Is respondents' inattention in online surveys a major issue for research?

Riccardo Vecchio[*,a], Gerarda Caso[a], Luigi Cembalo[a], Massimiliano Borrello[a]

[a] Università degli Studi di Napoli Federico II, Italy

Abstract

Participant attentiveness may represent a major concern for all researchers using online self-report survey data, as findings from non-diligent participants add noise and can significantly decrease results reliability. Therefore, attention checks have become a popular method in survey design across social sciences to capture careless or insufficient-effort of respondents, thus increasing quality of samples and the internal validity of the research. The aim of this note is to offer an overview and categorization of the different techniques adopted to flag inattentive respondents and present the potential drawbacks of not considering the issue in social sciences research.

* *Corresponding author*: Riccardo Vecchio - Associate Professor - Department of Agricultural Sciences - University of Naples Federico II - Via Università 100, 80055 - Portici (Naples) - E-mail: riccardo.vecchio@unina.it.

Riccardo Vecchio, Gerarda Caso, Luigi Cembalo, Massimiliano Borrello

Introduction

The Internet offers a great opportunity for researchers to collect large and relatively cheap samples compared with other techniques such as face-to-face interviews or laboratory samples (Buhrmester *et al.*, 2011). Online surveys rely on respondents' attentiveness to questions and treatments in order to obtain valid data. However, with the increasing use of online self-administered surveys, the quality of responses is progressively questioned due to careless respondents. Researchers define this phenomenon as *inattentive response behaviour* or *satisficing behaviour* or *careless responding* (Meade & Craig, 2012) or *insufficient effort responding* (Bowling *et al.*, 2016) or, in earlier literature, *random responding* (Beach, 1989); i.e.: the phenomenon occurs when respondents carelessly read instructions/items and provide thoughtless answers (Krosnick, 1991). Indeed, respondents' inattention is a severe threat for the internal validity of social science research, particularly when relying on stated preference data (Carlsson, 2012; Murphy *et al.*, 2005). Other issues that compromises the accuracy of the data, beyond careless responding, is when respondents deliberately give false answers, due to social desirability, linguistic incompetence, misunderstanding, difficulty of the questions or short reaction time. These latter topics are, however, beyond the scope of current note and will not be discussed hereafter.

Inattention can obscure both experimental manipulations and correlational results (Meade & Craig, 2012; Oppenheimer *et al.*, 2009). Among the many causes of inattention, scholars have highlighted some, such as: having a low level of motivation or interest in the survey, low personalization, difficulty of the questions, physical distance with the investigator, environmental distractions (Meade & Craig, 2012) and personality variables (Bowling *et al.*, 2016). A negative effect has also been confirmed by the length of the questionnaire (Gibson & Bowling, 2019; Meade & Craig, 2012). Indeed longer surveys require more efforts from respondents (Galesic, 2006), thus it is reasonable to expect that participants' attention decrease during compilation. This, in turn, leads to a higher chance to find inattentive responses.

To illustrate, different cognitive models have been proposed to account for how respondents answer to surveys. Among the most cited ones, we recall the one by Tourangeau *et al.* (2000: 7-14) that structures the process of responding a question in four consecutive steps: 1] comprehending the question, 2] retrieving relevant information, 3] integrating this information into a required judgment, and 4] selecting and reporting the appropriate answer. Performing any of these steps in a superficially manner results in non-optimal response behaviour (Krosnick, 1999). As a consequence, researchers have developed a plethora of attention checks, also called screeners, aimed to identify careless respondents and allow scholars to reduce noise prior

to conducting analyses (Maniaci & Rogge, 2014). Indeed, psychologists, sociologists, and political scientists nowadays strongly rely on screeners to identify inattentive respondents and, eventually, remove invalid responses from datasets (Berinsky *et al.*, 2014). Whereas, other researchers rarely apply this type of technique to detect careless respondents even if most of the studies undertaken in these fields of inquiry use self-administered online survey data to investigate individual-level attitudes and behaviours (Malone & Lusk, 2018).

Starting from these considerations, the overall objective of this note is to provide a complete and systematic overview of main techniques adopted in this field, as well as highlighting their potential in effectively flagging careless respondents. The contribution of current note: it summarizes empirical studies that have addressed the issue of respondents' inattention in online surveys; it categorizes the main techniques used in the literature to comprehend their potential to detect respondents' inattention.

