Fundusze Europejskie Wiedza Edukacja Rozwój
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Społeczny

University of Warsaw

mgr Marta Kabut

False Respondents in Web Human Resource Surveys

Fałszywi respondenci w ankietach internetowych w badaniach z zakresu zarządzania zasobami ludzkimi

Summary of the doctoral dissertation

Dissertation written under the supervision of
Prof. dr hab. Grażyna Wieczorkowska-Wierzbińska
Associate supervisor: dr. Anna Kuzminska
University of Warsaw, Faculty of Management
Managerial Psychology and Sociology Unit

Warsaw, 2021

projekt „Zarządzanie wielokulturowe w erze globalizacji" realizowany przez Wydział Zarządzania Uniwersytetu Warszawskiego na podstawie umowy nr POWR.03.02.00-00-I053/16-00 w ramach Programu Operacyjnego Wiedza Edukacja Rozwój 2014-2020 finansowanego z funduszy strukturalnych Unii Europejskiej

Online surveys have replaced other ways of conducting studies and have already had a dominant position among quantitative research methods[1]. Growth dynamics are impressive: in 2006, about 20%, and **in 2013 above 50%** of all data collection expenditures were spent on online surveys.[2]

Web self-administered surveys[3] have become a prevalent form of data collection in HRM, and research focused on the satisfaction of customers and employees[4], marketing[5], consumer preference and behaviour[6].

Types of internet research can be distinguished based on the following criteria[7]: (1) participants' awareness that they take part in research; (2) time of the research: real-time vs anytime; (3) level of participant's required engagement: active vs passive; (4) knowledge about participant's identity: anonymous vs identified. The advantages of online research include, among others: (1) higher availability of respondents; (2) easiness/fastness of reaching specific groups and hard-to-reach in other way persons; (3) time saving ; (4) lower price (no need to hire interviewers, pay travel costs); (5) flexibility (next question could be selected depending on the former answers).

Internet is also suitable for experimental research and enables the possibility of integrating qualitative and quantitative methods in one study. Easiness of getting respondents comes with limited or lack of control over their behaviour and environment.
Some respondents can choose one of the following (harmful to the research validity) strategies[8]:

1) **Selecting the first response alternative** that seems reasonable[9] without reading all possible options
2) **Selecting the most visible option**[10] choice dependent on the way of display: dropdown vs radio buttons vs scrollable dropdown)
3) **Speeding** – answering too fast without thinking about the answers[11].
4) **Acquiescence bias - Agreeing with any statement** regardless of its content [12].
5) **Endorsing the status quo**[13] – when a question asks about increasing or decreasing something, respondents often choose a base (starting) value when explicitly given to them.
6) **Non-differentiation in using rating scales**[14] – when using the same response options, in the same order, there is a danger that respondents will not differentiate between objects. Consequently, respondents will choose the same or almost the same options in each question.
7) **Preferring 'do not know'** answer - as 'do not know' is hard to interpret but also does not require much thinking; when that answer is presented, satisficing respondents will choose to pretend they do not have an opinion rather than trying to put effort into creating one, although research shows, that providing this answer option increases data quality[15].
8) **Mental coin-flipping**[16] – choosing randomly from among the response alternatives.

---

[1] ESOMAR, 2014
[2] Vehovar & Lozar Manfreda, 2008; *ESOMAR, 2013*
[3] Batorski & Olcoń-Kubicka, 2006
[4] Kasvi, 2017; Barakat et al., 2015; Mitchell et al., 2021
[5] ex. Queloz & Etter, 2019; Kumar Mishra et al., 2016
[6] Molenaar et al., 2018
[7] Batorski & Olcoń-Kubicka, 2006
[8] Krosnick, 1991
[9] Galesic et al., 2008
[10] Couper et al., 2004
[11] Conrad, et al., 2017; Michałowicz, 2016
[12] Krosnick, 1991
[13] Schuman & Pressner, 1981
[14] i.e. Krosnick & Alwin, 1989
[15] Albaum et al., 2011
[16] Converse, 1964

9) **Omitting a whole set of questions**, either by losing one's attention or by purpose, does not mean that answers are worthless, but there are difficulties with determining what to do with them – include or not.

FALSE responding[17] has been called in literature in many ways: random[18], insufficient effort[19], careless[20], satisficing[21], inattentive/participant inattention[22], and indiscriminate responding[23]. It can be defined broadly as happening when the respondent filling a survey does not behave cooperatively.

Such people may introduce random noise to data collected in surveys. However, they usually do not answer entirely randomly, which results in a systematic bias in responses, and, as a result, a change in obtained results (obtaining statistically significant effects or, on the contrary, no results)[24]. It is crucial to distinguish data from attentive and FALSE (inattentive) respondents.

While online surveys have become increasingly popular, new opinion polling companies have also sprung up. These companies bring together people in their research panels who view survey completion as an additional income easy job (often in the form of reward points).

Research Purpose and Research Tasks

There are many studies on inattentive respondents that have been done on English-speaking samples[25]; however, this phenomenon has not[26] been studied thoroughly for Polish samples. The research gap to be filled by this research is determining the level of inattention of respondents, the consequences of not excluding FALSE respondents from the analysed data, and devising an **FR procedure** to detect FALSE respondents in data sets.

Two research tasks were carried out: (1) estimation of the magnitude of the FALSE Respondents problem in 12 data sets by using a procedure based on 4 Warning Signs; (2) estimation of the consequences of ignoring the FALSE RESPONDENTS problem and testing the usability of the FLEXMIX[27] procedure for detecting FALSE respondents.

[17] Levi et al., 2021
[18] Credé, 2010
[19] Huang et al., 2012; Huang & DeSimone, 2021
[20] Meade & Craig, 2012; Bowling et al., 2020
[21] Krosnick, 1991
[22] McKibben & Silvia, 2017; Beck et al., 2019; Steedle et al., 2019
[23] Holden et al., 2019
[24] Alvarez et al., 2019
[25] ex. Nichols & Edlund, 2020, Schneider et al., 2018, Bowling & Huang, 2018, Alvarez et al., 2019
[26] The only exception is Wieczorkowska's work (1999, 2011, 2022), which has been carried out for many years.
[27] finite mixtures of generalized regression models

Key terms

A **FALSE respondent** (FR, careless respondent, inattentive respondent, flagged respondent) is a person who voluntarily participates in a survey and answers questions without thinking (e.g., chooses a random or first good enough answer).

A **web/online/internet survey** is a self-administered online questionnaire.

The **RATING STYLE** (RS, response style) is defined as the tendency to respond consistently to questionnaire items other than what the items were specifically designed to measure[28]. The rating style can manifest itself through: (1) too severe (or lenient) assessment[29], (2) lack of differentiation of partial dimensions of evaluation[30] , e.g. AGREE to almost all items on the scale.

DK – (Do not KNOW - Non-informative answers) answers – do not convey any information about the question's opinion/thinking/facts.

The BEHAVIOURAL cooperation level is operationalised by logical inconsistency, odd answers to open-ended questions answer analyses to attention check questions.

The DECLARATIVE cooperation level is operationalised by answers about respondents' engagement (i.e., would their answers change if it was a different day).

LOGICAL INCONSISTENCY is operationalised based on lack of congruency in answers (respondents respond 'I do not have a job currently' in one question but respond 'I like my job' instead of 'not applicable' later in the survey)

ODD ANSWERS to open-ended questions mean answers that are too short or cannot be interpreted concerning the question content (e.g. Answers "I need more financial rewards" to the question on satisfaction).

WARNING SIGN (WS) indicates that respondents do not follow the rules, and it could be useful to consider excluding them from the analyses. There were 4 Warning Signs:

1. WS1 is based on too short an answering TIME.
2. WS2 is based on the number of incorrect answers to **Attention Check Questions [ACQ].**
3. WS3 is based on the too big number of **Do not KNOW Answers** and **Low Differentiation Rating Style**
4. WS4 is based on low behavioural (logical inconsistency, odd answers to open-ended questions) and low declarative engagement

Exclusion criteria:

[28] Wieczorkowska, 1993, Harzing et al., 2011

[29] (Hoyt, 2000)

[30] (Landy, Vance & Barnes-Farrell and Steele, 1980)

- **STRICT** exclusion criterion means that all respondents flagged by any of the Warning Signs would be excluded
- **LENIENT** exclusion criterion means that all respondents flagged by at least TWO Warning Signs would be excluded from the data set.
- **GLOBAL** exclusion criterion- respondents are excluded from the whole data set
- **LOCAL** exclusion criterion- respondents are excluded only from the block of items when e.g. the number of DK answers is very big only for this part of a survey. We can accept **local inattention** when the respondent becomes lost in thought, pondering, or deliberately ignoring a specific block of questions, but answers other blocks with due diligence.

