Association for Information Systems

AIS Electronic Library (AISeL)

SIGHCI 2022 Proceedings

Special Interest Group on Human-Computer Interaction

12-12-2022

Too Fast? Too Slow? A Novel Approach for Identifying Extreme Response Behavior in Online Surveys

Manasvi Kumar
The University of Arizona, manasvik@email.arizona.edu

Joseph Valacich
The University of Arizona, valacich@arizona.edu

Jeffrey Jenkins
Brigham Young University, jjenkins@byu.edu

David Kim
The University of Arizona, davidkim7@arizona.edu

Follow this and additional works at: https://aisel.aisnet.org/sighci2022

Recommended Citation

Kumar, Manasvi; Valacich, Joseph; Jenkins, Jeffrey; and Kim, David, "Too Fast? Too Slow? A Novel Approach for Identifying Extreme Response Behavior in Online Surveys" (2022). *SIGHCI 2022 Proceedings*. 4.
https://aisel.aisnet.org/sighci2022/4

This material is brought to you by the Special Interest Group on Human-Computer Interaction at AIS Electronic Library (AISeL). It has been accepted for inclusion in SIGHCI 2022 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Too Fast? Too Slow? A Novel Approach for Identifying Extreme Response Behavior in Online Surveys

Manasvi Kumar
University of Arizona
manasvik@email.arizona.edu

Joseph S. Valacich
University of Arizona
valacich@arizona.edu

Jeffrey L. Jenkins
Brigham Young University
jjenkins@byu.edu

David Kim
University of Arizona
davidkim7@arizona.edu

ABSTRACT

Some participants of online surveys engage in extreme answering behavior while generating responses (i.e., they respond too fast or too slow) relative to population norms. Here, we demonstrate how participants' navigation behaviors can be used to potentially identify such responses. We administered an online survey where students (who were earlier instructed to complete a task) report lenience scores towards non-appropriate behavior while completing the task. We draw on cognitive dissonance theory to posit that failure to follow instruction predicts lenience scores. We then created different datasets by excluding data from participants flagged by our metrics and generated predictive models. We found that model performance improves by removing data from flagged participants, indicating a reduction in noise from the dataset. Despite demonstrating the effectiveness of our approach, we encourage researchers to exercise caution and elaborate on the limitations of our approach and future avenues of research.

Keywords

Data Quality, Survey Research, Self-Report Data, Online Survey

INTRODUCTION

Surveys are among the most widely used methods adopted by researchers and practitioners to collect human response data (Wright, 2005). While the appropriateness of the data collection method used in administering surveys largely depends on the context (Wyatt, 2000), online web surveys are very popular in most scenarios. Surveys have also featured prominently in Information Systems (IS) literature as a common method for collecting human response data for decades. The search term "survey" on the Association for Information Systems (AIS) eLibrary returns 29,389 results (May 2, 2022). Clearly, surveys are an important method used by researchers to collect data with online surveys being the prominent method today.

Despite widespread use, researchers who collect self-report data using surveys face several challenges. A key challenge with using online survey data is problematic data provided by certain survey respondents. How participants generate survey answers is a non-trivial process and several theories have been proposed to understand it (Tourangeau, Rips and Rasinski, 2000). While earlier theories proposed by Simon (1976) perceived it to be a scientific process that was logical and systematic, it was later established that participants may consider multiple routes to select an answer based on factors such as cognitive and motivational difficulties or other cues (Cannel et al., 1981). The availability of multiple routes in arriving at the answer directly manifests into the possibility of various kinds of participant responses, not all of which are useful.

While it may be difficult to capture the motivation, cognitive processes, intent, or other factors behind selecting a response, researchers who utilize online surveys have tried to unobtrusively obtain data while the response is being generated i.e., response times. Most researchers agree that it takes a minimum amount of time to complete a survey, and participants who take less time than that are likely to generate poor quality responses (Matjašic, Vehovar and Manfreda, 2018). The assumption here is that generating a survey response requires a certain amount of effort. However, while the effects of excluding "too fast" responses have been well studied in the literature, the same cannot be said for "too slow" responses, which may be equally problematic. In this study, we argue that both "too fast" and "too slow" responses generated by participants can be problematic and that including either response in our analysis is not ideal. We propose utilizing participants' navigation behavior to devise interpretable metrics that may help identify fast and slow responding participants. Finally, we posit that removing data from these participants may help reduce noise in the survey data. To demonstrate the efficacy of this approach, we report results from a survey we administered which indicate that participants who fail to follow instructions, are more likely to report lenient views on what constitutes appropriate behavior. Further, we demonstrate that utilizing the newly developed navigation behavior-based metrics help researchers select an appropriate threshold for identifying fast/slow participants. Finally, we created prediction models on datasets obtained by removing data from these participants and share these results. The results obtained indicate that removing data from the participants identified as too fast, too slow, or a combination of both helps reduce noise.