1. Empirical studies overview

In order to tackle the objective of current contribution, a review of studies that have performed original Internet-based data collection procedures to detect and measure the number of careless respondents was performed. To make such review complete and replicable, the standard requirements of systematic literature reviews as specified by PRISMA was performed (Liberati *et al.*, 2009). The sources were identified by searching peer-reviewed published literature in the databases Scopus® and Web of Science®. For both databases a structured query was developed by using twenty-two search terms in Boolean combination with 'survey'. Keywords were first selected by a random reading of some papers and then adding all the possible keywords as literature was covered. In the end, in Scopus® the search was conducted within the fields 'article title, abstract and keywords'. Results were limited to documents written in English and included in the fields 'article', 'review', 'short survey' and 'undefined'. Documents whose subject area was focused on Biochemistry, Genetics and Molecular Biology, Engineering or Mathematics were excluded. In Web of Science®, the search was conducted within the field 'topic', limiting results to those written in English, included in the field 'article' and not covering the research areas of Mathematical Methods in Social Sciences and Mathematics. The refined search produced 108 documents for the first database and 61 for the second one. After this procedure, 10 additional records relevant to the topic were selected by checking the bibliography of authors having at least two documents included in the stock of those identified previously. The identification process ended with removal of duplicate studies.

Documents resulting from identification were screened by two independent reviewers according to five eligibility criteria (one inclusion criterion and four exclusion criteria) – besides the exclusion of articles whose topic was completely off-track respect to the purpose of the study. Only articles useful to provide data to respond to the research question were included in the review (inclusion criterion). Results excluded were those that: 1) were based on non-online surveys (since the current study is focused on online surveys); 2) were exclusively devoted to a technical description of one of the methodologies at hand (since not useful to respond to the research question); 3) were aimed to identify inattentive responses connected to mental health disorders (since inattentive responses due to mental disorders would bias the validity of results); and 4) were mainly focused on personality traits driving careless behaviour (since the focus of these studies is unbalanced toward technical aspects related to psychological analysis). The screening process was carried out in two consecutive steps. First, hand searching, reading title, source and abstract of each document were used to check the meeting with eligibility criteria. Then, the entire articles were read when more information was required to determine correspondence with eligibility criteria. We finished the document selection process on the 24th of July 2019. A total of 179 documents were identified through databases searching yielding to 54 articles included for deep reading[1]. Nearly 63% of all the reviewed papers have been published in the last three years (2017-2019) and the overall time interval of publication is between 2012 and 2019. The studies were conducted in 20 countries, notably 31 out of 54 (57.4%) were performed in the USA (including those performed in more than one country), followed by Germany (5 studies) and Spain (4 studies). The average sample size of the analysis was slightly above 1.199 respondents (median = 690), with a minimum number of 73 observations and a maximum of 13.340.

1.1. *Categorization of techniques*

From the literature consulted, many are the methods researchers can implement to identify respondents who have failed to provide thoughtful responses (Table 1). One simple, direct technique to detect inattention is to include *self-report questions*, requiring respondents to rate their degree of care in taking the survey task. These indices generally appear in the form of a direct question on a self-reported measure of effort (as "I did not pay much attention to this questionnaire") at the end of a survey.

---

1. Besides articles cited in the text, all the documents resulting from the database search are reported in the reference section.

Table 1 - Implementation examples of different techniques to detect respondents' inattention

Technique	Definition (Source)	Implementation example (Source)
Ad-hoc questions		
Self-report questions	Items asking participants how much effort they applied in responding or how they judge the quality of their data (Curran, 2016)	"I never really put much thought into my answers on the evaluations but just bubble in answers to get done quickly" (Bassett *et al.*, 2017)
Attention checks	Items placed in scale with explicit correct response (Curran, 2016)	"Choose strongly agree if you are reading this" (Kostyk *et al.*, 2019)
Inconsistency scales	Pairs of items with highly redundant content, to which participants should provide similar responses (Meade & Craig, 2012)	"I am an active person" paired with "I have an active lifestyle" (Maniaci & Rogge, 2014)
Bogus items	Odd items placed in scale to solicit particular responses (Curran, 2016)	"I was born on February 30" (Huang *et al.*, 2015)
Trap questions	Participant is instructed to ignore the response format and select a specific answer (Berinsky *et al.*, 2014)	"Please ignore the question below about how you are feeling and instead check only the 'none of the above' option as your answer"; "Please click on the word that describes how you are currently feeling" (Malone & Lusk, 2019)
Post-hoc analysis		
Response time	Time of survey completion (Curran, 2016)	e.g. faster than 2 s/item (Huang *et al.*, 2012)
Longstring analysis	Measuring the tendency to choose identical answers in blocks of items (Meade & Craig, 2012)	e.g participants who indicate consecutive strings of at least seven "strongly disagrees," seven "disagrees," twelve "neither agree nor disagrees," ten "agrees," or eight "strongly agrees" should be flagged (Huang *et al.*, 2012)

