Studia i Materiały, 1/2023 (38): 85–95
ISSN 1733-9758,
https://doi.org/10.7172/1733-9758.2023.38.8

False Respondents in Web Surveys

*Marta Kabut**

Purpose: *The aim of the conducted analyses was to propose and test an FR procedure for detecting false respondents (who answer survey questions mindlessly) in online surveys.*
Design/methodology/approach: *Statistical analyses of data from 9 online surveys with a total of 4224 respondents, and 3 offline surveys (a total of 3169 respondents), aimed to identify false respondents using 4 warning signs (WS) based on: (WS1) too short answering time, (WS2) attention check questions, (WS3) rating style that considers, among others, the number of "Don't know", (WS4) logical consistency test of the answers and self-reported engagement of respondents.*
Findings: *The percentage of respondents flagged by any of 4 signs (strict criteria) ranged from 5.2% to 71% depending on the survey. With lenient criteria (allowing respondents to be flagged by one warning sign), the percentage of excluded respondents ranged from 0% to 45.9%. Respondents could be excluded from analyses locally (for a specific block of items) or globally.*
Research limitations/implications: *The surveys used in the analyses in this paper were of high quality (designed to minimize the participation of false respondents), which means that the percentages of false respondents for surveys made available to all interested parties will be higher. The analyzed data included respondents with at least secondary education.*
Originality/value: *The conducted analyses provide evidence for the necessity of cleaning data obtained in online surveys. The tested FR procedure proved to be useful. The utility of the FLEXMIX procedure for examining the logical consistency of respondents' answers was also demonstrated.*

Keywords: online survey, false respondents, survey research.

Submitted: 16.04.2023 | Accepted: 29.05.2023

Fałszywi respondenci w ankietach internetowych

Cel: *celem przeprowadzonych analiz było zaproponowanie i przetestowanie procedury FR do wykrywania fałszywych respondentów (odpowiadających bezmyślnie na pytania ankiety) w badaniach internetowych.*
Projekt/metodologia/podejście: *analizy statystyczne danych z 9 badań internetowych (w których w sumie uczestniczyło 4224 osób) i 3 badań nieinternetowych (łącznie 3169 respondentów) miały na celu identyfikację fałszywych respondentów za pomocą czterech sygnałów ostrzegawczych opartych na: (WS1) zbyt krótkim czasie odpowiedzi; (WS2) pytaniach sprawdzających uwagę; (WS3) stylu oceniania z uwzględnieniem liczby odpowiedzi DK (nie wiem); (WS4) teście spójności logicznej odpowiedzi oraz deklaracji zaangażowania respondentów.*
Wyniki: *procent respondentów oznaczonych przez wszystkie cztery sygnały (kryterium ostre) wahał się w granicach od 5,2 do 71% w zależności od badania. Przy kryterium łagodnym (akceptowani respondenci także z jednym sygnałem ostrzegawczym) procent wykluczonych wyniósł od 0 do 45,9%.*

* **Marta Kabut** – PhD, Faculty of Management, University of Warsaw. https://orcid.org/0000-0002-2870-0258. *Correspondence address:* Faculty of Management, University of Warsaw, Szturmowa 1/3, 02-678 Warsaw, Poland; e-mail: MartaKabut@wz.uw.edu.pl.

Edition of that article was financed under Agreement Nr RCN/SP/0323/2021/1 with funds from the Ministry of Education and Science, allocated to the "Rozwoj czasopism naukowych" programme.

Ograniczenia/implikacje: *badania w analizach użytych w tym artykule były wysokiej jakości (zaprojektowane w sposób minimalizujący udział fałszywych respondentów), co oznacza że procenty fałszywych respondentów dla ankiet udostępnianych dla wszystkich chętnych będą wyższe. W analizowanych danych respondenci mieli co najmniej wykształcenie średnie.*
Oryginalność/wartość: *przeprowadzone analizy dowodzą konieczności czyszczenia danych pozyskanych w badaniach internetowych. Przetestowana procedura FR wykazała swoją przydatność. Pokazano także użyteczność procedury FLEXMIX do badania spójności logicznej odpowiedzi respondentów.*

Słowa kluczowe: ankieta internetowa, fałszywi respondenci, badania ankietowe.

