also includes erroneous questionnaire design (question wording, format) and the interview environment (presence of third parties, interview location, noise) (Biemer, 2010). In general, these situations occur when there is a systematic breakdown in the cognitive processof providing an answer based on the R. Tourangeau's survey response model: perception, comprehension, retrieval, judgment, and editing (Tourangeau, Rips and Rasinski, 2000; Dillman, Smyth and Christian, 2009; Olson and Parkhurst, 2013).

Though there are some studies that show how certain types of paradata can be applied in certain computer-assisted data survey modes, research in this field is quite

fragmented, and there is no comprehensive scientific work that shows how different types of paradata (including such advanced paradata as GPS and pupil diameter) can be applied in different survey modes (CAWI, CAPI, CASI) for measurement error evaluation and correction. My thesis will fill this gap in the literature. It will be based on a CASI laboratory experiment (pupil diameter dynamics), CAWI experimental study (device type paradata), and tablet-based CAPI study (GPS paradata).

The paradata is indeed of great value and offers considerable promise for methodological research. I will show that it is essential to use information about the behavior of respondents and interviewers within the survey or during data collection process (i.e., paradata) to assess and mitigate measurement error (Mavletova, Couper and Lebedev, 2018: 665; Deviatko, Bogdanov and Lebedev, 2021: 45-46; Lebedev, 2020: 23-24; Lebedev, 2022: 25-27). As a result, our ability to address the growing and quite reasonable criticism of data quality in computer-assisted data collection survey modes will increase dramatically.

Development of the research question

The different types of paradata that are available for data collection and analysis depend on the survey mode. CAPI allows to capture timestamps (completion time), survey navigation paradata (on the level of an interviewer), general interface paradata (including GPS location), respondent and interviewer vocal characteristics (audio recordings), interviewer evaluations of respondent behavior (e.g., level of motivation and cooperation), and the survey environment in general (e.g., presence of third parties). Web survey certainly allows for the most advanced and wide set of paradata – timestamps, survey navigation (e.g., scrolling, changing an answer, going back and forward), mouse clicks and movements, as well as user interface data (e.g., type of device, operation system, screen size, and resolution).

Over the past 20 years, methodological researchers have developed a wide range of possible strategies for using paradata to estimate and reduce measurement error. *Completion time* is used to distinguish wording that imposes higher cognitive load on respondents (Lenzner et al., 2010). *Survey navigation* – scrolling, browser

focus, answer change, going back and forward in the questionnaire – is used to inquire on respondent or interviewer behavior during survey completion (Gummer and Kunz, 2019). *Mouse movements and clicks* were shown to be an indicator of respondents' cognitive load (Horwitz et al., 2017). *Vocal and physiological characteristics of respondents* are used to investigate how accurate is the data provided by respondents in CAPI interviews (Jans, 2010).

Although there are several detailed reviews of paradata (Kreuter and Casas-Cordero, 2010; Olson and Parkhurst, 2013; Nicolaas, 2011; McClain et al., 2019), and the field of paradata applications for data evaluation and quality improvement is actively developing, there is still no approach to using different types of paradata including indirect paradata (GPS paradata, eye tracker data), that considers different data collection modes and paradata quality itself. There is a lack of scientific work that shows how different types of paradata (including indirect types such as GPS and pupil diameter) can be applied in different survey methods (web surveys, CAPI, CASI) to estimate and reduce measurement error. This thesis will partially fill this gap in the scientific literature. It will be based on a CASI lab experiment (paradata on pupil diameter dynamics) (Devyatko, Bogdanov and Lebedev, 2021: 38-39), a web survey experiment (paradata on device type) (Mavletova, Couper and Lebedev, 2018: 651) and a CAPI study on a tablet (GPS paradata) (Lebedev, 2022: 18-20). Some examples will show how, in the context of using such data collection methods, paradata can be used to estimate and reduce measurement error.

Research question

The main research question of the thesis is as follows: How can different types of paradata (including indirect types such as GPS and pupil diameter) be used to evaluate and reduce measurement error in different computer-assisted data collection survey modes?

The ***theoretical object*** of the research is the paradata available in computer-assisted data survey modes (Groves and Lyberg, 2010). The ***subject*** of the research is the possibility of using paradata to assess and reduce measurement error in computer-assisted survey modes.

The degree of development of the research problem

The term paradata was introduced by M. Couper at American Statistical Association conference in 1998 (Couper, 1998). However, paradata as it is described now was present in the methodological and instrumental arsenal prior to that. Completion time has been used for measurement error investigation (Swanson, Brazer, 1959; Bassili and Fletcher, 1991; Bassili and Scott, 1996). A type of paradata such as call record data has been used as indicators of coverage error (Fazio, 1990; Swires–Hennessy and Drake, 1992; Eckman, 2013: 108) in CATI surveys. Since the introduction of the term, more research has emerged with a focus on paradata, especially in web surveys (Olson and Parkhurst, 2013). The main authors in this field over the past 20 years are M. Couper (Coupert, 1998; Couper, 2009; Couper and Wagner, 2011; Couper and Kreuter, 2013; Couper, 2017), F. Kreuter (Kreuter, 2013; Kreuter and Olson, 2013; Kreuter and Müller, 2015), G. Durrant (Durrant, D'Arrigo and Steele, 2011; Durrant and Kreuter, 2013; Durrant, Maslovskaya and Smith, 2017; Durrant and Maslovskaya, 2017), and M. Callegaro (Callegaro, 2013; Callegaro, 2014; Callegaro et al., 2017).