Table 1 - Continued

Technique	Definition (Source)	Implementation example (Source)
Even-odd consistency	Within-person correlation of odd numbered scale scores with even numbered scale scores (Curran, 2016)	e.g. we numbered the items of scales and split them by even and odd numbers to create the subscales. We then dichotomised the Even-Odd Consistency value by calculating the mean (M = .78) and standard deviation (SD = .17), and setting the cut score two standard deviations below the mean (.45), such that participants with values below .45 receiveda score of 1 which indicated they were flagged as careless (Francavilla *et al.*, 2019, p. 233)
Semantic/ Psychometric antonyms/ synonyms	Within-person correlations on sets of semantically/correlation-based matched pairs of items with opposite or similar meaning (Curran, 2016)	e.g. are computed as the within person correlation across pairs of items […] We sought to ensure item pairs with a negative correlation stronger than –.60 […] strong positive correlations exceeding +.60 threshold (Meade & Craig, 2012, p. 442)
Inter-item standard deviation	Degree of drift in responses from mean of individual response pattern (Curran, 2016)	e.g. this index calculates intrapersonal variation in responses using the calculation for the unbiased estimate of the standard deviation. The mean for the standard deviation calculation is that individual's scale mean, rather than the group's mean. An Inter-Item standard deviation closer to zero represents highly consistent responding (Francavilla *et al.*, 2019, p. 234)
Individual response variability	The standard deviation of a participant's responses to all items on a questionnaire (Dunn *et al.*, 2018)	e.g. we calculated this index over items from various scales and without recoding reverse scored items. Lower values of the index are indicative of a higher level of insufficient effort responding (Dunn *et al.*, 2018, p. 112)

Table 1 - Continued

Technique	Definition (Source)	Implementation example (Source)
Mahalanobis distance	Distance of response pattern from multidimensional center of all responses (Curran, 2016)	e.g. participants were flagged if their D^2 value placed them in the highest 5% of the chi-square distribution (DeSimone & Harms, 2018, p. 566)

Source: authors' elaboration.

While highly transparent for respondents, limitations of this technique are clearly related to its vulnerability to dishonesty (Desimone *et al.*, 2015) and to the fact that careless participants may likely respond inattentively also to self-reported attention questions. Whereas, in *bogus items*, also called *infrequency items*, questions are constructed to appear face-valid on a quick visual inspection, but obvious or absurd on deeper inspection (Meade & Craig, 2012). To illustrate, in bogus items respondents are required to agree or disagree with statements to which everyone should answer with the same response, flagging anyone who endorses aberrant options (for example: "I am paid biweekly by leprechauns" or "While watching TV, I had a fatal heart attack"). A known downside of infrequency scales is that counterfactual or improbable items that have clear correct responses can annoy and/or confuse participants (McKibben & Silvia, 2015), eventually leading to survey drop-out. Other scholars have also pointed-out that humorous or ambiguous content could influence item endorsement by attentive respondents (Meade & Craig, 2012). Another technique is called *instructional manipulation checks* (Oppenheimer *et al.*, 2009), also called *attention checks* or *screener questions*, where a key phrase is added to a longer set of instructions, generally instructing the respondent to ignore the question options and respond in some other manner. Individuals who pass the check can be defined as attentive respondents. Alternatively, respondents might be instructed to pick two or more unlikely choices from a list of answers, or to choose the 'other' option (eventually inserting an experimenter-determined response). This method is also defined as *instructed-response items*, or *trap questions*, or *validation questions* – e.g. "If you are paying attention select strongly disagree for this item" (Malone & Lusk, 2018). An important limitation of these checks is that they only detect whether respondents are attentive when reading instructions, not while answering survey questions (Edwards, 2019). A further technique to detect attentiveness is the adoption of *inconsistency scales*, in which respondents are faced

with paired items that are almost identical in meaning, located in different parts of the survey (generally at the beginning and then at the end). Thus, all pairs of items should be endorsed similarly; as "I am an active person" and "I have an active lifestyle". However, a core limitation of inconsistency measures is the use of Likert-type scales, and therefore participants may respond consistently around the midpoint of the scale across the pairs of items resulting in an unreliable detection of inattentiveness (Maniaci & Rogge, 2014). These different families of listed techniques use the inclusion of specific elements in the questionnaire before data collection, to verify the attention of respondents, and use them to make decisions on data quality.