JEL: C82

1. Introduction

Online surveys have replaced other ways of conducting studies and have had a dominant position among quantitative research methods (ESOMAR, 2014). The total data collection expenditures in online surveys increased from 20% in 2006 to above 50% in 2013 (Vehovar & Lozar Manfreda, 2008).

Web self-administered surveys (Batorski & Olcoń-Kubicka, 2006) have become a prevalent form of data collection in human resource management (HRM) and research focused on the satisfaction of customers and employees (Kasvi, 2017; Barakat et al., 2015; Mitchell et al., 2021), marketing (e.g., Queloz & Etter, 2019; Kumar Mishra et al., 2016), consumer preference and behavior (Molenaar et al., 2018).

2. Literature Review

Types of internet research can be distinguished based on the following criteria (Batorski & Olcoń-Kubicka, 2006): (1) participants' awareness that they are taking part in research; (2) time of the research: real-time vs. anytime; (3) level of participants' 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 persons hard to reach in other ways; (3) time saving; (4) lower cost; (5) flexibility (the next question could be selected depending on the former answers).

The internet is also suitable for experimental research and enables the possibility of integrating qualitative and quantitative methods in one study. The ease of recruiting respondents comes with limited or lack of control over their behavior and environment.

Some respondents can choose one of the following (harmful to the research validity) strategies (Krosnick, 1991):

1) Selecting the first response alternative that seems reasonable (Galesic et al., 2008);
2) Selecting the most visible option (Couper et al., 2004);
3) Speeding – answering too fast (Conrad et al., 2017; Michałowicz, 2016);
4) Acquiescence bias – agreeing with any statement regardless of content (Krosnick, 1991);
5) Endorsing the status quo (Schuman & Pressner, 1981) – when a question asks about increasing or decreasing something, respondents often choose a base (starting) value when explicitly given to them;
6) Lack of differentiation in using rating scales (Krosnick & Alwin, 1989) – 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 pretend they do not have an opinion rather than putting effort into forming one. However, research shows that providing this answer option increases data quality (Albaum et al., 2011);
8) Mental coin-flipping (Converse, 1964) – choosing randomly from among the response alternatives;

9) Omitting a whole set of questions, either by losing one's interest or on purpose. This does not mean that the answers are worthless, but there are difficulties in determining what to do with them.

False responding (Levi et al., 2021) has been named in literature in many ways: random (Credé, 2010), insufficient effort (Huang et al., 2012; Huang & DeSimone, 2021), careless (Meade & Craig, 2012; Bowling et al., 2020), satisficing (Krosnick, 1991), inattentive/participant inattention (McKibben & Silvia, 2017; Beck et al., 2019; Steedle et al., 2019), and indiscriminate responding (Holden et al., 2019). It can be defined broadly as happening when the respondent does not cooperatively fill in the survey. Such people may introduce random noise to data collected in surveys. However, they usually do not answer entirely randomly, which leads to systematic bias in responses, thus incurring a deviation from inferred results (Alvarez et al., 2019).

Based on the literature review, it can be stated that the percentage of false respondents varies and depends on the type and number of methods used in a particular study (Johnson, 2005; Kurtz & Parish, 2001; Meade & Craig, 2012; Curran et al., 2010; Baer et al., 1997).

3. Research Purpose and Research Tasks

Many studies on inattentive respondents have been done on English-speaking samples (e.g., Nichols & Edlund, 2020, Schneider et al., 2018, Bowling & Huang, 2018, Alvarez et al., 2019) but not for Polish samples. The research gap to be filled is determining the level of inattention of respondents, the consequences of including false respondents in analyses, and devising an **FR procedure** to detect false respondents in data sets.

Three 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 (3) testing the usability of the FLEXMIX (finite mixtures of generalized regression models) procedure for detecting false respondents.