Paradata are most often used as key indicators of various types of errors and biases within the Total Survey Error framework (TSE) (Korytnikova, 2018;McClain et al., 2019; Karaeva, 2015; Rogozin and Saponov, 2014; Ipatova, 2016; Sidorov, 2011; Kreuter, 2018):

(1) coverage error is assessed using general paradata regarding the time and location of the interview/survey (Kreuter and Casas-Cordero, 2010; Eckman, 2013)

(2) paradata on interviewer location and movement during data collection are used to estimate sampling error (West, 2011; Wagner, Olson and Edgar, 2017; Choumert-Nkolo et al., 2019; Elevelt et al. 2019)

(3) interviewer observations and paradata about the process of establishing contact/communication with the respondent are used for measuring non-response and data adjustment error (Kreuter, 2018; Kreuter, 2017; Nicolaas, 2011)

(4) keyboard clicks, mouse movements, and respondent's and interviewer's behaviours during survey completion are used to assess measurement error and data

validity (Kreuter, 2018; Nicolaas, 2011; Smith, 2011; Lynn and Nicolaas, 2010).

The thesis will focus on the fourth approach, which involves the use of paradata to assess and reduce measurement error. This includes such uses of paradata as assessing respondents' cognitive load, identifying respondents' survey completion patterns (e.g., models such as "satisficing", "optimizing", "speeders" are identified in the literature), assessing questionnaire design (question wording, presentation on different devices), assessing interviewer behavior during data collection (identifying falsifications and fabrications), assessing interviewer effect and interview environment (social desirability of respondent answers) (Biemer, 2010). This dissertation study will focus on the most common sources of measurement error in computerized data collection methods: respondents' cognitive load, questionnaire design, and interviewer behavior during the data collection process (namely, identifying falsifications and fabrications through controlling the data collection process). Below is a summary of the relevant findings in the highlighted areas.

Speaking of measuring cognitive load, it is worth noting that there are different methods for measuring both using subjective assessments (Paas et al., 2003; Hart and Staveland, 1988) and using paradata - mouse movements, completion time (Stieger and Reips, 2010), etc. In recent years, the use of neurophysiological measures to measure cognitive load has become increasingly easier with the increasing availability of eye-trackerts and software to process such data, but specific studies and examples of the use of such paradata in a methodological context are still rare (Deviatko and Lebedev, 2017). For this reason, this dissertation study will assess the feasibility of measuring respondents' cognitive load using paradata (data on pupil diameter dynamics).

Overall, eye movement and pupil dilation paradata were shown to be a valid research tool for identifying respondent burden (Neuert, 2020). Survey navigation paradata (change of the response, clicking back and forward, and mouse movements) and respondent interface paradata were found to identify problematic questions with high measurement error (Stieger and Reips, 2010). Browser focus changes were found to be an indicator of multitasking and “cheating” behavior during web survey

completion (Diedenhofen and Musch, 2017; Höhne et al., 2020). Paradata was also found to identify design features that decrease survey error in web surveys. Completion time, browser focus change, and screen size/orientation were used to compare scale format (vertical vs horizontal), label format, and gamification features (Link, Lai and Bristol, 2014; Revilla and Couper, 2018; Keusch and Yan, 2019; Gummer and Kunz, 2021). GPS paradata were found to detect fraud or fabrications by interviewers in CAPI (Cecchi and Marquette, 2010; Murphy et al., 2016: 314). GPS paradata in combination with survey completion time, voice data, and call recording data proved to be an effective tool for monitoring interviewer performance as well as reducing measurement error (Mohadjer and Edwards, 2018).

Interest in the topic of paradata is growing among Russian sociologists as well. They used paradata to evaluate and increase data quality in CATI surveys (Ipatova, 2014; Ipatova, 2016; Ipatova and Rogozin, 2014; Turchik, 2010), to compare data quality between face-to-face and telephone interviews (Karaeva, 2015), in web surveys (Mavletova, 2017; Maloshonok and Terentev, 2014), and in CAPI surveys (Terentev, Mavletova and Kosolapov, 2018).

Though a lot has been studied in the field of using paradata for measurement error evaluation and reduction, the research lacks an approach summarizing the possibilities and limitations of paradata in different computer- assisted data collection modes (CAPI, CASI, web surveys). I will focus on that approach in the thesis to show the possibilities of using paradata for measurement error evaluation and reduction in different research approaches (cognitive effort evaluation, web survey design, and fieldwork monitoring) and across different data collection modes (CAPI, CASI, web surveys).

Aim and objectives of the research

The aim of the study is to evaluate the possibilities of using paradata to estimate and reduce measurement error in computer-assisted survey modes such as web surveys, CAPI, and CASI.