A different family of methods to detect inattentiveness also exists in the literature, which uses archival or statistical screening. These techniques are mainly based on post-hoc analyses of the data of substantive interest and do not introduce dedicated additional elements to the questionnaire (and thus are sometimes defined as non-invasive). The most frequently used technique of this family is the *response time*, also called *page timing* (i.e.: the time taken by an individual to respond to a set of elements (Curran, 2016). Response time can be calculated on an entire questionnaire or page-by-page basis, and in this case being more useful when attempting to identify sporadic or local random responding (DeSimone & Harms, 2018). The response time will obviously be different for the different surveys depending on the number of elements and difficulty in understanding the question, but it is an immediate technique to detect answers that are too fast and potentially random. Another intuitive technique for identifying random answers is the *longstring analysis* which detects those who do little or no effort to change their response during a survey, (almost) always using the same response option. Similar to the response time, long-string analysis can vary depending on the type of scale and length of options, so that the technique can be difficult to compare among different data collections without first making an appropriate scaling. A set of post-hoc techniques are used to measure internal consistency such as the *even-odd consistency* ("within-person correlation of odd numbered scale scores with even numbered scale scores"), *semantic* or *psychometric antonyms/synonyms* ("within-person correlations on sets of semantically matched pairs of items with opposite or similar meaning"), *inter-item standard deviation* ("degree of drift in responses from mean of individual response pattern") (Curran, 2016: 17) and – as an extension of the long string index – the *individual response variability* ("the standard deviation of a participant's responses to all items on a questionnaire")*,* a technique flexible and easy to calculate (Dunn *et al.*, 2018). The underlying assumption of these methods of interpersonal coherence is that an attentive responder provides a pattern of responses that is internally coherent and therefore any random data should be easily identifiable as noise against a valid response (Curran,

2016). A direct consequence of the inattentive response may be the presence of outliers. Among post-hoc indirect techniques, the *distance of Mahalanobis* is used in several studies (Maniaci & Rogge, 2014; Meade & Craig, 2012). It measures the distance of the response model from multidimensional centre of all answers identifying, in this way, random answers.

Since literature has not established a commonly accepted categorization of techniques, in the current review we propose the simple distinction between studies applying in their research one methodology (single technique) or more than one (multiple techniques).

Overall, there are several ways to identify low quality data in surveys. The removal of these invalid respondents has been shown to reduce the error and provide more valid results; in addition, the use of these techniques can help explain any anomalies in the results. In general terms, some scholars have also pointed-out that all types of attention check questions can "wake up" individuals as when a respondent realizes that some queries are made to detect inattentiveness, he/she will be more careful in avoiding errors in subsequent questions (Mancosu *et al.*, 2019). However, it is worth noting that choosing a particular attention check among the others, as well as identifying the best way to treat inattentive respondents, are not processes free of pitfalls. As highlighted in the next sections of this note, the power of different techniques to detect careless responses is different; therefore, techniques should be chosen according to the best trade-off between the appropriateness of the technique in the study design, the cognitive effort required to participants in the questionnaire and a minimum threshold of the expected number of detected individuals. Furthermore, management of

Figure 1 - Attempted categorization of techniques to detect respondents' inattention
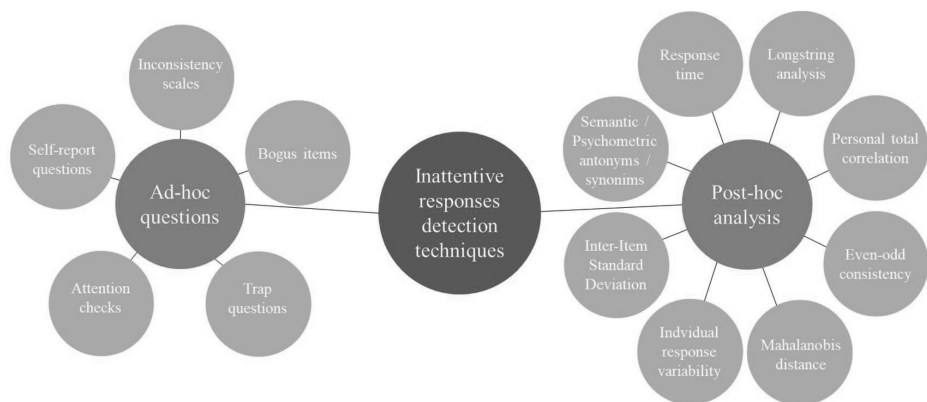
Source: authors' elaboration.

inattentive responses (*e.g.* dropping *versus* retaining flagged respondents) should take into account that a particular choice will influence the study outcomes in terms, for example, of statistical power, inference accuracy or estimation reliability.

In Figure 1, a categorization of techniques is proposed according to whether their logic to detect respondents' inattention is based on the introduction of specific elements in questionnaires (ad-hoc questions) or on the adoption of different types of analysis to perform after the survey (post-hoc analysis).