BACKGROUND AND THEORY

In this section, we describe the theory behind the central predictive model used in the study. We then illustrate how the metrics developed in this study could help identify participants who provide extreme answering behaviors (i.e., too fast or too slow).

Cognitive Dissonance Theory

Cognitive Dissonance Theory has been widely used in social psychology for over 70 years (Festinger, 1957). It discusses the state of cognitive dissonance which occurs within every individual when he/she perceives that a pair of cognitions are inconsistent. This dissonance serves as an unpleasant drive that needs to be reduced by changing cognition. We apply this theory in the context of an individual assigned to complete a task by following a set of instructions, specifically when they are later asked to answer a few questions about what constitutes appropriate behavior while completing the task. An individual who performs the task diligently will have no difficulty in rating the appropriateness of behaviors accordingly. In contrast, an individual who failed to follow instructions while completing the task will not generate straightforward responses. If such an individual is asked questions about certain negative behaviors, they are likely to experience cognitive dissonance which is reduced by modifying his/her belief about the appropriateness of negative behaviors to remain consistent, becoming more lenient in the process. To summarize, we expect individuals who didn't follow instructions to have more lenient answers to questions about what constitutes appropriate behavior, especially when asked about negative behaviors. This suggests our first hypothesis.

H1: Failure to follow instructions will result in more lenient answers to questions about what constitutes appropriate behavior.

Answering Behavior

Answering a survey question requires the participant to expend a certain amount of effort to respond accurately. However, not all participants put in the effort required to do so. These participants have been referred to as Careless/ Insufficient Effort (C/IE) participants (Curran, 2016), and there exist several techniques designed to identify and remove these participants from a dataset (Meade and Craig, 2012). The rationale for excluding these data is clear: including data from C/IE participants in any analysis acts as an "insidious confound in survey data" and may affect the relationships being examined (Huang, Liu and Bowling, 2015).

Intuitively, one may consider the total time taken by a participant to complete the survey for identifying whether a participant put in enough effort while completing a survey. This metric, known as Response Time for a survey, is one of the most used approaches to identify C/IE participants. Most online survey administration platforms like Qualtrics provide researchers with survey completion times making it easier for researchers to identify C/IE participants by setting an "educated" threshold for the minimum time needed to complete a survey. Surveys may, however, consist of several types of questions (demographic questions, construct-related questions, feedback questions, etc.), and it may be more useful to focus on important portions of the survey and collect more nuanced metrics while a participant answers these questions.

While utilizing item level response times could be more useful in finding C/IE participants as opposed to survey level response times, there are several challenges in determining the exact C/IE participants even after item level response time is known. One approach for identifying C/IE participants could be to use an anecdotally observed minimum response time which is 2 seconds per item (Huang, Curran, Keeney, Poposki, and DeShon, 2012), though this may be tweaked based on the question. Another approach could be to utilize a traditional outlier analysis, but these may favor certain responses (closer vs. away from the next button) depending on the survey layout. In this study, we propose the use of the metric "Speed" which is defined as the total distance travelled by the mouse cursor divided by the total time spent generating the response. This metric is not as sensitive to layouts but is equally effective in identifying C/IE participants who generate minimum effort responses. To determine thresholds for maximum Speed, we propose creating a metric "Speed Ratio" which is the ratio of the Speed of the participant's mouse cursor while generating the response to the median Speed of all participants while generating the response. The researcher may decide the suitability of the threshold for this metric by observing the distribution of the "Speed Ratio" metric and may interpret the threshold in a straightforward manner (i.e., if the threshold is 3, then the maximum allowed Speed of the participant is thrice that of the median participant). Thus, removing the participants with high relative speeds as determined by the "Speed Ratio" metric should more likely remove the noisy data from C/IE participants.