We had the following objectives:

1. Describe the major research directions of using paradata for evaluating and

decreasing measurement error in computer-assisted survey modes such as CAWI, CAPI, and CASI.

2. Propose typology of paradata that incorporates the physiological data and considers the quality of paradata. Based on this typology, illustrate how different types of paradata can be used for evaluating and decreasing measurement error in computer-assisted survey modes such as web surveys, CAPI, and CASI.

3. Evaluate the possibilities of using device type paradata to decrease measurement error in web surveys.

4. Evaluate the possibilities of eye tracker data (pupil diameter dynamics) in a CASI-based laboratory setting for cognitive effort evaluation to decrease measurement error.

5. Evaluate the possibilities of collecting and using GPS paradata in CAPI surveys as part of methods of fieldwork monitoring process to evaluate and decrease measurement error.

6. Evaluate paradata quality and factors associated with its change, using the example of GPS paradata quality in CAPI study.

Hypotheses

Based on our review of the literature, the following three hypotheses were developed. They are all based on our key assumption that paradata can be used to estimate and reduce measurement error in studies using computerized data collection methods such as web surveys, CAPI, and CASI.

1. Paradata about the type of device (smartphone or PC) can be used to assess and increase the equivalence of measurement between different devices and different formats of tabular question presentation in web surveys.

2. Measuring pupil diameter dynamics can be a valid method for assessing the cognitive load of both respondents and interviewers in CAPI and CASI surveys.

3. GPS paradata is a valid tool for fieldwork monitoring process in CAPI surveys. However, GPS paradata are themselves subject to measurement error, arising mainly from missing values. Consistent with the so-called "urban canyon" hypothesis, we expect more GPS paradata missing data in highly urbanized regions

and among interviewers who are less confident with the tablet and have lower success rates when switching survey methods from PAPI to CAPI.

THEORETICAL AND METHODOLOGICAL FOUNDATIONS OF THE RESEARCH

Personal contribution of the author to the development of the problem and data collection

The results of this dissertation research are cited in four published articles (Mavletova, Couper and Lebedev, 2018; Lebedev, 2020; Deviatko, Bogdanov and Lebedev, 2021; Lebedev, 2022). Thesis author took an active participation in all four papers and all research projects, on the basis of whichthese articles were produced.

In (Mavletova, Couper and Lebedev, 2018), the author participated and contributed to all research stages. A two-wave within-subject cross-over CAWI survey experiment was designed. The author programmed all versions of the web questionnaire. Overall, 1,678 respondents completed the first wave and 1,079 respondents participated in the second wave. Data were analyzed using chi-square tests, ordinary least squares regression modelling, logistic regression modelling, t-tests, negative binomial regression models, z-tests, and multigroup confirmatory factor analysis.

To measure the dynamics of pupil diameter as a proxy for cognitive load assessment, a laboratory-based CASI experiment using the eye tracker was conducted in collaboration with other researchers (Deviatko, Bogdanov and Lebedev, 2021). The author was involved in all research stages. The laboratory experiment design required accounting for possible light and noise interferences in the dynamicsof pupil diameter. We also had to develop full and concise experiment plan to avoiddifferences between experiments conducted by different moderators. The author programmed the CAPI questionnaire and provided an equivalent paper version withrandomization of question blocks. Data collection, each experiment taking up to 40minutes, was conducted in collaboration with three other moderators.

As a result, 52experiments were conducted. The collected data needed preprocessing and linkage with survey data. We manually marked the data on pupil diameter dynamics(approximately 120 000 measures for 20-minute experiment) for each experiment, marking the time when the respondent proceeded to the next page of the survey, using a video recording of a tablet screen or eye tracker (in case of Paper and Pencil Self-Interview - P&PSI condition). We also aggregated data at the respondent level and calculated baseline values for all participants. Data were analyzed by two-way repeated measures ANOVA method for the dynamics of mean pupil diameter, adjusted for baseline between the modes and questions.

Two papers were written by myself without co-authors (Lebedev, 2020, 2022). The paper (Lebedev, 2020) provides an extensive literature review of paradata, its types, possible uses in survey methodology, limitations, specifics of use, possibilities of collection and practical recommendations regarding paradata use. The paper (Lebedev, 2022) includes an extensive literature review of GPS paradata opportunities in CAPI surveys for fieldwork monitoring and analysis of GPS paradata quality. In addition, GPS paradata collected within the 26th wave of the Russian Longitudinal Monitoring Survey were analyzed (N=448) to assess their quality. Binary logistic regression and ordinary least squares modelling were used to analyze missing GPS data and their accuracy.