1.2. *Techniques' ability to detect inattentive respondents*

Figure 2 reports the frequency of the methodology adopted to detect careless responses. The most common method used in the reviewed articles is to adopting more than one technique at the same time, with 20 studies out of 54 (37%) classified in this category, mostly adopting 2, 3 or 5 techniques (respectively 30%, 20% and 20% of multiple checks articles). If one excludes multiple checks articles, the most used methodologies are those included in the category "Attention check/Instructional Manipulation Checks/Screener questions" (11 articles; 20.4%), followed by "Validation question/Trap question/ Instructed-Response Items" (9 articles; 16.7%) and "Post hoc analysis" (5 articles; 9.3%). Whereas, the least represented categories are "Self-report question", "Individual response variability" and "New technique"[2] (1 article; 1.9% each), while "Longstring", "Inter-Item Standard Deviation", "Mahalanobis D2", "Psychometric/Semantic antonyms/synonyms" are never used as single check strategy (0 articles). Whereas, considering also multiple checks articles for the count of individual techniques implementation, the most adopted techniques are those in the category "Response time/Page timing" (21 articles; 38.9%), followed by "Multiple checks" (20 articles; 37%), "Attention check/ Instructional Manipulation Checks/Screener questions" (16 articles; 29.6%) and "Validation question/Trap question/Instructed-Response Items" (15 articles; 27.8%). In this case, the least represented categories are "Inter-Item Standard Deviation" and "New technique" (1 article; 1.9% each). Lastly, among the multiple checks studies, the most represented category are "Response time/ Page timing" and "Longstring", with 18 (90%) and 11 (55%) articles out of 20, respectively; whereas the least are "New technique", "Other Types of Post

---

2. Innovative techniques to detect careless respondents from careful respondents such as: a specific cut-off value of response time (floodlight detection of careless respondents) (Dogan, 2018); a virtual presence and animated shape in the survey (Ward *et al.*, 2017); use of termed person temporal consistency (D2 ptc) in empirical data (Kerry, 2018) or implementation of indicators/scales to measure specific dimensions (Wood *et al.*, 2017).

Figure 2 - Frequency of techniques identified in the reviewed articles
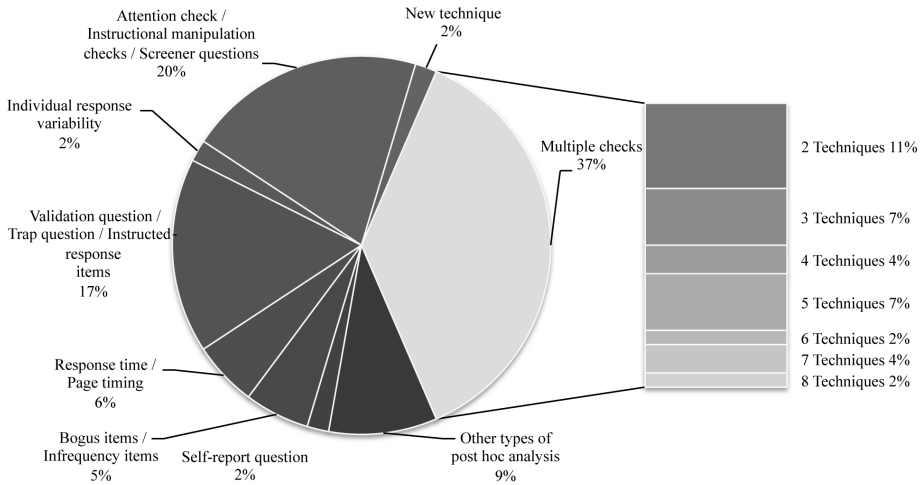
Source: authors' elaboration.

Figure 3 - Percentage of inattentive responses detected by technique
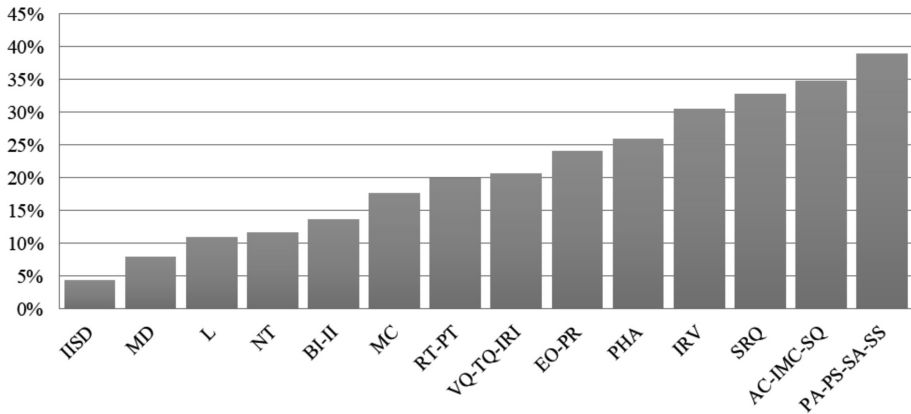
* IISD = Inter-item standard deviation; MD = Mahalanobis D2; L = Longstring; NT = New technique; BI-I I= Bogus items/Infrequency items; MC = Multiple checks; RT-PT = Response time/Page timing; VQ-TQ-IRI = Validation questions/Trap questions/Instructed-response items; EO-PR = Even-odd/Personal reliability; PHA = Other types of post hoc analysis; IRV = Individual response variability; SRQ = Self-report questions; AC-IMC-SQ = Attention checks/Instructional manipulation checks/Screener questions; PA-PS-SA-SS = Psychometric/Semantic antonyms/synonyms.