H2: Removing data obtained from participants who generate responses by moving the mouse cursor at high relative speeds reduces noise in the dataset.

As discussed earlier, surveys suffer from not only participants who have abnormally short response times but also those whose response times are longer. Participants who have longer response times are likely to be distracted and very few studies have attempted to consider their behaviors while accounting for participant attentiveness (Read, Wolters, and Berinsky, 2021). Zwarun and Hall (2014) found that participants of online surveys engage in several activities (multitask) while completing the survey, especially younger participants. However, as several factors could impact why a participant may have long response times (e.g., linguistic incompetence, misunderstanding, misrepresentation etc.), there is scant prior research on this topic.

In this study, we attempt to identify participants who are distracted using the Speed Ratio metric. The researcher may utilize the Speed Ratio metric and select a threshold for low Speed after looking at the distribution. The interpretation here remains the same (i.e., if the threshold is 1/3, then the minimum Speed allowed of the participant is a third of the Speed of the median participant). This approach helps in capturing distracted participants as they spend a lot of time performing other tasks while keeping the cursor still, in turn reducing the Speed significantly. We posit that removing participants with low relative speeds as determined by the Speed Ratio metric should remove noisy data from distracted participants.

H3: Removing data obtained by participants who generate responses by moving the mouse cursor at low relative speeds reduces noise in the dataset.

We utilized the Speed Ratio metric to distinguish noisy data from C/IE participants (high relative speeds) and noisy data from potentially distracted participants (low relative speeds). The characteristics of these participants are different, and these correspond to different types of responses in the survey. For example, C/IE participants may be assumed to be satisficing and generate predetermined responses, while those who are distracted may generate random responses. Removing both types of data should reduce the noise in the dataset.

H4: Removing data obtained by participants who generate responses by moving the mouse cursor at either low or high relative speeds reduces noise in the dataset.

METHODOLOGY

To test our hypotheses, we conducted a study where university students, who had completed a class assignment, provided self-reported responses towards the appropriateness of hypothetical behaviors while completing the assignment. Students enrolled in an undergraduate-level operations management course at a public US-based university were recruited as participants for the study. All students enrolled in the class were part of a five-member group that submitted a 10-minute online recorded presentation about a case study. Each student was then required to individually evaluate five randomly assigned presentations created by other groups and provide necessary feedback. Feedback was collected through an online survey where students rate various facets of the presentation on a five-point scale and briefly summarize the contents of the video. Before providing feedback, students were expected to view each group's presentation video and gain an understanding of the strengths/weaknesses of various facets of the presentation. This reviewing and evaluation process is time-consuming as the time taken by the student to complete each evaluation may depend on various factors such as the students' understanding of the case and the overall quality of the presentation being evaluated. However, the lack of ample incentive to provide high-quality feedback may entice some students to provide poor quality feedback.

After completing the reviewing assignment, students were recruited for this follow-up study and incentivized to complete an online survey to receive extra course credit. In the survey, we asked the students to rate the appropriateness of hypothetical behaviors exhibited by students while providing feedback. Students participating in this follow-up study could then optionally provide consent to connect their survey responses and educational records for the class for the purpose of the study. As the presentations for all cases were hosted on an online learning platform that maintains records of each student's access behavior and viewing duration, we were able to record the file access behavior and viewing times of all videos for those students who consented to allow this analysis. Comparing student responses towards the appropriateness of various behaviors to actual student behavior (i.e., following instruction) enables us to test H1. By capturing fine grained mouse cursor data, we then compute the mouse speed of participants while answering survey questions and, consequently, their Speed Ratio. Placing upper and lower thresholds on this metric enables us to test H2 and H3, respectively. Finally, we utilize both thresholds on the Speed Ratio metric to test H4.

Survey Design

The survey consists of two types of questions: demographic questions and questions about the appropriateness of certain hypothetical behaviors. After providing answers to demographic questions, students were asked to "rate the appropriateness of the following behaviors exhibited by students while viewing the presentation before providing feedback." The scenarios presented included positive behavior related questions (i.e., complete viewing) and negative behavior related questions (i.e., Not viewing). All responses were collected on a 5-point Likert scale. All survey items used in the study are available upon request. In this study, we only consider the response to the negative behavior question "student minimally viewed the presentation video before providing feedback" as a measure of "lenience."