4. Key Terms

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

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 answering TIME. The overall answering time (OAT) is the time that passed from the first load of the first survey page to the end page shown. The partial answering time (PAT) is the time spent on answering blocks of the survey.
2) WS2 is based on the number of incorrect answers to attention check questions (ACQ).
3) WS3 is based on too big a number of Do not KNOW answers and low differentiation rating style.
   DK – (Do not KNOW – Non-informative answers) **–** answers that do not convey any information about the respondent's opinion/thinking/facts.
   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 (Wieczorkowska, 1993, Harzing et al., 2011). The rating style can manifest itself through: (1) too severe (or lenient) assessment (Hoyt, 2000), (2) lack of differentiation of partial dimensions of evaluation (Landy et al., 1980), e.g., AGREE to almost all items on the scale.
4) WS4 is based on a low BEHAVIORAL cooperation (BC) level (logical inconsistency, odd answers to open-ended questions) and a low DECLARATIVE cooperation (DC) level (answers to direct questions about respondents' engagement, i.e., would their answers have changed if it had been a different day?).
   LOGICAL INCONSISTENCY is operationalized based on a 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 a question on satisfaction).

Exclusion criteria:

- 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.

5. Materials and Methods

The distribution of warning signs was analyzed in **9 web surveys conducted by our doctoral team** at the Academic Unit for Organizational Psychology and Sociology, Faculty of Management, University of Warsaw, between 2020 and 2022 (WS1 tested on 9 data sets, WS2 tested on 8 data sets, WS3 and WS4 tested on 12 data sets):

- two data sets (A1–A2) consisting of 2,918 employees (commercial panel participants);
- six data sets (B1–B6) based on responses from 2,399 participants who, in the overwhelming majority, combined 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, 1,203 Polish employees;
- data sets E1–E2, World Values Survey, two waves (5 + 6), 1,966 Polish respondents.

IBM SPSS software was used to analyze all data sets.

6. Results

The analysis showed that the percentage of respondents flagged as “false” depended on the survey and the type of WS.

As we can see in Figure 1, for 7 web surveys: C and B1–B6, the more discriminating criterion was WS4.

Figure 1. Comparison of Different Rates of Exclusion for Data Sets B1–B6 and C, in Division by WS

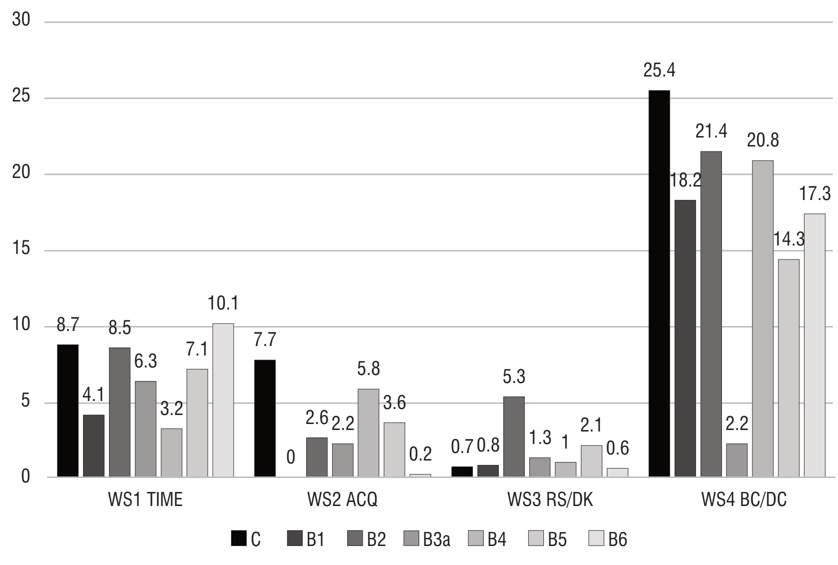
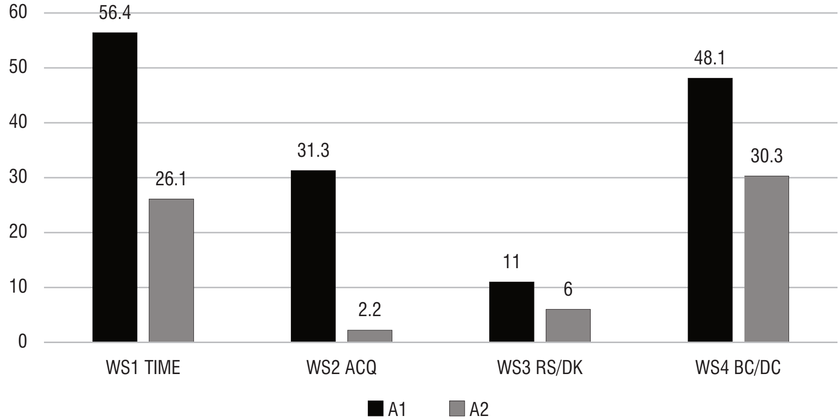