Source: authors' elaboration.

hoc analysis" and "Inter-Item Standard Deviation", with 0, 1 (5%) and 1 (5%) articles out of 20, respectively.

To estimate the share of careless respondents detected, each study was scanned calculating the percentage of inattentive respondents with respect to total sampled individuals. For each technique, an average value was calculated (Fig. 3). For articles including more than one result for each technique (e.g. those composed of more than one study), the average percentage value was computed. In those articles adopting a multiple checks strategy, we accounted the overall pervasiveness value as well as the percentage associated to each technique used. However, some articles reported either the former or the latter information. Only two articles (Borger, 2016; Huang *et al.*, 2015), initially included in the review, passing all eligibility criteria, lacked percentages to incorporate in the calculation; these documents were not excluded since still suitable to answer the research question.

Adopting this method, techniques having the highest percentages of flagged respondents are those belonging to the category "Psychometric/ Semantic antonyms/synonyms" (38.9% of inattentive responses identification), followed by "Attention check/Instructional Manipulation Checks/Screener questions" (34.9%) and "Self-report question" (32.8%). The lowest shares are found for the Inter-Item Standard Deviation methodology, for which a value of 4.35% was calculated. Those researches, in which a multiple check strategy was adopted, were able to detect, on average, 17.7% of inattentive responses, a higher share compared to the categories "Inter-Item Standard Deviation", "Mahalanobis D2", "Longstring", "New Technique" and "Bogus item/Infrequency items". However, this average is lower than the median value of the overall distribution (20.4%) and the average percentage value of single check strategies (21.3%).

Following the categorization proposed in Section 2, and according to the techniques included in the studies identified in the overview, ad-hoc questions detect higher shares of careless respondents (25.5% flagged on average) than post-hoc methodologies (20.4%).

2. Discussion and Conclusions

Present note revealed that the amount of research investigating careless responding in online surveys is still quite limited in the academic literature (56 identified articles), with the majority of papers published in the last three years and 57% performed in the USA. Most studies adopt one, single technique to detect respondents' inattention (63%), while the remaining studies apply two or more flagging techniques. Among the various available methodologies, the most commonly used are the categories: Attention

check/Instructional Manipulation Checks/Screener questions and Validation question/Trap question/Instructed-Response Items. Considering all the reviewed studies, the median share of inattentive respondents detected in the online surveys was 20.4%, revealing the strong pervasiveness of the issue and thus its detrimental potential for the validity of research findings. Furthermore, the percentage of flagged responses ranges widely among the studies and the different techniques applied by scholars. Starting from as low as a 4.3% share, estimated in studies using Inter-Item Standard Deviation, up to 38.9% computed in researches applying Psychometric/ Semantic antonyms/synonyms. The great variability of outcomes stemming from diverse techniques that flag inattentive respondents powerfully suggests the need to better tune these methodologies to common standards and consolidated thresholds. A further reason of concern is that adopting different techniques may thus influence research findings, leading to reject or validate the tested hypothesis depending on the technique. Further studies should therefore aim to target and precisely represent defined sub-populations (as, for example, household primary responsible for shopping or demarcated generational cohorts). In addition, research developments should focus on the potential biases residing in the study design and in the data gathering process to assure higher robustness of overall outcomes. Results of the following three typologies of studies should be also compared by future literature: researches adopting non online data collection methods, researches adopting online data collection methods without attention checks and researches adopting online data collection methods with different types of attention checks.

Ultimately, it is crucial for scholars to effectively manage data that includes inattentive responses. Literature suggests that the possible approaches to deal with carelessness issue in survey research are: *i*) simply remove respondents flagged as inattentive from dataset; *ii*) drop flagged respondents and reweight the rest of the data; *iii*) keep all data and account for attentiveness through statistical adjustments; *iv*) retain all respondents and ignore attentiveness in data analysis. Clearly all the four methods face several limitations (Alvarez *et al.*, 2019; Berinsky *et al.*, 2014). For example, dropping inattentive respondents from data leads to a loss in power due to sample size reduction. In addition, inattentiveness may be related to respondents' individual characteristics associated with the research topic and thus removing careless responses from the data may alter estimations. On the contrary retaining all responses, neglecting inattention, adds noise to data and can likely prompt inaccurate inferences. Consequently, scholars should cautiously evaluate the best solution to manage inattentiveness based on the specific characteristics of their research questions, survey design and sample (always bearing in mind that data collection is costly).