Participants

677 of the 931 students enrolled in the class over two separate semesters participated in the study. The same survey was used to collect data from both cohorts after they provided feedback for the assignment, which was issued through the same modality. While the case studies assigned to the two cohorts were different, they were of comparable levels of difficulty and the recorded presentations created by student groups required similar duration limits (10 minutes). Of the 677 students who participated in the study, 450 students provided consent to allow the use of their educational records (relating to the course) for this study. As retrieving view times requires access to educational records and is essential to determine whether a student

followed instructions while completing the assignment, only data from 450 students can be used for the study.

Mouse Cursor Data and Metrics

The survey collecting participants' self-reported behavior was hosted online in the Qualtrics survey system. The research team developed a custom JavaScript library that was embedded in the Qualtrics surveys to collect raw mouse-cursor movements (e.g., x-y coordinate positions, timestamps, etc.) and related behavioral data (e.g., clicks, HTML elements, etc.). The library captured this behavioral data at a millisecond precision rate and sent this data to a web service developed by the research team for processing.

This raw data was later processed through code developed by the team to obtain metrics. All scripts required to generate these metrics are available upon request. Particularly, we use the metric Speed in this study. We compute speeds for all survey responses provided by a participant and analyze them at a question level to account for question difficulty. We propose the following approach to identify extreme response behavior from participants using the Speed metric: Speeds utilized by all participants to answer a survey question may be used to compute median Speed for the question. We then divide all values of Speed by the median Speed to compute the Speed Ratio, which is a useful metric in the context of this study. Having extreme values of Speed Ratio could imply that a participant engaged in extreme answering behavior. By selecting a threshold for the value of Speed Ratio, researchers may utilize this metric to identify participants engaging in extreme answering behavior and may choose to exclude their responses from the analysis.

ANALYSIS AND RESULTS

Before analyzing our hypotheses, we define how a student may classify him/her as one who "failed to follow instruction." While any student who didn't view each video in its entirety technically failed to follow instruction, there may be several reasons why this occurs. We conservatively state that a student failed to follow the instruction if he/she provided feedback to a group without viewing their video for at least a minute. Of the 450 students whose data was used for the study, 136 were classified as those who failed to follow instructions (30.22%).

We created a linear regression equation to test our four hypotheses. We use the survey response to the negative behavior question about what constitutes appropriate behavior, i.e., "student minimally viewed the presentation video," as the dependent variable for Hypothesis 1 (Lenience). The independent variable for these equations is the binary variable for whether a student failed to follow instructions (i.e., 1 if the student failed to follow instructions and 0 if the student followed instructions). The equation is shown in Figure 1.

$$\text{Lenience}_i = \beta_0 + \beta_1 \text{ Failure}_i + \varepsilon_i$$

Figure 1. Linear regression equation

We created four different datasets to train models based on the equation in Figure 1. The models generated by training the four datasets are used to test the four hypotheses. The first dataset is the original dataset (n=450) which comprises of data from all participants. The second dataset (fast speed dataset, n=421) is the subset of the original dataset obtained by removing responses from participants with high mouse speeds (Speed Ratio greater than 3). The third dataset (slow speed dataset, n=418) is created by removing responses from participants with low mouse speeds from the original dataset (Speed Ratio less than 0.33). Finally, the fourth dataset (combined dataset, n=389) is obtained by removing data from participants with either low or high mouse speeds from the original dataset (Speed Ratio less than 0.33 or greater than 3). Table 1 provides summarized statistics of participants' speeds while generating the response. It suggests that the original dataset consists of participants who were not only moving the mouse at very high speeds but also moving at low speeds. The irregularity is vast, as the fastest participant was ~8700 times faster than the slowest one. The use of the Speed Ratio helps us regulate the irregularity, with the fastest participant being no more than 9 times faster than the slowest one.

	Original Dataset	Fast Speed Dataset	Slow Speed Dataset	Combined Dataset
Min.	0.0002	0.0002	0.0672	0.0672
1st Quartile	0.1197	0.1159	0.1425	0.1354
Median	0.2008	0.1884	0.2127	0.2023
3rd Quartile	0.2961	0.2699	0.3052	0.2804
Max.	1.7471	0.5838	1.7471	0.5838
Mean	0.2505	0.2057	0.2661	0.2188

Table 1. Summary Statistics of participant speeds for different datasets

We analyzed H1 by training the linear regression model on the original dataset. The results are summarized in Table 2. The results show that H1 is supported, indicating that failure to comply with instruction resulted in more lenient answers to questions about what constitutes appropriate behavior. The r-squared value of the model is 0.0259.