ANSWERING TIME: Overall answering time **(OAT)** is the time that passed from the first load of the first survey page to the end page shown. **Partial** answering time **(PAT)** is the time spent on answering blocks of the survey. Words per minute **(wpm)** is an indicator of reading speed, calculated by dividing the number of words on a single survey page by the time that part was seen (in minutes for wpm and the seconds in wps).

Dissertation structure

The dissertation consists of two parts: theoretical and empirical.
The first part of the dissertation focuses on the advantages and disadvantages of internet research and using online panels, a cognitive model of answering survey questions, FALSE respondent description, and the impact of FALSE respondents on data quality. Then the problem of operationalisation of Warning Signs is presented.

Based on the literature review,[31] we can see that percentage of FALSE (careless) respondents varies from study to study.
As the studies are not consistent in their inattentiveness indicators' use, a short list of the examples of studies is presented in the **Błąd! Nie można odnaleźć źródła odwołania.**.

[31] Johnson, 2005; Kurtz & Parish, 2001; Meade & Craig, 2012; Curran et al., 2010 ; Baer et al., 1997

Year & Exclusion Criterion	Methods of detection used in the study	N, Sample	% excluded
2009[32], One WS Warning Signal	ACQ (IMC)	144, Students	35%
2015[33], at least one WS	SR, ACQ	400, MTurk respondents	5.5%
2016[34], No exact cut-off points	Respondent's Goodness of Fit	205, Purposive sample	10.81%[35]
2016[36], Failed ACQ	ACQ (IMC)	396, MTurk	5%
		85, Students	61%
	ACQ (Novel IMC – long instruction with the hidden correct answer)	185, MTurk	4%
		245, Students	74%
	ACQ (more difficult novel IMC – short instruction to mark two answers)	239, MTurk	74.5%
		90, Students	97.8%
2020[37], at least one WS	OAT, IRV, psychometric synonyms, OEC	3 groups of students (N1=278, N2=281, N3=268)	12.8%
			12.5%
			15.7%
2017[38], Faster than 1 spi, consistency measure< 0.5	OAT, Response consistency (correlations between related items)	421, MTurk	5 - 24%
Faster than 1 spi, consistency measure<0.43		296, Students	12%
2018[39], 10% of the sample on each measure	ACQ (infrequency type), OAT, OEC, LSI, Intra-Individual Response Variability	199, Students	30.2%
2018[40], WS for each measure separately	Mahalanobis distance, Psychometric synonyms, antonyms, Maximum LSI, OAT, SR, ACQ	3 groups of Students (N1=274, N2=614, N3=394)	5.9% (per method)
	The same as 1st study + OEC		2.9%
			4.3%
2019[41], at least one WS	Contradicting answers to reversed items, OAT	129, Students (online)	23%
		101, Students (paper)	27%
		110, MTurk	46%
2019[42], Wrong answers for both ACQ	ACQ (instructed response), OAT, Straight lining, Item nonresponse	5205, GLES (panel)	6.1%
2020[43], Based on Latent Profile (Class) Analysis	Open-ended questions, Resampled Individual Reliability, Person-Total Correlation, SR, ACQ, OAT, LSI, OEC	394, Crowdsourcing research platform	45.9%

ACQ- Attention Check Questions; IMC - Instruction Manipulation Check
LSI – Long-String Index; GLES - German Longitudinal Election Study
OEC – Odd-Even Consistency; SR – Self-reported low level of cooperation

The empirical part of this dissertation begins with defining the operationalisation of 4 warning signs and ends with a **description of the FR procedure for detecting FALSE respondents.**

The distribution of warning signs was analysed in **9 web surveys conducted by our doctoral team** at the Department of Managerial Psychology and Sociology at WZUW between 2020-2022[44].

- two data sets consisting of **2918 employees** [commercial panel participants]
- six data sets B1- B6 based on responses from **2399** participants who, in the overwhelming majority, combine studies at the Faculty of Management with professional work
- one data set C, based on responses from **287 employees** with at least three years of work experience

and 3 pre-existing data files:

- Data set D, European Working Conditions Survey, personal interviews, **1203 Polish employees**
- Data sets E1 + E2, World Values Survey, two waves (5+6), **1966 Polish respondents**.

Research task #1: FALSE respondent scope

The first research task was to determine the scope of respondents' inattention in 12 analysed data sets.
Analysis showed that the percentage of respondents flagged as "FALSE" depended on the survey and the type of Warning Sign.
The graph below shows that for 6 web surveys C, B1 to B6, the more discriminating criterion was WS4.

[32] Oppenheimer et al., 2009
[33] Rouse, 2015
[34] Kountur, 2016
[35] exclusion by design of the research – group of 20 respondents was instructed to behave inattentive when responding
[36] Hauser & Schwarz, 2016
[37] Iaconelli & Wolters, 2020
[38] Wood et al., 2017
[39] Dunn et al., 2018
[40] Ward & Meade, 2018
[41] Aruguete et al., 2019
[42] Silber et al., 2019
[43] Brühlmann et al., 2020
[44] WS1 tested on 9 datasets, WS2 tested on 8 datasets, WS3 and WS4 tested on 12 datasets

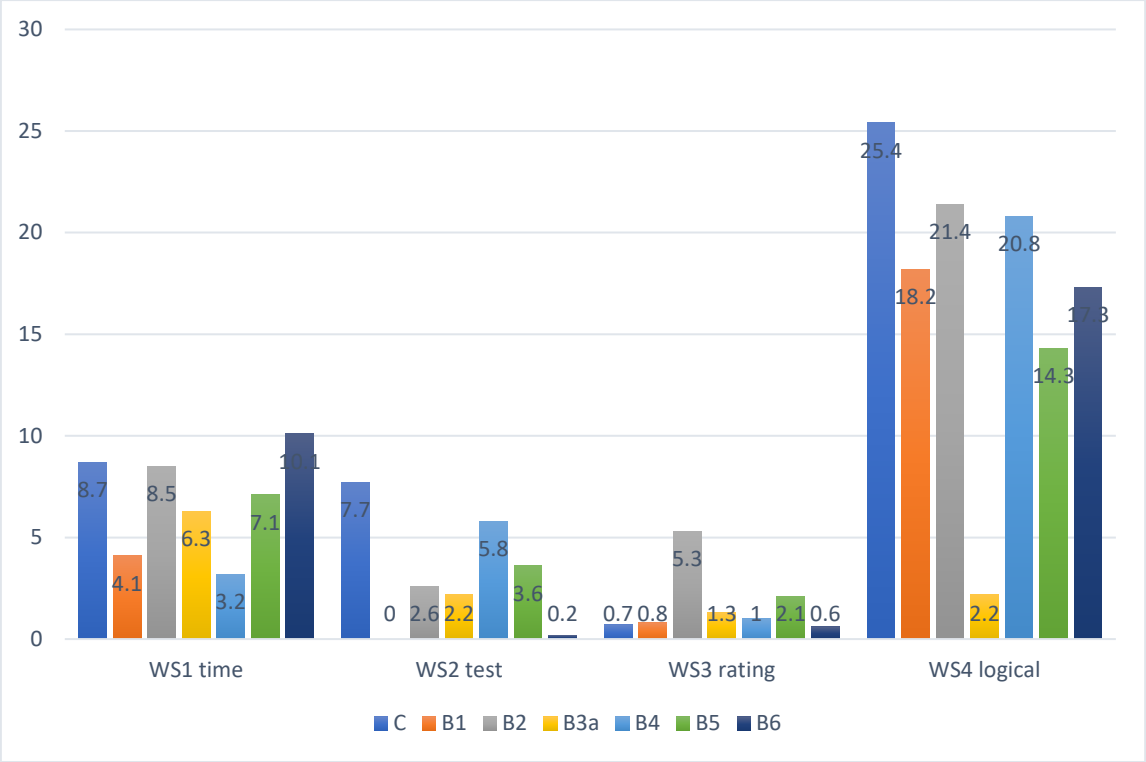

Figure 1 Comparison of different rates of exclusion for data sets B1-B6, C, presented in division by WS

For two panel data sets, A1 and A2, the most exclusionary criterion was WS1 (time), and WS4 was the second.