Figure 2. Rates of Exclusion for 4 WS for Two Commercial Panel Data Sets [A1 and A2]



As shown in Figure 2, for two panel data sets: A1 and A2, the most exclusionary criterion was WS1 (time), and WS4 was the second. Data sets A1 and A2 were two different questionnaires – data set A1 contained more longer paragraphs, data set A2 contained mainly short questions.

There are statistically significant differences between the distribution of WS in two paid panel studies. Both surveys were conducted by the same company that sells its services to researchers.

The difference in WS2 between A1 and A2 can be explained by the different types 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 without providing the respondent with a reason. This unexplained order could make some respondents angry and reactant. In A2, five arithmetic questions (i.e., “Choose the correct result of this operation 23 + 5 = ”) were used, and it was justified as a break to alleviate 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 a mistake.

In Table 1, the median of answering time with FR excluded for A1 is shorter than for A2 by 3 minutes, contradicting the slogan that ‘the shorter survey, the better’. With exclusion by WS 2–4, but without WS1, OAT differs by around 13 minutes – showing that using WS1 is necessary.

Table 1. Comparison of Median Time for Data Sets A1 and A2

Data set	A1	A2
False respondents (based on 4 WS – strict criterion)	**71.0%**	**46.6%**
OAT median (for attentive respondents not excluded by WS1)	14:06 min	26:25 min
Number of words (in all questions, without repeated rating scales)	3383 words	3628 words
Median time without FR	27:17 min	30:17 min

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 WS were available.

Table 2. WS 3 and 4 for Three Offline Data Sets

Warning signs	D	E1	E2
WS3 rating	3.7	2.8	3.1
WS4 logical	5.7[a]	13.2[b]	5.7[b]

[a] based on assessed cooperation – 2 questions,
[b] based on assessed interest – 1 question.

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

(1) the procedure based on 4WS (4 warning signs),

(2) the FLEXMIX model (combining cluster and regression analysis).

FLEXMIX allows respondents to be divided into subgroups based on their fit to different regression lines – it divides respondents into two groups based on the 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 they know. 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 inattentive in reading the questions.

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

The next step was a comparison of Cronbach's alphas in the group of FR flagged by each procedure and in the group that passed the test. To compute Cronbach's alpha, two indicators from SSA (Wieczorkowska, 2022) were used: in data set A1 – METHODICALITY index and in data set A2 – EXTRAVERSION index.

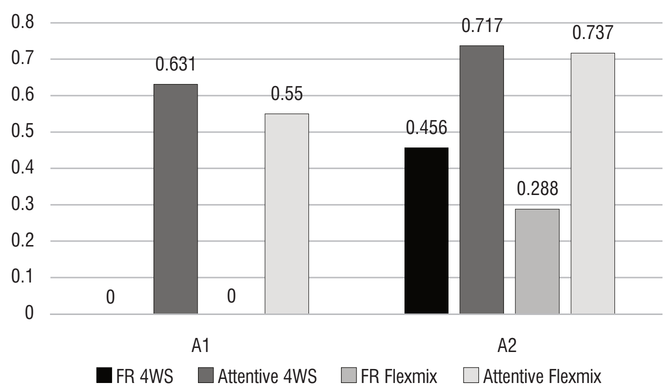
In Figure 3, a comparison of 4 Cronbach's alphas in study A1 (left) and study A2 (right) is shown.