The findings of the empirical studies overview must be interpreted with some caution. Specifically, we are aware that some articles might be missing since numerous terms have been used to describe responses to questionnaire items that are made without consideration of the content (e.g. insufficient effort responding, careless responding, inconsistent responding). Future studies should try to standardise keywords and vocabulary to assist the development of knowledge on this topic.

Online surveys are easy to administer, relatively cheap, not disposed to data entry errors, and effectively integrate with statistical software programs (Francavilla *et al.*, 2019). However, the reliability and validity of the generated data is often impacted by inattentive respondents. While scholars are well aware that not all respondents pay sufficient attention when completing self-report surveys, the issue is quite overlooked in many fields of research (Gao *et al.*, 2016; Malone & Lusk, 2018; Maniaci & Rogge, 2014). Therefore, researchers using online surveys are encouraged to become familiar with techniques for detecting careless responding, providing evidence of findings validity and ruling out potential confounds. Furthermore, scholars are increasingly urged to carefully design online surveys in order to keep respondents interested, attentive, focused, and motivated. In addition, useful procedures are also available for researchers interested in treating effectively the flagged data, without restricting the study's representativeness and external validity (Abbey & Meloy, 2017). Today there are several guidelines researchers can follow to determine the most appropriate screening technique for their specific study depending on the survey design and methodology (e.g. Curran, 2016; Desimone *et al.*, 2015). The findings of current note aim to increase this literature, providing an overview and categorization of the different detection techniques and reporting the available indicators of careless responses.

Acknowledgments

References

Abbey, J.D. & Meloy, M.G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, 53-56, 63-70, doi: 10.1016/j.jom.2017.06.001.

Alvarez, R.M., Atkeson, L.R., Levin, I. & Li, Y. (2019). Paying attention to inattentive survey respondents. *Political Analysis*, *27*(2), 145-162, doi: 10.1017/pan.2018.57.

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, doi: 10.1080/02602938.2015.1119801.

Beach, D.A. (1989). Identifying the random responder. *Journal of Psychology: Interdisciplinary and Applied, 123*(1), 101-103, doi: 10.1080/00223980.1989.10542966.

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*(3), 739-753, doi: 10.1111/ajps.12081.

Börger, T. (2016). Are Fast Responses More Random? Testing the Effect of Response Time on Scale in an Online Choice Experiment. *Environmental and Resource Economics, 65*(2), 389-413, doi: 10.1007/s10640-015-9905-1.

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, doi: 10.1037/pspp0000085.

Buhrmester, M., Kwang, T. & Gosling, S.D. (2011). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5, doi: 10.1177/1745691610393980.

Carlsson, F. (2012). Non-Market Valuation: Stated Preference Methods. *In The Oxford Handbook of the Economics of Food Consumption and Policy*.

Curran, P.G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19, doi: 10.1016/j.jesp.2015.07.006.

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, doi: 10.1007/s10869-017-9514-9.

Desimone, J.A., Harms, P.D. & Desimone, A.J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior, 36*(2), 171-181, doi: 10.1002/job.1962.

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, doi: 10.1007/s10869-016-9479-0.

Edwards, J.R. (2019). Response invalidity in empirical research: Causes, detection, and remedies. *Journal of Operations Management*, doi: 10.1016/j.jom.2018.12.002.

Francavilla, N.M., Meade, A.W. & Young, A.L. (2019). Social Interaction and Internet-Based Surveys: Examining the Effects of Virtual and In-Person Proctors

on Careless Response. *Applied Psychology-an International Review-Psychologie Appliquee-Revue Internationale, 68*(2), 223-249, doi: 10.1111/apps.12159.

Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics*, *22*(2), 313.

Gao, Z.F., House, L. & Bi, X. (2016). Impact of satisficing behavior in online surveys on consumer preference and welfare estimates. *Food Policy, 64*, 26-36, doi: 10.1016/j.foodpol.2016.09.001.

Gibson, A.M. & Bowling, N.A. (2019). The effects of questionnaire length and behavioral consequences on careless responding. *European Journal of Psychological Assessment*, doi: 10.1027/1015-5759/a000526

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

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

Kostyk, A., Zhou, W.K. & Hyman, M.R. (2019). Using surveytainment to counter declining survey data quality. *Journal of Business Research, 95*, 211-219, doi: 10.1016/j.jbusres.2018.10.024.

Krosnick, J.A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236, doi: 10.1002/acp.2350050305.

Krosnick, J.A. (1999) Survey research. *Annual Review of Psychology*, 50, 537-567.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: *Explanation and elaboration. Italian Journal of Public Health, 6*(4), 354-391. -- Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-76549089252&partnerID=40&md5=dc8847f1671dc34e0b515045479a8b14.