Theoretical basis of the study

The theoretical framework of the total survey error (TSE) was used to identify an increase in measurement error (Groves and Lyberg, 2010). Measurement error can be defined as the difference between the true value of the studied parameter and the obtained value. It includes random and systematic error. We focused on using paradata as a tool to evaluate and reduce possible biases (systematic errors) (Mavletova, Couper and Lebedev, 2018: 665; Deviatko, Bogdanov and Lebedev, 2021: 45-46; Lebedev, 2020: 23-24; Lebedev, 2022: 25-27). The term “measurement error” is included in the concept of data quality, although it is not exhaustive (non-response error, coverage error, sampling error, processing error,etc. can also affect data quality and are a part of TSE framework). Measurement biases appear in

situations of systematic breakdown in the cognitive process of providing an answer (Olson and Parkhurst, 2013). There are 5 main cognitive stepsof providing an answer for attitudinal question: comprehension, retrieval, judgment, response, editing (Tourangeau et al., 1984; Callegaro, 2005; Olson and Parkhurst, 2013).

Most importantly, the work relied on empirical research presenting ways of paradata employment for measurement error evaluation and reduction in different survey modes (Olson and Parkhurst, 2013; McClain et al., 2019; Kreuter and Casas-Cordero, 2010; Ipatova, 2014; Rogozin and Saponov, 2014; Neuert, 2020; Revilla and Couper, 2018; Stieger and Reips, 2010; Cecchi and Marquette, 2010).

Methodology and research methods

We used web surveys, laboratory-based CASI, and CAPI survey modes. It is worth noting that different sets of sources of measurement error may be relevant in the context of different data collection methods (for example, the interviewer effect will not be relevant in case of web surveys, because interviewers and their possible influence on the respondent are absent in this method of data collection).We collected paradata such as device type, GPS, and pupil diameter.

For the 1st and 2nd tasks extensive literature review is used focusing on paradata in general. The structure of the review (Lebedev, 2020) is as follows: definition of the term "paradata"; formation of a typology, taking into account different grounds for separating various types of paradata; description of areas of paradata use in research practice; description of the specifics of paradata and the difficulties of their collection, analysis and interpretation; analysis of promising areas of research related to the development of methods for the analysis and application of paradata in practice and methodological theory; a description of the possibilities for collecting paradata. The conclusion is a list of practical recommendations that will be useful to researchers planning to use paradata in practice. Review process was also focused on GPS paradata opportunities within CAPI surveys for fieldwork monitoring with description of the structure of existing methods for fieldwork monitoring, the possibilities that GPS paradata provide for data collection monitoring. process and the limitations associated with the use of

such data (Lebedev, 2022).

The 3d objective implied collecting paradata such as the type of device used by respondents to complete the web survey. The purpose was to measure differences between question sets presented in different formats (item-by-item or grid) on different devices (PC or smartphone) in a web survey experiment (Mavletova, Couper and Lebedev, 2018). We conducted a two-wave experiment with crossover design in which we varied question format and type of device. The item-by-item format included presentation of the question items on the same screen with scrolling design. Respondents were randomly allocated to one of the following four conditions in the first wave: grids/smartphone, grids/PC, item-by-item/smartphone, item-by-item/PC. In the second wave, they were asked to complete the survey on a different device. This experimental design allowed to investigate the effect of paradata such as type of device on reliability and other data quality indicators given the different question formats.

The 4th objective was presented as the collection and analysis of pupildiameter dynamics paradata in the laboratory-based CASI experiment aimed at comparing cognitive load imposed by different data collection modes. All participants were randomly assigned to one of the conditions – CASI (self- completion on a tablet) or P&PSI (self-completion in the paper survey mode). Datawas collected from 14 to 21 December, 2019. A total of 52 observations were collected (28 by CASI and 24 by P&PSI). In the final analysis for various reasons, 27 experiments were not included. The final analysis included 25 subjects (15 CASIand 10 P&PSI).

The questionnaire was developed based on the RLMS HSE questionnaire, which provided the validity of the instrument. To reduce the effect of the question order, two versions of the questionnaire were developed with a direct and reverse order of the questions. We used a simple counterbalancing of the order of questions presentation since complete randomization of the questionnaire blocks was hardly possible in the paper questionnaire. The experiment took place in a room, in which one or two researchers (including a moderator) were present but were out of the subject's field of vision not to distract respondents from completing the

questionnaire. That was necessary to reduce the influence of the external environment on pupil diameter dynamics. Lightning was also controlled to reduce external “noise” and its influence on participants’ vision. A Pupil Labs monocular eye-tracker with a sampling frequency of 200 Hz was used to collect data on pupil size and to video record survey completion on a tablet / paper (for subsequent data marking). The Samsung Galaxy Tab A 16 SM-T355 model was used in the CASI condition with the open-access software Survey Solutions.

GPS paradata (5th objective) were collected within the RLMS HSE longitudinal panel survey (Lebedev, 2022). In the 26th wave, 36 interviewers collected some of the interviews in tablet-based CAPI survey mode (vs. PAPI traditional survey mode used in the RLMS HSE). That provided the opportunity to collect GPS paradata to monitor the data collection process. A total of 491 CAPI interviews were conducted from November 2017 to February 2018. The interviewers used Samsung Galaxy Tab A 16 SM-T355. Open-source free software Survey Solutions developed by The World Bank was used to record the data.

The analysis checked missing measures in GPS paradata (either at the beginning of the interview or at the end) and accuracy of the GPS measures. The accuracy of GPS measures was represented by the average Horizontal Dilution of Precision (HDOP) score between the two values (at the beginning and at the end) for each individual CAPI interview.