In both datasets and both procedures, the value of Cronbach's alpha is acceptable in the group of attentive respondents and NOT acceptable in the group of FR. A negative alpha value indicates that FR

Table 3. Comparison of FR Flagged by 4WS and FLEXMIX

	# 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%)

Figure 3. Cronbach's Alphas for Groups of FR and Attentive Respondents



Note. This is a division by data set [A1, A2] and method [WS, FLEXMIX; negative Cronbach's alphas were converted to 0 (-2.8 for A1, and -0.38 for A2).

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 FR. This can be explained as 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 analyzes the respondent's behavior throughout the survey.

Therefore, the FLEXMIX procedure can only be recommended to help examination for WS4. Automation of this process using the FLEXMIX procedure is advisable, but we must use more than two questions.

Procedure for Detecting False Respondents

The values of 4WS should be computed for each respondent. The procedure itself starts BEFORE the study commences – all questions (metadata) for WS need to be planned and placed in the survey with an assumed version of the procedure already in mind.

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

The threshold for WS1 means the minimal time needed to read the questions. It can be set by testing the survey on a small sample of trusted respondents or using the reading speed (words per minute; reading speed depends on the characteristics of the sample, but default threshold for my studies was 300 word per minute, which is a conservative approach).

The threshold for WS2 means the acceptable number of errors in ACQ. This depends on whether a lenient or strict criterion was chosen for this WS – lenient means that 1 error is acceptable, strict means that no errors are acceptable.

The threshold for WS3 means the lowest acceptable variance in answering a series of questions with the same rating scale (variance is below -2σ), the biggest acceptable number of DK answers (usually less than 50% – based on SSA questionnaire).

The threshold for WS4 means an acceptable level of logical inconsistency in closed- and open-ended questions, an acceptable level of declared engagement in the survey, etc. This is highly context- and survey-dependent.

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

From 0 – means NO WS, to 4 – means that all 4 WS flagged the respondent.

Step 2. Decide on a STRICT or LENIENT criterion.

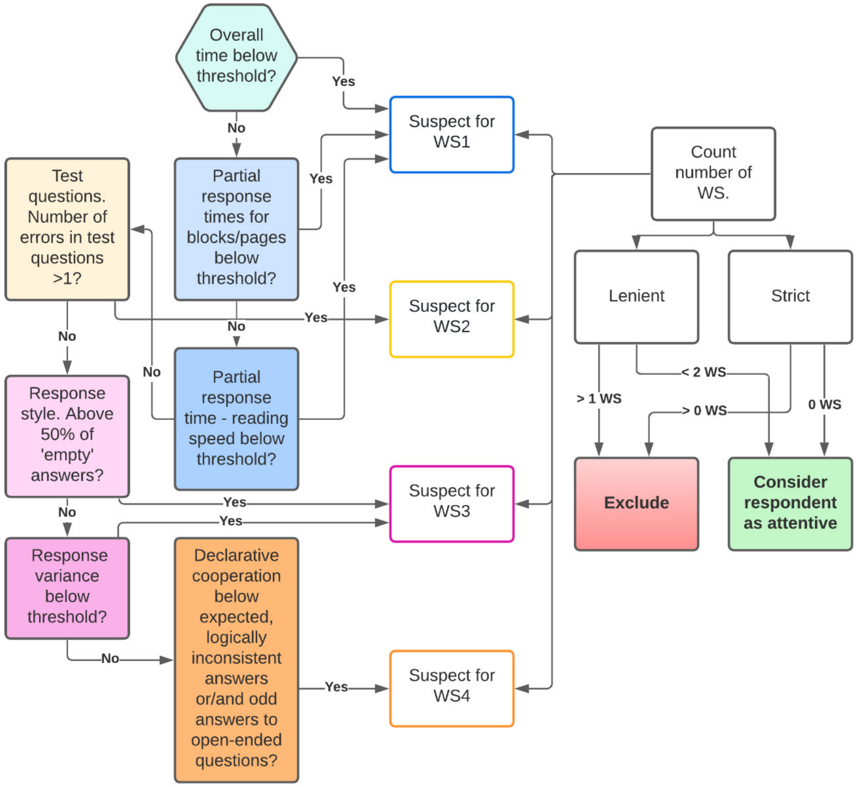
The comparison of the consequences of this decision can be seen in Table 4.