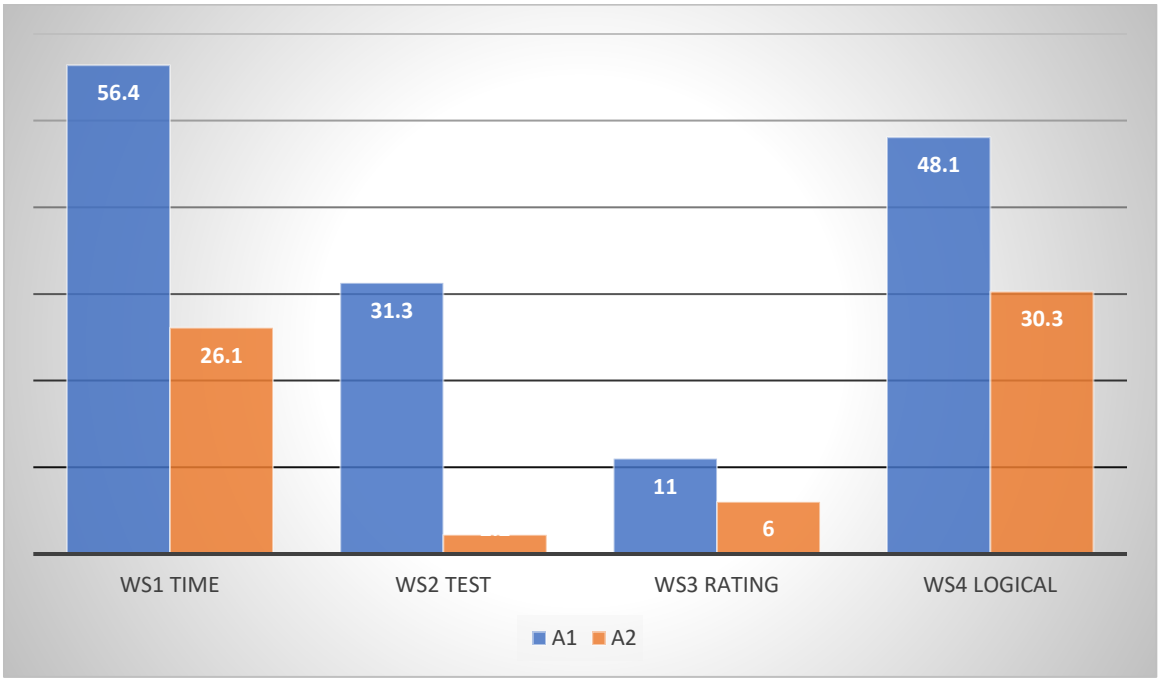

Figure 2 Rates of exclusion for 4 WS for two commercial panel data sets [A1 and A2]

There are shocking differences between the distribution of Warning Signs in two paid panel studies. Our doctoral team designed both surveys conducted by the same commercial company that sells its services to researchers.

The difference (29 p.p.) in Warning Sign #2 between A1 and A2 can be explained by the different type of attention check questions used in both surveys. In A1, three instructed response items (i.e., "Please choose <<Rather A>> in this question") were used and it has not been explained why a respondent should do that. This unexplained order could make some respondents angry and reactant. In A2, five arithmetic questions (i.e., "Choose correct result of this operation 23+5=") were used, and it was justified as a fight with the monotony of other questions. The software change can also explain the A1-A2 difference. In A1, respondents could not return to the previous question and change their answers. In A2, respondents could change their answers if they noticed that they had made a mistake.

If we compare the answering time in the table below- we can see that A1 was a little shorter than A2. So, it contradicts the slogan 'that the shorter survey, the better.

Data set	**A1**	**A2**
False respondents (based on 4 WS- strict criterion)	**71.0**	**46.6**
OAT median[45]	14:06	26:25
Number of words	3383	3628
Median time without FALSE respondents	27:17	30:17

The lowest percentage of FALSE respondents was for offline data files (from high-budget international surveys that were carefully designed and cleaned by international teams of researchers before they were made available to the public) because, in this case, only 2 Warning Signs were available.

Warning signs	**D**	**E1**	**E2**
WS1 time	-	-	-
WS2 test	-	-	-
WS3 rating	3.7	2.8	3.1
WS4 logical	5.7[g]	13.2[h]	5.7[h]

g. based on assessed cooperation – 2 questions
h. based on assessed interest – 1 question

Research task #2: Consequences of not excluding FALSE respondents from data sets

The second research task was to show the consequences of ignoring the problem of FALSE respondents. For this purpose, the **reliability of the measurement** (operationalised by **Cronbach's alpha)** was compared in groups of excluded respondents (FALSE) and not excluded (attentive) respondents.

[45] for attentive respondents not excluded by WS1

Two procedures to divide survey samples into groups of FALSE and attentive respondents were checked for their utility:

(1) described in the dissertation procedure based on **4WS [4 Warning Signs].**
(2) the **Flexmix**[46] model (combining cluster and regression analysis)

Flexmix allows us to divide respondents into subgroups based on their fit to different regression lines. FLEXMIX divides respondents into two groups based on correlation between their answers to 2 questions in the simplest version. If the correlation in both groups differs in sign and we know that **theory predicts a negative correlation** between the answers to 2 questions with a rating scale <1- *like person A* to 4- *like person B*>:

1. *People say that at business dinners or social gatherings,* ***person A often dominates*** *the conversation.* ***Person B says little,*** *so others have to keep the conversation going.*
2. *Being in a large group of people,* ***person A typically*** *talks to a few people, primarily those he knows.* ***Person B talks to many people****, including strangers.*

Respondents classified by the Flexmix algorithm as the group with a **positive correlation** are potentially suspected to be not attentive in reading the questions.

The two procedures excluded different percentages of samples. 4WS procedure showed a better quality of the A2 data set (only 17% false). Flexmix excludes in both data sets a similar number of respondents.

	# Of FALSE respondents Flagged by 4 WS Procedure	**# Of FALSE respondents Flagged by Flexmix**
Data set A1 N=1421	652 (46%)	456 (32%)
Data set A2 N=1497	261 (17%)	509 (34%)

The next step was a comparison of Cronbach's alphas in the group of FALSE respondents flagged by each procedure and in the group that passed the test. To compute Cronbach's alpha, two indicators from SSA[47] were used: in data set A1 METHODICALITY index and in data set A2 EXTRAVERSION index.

The graph below compares 4 Cronbach's alphas in study A1 (on the left) and study A2 (on the right).

[46] a general framework for finite mixtures of regression models

[47] Wieczorkowska, 2022

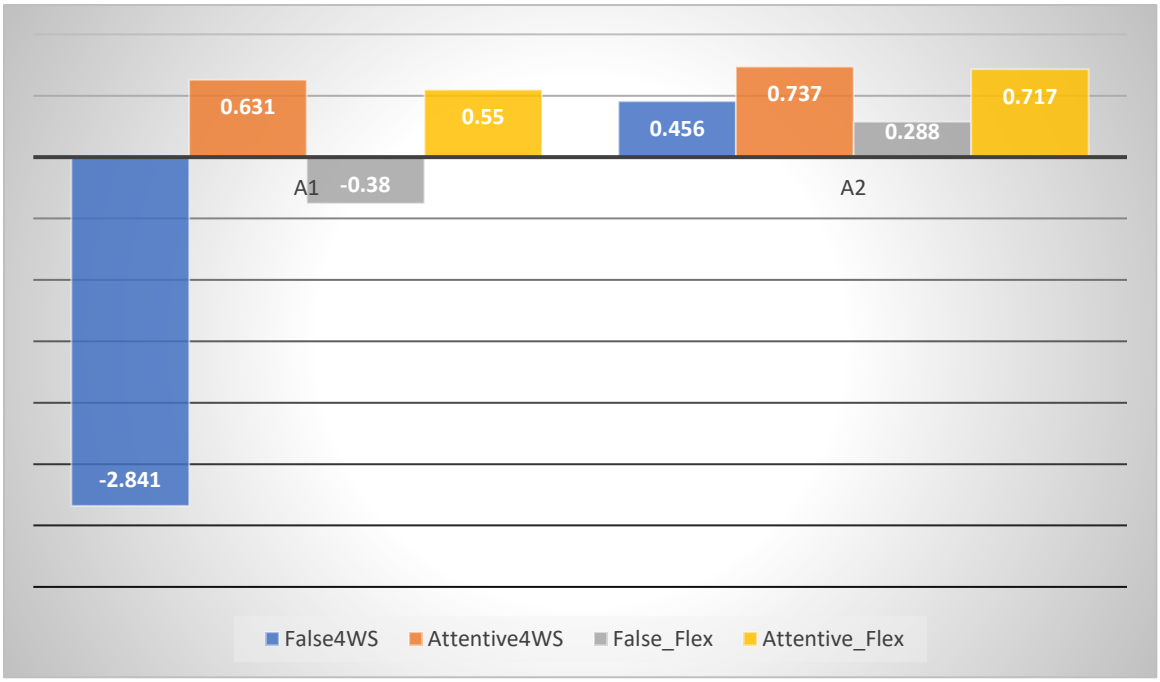

Figure 3 Cronbach alphas for groups of false and attentive respondents in a division by data set [A1, A2] and methods [WS, Flexmix]

In both datasets and both procedures, the value of Cronbach alpha is acceptable in the group of attentive respondents and NOT acceptable in the group of FALSE respondents.
A negative alpha value indicates that false respondents did not read the questions because the index should not include negatively correlated items.