Table 3 summarizes the results of training the linear regression model on the fast speed dataset. The results are significant, indicating that failure to follow instructions is an indicator of lenient answers. In comparison to the results for H1, we observe two major differences. Firstly, the coefficient for the variable of interest (Failure) increases by nearly half a standard error when the model is trained on the fast speed dataset (0.33368-0.28981=0.04387). Secondly, the r-squared value of the model is 0.0336, which is greater than that for the model trained on the original dataset. These results suggest that removing participants with larger values of Speed Ratio reduces noise in the dataset, and hence H2 is supported.

	Estimate	Std. Error	t-value	p-value
Intercept	2.21019	0.04619	47.850	<0.001
Failure	0.28981	0.08402	3.449	<0.001

Table 2. Training the model on the Original dataset

	Estimate	Std. Error	t-value	p-value
Intercept	2.19388	0.04798	45.72	<0.001
Failure	0.33368	0.08736	3.82	<0.001

Table 3. Training the model on the Fast Speed dataset

We trained the linear regression model on the slow speed dataset. The results obtained are significant, as summarized in Table 4, and the r-squared value of the model is 0.0290. We observe similar results here as we did with training the model on the fast speed dataset (increase in r-squared value & increase in the coefficient of interest compared to the model trained on the original dataset). Thus, our results suggest that removing participants with smaller values of Speed Ratio reduces noise in the dataset. Hence, H3 is supported.

	Estimate	Std. Error	t-value	p-value
Intercept	2.19377	0.04769	46.003	<0.001
Failure	0.30235	0.08584	3.522	<0.001

Table 4. Training the model on the Slow Speed dataset

Finally, Table 5 summarizes the results of training the linear regression model on the combined dataset. The results obtained are significant, and the r-squared value of the model is 0.0381. We observe the greatest increase in both the coefficient of interest and the r-squared value (relative to the model trained on the original dataset) compared to the results obtained from other models. As the results suggest a reduction of noise in the dataset by removing data points from participants with extreme values of Speed Ratio (low or high), H4 is supported.

The summarized results of all models are shown in Table 6. We observe that the greatest increase in the r-squared value is obtained by training the model on the combined dataset (47.10%). This dataset also discards the most datapoints (13.56%) from the original dataset. Utilizing the high-speed dataset alone increases the r-squared value (29.73%) by discarding a relatively low percentage of datapoints (6.44%). While the loss of 6 to 14% of data points may seem high, it is consistent with results from previous studies where a similar proportion of data points are discarded as C/IE responders when data was collected from undergraduate students (DeRight and Jorgensen, 2015), similar to this study.

	Estimate	Std. Error	t-value	p-value
Intercept	2.17472	0.04966	43.794	<0.001
Failure	0.35028	0.08941	3.918	<0.001

Table 5. Training the model on the Combined dataset

Dataset	Dataset Size	% Data Excluded	R^2	$\%R^2$ Increase
Original dataset	450	--	0.0259	--
Fast Speed Dataset	421	6.44%	0.0336	29.73%
Slow Speed Dataset	418	7.11%	0.0290	11.97%
Combin ed Dataset	389	13.56%	0.0381	47.10%

Table 6. Summarized results of models trained on different datasets

DISCUSSION

In this paper, we demonstrate how participants who complete a survey exhibit different behaviors and how removing data from participants likely exhibiting extreme answering behavior can help reduce noise in the dataset. To identify and account for these differences, we devised a metric called Speed Ratio which was created by dividing the mouse speed of the participant by the median mouse speed for that question. By removing data from participants whose Speed Ratio exceeded or was less than a particular threshold, we created datasets that likely excluded participants who engaged in extreme answering behavior. We then created several linear regression models on the different datasets, which examined the relationship between failure to follow instructions and lenience towards negative behavior. We found that removing data from participants who engaged in extreme answering behaviors helped improve the r-squared value of the models and increased the coefficient of the variable of interest, indicating a reduction in noise from the original dataset.