The graphical form of the 4 WS procedure is shown in Figure 4.

Table 4. Percent of FR Depending on Study and Criterion

Data set	Year	Sample	% of respondents excluded	
			Lenient criterion	Strict criterion
A1	2018	1,421 + 1,497 panel employed respondents	45.9	71.0
A2	2021		14.2	45.4
C	2020	287 employees, convenience sample	6.6	33.8
B1–B6	2018–2021	2,440 respondents (in the overwhelming majority, combine studies with work)	1.2–5.0	6.1–13.8
D, E2, E3	2005/2010/2015	3,169 respondents (personal interviews, offline)	0–0.9	5.2–7.3

Figure 4. Visual Scheme of the Procedure for Detecting False Respondents



7. Conclusions

Comparing percentages of FR in the analyzed 12 data sets from **7.3% to 71.0%**, with literature review points in the range between **4% and 97.8%**, we conclude that they are study-dependent. The magnitude of the problem could be enormous.

The analysis of WS does not show any general patterns that allow us to rank their relative importance. That means that all should be calculated, but **we need to plan it before data collection.**

Using only **WS1 <too short answering time>** to detect false respondents is not enough. Speeding respondents can take a coffee break and stay undetected.

Some researchers claim that a single ACQ can be effective (Maniaci & Rogge, 2014), while others recommend using more than one (Liu & Wronski, 2018; Berinsky et al., 2014) because of the dynamics of the respondent's attention.

More ACQ are a better choice, but we need to provide justification for the respondents, so arithmetic questions are recommended.

WS2 and WS4 can be used only globally, but for WS1 and WS3, a local analysis is recommended: measuring the answering time and the number of "Do not know" answers for the survey blocks. If the value is above the threshold, all answers for that block could be converted into missing values.

8. Limitations

The limitations of the research presented in the paper come from the type of analyzed data.

High-quality surveys. Offline data files consist of publicly available high-budget international surveys. Online data files consisted of research conducted by the doctoral team at the Academic Unit for Organizational Psychology and Sociology, Faculty of Management, University of Warsaw, where measurement tools were constructed with great concern about respondents' motivation.

Limited-access survey. The invitation to participate in this research was sent to selected groups of respondents who were motivated by different means. We can predict that the number of false respondents will be much bigger in open-access surveys.

Restricted education level of respondents. All respondents in the online survey were at least high school graduates.

9. Directions for Future Research

Parts of the process can be automatized – currently, the proposed FR procedure must be executed mostly manually.

The proposed procedure should be compared with the results of machine learning algorithms (Schroeders et al., 2022; Gogami et al., 2021).

It should be checked whether the **FR procedure** could be used to detect bots (Dennis et al., 2018; Buchanan & Scofield, 2018) and to investigate how efficient it possibly would be in that application.

Another issue is to test the impact of immediate feedback and feedback in general, which seems to motivate respondents to give more thought-out responses.

What else might be checked is 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 (Maniaci & Rogge, 2014) indicating that older respondents are more attentive than younger ones.

More experimental studies are needed. All presented analyses are correlational, hence their internal validity is limited.

The first experiment has already been conducted (see: Kabut, 2021). 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 significantly more errors (operationalized as more than two clicks on the arithmetic question) in group E1 than in E2. Both groups did not differ concerning other warning signs. Forcing respondents to correct the wrong answer **did not improve** their attention.

10. Contribution

The work presented here (see also: Kabut, 2021) has a cognitive, methodological and applicative contribution. The presence of FR in data files drastically reduced the reliability of the measurement. Unreliable data from false respondents may change correlations, render the analysis and evaluation of research results difficult (Maniaci & Rogge, 2014), decrease the statistical power and effect size (Brühlmann et al., 2020), and lower internal consistency. HRM theories confirmed by biased data are not valid, so the detection of false respondents is an important pre-analysis task.