The following paradata types were used in the thesis:

- Device type paradata (objective 3) was measured automatically by Enjoy Survey software as user interface data. This data was used both as a basis foranalysis and as a control for whether the assigned device was used for surveycompletion. This type of paradata showed which device (smartphone or PC) respondent used to complete the web survey.

- The pupil diameter dynamics (objective 4) was measured by the Pupil Labs eye tracker. The following settings were specified as the parameters for data collection using eye tracker in the Pupil Capture application, which allows torecord data from the device to a computer and save for subsequent analysis: World camera’s

recorded the survey completion with a resolution of 800 × 600, a frame rate of 60, and an absolute exposure time of 157; The Pupil camera (pupil camera) recorded pupil size and gaze direction at a resolution of 192×192, a frame rate of 120, and an absolute exposure time of 32. Each second of the experiment after the eye tracker was turned on, 120 values wererecorded with corresponding accuracy measures.

- GPS paradata (objective 5, 6) – a record of the tablet's GPS coordinates at the beginning and at the end of the interview was measured "actively" by the interviewer (it was necessary to click on the "Measure location" button whenthe corresponding GPS question appeared on the screen). The use of Survey Solutions software also made it possible to automatically obtain GPS measurement accuracy (HDOP) values for each of the GPS paradata measurements. The GPS paradata measurements consisted of a set of three measures: longitude, latitude, and altitude. Missings and accuracy of longitude and latitude were analysed as GPS paradata quality indicators.

The scientific contribution of research to the development of the subject field

There are five major scientific contributions of the thesis into the field of using paradata for evaluating and reducing survey errors in computer-assisted survey modes. Below is a diagram reflecting the proposed approach, considering the method of data collection, available types of paradata, their quality, and the relationship to the main sources of increasing measurement error and the concomitant decline in the quality of survey data (Figure 1).

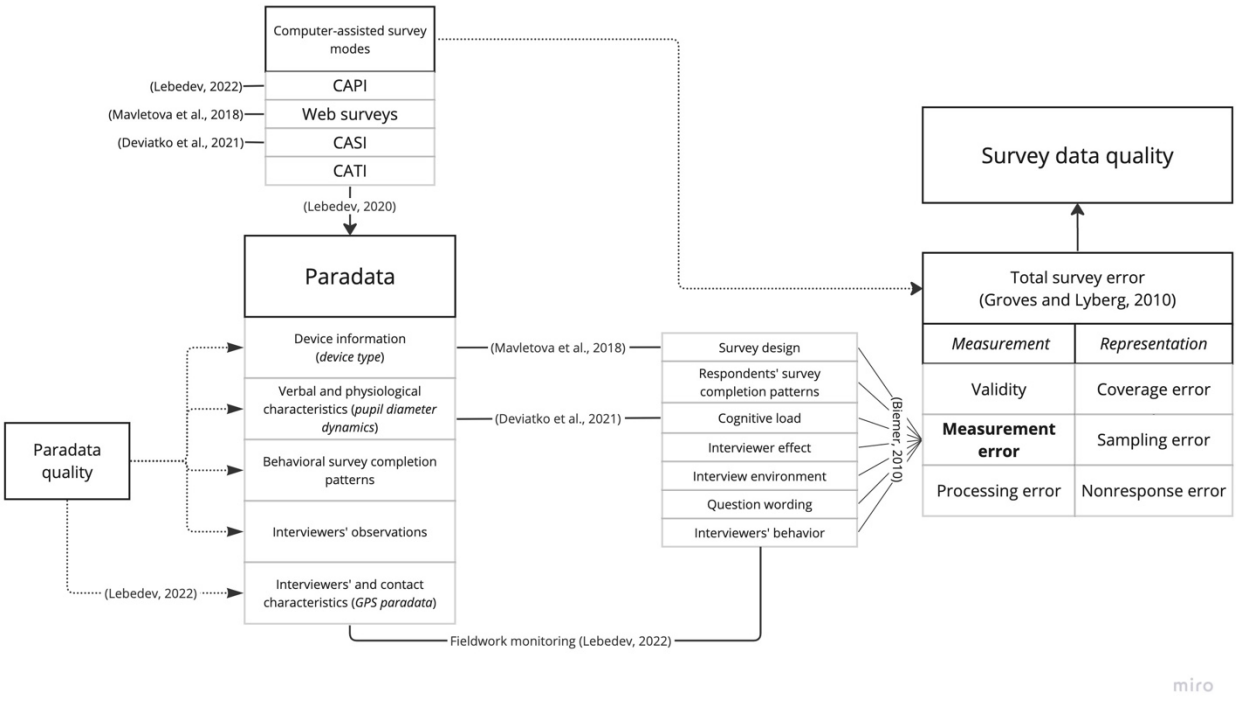

Figure 1: Schematic representation of the proposed approach of using paradata to estimate and reduce measurement error in surveys with application of computer-assisted survey modes

First, the comprehensive theoretical and methodological foundations for collecting and analyzing paradata in web surveys, CAPI, and CASI were proposed (Figure 1). We suggested a typology of paradata and described its types, limitations, specifics of use, as well as provided practical recommendations for possible application in survey methodology (Lebedev, 2020: 10-23). We also gave theoretical foundations for using GPS paradata in CAPI surveys for fieldwork monitoring (Lebedev, 2022: 12-18).