The comparison of who was flagged by each procedure shows low contingency between the two procedures in detecting FALSE Respondents. This can be explained by the fact that the **FLEXMIX procedure** is a **local** one – it was based on correlation analysis between answers to two questions ONLY. The **4WS procedure** is global because it analyses the respondent's behaviour throughout the survey.

Therefore, the Flexmix procedure can be recommended to help examination for Warning Sign #4 only. Judging logical inconsistency in the responses is very difficult due to the flexibility of cognitive processes described in the model. Automating this process using the FLEXMIX procedure is advisable, but we must use more than two questions.

Procedure for detecting FALSE respondents

In the next step, after the standard procedures (checking data completeness and or analyses focused on variables distributions), the values of 4 WS should be computed for each respondent.

Step 1. Set the thresholds for all Warning Signs. Check the univariate distributions of Warning Signs

The threshold for WS1 means the minimal time needed to read the questions
The threshold for WS2 means the acceptable number of errors in Attention Check Questions
The threshold for WS3 means the lowest acceptable variance in answering a series of questions with the same rating scale, the biggest acceptable number of Do not know answers (usually less than 50%)
The threshold for WS4 means the acceptable level of logical inconsistency in closed and open-ended questions, an acceptable level of declared engagement in the survey etc.

Computation of:

WS1 is based on too short an answering TIME.
WS2 is based on the number of incorrect answers to Attention Check Questions.
WS3 is based on a large number of Do not Know Answers and Low Differentiation Rating Style
WS4 is based on low behavioural engagement (logical inconsistency, odd answers to open-ended questions) and low declarative engagement

Based on each threshold, "1 "(means above threshold) or "0 "(mean below) will be assigned to every respondent. So the sample will be divided into five categories :

From 0 – means NO warning signs to 4 -means that all 4 Warning Signs flagged the respondent.

Step 2. Decide on a STRICT or LENIENT criterion

The comparison of the consequences of this decision can be seen in the table below:

Data set	**Year**	**Sample**	**% of respondents excluded**	
			Lenient criterion	**Strict criterion**
A1	2018	1421+ 1497 panel employed respondents	45.9	71.0
A2	2021		14.2	45.4
C	2020	287 employees, convenience sample	6.6	33.8
B1,B2,B3 B4,B5,B6	2018-2021	2440 respondents (in the overwhelming majority, combine studies with work	1,2-5,0	6,1-13,8
D,E2,E3	2005/2010 /2015	3169 respondents [personal interviews, offline]	0- 0.9	5,2-7,3

The graphical form of the 4 WS procedure is below:

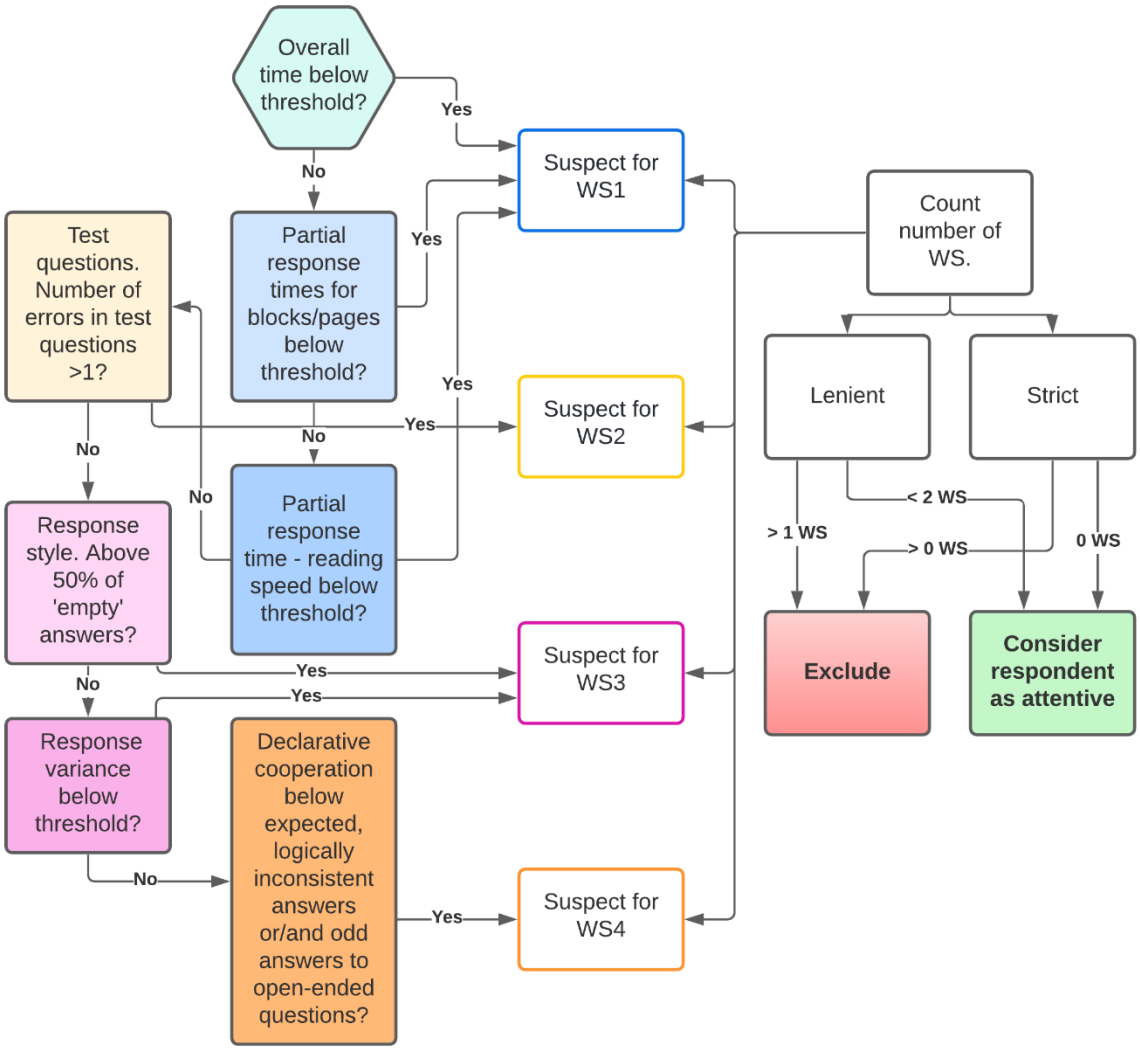

Figure 4 Visual scheme of the procedure for detecting FALSE responde

Conclusion

Comparing percentages of false respondents in analysed 12 data sets from **7.3 to 71.0%** with literature review points the range between **4 and 97.8%**, so we can say much (almost all) depends on the study. The magnitude of the problem could be enormous.

The analysis of WS in 12 studies does not show any general patterns allowing us to say which Warning Signal is the least important. That means that all should be calculated, but **we need to plan it before data collection.**

The filter most used by researchers is WS1 (time)[48]. Some studies use the time that the respondent spent on a specific page[49]. Using only **WS1 <too short answering time>** to detect FALSE respondents is not enough. False speeding respondents can take a coffee break and stay undetected.

The second most used filter is WS2 (attention check questions)[50]. Some researchers claim that a single attention check question can be effective,[51] while others recommend using more than one attention check question[52]. For sure, 1 ACQ item is not enough because of the dynamics of the respondent's attention.

More attention check questions are a better choice, but we need to explain their role to respondents, so arithmetic questions are recommended.

Due to the dynamic nature of respondents' attention, it is worth analysing Warning Signals locally (separately for sets of items). WS2 and WS4 can be used only globally, but for WS1 and WS3 local analysis is recommended: measuring answering time and the number of Do not Know answers for the survey blocks. If the measured value is above the threshold for the block, all answers for the block could be converted into missing values. The local exclusion method is a standard procedure we use at the Department of Managerial Psychology and Sociology, and it yields very good results.

Limitations

The limitations of the research presented in the doctoral dissertation come from the type of analysed data.