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

The applicative contribution consists in developing a procedure for detecting false respondents in HRM studies that other researchers could use.

The proposed 4 WS procedure could be used to increase the quality of data as well as for analysis and conclusions.

References

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

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

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.

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

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

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

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

Berinsky, A.J., Margolis, M.F., & Sances, M.W. (2016). Can we turn shirkers into workers? *Journal of Experimental Social Psychology*, *66*, 20–28.
https://doi.org/10.1016/j.jesp.2015.09.010

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

Brühlmann, F., Petralito, S., Aeschbach, L.F., & Opwis, K. (2020). The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology*, *2*, 100022.

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

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* (pp. 206–261). 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* [Paper presentation]. 25th Annual Conference of Society for Industrial and Organizational Psychology, Atlanta, GA.

Dennis, S., Goodson, B., & Pearson, Ch. (2018). Mturk workers' use of low-cost 'virtual private servers' to circumvent screening methods: A research note. *SSRN Electronic Journal*, 10.2139/ssrn.3233954.

European Society for Opinion and Market Research (ESOMAR). (2013). 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*. Retrieved September 5, 2018, form https://www.esomar.org/uploads/industry/reports/global-market-research-2014/ESOMAR-GMR2014-Preview.pdf

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.

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

Harzing, A.W., Köster, K., & Magner, U. (2011). Babel in business: The language barrier and its solutions in the HQ-subsidiary relationship. *Journal of World Business*, *46*(3), 279–287.

Holden, R., Marjanovic, Z., & Troister, T. (2019). Indiscriminate responding can increase effect sizes for clinical phenomena in nonclinical populations: A cautionary note. *Journal of Psychoeducational Assessment*, *37*(4), 464–472.

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., 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.

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

Kabut, M. (2021). False respondents in web human resource surveys [Unpublished doctoral dissertation]. Uniwersytet Warszawski.

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

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

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.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.

Kumar Mishra, M., Kesharwani, A. & Das, D. (2016). The relationship between risk aversion, brand trust, brand affect and loyalty: Evidence from the FMCG industry. *Journal of Indian Business Research*, *8*(2), 78–97.

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.

Landy, F.J., Vance, R.J., Barnes-Farrell, J.L., & Steele, J.W. (1980). Statistical control of halo error in performance ratings. *Journal of Applied Psychology*, *65*(5), 501–506.

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

Liu, M., & Wronski, L. (2018). Trap questions in online surveys: Results from three web survey experiments. *International Journal of Market Research*, *60*(1), 32–49.

Maniaci, M., & Rogge, R. (2014, February). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83.

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.

Michałowicz, B. (2016) Ankiety ewaluacyjne w szkolnictwie wyższym: wpływ wyboru ewaluatorów [Doctoral dissertation, access at 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

Molenaar, D., Bolsinova, M., & Vermunt, J. (2018). A semi-parametric within-subject mixture approach to the analyses of responses and response times. *British Journal of Mathematical and Statistical Psychology*, *71*, 205–228.

Nancarrow, C., & Cartwright, T. (2007). Online access panels and tracking research: The conditioning issue. *International Journal of Market Research*, *49*, 573–594.

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

Pratt, J.W., Raiffa, H., & Schlaifer, R. (1995). *Introduction to statistical decision theory*. MIT Press.

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

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. Academic Press.

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

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

Vehovar, V., & Lozar Manfreda, K. (2008). Overview: Online surveys. In N. Fielding, R.M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 177–194). SAGE. https://doi.org/10.3102/0013189X211040054

Wieczorkowska, G. (1993). Pułapki statystyczne. In M.Z. Smoleńska (Ed.), *Badania nad rozwojem w okresie dorastania*. Instytut Psychologii PAN.

Wieczorkowska-Wierzbińska, G. (2011) *Psychologiczne ograniczenia*. WN WZ UW.

Wieczorkowska-Wierzbińska, G. (2023). *Zarządzanie ludźmi – z psychologicznego i metodologicznego punktu widzenia*. Wydawnictwa Uniwersytetu Warszawskiego.