Second, to the best of our knowledge, for the first time in existing literature it has been shown how various types of paradata can be used in different survey modes (CAPI, CASI, web surveys). Overall, the articles present an approach to the use of paradata for evaluation and reduction of measurement error. GPS paradata can be used in CAPI studies to monitor the data collection process and to identify "suspicious" interviews that may be indicative of falsification and fabrication by interviewers (Lebedev, 2022: 12-18). Pupil size dynamics data have also been shown to be a valid measure of cognitive load and have allowed the equivalence of CASI

and PAPI data collection methods to be traced (Devyatko, Bogdanov, & Lebedev, 2021: 44-46).

Third, the use of device type paradata (smartphone or PC) shows how it can be used to evaluate and increase measurement equivalence between different devices and across different question formats (grid vs. item-by-item) in web surveys. It contributes to the literature on smartphone web surveys and echoes earlier results that showed the critical importance of collecting this type of paradata in all web surveys. The results showed that in the case of matrix questions with a number of scale values of 7 or more, the equivalence of the results is higher when the item-by-item format is used on both types of devices compared to using the grid format on both devices or using mobile optimization (item-by-item on a mobile device and matrix format on a PC) (Mavletova, Couper and Lebedev, 2018: 663-665).

Fourth, the use of eye-tracker measuring pupil diameter dynamics is quite innovative for that field. In most articles, researchers use eye fixations to evaluate cognitive load, which is relatively easy to measure and analyze since it does not require a lot of researcher's efforts. We used a more sophisticated and innovative technique as a proxy for difference in cognitive effort in CAPI and PAPI survey modes. Measuring the dynamics of pupil diameter allows us to assess cognitive load neurophysiologically and is the most valid metric compared to common subjective assessments and at the same time easier to implement compared to measuring gaze fixations and measuring skin-galvanic response (Devyatko, Lebedev, 2017). Our results contribute to the literature on cognitive load in CAPI, web surveys, and CASI, as well as to the literature that shows how paradata can be used to evaluate cognitive load of the respondents and interviewers. We also suggest using the pupil diameter dynamics as an indicator of cognitive load in conjunction with other types of paradata (mouse movements, completion time, browser focus change, alerts, etc.) (Deviatko, Bogdanov and Lebedev, 2021: 44-46).

Finally, we showed how GPS paradata can be used as an indicator of data quality in computer-assisted personal interviews when data is collected by interviewers. We suggest that it is not only important to consider the region in which

the interview takes place (e.g., Moscow, Samara, etc.), but also to focus on interviewer trainings to collect more reliable GPS paradata. Such results contribute to the literature on GPS paradata use for fieldwork monitoring purposes (Лебедев, 2022: 26-27).

Scope and limitation of the research

There are four main limitations of the thesis which can also be traced in the scheme (Figure 1).

First, nonresponse error was not included in the focus of the thesis since it is usually analyzed based on more conventional paradata types (interviewer observations, completion time, and call record data), leaving out advanced paradata types. In addition, the focus of the paper also does not include the other components of TSE - validity, processing error, coverage error, and sampling error - because it is beyond the scope of what can be included in a given thesis, and paradata are used less frequently than for measurement error and non-response error to estimate the bias and variation associated with these elements of TSE.

Second, CATI surveys were not included in the research study due to a lack of sources to collect and analyze indirect type of CATI paradata, such as vocal characteristics. This is due to the fact that this topic is the most developed among Russian-based sociologists with a focus on question-answer communication between interviewers and respondents (Ipatova, 2016; Ipatova and Rogozin, 2014; Turchik, 2010) and the possibilities of using interviewer observations (comments) to adjust data and methodological audit/refine the instrument (Ipatova, 2014), which within the framework of the above approach can be used to assess respondent completion patterns, interview space and interviewer behavior during questioning (Figure 1).

Third, though web surveys have a number of opportunities to collect different types of paradata related to respondent patterns of survey completion, these types of paradata (e.g., mouse movements, going back and forward in the web questionnaire, browser focuschange, answer option change) were not included in the study due to limitations of the software used to collect data. Paradata is still an emerging topic,

and only a fewprograms collect such advanced types of paradata. Finally, the study does not illustrate how various types of paradata can be used in combination with each otherto decrease survey errors.

Despite the limitations, the thesis showed how to incorporate the use of different types of paradata into the various research directions on measurement error evaluation and reduction within the most prominent data collection modes, such as web surveys, CAPI, and CASI.