[48] i.e. Skarżyńska et al., 2021
[49] Greszki et al., 2015
[50] Kuźmińska & Pazura, 2018; Kuźmińska et al., 2019
[51] Maniaci & Rogge, 2014
[52] Liu & Wronski, 2018; Berinsky et al., 2014

High-quality surveys

Offline data files consist of publicly available high-budget international surveys (World Values Survey and European Working Conditions Survey) that are carefully designed and cleaned by international researchers.
Online data files consisted of research conducted by the doctoral team at the Department of Managerial Psychology and Sociology WZ UW, where measurement tools were constructed with great concern about respondent's motivation, encouraged taking breaks, and carefully prepared instructions and information about the topic and content of the questions. In our research, the respondent has the freedom not to answer a question or say 'Do not know', which means they can choose a non-informative answer.

Limited-Access Survey

An invitation to participate in a typical Internet survey is posted on the Internet, where everyone has access to it. Invitation to participate in the research analysed in the dissertation was sent to selected groups of respondents ONLY who were motivated by different means (e.g., paid, getting bonus points for MBA and other students). We can predict that the number of FALSE respondents will be much bigger in Open-access Surveys.

Restricted respondents' education level

All of the respondents in the online survey were at least high school graduates, which means that the studies on groups of less-educated respondents are needed.

Directions for future research

There are at least five possible directions for future research:

First, automatisation of the process - the FR procedure proposed in this dissertation must be executed mostly manually, with the researcher making decisions about which thresholds are suitable for a particular dataset at hand.

Second, the proposed procedure should be compared with the results of machine learning algorithms[53].

Third, it would be interesting to check whether the **FR procedure** could be used to detect bots[54] (machines that fill questionnaires without human intervention) and, if it could, how efficient it is in doing so.

Fourth, it would be interesting to test the impact of immediate feedback and feedback in general, which seems to motivate respondents to give more thought-out responses.

[53] Schroeders, et al., 2022; Gogami et al., 2021
[54] Dennis et al., 2018; Buchanan & Scofield, 2018

Fifth, it would be interesting to study further the relationship between respondents' age and the number of warning signs they were flagged by. The negative correlation we found in A2 is consistent with previous research[55] indicating that older respondents are more attentive than younger respondents.

Sixth, experimental studies are needed. All analyses presented in the dissertation are correlational – so their internal validity is limited, as in all correlational studies. We have just started the series of experimental research on the impact of manipulated values. The dependent variable is the frequency of Warning Signs that differ in the values of independent variables, e.g. type of feedback.

The first experiment has been conducted and is described in Attachment 11.
Respondents were **randomly** divided into two groups that differed in the type of feedback in the test questions (arithmetic questions). In group E1 (N = 191), the respondent chose the wrong answer, e.g., '25' in the question '18 + 4 = 'Got the signal 'incorrect' and was forced to choose again, in group E2 (N = 223) the wrong answer was accepted. There were paradoxically and significantly more errors (operationalised as more than two clicks on the arithmetic question) in group E1 than in E2. Both groups did not differ concerning other warning signs. Contrary to the hypothesis, forcing respondents to correct the wrong answer **did not improve** their attention.

[55] Maniaci & Rogge, 2014

Contribution

The doctoral dissertation has a cognitive, methodological, and application contribution. It tries to estimate the scale of the occurrence of FALSE respondents in 9 well-prepared surveys – it was shown that the presence of the group of FALSE respondents in the data files drastically reduced the reliability of the measurement. Unreliable data from FALSE respondents may change correlations, make the analysis and evaluation of research results difficult[56], decrease statistical power and effect size[57], and lower internal consistency. HRM theories confirmed by biased (unreliable) data are not valid, so FALSE respondents' detection is an important pre-analysis task.

The original methodological contribution is the 4 Warning Signs procedure for detecting False respondent and the empirically tested proposal of using the FLEXMIX procedure (combination of regression with cluster analyses) to check logical inconsistency in respondents' answers.

The application contribution consists of developing a procedure for detecting FALSE respondents in HRM studies that other researchers could use.

To sum up: the ease of data collection in web HRM surveys does not accompany methodological diligence in data analyses.
Analyses performed on uncleaned data could lead to FALSE conclusions, which, if incorporated into scientific circulation, harm the development of management research. There is a need for data cleaning techniques that allow for control of respondents' engagement, which was less problematic in paper surveys[58]. HRM theories confirmed by biased (not reliable) data are not valid, so FALSE respondents' detection is a vital pre-analysis task. The proposed 4 Warning Signs Procedure could be used to increase the quality of data, analyses, and conclusions.

[56] Maniaci & Rogge, 2014
[57] Brühlmann et al., 2020
[58] Kiesler & Sproull, 1986.

References cited in the dissertation

Albaum, G., Wiley, J., Roster, C., & Smith, S. M. (2011). Visiting Item Non-responses in Internet Survey Data Collection. *International Journal of Market Research*, 53(5), 687–703. https://doi.org/10.2501/IJMR-53-5-687-703

Allen, M. (2017). The sage encyclopedia of communication research methods (Vols. 1-4). Thousand Oaks, CA: SAGE Publications. https://doi.org/10.4135/9781483381411

Alvarez, M. R., Atkeson, L. R., Levin, I., & Li, Y. (2019). Paying Attention to Inattentive Survey Respondents. *Political Analysis, 27*(2), 145–162. https://doi.org/10.1017/pan.2018.57

Anduiza, E., & Galais, C. (2017). Answering Without Reading: IMCs and Strong Satisficing in Online Surveys. *International Journal of Public Opinion Research*, 29(3), 497–519. https://doi.org/10.1093/ijpor/edw007

Anseel, F., Lievens, F., Schollaert, E. & Choragwicka, B. (2010). Response Rates in Organizational Science, 1995–2008: A Meta-analytic Review and Guidelines for Survey Researchers. *Journal of Business and Psychology, 25*(3), 335-349. http://dx.doi.org/10.1007/s10869-010-9157-6

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*, 527–535. https://doi.org/10.3758/s13428-012-0265-2

Baer, R. A., Ballenger, J., Berry, D. T. R. & Wetter, M. W. (1997) Detection of random responding on the MMPI-A. *Journal of Personality Assessment, 68*, 139-151.

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D., Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J., Lavrakas, P. J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R. K., & Zahs, D. (2010). Research synthesis: AAPOR report on Online Panels. Public Opinion Quarterly, 74(4), 711–781. https://doi.org/10.1093/poq/nfq048

Barakat, L. L., Lorenz, M. P., Ramsey, J. R., & Cretoiu, S. L. (2015). Global managers: An analysis of the impact of cultural intelligence on job satisfaction and performance. *International Journal of Emerging Markets*, *10*(4), 781–800. https://doi.org/10.1108/IJOEM-01-2014-0011

Bassett, J., Cleveland, A., Acorn, D., Nix, M., & Snyder, T. (2017). Are they paying attention? Students' lack of motivation and attention potentially threaten the utility of course evaluations. *Assessment & Evaluation in Higher Education*, *42*(3), 431–442. https://doi.org/10.1080/02602938.2015.1119801

Batorski, D. & Olcoń-Kubicka, M. (2006). Prowadzenie badań przez Internet – podstawowe zagadnienia metodologiczne. *Studia Socjologiczne, 3*(182), 99-132.

Beatty, P. & Herrmann, D. (2002). To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse. In Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little R. J. A. (Eds.), Survey Nonresponse (pp. 71-85). New York: Wiley.

Beck, M. F., Albano, A. D., & Smith, W. M. (2019). Person-Fit as an Index of Inattentive Responding: A Comparison of Methods Using Polytomous Survey Data. *APPLIED PSYCHOLOGICAL MEASUREMENT, 43*(5), 374–387. https://doi.org/10.1177/0146621618798666

Bell, D. E., Raiffa, H. & Tversky, A. (1988). Descriptive, Normative, and Prescriptive Interactions in Decision Making. In D. E. Bell, H. Raiffa, and A. Tversky (eds.), Decision Making — Descriptive, Normative, and Prescriptive Interactions (pp. 9–30). New York: Cambridge University Press.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science, 58*, 739–573. https://doi.org/10.1111/ajps.12081

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

Bowling, N. A., & Huang, J. L. (2018). Your Attention Please! Toward a Better Understanding of Research Participant Carelessness. *Applied Psychology: An International Review, 67*(2), 227–230. https://doi.org/10.1111/apps.12143

Bowling, N. A., Gibson, A. M., Houpt, J. W., & Brower, C. K. (2020). Will the Questions Ever End? Person-Level Increases in Careless Responding During Questionnaire Completion. *Organizational Research Methods, 24*(4), 718–738. https://doi.org/10.1177/1094428120947794

Brown, R. V. (1968). Evaluation of total survey error. *Statistician, 17*(4), 335-343.

Brown, R. V. (1969). *Research and the credibility of estimates*. Boston, MA: Harvard University, Graduate School of Business Administration, Division of Research.