Malone, T. & Lusk, J.L. (2018). Consequences of Participant Inattention with an Application to Carbon Taxes for Meat Products. *Ecological Economics, 145*, 218-230, doi: 10.1016/j.ecolecon.2017.09.010.

Malone, T. & Lusk, J.L. (2019). Releasing the trap: a method to reduce inattention bias in survey data with application to us beer taxes. *Economic Inquiry, 57*(1), 584-599, doi: 10.1111/ecin.12706.

Mancosu, M., Ladini, R. & Vezzoni, C. (2019). 'Short is Better'. Evaluating the Attentiveness of Online Respondents Through Screener Questions in a Real Survey Environment. *BMS Bulletin of Sociological Methodology/ Bulletin de Methodologie Sociologique, 141*(1), 30-45, doi: 10.1177/0759106318812788.

Maniaci, M.R. & Rogge, R.D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*(1), 61-83, doi: 10.1016/j.jrp.2013.09.008.

McKibben, W.B. & Silvia, P.J. (2015). Inattentive and Socially Desirable Responding: Addressing Subtle Threats to Validity in Quantitative Counseling Research. *Counseling Outcome Research and Evaluation, 7*(1), 53-64, doi: 10.1177/2150137815613135.

Meade, A.W. & Craig, S.B. (2012). Identifying Careless Responses in Survey Data. *Psychological Methods, 17*(3), 437-455, doi: 10.1037/a0028085.

Murphy, J.J., Allen, P.G., Stevens, T.H. & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics, 30*(3), 313-325, doi: 10.1007/s10640-004-3332-z.

Oppenheimer, D.M., Meyvis, T. & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867-872, doi: 10.1016/j.jesp.2009.03.009.

Tourangeau, R., Rips, L.J. & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Riccardo Vecchio, Gerarda Caso, Luigi Cembalo, Massimiliano Borrello

Riccardo Vecchio
Department of Agricultural Sciences, University of Naples Federico II, Via
Università 96, 80055 Portici, Naples, Italy
Tel: +39 0812539053, E-mail: riccardo.vecchio@unina.it
Riccardo Vecchio is associate professor of Agricultural and Resource Economics at
the Department of Agricultural Sciences, University of Naples Federico II (Italy).
His scientific interests focus on consumer behaviour towards food products and
wine. He is author of 64 scientific papers in ISI/Scopus journals.

Gerarda Caso
Department of Agricultural Sciences, University of Naples Federico II, Via
Università 96, 80055 Portici, Naples, Italy
Tel: +39 0812539053, E-mail: gerarda.caso@unina.it
PhD Candidate, Gerarda Caso helds a Master's degree in Food science and
technology conceived at the University of Naples Federico II, Italy. She is a first-
year PhD student in Food Science at the Department of Agricultural Sciences of
the University of Naples Federico II (Italy). Her doctoral research focuses on the
application of behavioural and experimental economics in the food domain.

Massimiliano Borrello
Department of Agricultural Sciences, AgEcon and Policy Group, University of
Naples Federico II, Via Università 96, 80055 Portici, Naples, Italy
Tel: +39 0812539083, E-mail: massimiliano.borrello@unina.it
Post-doc researcher, Massimiliano Borrello helds a post-doc fellowship at the
Department of Agricultural Sciences, University of Naples Federico II (Italy). Borrello
has a PhD degree in Agricultural Economics and Policy held at the University of Naples
Federico II, Italy. His scientific interests focus on radical innovation implementation
and management in the agro-food sector and sustainability with particular emphasis on
circular economy. He is author of 12 scientific papers reported in ISI/Scopus journals.

Luigi Cembalo
Department of Agricultural Sciences, AgEcon and Policy Group, University of
Naples Federico II, Via Università 96, 80055 Portici, Naples, Italy
Tel: +39 0812539065, E-mail: cembalo@unina.it
Luigi Cembalo, PhD is full professor of Agricultural and Resource Economics at
the Agricultural Economics and Policy Group of the Department of Agricultural
Sciences, University of Naples Federico II (Italy). Cembalo has two MSc degree
in Agricultural Economics and Policy (held at the Centre of Advanced Training,
Naples, Italy) and Agricultural and Resource Economics (held at the AREC
Department of the University of Arizona, USA) and a PhD in Agricultural Economics
and Policy (held at the University of Naples Federico II, Italy). His scientific interests
focus on innovation and sustainability in high value supply chain, consumer studies,
demand estimations, and agro-energy economics. His expertise also includes
microeconometrics and operational research modeling procedures with a focus on
multicrieria approaches. Cembalo's works comprise about 100 publications of which
58 in ISI/Scopus journals.