It is necessary to restate the key intent of this study: to remove data from participants who are likely to be engaging in C/IE responding or are distracted. This may therefore be viewed as an exercise for reducing Type I and Type II errors while identifying problematic participants. The extent to which we reduce Type II errors while accounting for Type I errors in this study largely depends on the threshold we set for the value for the Speed Ratio. While removing data from participants who engaged in C/IE responding or were distracted is important, it may not be as important as ensuring that data from those participants who failed to follow instruction during the class assignment is not lost disproportionally. This is because responses collected from participants who didn't follow instruction help in determining the fundamental relation between the constructs being analyzed. Unfortunately, accounting for this while deciding on the threshold for the Speed Ratio may be seen as p-hacking (Head, Holman, Lanfear, Kahn and Jennions, 2015). While deciding on the threshold for the Speed Ratio, we only considered the Speed distribution of the original dataset (Table 1). It was evident that the distribution had a long tail (a few extremely fast-moving participants) and the Speed around the 1st quartile was half the Speed of the median participant. As we didn't want to risk losing a quarter of our participants, we didn't choose a threshold of 2 and instead chose the threshold to be 3. Therefore, we defined participants who moved slower than 3 times the median Speed as too slow, while those who were faster than 3 times the median Speed were considered too fast. As a robustness check, we examine the scenarios where different thresholds was chosen for the upper threshold for fast speed and different thresholds were chosen for the lower threshold of slow speed. Table 7 summarizes the results of varying the upper threshold for Speed Ratio in the fast speed dataset, while Table 8 summarizes the results for varying the lower threshold for Speed Ratio in the slow speed dataset.

Speed Threshold	Dataset Size	% Data Excluded	R^2	$\%R^2$ Increase
None	450	--	0.0259	--
Twice	393	12.67%	0.0445	71.81%
Thrice	421	6.44%	0.0336	29.73%
Four times	436	3.11%	0.0300	15.83%
Five times	442	1.78%	0.0291	12.36%

Table 7. Effect of varying the upper threshold for Speed Ratio – Fast Speed Dataset

Table 7 reveals that as thresholds for the tolerable limit for the maximum speed increase, fewer data points are removed, and the improvement in the r-squared value of the model. This is very much in line with what we expect; removing data from participants who had high mouse speeds reduces noise in the data and excluding more of them improves model performance. The results are different in Table 8, where any threshold for the lower speed limit other than the third result in a decrease in model performance. While it is impossible to know whether this is due to an increase in Type I error or not (as we cannot know for certain whether a participant was distracted), we can examine the characteristics of the removed data, especially the percentage of participants who failed to follow instruction. Note that the base rate of participants who failed to follow the instruction in the original dataset is 30.22% (136 of 450). We found that the percentage of participants who failed to follow instructions among excluded participants is least when the threshold is a third (21.88%). The highest percentage is 33.33%, slightly greater than the baseline. We acknowledge that in cases where the behavior being studied (failure to follow instruction) may be related to the undesirable characteristics of extreme answering behavior in surveys, researchers must exercise caution in selecting thresholds and report disagreeing analyses (if found) as recommended in previous literature (Curran, 2016).

Speed Threshold	Dataset Size	% Data Excluded	R^2	$\%R^2$ Increase
None	450	--	0.0259	--
Twice	367	18.44%	0.0231	-10.81%
Thrice	418	7.11%	0.0290	11.97%
Four times	436	3.11%	0.0223	-13.90%
Five times	438	2.67%	0.0234	-9.65%

Table 8. Effect of varying the lower threshold for Speed Ratio – Slow Speed Dataset

LIMITATIONS

In this study, we collect data from undergraduate students who provided consent to connect their survey responses to their educational records. As we only have access to student data for those students who provide consent, we may have sampling issues as we do not know anything about the characteristics of students who failed to follow instruction but didn't provide consent. We also acknowledge that an arbitrary choice of setting a lower threshold for the Speed Ratio resulted in H3 being significant. Finally, the loss of power in results from removing data is not fully accounted for in the current analysis. We acknowledge that to utilize our suggested technique, the data collection process could be longer to achieve optimal results with adequate power.