Brown, R. V. (1992). The state of the art of decision analysis: A personal perspective. *Interfaces, 22*, 5-14.

Brown, R. V. & Vari, A. (1992). Towards an agenda for prescriptive decision research: The normative tempered by the descriptive. *Acta Psychologica, 80*, 33-47.

Buchanan, E. A., & Hvizdak, E. E. (2009). Online Survey Tools: Ethical and Methodological Concerns of Human Research Ethics Committees. Journal of Empirical Research on *Human Research Ethics, 4*(2), 37–48

Buchanan, E., & Scofield, J. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods, 50*(6), 2586–2596. https://doi.org/10.3758/s13428-018-1035-6

Carver, R.P. (1992), Reading rate: Theory, research and practical implications. *Journal of Reading, 36*, 84–95.

Christensen, A. I., Ekholm, O., Glümer, C., & Juel, K. (2014). Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys. *European Journal of Public Health, 24*(2), 327–332.

Cichomski, B., & Morawski, P. (1996). *Polski Generalny Sondaż Społeczny: Struktura skumulowanych danych, 1992-1995*. Warszawa: Instytut Studiów Społecznych UW.

Conijn, J. M., van der Ark, L. A., & Spinhoven, P. (2020). Satisficing in Mental Health Care Patients: The Effect of Cognitive Symptoms on Self-Report Data Quality. *Assessment, 27*(1), 178–193. https://doi.org/10.1177/1073191117714557

Conrad, F. G., Tourangeau, R., Couper, M. P., Zhang, C. (2017). Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods, 11*(1), 45-61.

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. Apter (Ed.), *Ideology and discontent* (206-261). New York: Free Press

Couper, M. P., Tourangeau, R., Conrad, F. G., Crawford, S. D. (2004). What They See Is What We Get: Response Options for Web Surveys. *Social Science Computer Review, 22*(1), 111-127.

Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement, 70*, 596–612.

Curran, P. G., Kotrba, L., Denison, D. (2010) Careless responding in surveys: applying traditional techniques to organizational settings. 25th annual conference of Society for Industrial and Organizational Psychology, Atlanta, GA.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66,* 4-19.

Dennis, S., Goodson, B., & Pearson, C. (2019). Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3233954

DeSimone, J. A., & Harms, P. D. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology, 33*(5), 559-577.

Dewberry, C., Davies-Muir, A., & Newell, S. (2013). Impact and Causes of Rater Severity/Leniency in Appraisals Without Postevaluation Communication between Raters and Ratees. *International Journal of Selection and Assessment*, *21*(3), 286-293.

Dodou, D. & de Winter, J. C. F. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior, 36*, 487-495. https://doi.org/10.1016/j.chb.2014.04.005

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology, 33*(1), 105-121.

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020). The Effects of Sampling Frequency and Questionnaire Length on Perceived Burden, Compliance, and Careless Responding in Experience Sampling Data in a Student Population. *Assessment*. https://doi.org/10.1177/1073191120957102

European Society for Opinion and Market Research (ESOMAR). (2013). Global Market Research 2013 [online]. Available at: https://www.esomar.org/uploads/industry/reports/global-market-research-2013/ESOMAR-GMR2013-Preview.pdf

European Society for Opinion and Market Research (ESOMAR). (2014). Global Marketing Research 2014: an ESOMAR industry report [online]. Available at: https://www.esomar.org/uploads/industry/reports/global-market-research-2014/ESOMAR-GMR2014-Preview.pdf [Accessed: September 5, 2018]

Eysenbach, G. & Wyatt, J. (2002). Using the Internet for Surveys and Health Research. J Med Internet Res., 4(2), e13. Published online 2002 Nov 2 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1761932/ [Accessed: September 5, 2018].

Fazio, R. H. (1986) How do attitudes guide behavior? In Sorrentino & Higgins (Eds.) *Handbook of motivation and cognition* (pp. 204-243). New York: Guilford.

Fiske, S. T. & Kinder, D. R. (1981). Involvement, expertise, and schema use: evidence from political cognition. In: N. Cantor & J. Kihlstrom (Eds.), *Personality, cognition, and social interaction* (pp. 171-190). Hilsdale, NJ: Erlbaum.

Forgas, J.P., Vargas, P.T. (2005). Wpływ nastroju na społeczne oceny i rozumowanie. In: M. Lewis, J.M. Haviland-Jones (eds.). *Psychologia emocji*. Gdańsk: GWP.

Fronczyk, K. (2014) The identification of random or careless responding in questionnaires: The example of the NEO-FFI [Identyfikacja odpowiadania losowego lub nieuważnego w kwestionariuszu na przykładzie NEO-FFI]. *Roczniki Psychologiczne, 17*(2), 439-473.
Galesic, M., Tourangeau, R., Couper, M. P. & Conrad, F. G. (2008). Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly, 72*(5), 892-913.
Gibson, A. M. & Bowling, N. A. (2020). The Effects of Questionnaire Length and Behavioral Consequences on Careless Responding. European Journal of Psychological Assessment, 36(2), 410-420. https://doi.org/10.1027/1015-5759/a000526
Gigliotti, L. & Dietsch, A. (2014). Does Age Matter? The Influence of Age on Response Rates in a Mixed-Mode Survey. *Human Dimensions of Wildlife, 19*(3), 280-287. https://doi.org/10.1080/10871209.2014.880137
Gittelman, S., & Trimarchi, E. (2009). Variance between purchasing behavior profiles in a wide spectrum of online sample sources. White Paper, Marketing. Inc.
Gogami, M., Matsuda, Y., Arakawa, Y. & Yasumoto, K. (2021) Detection of Careless Responses in Online Surveys Using Answering Behavior on Smartphone. IEEE Access(99), 1-1. https://doi.org/10.1109/ACCESS.2021.3069049
Goldammer, P., Annen, H., Stöckli, P. L. & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *The Leadership Quarterly, 31*(4). https://doi.org/10.1016/j.leaqua.2020.101384
Górecki, M. A. (2011). Electoral Salience and Vote Overreporting: Another Look at the Problem of Validity in Voter Turnout Studies. *International Journal of Public Opinion Research, 23*(4), Winter 2011, 544–557. https://doi.org/10.1093/ijpor/edr023
Göritz, A. S., Reinhold, N., & Batinic, B. (2002). Online panels. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online Social Sciences* (pp. 27–47).
Göritz, A. S. (2010). Using online panels in psychological research. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.), Oxford Handbook of Internet Psychology.
Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the Effects of Removing 'Too Fast' Responses and Respondents from Web Surveys. *Public Opinion Quarterly, 79*(2), 471–503. https://doi.org/10.1093/poq/nfu058
Grice, P. H. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Vol. III, Speech acts* (pp. 41–58). Academic Press.
Gummer, T., Roßmann, J., & Silber, H. (2018). Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementation. *Sociological Methods and Research*, *50*(1), 238–264. https://doi.org/10.1177/0049124118769083
Hauser, D. J., Sunderrajan, A., Natarajan, M., & Schwarz, N. (2017). Prior Exposure to Instructional Manipulation Checks does not Attenuate Survey Context Effects Driven by Satisficing or Gricean Norms. *Methods, Data, Analyses*, *10*(2), 195–220. https://doi.org/10.12758/MDA.2016.008
Hensel, P. G. (2021). Reproducibility and replicability crisis: How management compares to psychology and economics – A systematic review of literature. European Management Journal, article in press. https://doi.org/10.1016/j.emj.2021.01.002
Hillygus, D. S., Jackson, N., & Young, M. (2014). Professional respondents in non-probability online panels. Online panel research: A data quality perspective, 1, 219-237.
Holland, J. L., & Christian, L. M. (2009). The Influence of Topic Interest and Interactive Probing on Responses to Open-Ended Questions in Web Surveys. *Social Science Computer Review, 27*(2), 196–212. https://doi.org/10.1177/0894439308327481
Hoyt, W.T. (2000). Rater bias in psychological research: when is it a problem and what can we do about it? *Psychological Methods, 5*, 64-86.
Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015a). Detecting Insufficient Effort Responding with an Infrequency Scale: Evaluating Validity and Participant Reactions. Journal of Business and Psychology, 30(2), 299–311. https://doi.org/10.1007/s10869-014-9357-6
Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*, 99–114.
Huang, J. L., & DeSimone, J. A. (2021). Insufficient effort responding as a potential confound between survey measures and objective tests. *Journal of Business and Psychology, 36*(5), 807-828.
Huang, J. L., Liu, M., Bowling, N. A. (2015b) Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology, 100*(3), 828-845.
Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014. World Values Survey: Round Five - Country-Pooled Datafile Version: www.worldvaluessurvey.org/WVSDocumentationWV5.jsp. Madrid: JD Systems Institute.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014. World Values Survey: Round Six - Country-Pooled Datafile Version: www.worldvaluessurvey.org/WVSDocumentationWV6.jsp. Madrid: JD Systems Institute.