Provisions submitted for defense

1. Based on the extensive literature review on paradata use for data quality evaluation and mitigation, the following typology of paradata was proposed: 1) interviewer observations; 2) device information; 3) behavioral characteristics related to survey completion; 4) interviewer characteristics and contact/communication characteristics with the respondent; 5) verbal and physiological characteristics that appear during the interview. An innovative feature of the proposed typology is that it incorporates the physiological data (e.g., pupil diameter dynamics, fixations, heart rate) and takes into account the data quality of all types of paradata (i.e., not only interviewer observations, which is quite conventional in the methodological literature).

2. Direct paradata, such as the type of device (smartphone vs. PC), can be used for evaluating and reducing measurement error in CAWI surveys. Based on a two-wave cross-over survey experiment, we showed how this type of paradata can be used to evaluate and decrease measurement error between different devices and different questionnaire formats (grid vs. item-by-item) in web surveys.

3. The paradata of pupil diameter dynamics can be used as a valid method for assessing cognitive load during survey completion among both respondents and interviewers in the CAPI and CASI survey modes. Based on an experimental laboratory-based study, we showed that it is a useful tool for measurement error evaluation. Some difficulties are associated with the data collection process since this type of paradata is prone to measurement error itself. At the same time, it provides new opportunities to reduce measurement error in CAPI and CASI surveys.

4. GPS paradata can be used to reduce nonresponse and measurement error in CAPI surveys when they are used for fieldwork monitoring. This can be done by analyzing individual interview GPS points or interviewer routes during data collection.

5. The quality of the paradata themselves must be considered when using them to estimate and reduce measurement error. For example, GPS paradata are subject to measurement error and non-response error. Consistent with the "urban canyons" hypothesis, more GPS missing data in RLMS HSE data were found in highly urbanized regions. In addition, interviewer characteristics (e.g., subjective evaluation of tablet confidence) were also associated with the quality of GPS paradata obtained.

General conclusions of the research

Paradata has a wide range of opportunities to be used in methodological research. This thesis has shown that paradata can be used in web surveys to evaluate survey design decreasing measurement error, in CAPI to monitor fieldwork, and CASI and CAPI survey to estimate cognitive effort during questionnaire completion. All of these implications are aimed at evaluating and reducing measurement error.

Device used to complete the web survey showed that it can be important for survey design research. The type of device paradata was found to be useful as a basis for the analysis, as it allowed to compare data quality between different question formats (item-by-item or grid) depending on the type of device (PC and smartphone) used for survey completion. It was shown that using a grid format for matrix questions on mobile devices leads to higher measurement error as well as decreased subjective respondent satisfaction with completing the survey. For matrix questions with more than 7 scale values, using an item-by-item format on both types of devices (PCs and mobile devices) leads to higher measurement equivalence. In addition, using item-by-item format for PC devices increases the concurrent validity of results and reduces the undifferentiated responses and probabilities of choosing the same answers for all items of tabular questions (straightlining) (Mavletova, Couper and Lebedev, 2018: 663-665).We argue that device type paradata provides opportunities

not only to investigate data quality, but also to conduct web survey experiments aimed at finding optimal design features that can vary depending on the differences in paradata (e.g., type of device).

Pupil diameter dynamics were shown to be a valid measure of respondents' cognitive load based on the laboratory-based CASI and PAPI experimental study. This type of paradata can be used as an indicator of cognitive effort, which evaluates and reduces measurement error, as well as increases data quality overall. Consistent with previous studies, the equivalence of the CASI and PAPI methods in terms of cognitive load of respondents in the 18-25 age group has been demonstrated. In addition, the construct validity of using data on pupil diameter dynamics as a measure of cognitive load was confirmed (Devyatko, Bogdanov, & Lebedev, 2021: 44-46). It was also shown how to implement an eye-tracking device in a laboratory setting, which can be burdensome as many factors (light, noise, place of the stimulus, position of experiment moderators etc.) should be considered. The results of the experiment showed that there is a difference in cognitive load depending on the type of questions and survey modes (CASI vs. PAPI) (Devyatko, Bogdanov, & Lebedev, 2021: 45-46).

GPS paradata have been shown to have much potential for evaluating measurement error in CAPI and fieldwork monitoring by analyzing interviewer behavior (routes and interview locations). That can improve data collection efficiency and reduce fabrications/falsifications using prompts preventing interviews from "cheating" or "erroneous" behavior during the fieldwork. Analysis of the quality of GPS paradata showed that paradata themselves can be the subject to nonresponse and measurement error, indicating the need for more interviewer training (if they are involved in data collection process) and the need to consider the quality of paradata as a whole (Lebedev, 2022: 12-18, 25).

An analysis of GPS paradata quality revealed that the paradata themselves may be subject to non-response and measurement error, indicating the need for additional interviewer training and the need for greater scrutiny of paradata quality in general. It has been shown that an increase in the proportion of missed GPS

paradata measurements at the beginning and end of the 448 CAPI interviews in the 26th wave of the RLMS HSE survey was related to conducting interviews in more urbanized areas as well as to the characteristics of the interviewers themselves (increased confidence in the tablet was significantly associated with a lower probability of missing GPS paradata measurements at the beginning or end of the survey) (Lebedev, 2022: 26-27).