FUTURE RESEARCH

In this paper we attempt to identify participants who likely engaged in extreme answering behavior while generating responses. We contribute to the growing literature that examines the issue of poor quality data in surveys. Several techniques are reported that attempt to address the problem with C/IE responding in online surveys (Curran, 2016). These include the use of time-based measures, long-string analysis (Johnson, 2005) and odd-even consistency (Meade and Craig, 2012). Other approaches utilize cognitive approaches (Reading rate in Healey, 2007) and combined approaches (Greszki, Meyer and Schoen, 2015). Approaches that examine the other end of the spectrum (long durations) are relatively few. Studies that do so have attributed it to distraction and suggest innovative techniques to measure attention (Read et al., 2021). In this study, we introduce the speed metric and the idea of a Speed Ratio threshold to identify extreme answering behavior among participants. There are, however, other kinds of metrics that haven't been analyzed yet. More importantly, it has been widely accepted that a single metric is never adequate to completely identify problematic participants but rather a sequential use of several techniques has been suggested (Curran, 2016).

These combinations of metrics, their sequences, and their dependence on the types of constructs studied raise several research questions that require further study.

CONCLUSION

In this study, we address the issue of extreme answering behavior in online surveys by identifying and excluding participants who appear to be "too fast" (C/IE responding) or are "too slow" (distracted). We utilize participants' navigation behavior to devise metrics that help researchers identify and interpret thresholds for fast and slow responding behaviors. Researchers may choose to remove data likely generated from such extreme responding participants. To test the efficacy of this approach, we conducted a survey where students who were recently instructed to complete a repetitive assignment were recruited. In the survey, they self-reported lenience scores towards non-appropriate behavior while completing the assignment – scores which are known to be greater among participants who engaged in non-appropriate behavior. We generated different datasets by removing data from participants that appeared to engage in extreme behavior as determined by our newly devised metrics and ran predictive models that estimate the lenience scores based on whether participants engaged in non-appropriate behavior while completing the assignment. We found that model performance improves by removing data from flagged participants, indicating a reduction in noise from the dataset. We encourage researchers to exercise caution and elaborate on the limitations of the approach.

REFERENCES

1. Curran, P. G. (2016) Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4-19.
2. DeRight, J. and Jorgensen, R. S. (2015) I just want my research credit: Frequency of suboptimal effort in a non-clinical healthy undergraduate sample. *The Clinical Neuropsychologist*, 29(1), 101-117.
3. Festinger, L. (1957) A theory of cognitive dissonance (Vol. 2). Stanford University Press.
4. Greszki, R., Meyer, M. and Schoen, H. (2015) "Exploring the effects of removing "too fast" responses and respondents from Web surveys", *Public Opinion Quarterly*,79, 471–503.
5. Head, M. L., Holman, L., Lanfear, R., Kahn, A. T. and Jennions, M. D. (2015) The extent and consequences of p-hacking in science. *PLoS biology*, 13(3), e1002106.
6. Healey, B. (2007) "Drop downs and scroll mice: The effect of response option format and input mechanism employed on data quality in Web surveys", *Social Science Computer Review*, 25, 111–128.
7. Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M. and DeShon, R. P. (2012) Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99-114.
8. Huang, J. L., Liu, M. and Bowling, N. A. (2015) Insufficient effort responding: examining an insidious confound in survey data. *Journal of Applied Psychology*, 100(3), 828.
9. Johnson, J. A. (2005) Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, 39(1), 103-129.
10. Matjašic, M., Vehovar, V. and Manfreda, K. (2018) Web survey paradata on response time outliers: A systematic literature review. Metodološki zvezki, 15(1), 23-41.
11. Meade, A. W. and Craig, S. B. (2012) Identifying careless responses in survey data. Psychological Methods, 17(3), 437.
12. Read, B., Wolters, L. and Berinsky, A. J. (2021) Racing the clock: Using response time as a proxy for attentiveness on self-administered surveys. *Political Analysis*, 1-20.
13. Simon, H. (1976) Administrative Behavior.(3d Ed.) Free Press. New York.
14. Tourangeau, R., Rips, L. J. and Rasinski, K. (2000) The psychology of survey response.
15. Weigold, A., Weigold, I. K., Jang, M. and Thornton, E. M. (2021) College students' and Mechanical Turk workers' environmental factors while completing online surveys. *Quality & Quantity*, 1-24.
16. Wright, K. B. (2005) Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10(3), JCMC1034.
17. Wyatt, J. C. (2000) When to use web-based surveys. *Journal of the American Medical Informatics Association*, 7(4), pp. 426-430.
18. Zwarun, L. and Hall, A. (2014) What's going on? Age, distraction, and multitasking during online survey taking. *Computers in Human Behavior*, 41, 236-244