Ioannidis J. P. (2005). Why most published research findings are false. PLoS medicine, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Jerzyński, T. (2009). Wybrane korelaty liczby odpowiedzi beztreściowych w badaniach sondażowych [Unpublished doctoral dissertation]. University of Warsaw.

Johnson, J. A. (2005) Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality, 39*(1), 103-129.

Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kam, C. C. S., Meyer, J.P. (2015) How Careless Responding and Acquiescence Response Bias Can Influence Construct Dimensionality: The Case of Job Satisfaction. *Organizational Research Methods, 18*(3), 512-541.

Kane, J. V., & Barabas, J. (2019). No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments. *American Journal of Political Science, 63*(1), 234–249. https://doi.org/https://doi.org/10.1111/ajps.12396

Kasvi, A. (2017). *Employee satisfaction survey: Reippailuhalli Huimala* [Haaga-Helia ammattikorkeakoulu]. http://www.theseus.fi/handle/10024/128824

Keeley, J. W., Webb, C., Peterson, D., Roussin, L., & Flanagan, E. H. (2016). Development of a Response Inconsistency Scale for the Personality Inventory for DSM – 5 Development of a Response Inconsistency Scale for the Personality Inventory for. *Journal of Personality Assessment*, *98*(4), 351–359. https://doi.org/10.1080/00223891.2016.1158719

Kiesler, S., Sproull, L. S. (1986) Response Effects in the Electronic Survey. *Public Opinion Quarterly 50*(3), 402–413

Kordos, J. (1988). Jakość danych statystycznych. Warszawa: PWE.

Kountur, R. (2016) Detecting careless responses to self-reported questionnaires. *Egitim Arastirmalari - Eurasian Journal of Educational Research, (64)*, 307-318.

Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology, 5*, 213-236.

Krosnick, J. (2002). The Causes of No-Opinion Responses to Attitude Measures in Surveys: They are Rarely What They Appear to be. In Groves, Dillman, Eltinge and Little (Eds.) *Survey Nonresponse* (pp. 87-100). New York: Wiley.

Krosnick, J.A., Holbrook, A.L., Berent, M.K., Carson, R.T., Hanemann, W.M., Kopp, R.J., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K., Moody, W.R., Green, M.C., Conaway, M. (2002). The impact of „no opinion" response options on data quality: non–attitude reduction or an invitation to satisfice? *The Public Opinion Quarterly, 66*(3), 371–403.

Krosnick, J. A. & Alwin, D. F. (1989). A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. *Public Opinion Quarterly, 52*, 526-538.

Krosnick, J. A. & Presser, S. (2010). Questionnaire design. In: J. D. Wright & P. V. Marsden (Eds.), *Handbook of survey research* (2nd ed.). West Yorkshire, England: Emerald Group.

Król, G., & Kowalczyk, K. (2014). Ewaluacja projektów i abstraktów - wpływ indywidualnego stylu ewaluacji na oceny. *Problemy Zarządzania, 1/2014*(45), 137-155.

Kuha, J., Butt, S., Katsikatsou, M. & Skinner, C. J. (2017). The Effect of Probing "Don't Know" Responses on Measurement Quality and Nonresponse in Surveys. *Journal of the American Statistical Association*, *113*(521), 26-40. https://doi.org/10.1080/01621459.2017.1323640

Kung, F. Y. H., Kwok, N., & Brown, D. J. (2018). Are Attention Check Questions a Threat to Scale Validity? *Applied Psychology*, *67*(2), 264–283. https://doi.org/10.1111/apps.12108

Kurtz, J. E., Parrish, C. L. (2001) Semantic response consistency and protocol validity in structured personality assessment: the case of the NEO-PI-R. *Journal of Personality Assessment, 76*(2), 315-32.

Kuźmińska, A. O. & Pazura, D. (2018). The impact of Control Preferences Fit Between Employees and Their Supervisors on Employee Job Satisfaction. *Studia i Materiały, 29*(2), 18-32.

Kuźmińska, A. O., Schulze, D. & Koval, A. (2019). Who Doesn't Want to Share Leadership? The Role of Personality, Control Preferences, and Political Orientation in Preferences for Shared vs. Focused Leadership in Teams. In Kuźmińska, A. (ed.), Management Challenges in the Era of Globalization (pp. 11–27). Warsaw: University of Warsaw Faculty of Management Press.

Kwak, N. & Radler, B. (2002). A Comparison Between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. *Journal of Official Statistics, 18*(2), 257-273.

Lake, C. J., Withrow, S., Zickar, M. J., Wood, N. L., Dalal, D. K., Bochinski, J. (2013) Understanding the Relation Between Attitude Involvement and Response Latitude Using Item Response Theory. *Educational and Psychological Measurement, 73*(4), 690-712.
Levi, R., Ridberg, R., Akers, M., & Seligman, H. (2021). Survey Fraud and the Integrity of Web-Based Survey Research. *American Journal of Health Promotion*. Advance online publication. https://doi.org/10.1177/08901171211037531
Lu, J., Wang, F., Wang, X., Lin, L., Wang, W., Li, L., & Zhou, X. (2019). Inequalities in the health survey using validation question to filter insufficient effort responding: Reducing overestimated effects or creating selection bias? *International Journal for Equity in Health, 18*(1). https://doi.org/10.1186/s12939-019-1030-2
McGrath, R. E., Mitchell, M., Kim, B. H., Hough, L. (2010) Evidence for Response Bias as a Source of Error Variance in Applied Assessment. *Psychological Bulletin, 136*(3), 450-470.
McKay, A. S., Garcia, D. M., Clapper, J. P., & Shultz, K. S. (2018). The attentive and the careless: Examining the relationship between benevolent and malevolent personality traits with careless responding in online surveys. *Computers in Human Behavior, 84*, 295–303. https://doi.org/10.1016/J.CHB.2018.03.007
McKibben, W. B., & Silvia, P. J. (2017). Evaluating the Distorting Effects of Inattentive Responding and Social Desirability on Self-Report Scales in Creativity and the Arts. *The Journal of Creative Behavior, 51*(1), 57–69. https://doi.org/10.1002/jocb.86
Meade, A. W., Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437-455.
Merckelbach, H., Giesbrecht, T., Jelicic, M., Smeets, T. (2010) The problem of careless respondents in surveys [Kretenzers in surveys: Het probleem van onzorgvuldige respondenten]. *Tijdschrift voor Psychiatrie, 52*(9), 663-669.
Merritt, S. M. (2012) The Two-Factor Solution to Allen and Meyer's (1990) Affective Commitment Scale: Effects of Negatively Worded Items. *Journal of Business and Psychology, 27*(4), 421-436.
Meyer, J. F., Faust, K. A., Faust, D., Baker, A. M., Cook, N. E. (2013) Careless and Random Responding on Clinical and Research Measures in the Addictions: A Concerning Problem and Investigation of their Detection. *International Journal of Mental Health and Addiction, 11*(3), 292-306.
Michałowicz, B. (2016) Ankiety ewaluacyjne w szkolnictwie wyższym: wpływ wyboru ewaluatorów (Doctoral dissertation). [Access on the author's request] https://depotuw.ceon.pl/handle/item/1532
Mitchell, A. L., Hegedüs, L., Žarković, M., Hickey, J. L., & Perros, P. (2021). Patient satisfaction and quality of life in hypothyroidism: An online survey by the British thyroid foundation. *Clinical Endocrinology*, *94*(3), 513–520. https://doi.org/10.1111/cen.14340
Moxey, L. M. & Sanford, A. J. (1986). Quantifiers and Focus. *Journal of Semantics, 5*(3), 189-206. https://doi.org/10.1093/jos/5.3.189
Moxey, L. M. & Sanford, A. J. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology 14*(3),237–55.
Moździerz, T. (2019). Tempo czytania rodzimych i nierodzimych użytkowników polszczyzny. *Języki Obce W Szkole, 63*(4), 79–85. http://jows.pl/sites/default/files/wydania/jows_4_2019_online-calosc_0.pdf
Murphy, F., Macpherson, K., Jeyabalasingham, T., Manly, T., & Dunn, B. (2013). Modulating mind-wandering in dysphoria. *Frontiers in Psychology, 4*(NOV). https://doi.org/10.3389/fpsyg.2013.00888
Nancarrow, C., & Cartwright, T. (2007). Online access panels and tracking research: The conditioning issue. *International Journal of Market Research, 49*, 573–594.
Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. PLoS One, 11(6), e0157732. https://doi.org/10.1371/journal.pone.0157732.
Nichols, A. L., & Edlund, J. E. (2020). Why don't we care more about carelessness? Understanding the causes and consequences of careless participants. *International Journal of Social Research Methodology, 23*(6), 625–638. https://doi.org/10.1080/13645579.2020.1719618
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.
Pratt, J. W., Raiffa, H., and Schlaifer, R. (1995). *Introduction to statistical decision theory*. Cambridge, MA: MIT Press.
Revilla, M. (2012). Impact of the Mode of Data Collection on the Quality of Answers to Survey Questions Depending on Respondent Characteristics. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 116*(1), 44–60. https://doi.org/10.1177/0759106312456510
Robert Zajonc Institute for Social Studies. (2013). Polish General Social Surveys 1992-2010. Study documentation. Retreived July 23, 2021, from http://www.ads.org.pl/pobieranie-zbioru-danych.php?id=91