In the thesis, we developed an approach to the collection and analysis of paradata in computer-assisted survey modes. This approach includes consideration of the survey mode (CAPI, web surveys, CASI), the fundamental purpose of paradata collection, and its quality. We demonstrated this approach based on the analysis of different types of paradata, such as the type of device, GPS, and pupil diameter. We used different types of paradata for illustrating various implications such as measuring cognitive load, monitoring fieldwork, and finding optimal design features in a survey. Considering arising interest in immediate feedback (Conrad et al., 2005; Conrad et al., 2017; Kühne and Kroh, 2018), paradata is seen as having the potential not only to assess and reduce measurement error after data collection, but also to prevent from occurring during data collection.

It is crucial to continue research on the possible use of paradata in survey methodology, as this can reduce different survey errors and provide an approach to errors' evaluation and reduction. In the future, this approach should be supplemented by the inclusion of the CATI method, other types of paradata, and other uses of paradata to estimate and reduce measurement error. While this is beyond the scope of this thesis, it would complement the approach and cover all the main directions of paradata use to evaluate and decrease survey errors, as well as all major survey modes that allow paradata to be collected.

List of publications of the author of the dissertation, which reflect the main scientific results of the dissertation

The results of the work carried out on the topic are reflected in the author's publications:

- Lebedev D. V. (2020) Paradata: definition, types, collection, and possible uses. *Monitoring of Public Opinion: Economic and Social Changes*. No. 2. P. 4—32. https://doi.org/10.14515/ monitoring.2020.2.915. (In Russian) (Лебедев, Д. В. (2020). Параданные: определение, типы, сбор и возможное применение. *Мониторинг общественного мнения: экономические и социальные перемены*, (2 (156)), 4-32.)

- Mavletova A., Couper M. P., Lebedev D. Grid and item-by-item formats in PC and mobile web surveys //Social Science Computer Review. – 2018. – Т. 36. – №. 6. – С. 647-668. DOI: https://doi.org/10.1177/0894439317735307

- Deviatko I.F., Bogdanov M.B., & Lebedev D.V. (2021). Pupil diameter dynamics as an indicator of the respondent's cognitive load: Methodological experiment comparing CASI and P&PSI. RUDN Journal of Sociology, 21(1), 36-49. (In Russian) (Девятко, И. Ф., Богданов, М. Б., & Лебедев, Д. В. (2021). Динамика диаметра зрачка как индикатор когнитивной нагрузки респондента: методический эксперимент по сравнению CASI и P&PSI вопросников. *Вестник Российского университета дружбы народов. Серия: Социология*, 21(1), 36-49.)

- Lebedev, D. V. (2022). Using GPS-paradata to control the data collection process: Review of existing methods and analysis of GPS-paradata quality. *Sociological journal*, (4), 8-33. (In Russian) (Лебедев, Д. В. (2022). Возможности использования GPS-параданных для контроля процесса сбора данных: Обзор существующих методов и анализ качества данных. Социологический журнал, (4), 8-33.)

Approbation of the research results

The results of the work carried out on the topic were presented in the

following scientific events:

- 7th Conference of the European Survey Research Association (ESRA Conference, 17-21 July 2017, Lisbon). Presentation: “PC and mobile web surveys: grids or item-by item format?”

- 8th Conference of the European Survey Research Association (ESRA Conference, 15-19 July 2019, Zagreb). Presentation: “GPS-Paradata in Computer-Assisted Personal Interviews: Additional Opportunities for Monitoring Fieldwork Interviewers”

- 23rd Global Online Research conference (GOR Conference, 8-10 September, 2021, Cologne). Presentation: “GPS paradata: methods for CAPI interviewers fieldwork monitoring and data quality”

- 15th European Sociological Association conference (ESA Conference, 31 August – 3 September, 2021, Barcelona). Presentation: “The Multimodal Assessment Of Respondent's Cognitive Load: A Methodological Experiment Comparing CASI And P&PSI Modes”

- Mobile Apps and Sensors in Surveys Workshop (MASS Workshop, 22-23 April, 2021, Utrecht). Presentation: “Comparison of different methods of GPS paradata usage in CAPI surveys for interviewers’ monitoring”

- 9th Conference of the European Survey Research Association (ESRA Conference, 2-23 July 2021). Presentation: “Pupil diameter dynamics as an indicatorof the respondent's cognitive load in CASI and P&PSI modes”

- 9th Conference of the European Survey Research Association (ESRA Conference, 2-23 July 2021). Presentation: “Using paradata to measure and improve data quality in web surveys: experimental assessment of difference between satisficing and optimizing behaviour”

- European Sociological Association Research Network 21 Midterm Conference (ESA RN21 Midterm Conference, 5-8 October, 2022, Salamanca). Presentation: “Pupil Diameter and Mouse Movements As Indicators of a Respondent's Multimodal Cognitive Load: a Comparison of Traditional and Gamified Online Survey Designs”

- European Sociological Association Research Network 21 Midterm Conference (ESA RN21 Midterm Conference, 5-8 October, 2022, Salamanca). Presentation: “Informed (non)consent to Paradata Collection in a Web Survey: Experimental Assessment of the Impact on Data Quality”