Roivainen, E., Veijola, J. & Miettunen, J. (2016). Careless responses in survey data and the validity of a screening instrument. *Nordic Psychology, 68*(2), 114-123. https://doi.org/10.1080/19012276.2015.1071202
Roßmann, J., Gummer, T., & Silber, H. (2018). Mitigating satisficing in cognitively demanding grid questions: Evidence from two web-based experiments. *Journal of Survey Statistics and Methodology, 6*(3), 376–400. https://doi.org/10.1093/JSSAM/SMX020
Queloz, S., & Etter, J.-F. (2019). An online survey of users of tobacco vaporizers, reasons and modes of utilization, perceived advantages and perceived risks. *BMC Public Health, 19*(1), 1–11. https://doi.org/10.1186/S12889-019-6957-0
Sammelman, K., Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior research methods, 50*(2), 451–465. https://doi.org/10.3758/s13428-017-0913-7
Schlaifer, R. (1969). *Analysis of decisions under uncertainty*. New York: McGraw-Hill.
Schneider, S., May, M., & Stone, A. A. (2018). Careless responding in internet-based quality of life assessments. *Quality of Life Research, 27*(4), 1077–1088. https://doi.org/10.1007/s11136-017-1767-2
Schroeders, U., Schmidt, C., & Gnambs, T. (2022). Detecting Careless Responding in Survey Data Using Stochastic Gradient Boosting. Educational and Psychological Measurement, 82(1), 29–56. https://doi.org/10.1177/00131644211004708
Schuman, H. & Presser, S. (1981). Questions and answers in attitude surveys. New York: Academic Press
Shin, E., Johnson, T. P., & Rao, K. (2012). Survey Mode Effects on Data Quality: Comparison of Web and Mail Modes in a U.S. National Panel Survey. *Social Science Computer Review, 30*(2), 212–228. https://doi.org/10.1177/0894439311404508
Silber, H., Danner, D., & Rammstedt, B. (2019). The impact of respondent attentiveness on reliability and validity. *International Journal of Social Research Methodology, 22*(2), 153–164. https://doi.org/10.1080/13645579.2018.1507378
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22(11), 1359–1366. https://doi.org/10.1177/0956797611417632*
Skarżyńska, K., Urbańska, B., & Radkiewicz, P. (2021). Under or Out of Government Control? The Effects of Individual Mental Health and Political Views on the Attribution of Responsibility for COVID-19 Incidence Rates. *Social Psychological Bulletin, 16*(1), e4395. https://doi.org/10.32872/spb.4395
Smith, R., & Brown, H. H. (2006). Data and panel quality: Comparing metrics and assessing claims. In: Proceedings of the ESOMAR Panel Research Conference.
Steedle, J. T., Hong, M., & Cheng, Y. (2019). The Effects of Inattentive Responding on Construct Validity Evidence When Measuring Social–Emotional Learning Competencies. *Educational Measurement: Issues and Practice*, *38*(2), 101–111. https://doi.org/https://doi.org/10.1111/emip.12256
Sułek, A. (2001). *Sondaż polski: Przygarść rozpraw o badaniach ankietowych*. Warszawa: Wydawnictwo Instytutu Filozofii i Socjologii PAN.
Sztabiński, F., Sztabiński, P. B. & Sawiński, Z. (2005). *Fieldwork jest sztuką : jak dobrać respondenta, skłonić do udziału w wywiadzie, rzetelnie i sprawnie zrealizować badanie: praca zbiorowa*. Warszawa: Wydawnictwo Instytutu Filozofii i Socjologii PAN.
Toepoel, V., Das, M., & Van Soest, A. (2008). Effects of design in web surveys: Comparing trained and fresh respondents. *Public Opinion Quarterly, 72*, 985–1007.
Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013). The Science of Web Surveys. In *The Science of Web Surveys*. Oxford University Press.
Tourangeau, R., Rasinski, K. A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin, 103*(3), 299-314. https://doi.org/10.1093/acprof:oso/9780199747047.001.0001
Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
Ulrich, D. (1998) Measuring human resources: An overview of practice and a prescription for results. *Human Resource Management, 36*(3), 303-320.
Vehovar, V., Lozar Manfreda, K. (2008). Overview: Online Surveys. In Fielding, N.; Lee, R. M.; Blank, G. The SAGE Handbook of Online Research Methods. London: SAGE. pp. 177–194. https://doi.org/10.3102/0013189X211040054
Verbree, A.-R., Toepoel, V., & Perada, D. (2020). The Effect of Seriousness and Device Use on Data Quality. Social Science Computer Review, 38(6), 720–738. https://doi.org/10.1177/0894439319841027
Weathers, D., Bardakci, A. (2015) Can response variance effectively identify careless respondents to multi-item, unidimensional scales? *Journal of Marketing Analytics, 3*(2), 96-107.
Weijters, B., Baumgartner, H., Schillewaert, N. (2013) Reversed item bias: An integrative model. *Psychological Methods, 18*(3), 320-334.

Wieczorkowska, G (1993). Pułapki statystyczne. W: M.Z. Smoleńska (red.). Badania nad rozwojem w okresie dorastania. Warszawa: Instytut Psychologii PAN.
Wieczorkowska-Nejtardt, G. (1998). Inteligencja motywacyjna: mądre strategie wyboru celu i sposobu działania. Warszawa: Wydawnictwa Instytutu Studiów Społecznych UW.
Wieczorkowska-Wierzbińska, G. (2011) Psychologiczne ograniczenia. WN WZ UW, Warszawa.
Wieczorkowska-Wierzbińska, G. (2014). Diagnoza psychologiczna predyspozycji pracowników. *Problemy Zarządzania, 12*(1), 81-98.
Wieczorkowska-Wierzbińska, G. (2022). Zarządzanie ludźmi – z psychologicznego i metodologicznego punktu widzenia. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego
Wieczorkowska, G., & Kowalczyk, K. (2021). Ensuring Sustainable Evaluation: How to Improve Quality of Evaluating Grant Proposals? *Sustainability, 13*, 2842.
Wieczorkowska, G. & Król, G. (2016). Ten pitfalls of research practices in management science. *Problemy Zarządzania, 2*(2), 173-187.
Wieczorkowska, G., Król, G. Wierzbiński, J. (2015). Metody Ilościowe. W: Kostera, M. (red.) Metody badawcze w zarządzaniu humanistycznym, WA Sedno, Warszawa.
Wieczorkowska, G., Król, G., Wierzbiński, J. (2016). Cztery metodologiczne zagrożenia w naukach o zarządzaniu. *Studia i Materiały, 2*(2), 146-156.
Wieczorkowska, G., & Wierzbiński, J. (2005). *Badania sondażowe i eksperymentalne. Wybrane zagadnienia.* Warszawa: Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego.
Wieczorkowska, G. & Wierzbiński, J. (2011). Statystyka. Od teorii do praktyki. Warszawa: SCHOLAR.
Wieczorkowska-Wierzbińska, G., Wierzbiński, J., Kuźmińska, A. (2014). Porównywalność danych zebranych w różnych krajach. *Psychologia Społeczna 9*(2), 128-143.
Wood, D., Harms, P. D., Lowman, G. H., DeSimone, J. A. (2017) Response Speed and Response Consistency as Mutually Validating Indicators of Data Quality in Online Samples. *Social Psychological and Personality Science, 8*(4), 454-464.
Yan, T., Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology, 22*(1), 51-68.
Zhang, C., & Conrad, F. G. (2014). Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127–135. https://doi.org/https://doi.org/10.18148/srm/2014.v8i2.5453
Zieliński, M. (2009). Analiza hierarchiczna w szacowaniu wpływu ankietera na odpowiedzi respondenta. *Problemy Zarządzania, 4*, 238–253.