SIGHCI 2021 Proceedings

Special Interest Group on Human-Computer Interaction

12-12-2021

# Improving the Quality of Survey Data: Using Answering Behavior as an Alternative Method for Detecting Biased Respondents

manasvi kumar

David Kim

Joe Valacich

Jeffrey Jenkins

Alan Dennis

# Improving the Quality of Survey Data: Using Answering Behavior as an Alternative Method for Detecting Biased Respondents

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

**David Kim**
University of Arizona
davidkim7@email.arizona.edu

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

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

**Alan R. Dennis**
Indiana University
ardennis@indiana.edu

## ABSTRACT

Online surveys are used for collecting self-report data. Despite their prevalent use, data quality problems persist due to various response biases. Here, we demonstrate how participant answering behaviors can be used to identify biased responses. We administered an online survey where participants reported their personality dimensions of neuroticism and extraversion—two personality dimensions that have been previously shown to be correlated with a propensity to deceive—and were later presented with a scenario to exhibit deceptive behavior. We then generated models to predict deception using the neuroticism and extraversion constructs. Using respondents' fine-grained mouse movement data when answering these questions, we generated time, behavior, and navigation-based metrics to identify biased participants. By removing these outliers, model performance improved by 93% for neuroticism and 10% for extraversion. This approach aids in gaining a clearer understanding of how some types of response biases influence model performance.

### Keywords

Survey research, self-report data, online survey, response bias, data quality, outlier.

## INTRODUCTION

Surveys—a research instrument that asks a sample population one or more questions—are among the most common methods for collecting human response data in both academic and industry settings. Using the search term "survey" in Google Scholar returned 6.95 million results and the Association for Information Systems (AIS) eLibrary returned 27,379 results. Clearly, the use of survey response data in research is widespread.

A critical threat to the validity of survey results, both in academic and industry research, is a category of factors referred to as response biases, i.e., a tendency of responding to questions on some basis other than the question content. Response biases can have a detrimental effect on the quality of the results of a survey study (see Navarro-Gonzalez, Vigil-Colet, Ferrando, and Lorenzo-Seva, 2019 for an example). In this paper, we explain and demonstrate how answering behavior can help identify behavioral outliers when participants complete online surveys. We use computer mouse cursor data to generate a variety of continuous metrics which distinguish "normal" and "abnormal" answering behavior. We propose these behavioral outliers may indicate the presence of a bias and can be used to remove low quality data in a manner akin to attention check questions.

We report the results of a study that predicts if someone is being deceptive based on two widely used personality constructs that have been suggested previously to predict one's propensity to deceive—neuroticism and extraversion (Conrads, Irlenbusch, Rilke, and Walkowitz, 2013; Michikyan, Subrahmanyam, and Dennis, 2014). We found that removing behavioral outliers did improve how well neuroticism and extraversion predict deceptive behavior. Specifically, the R-squared of predicting deception based on neuroticism and extraversion improved by 93% and 10% respectively when removing behavioral outliers. In both analyses, the r-squared for these models significantly improved by removing low quality data as identified through analyzing respondents' behavioral data.

## BACKGROUND AND THEORY

While there is widespread and global use of online surveys, there is a large and growing body of literature related to various data quality concerns (Barge and Gehlbach, 2012). Many factors can cause poor data quality. A threat to the validity of survey results includes a category of factors referred to as response biases. A response bias is the tendency of people to respond to questions on some basis other than the question content (Paulhus, 1991). In general, people have the tendency to portray themselves in the best light particularly when asked about personal traits, attitudes, and behaviors, which often causes respondents to falsify or exaggerate answers. In other situations, a person might not be sure how to answer a question because of a lack of understanding of the question. Thus, there are several types of factors that can bias survey responses. For a summary of common types of response biases see Jenkins, Valacich, and Williams, 2017.

Response biases can have a detrimental effect on the quality of the results of a survey study. For example, the significant results of a study might be due to a systematic response bias rather than the hypothesized effect (Gove and Geerken, 1977). On the other hand, a hypothesized effect might not be significant because of a response bias. For example, the intention-behavior gap—a phenomenon that describes why intentions do not always lead to behaviors (Sheeran 2002)—may be attributed to response biases in some situations (Chung and Monroe, 2003). Thus, to increase the validity of many types of survey studies, better methods for detecting biased responses are needed. Response biases can lead to both type 1 errors (i.e., detecting an effect that isn't present) and type 2 errors (i.e., failing to detect an effect that is present).

### Biased versus Non-Biased Responses

Responding to a survey question is, in essence, similar to making a decision. Simon (1976) proposed a simple and elegant, three-step decision-making process – referred to as "Intelligence, Design and Choice" – that is applied here to explain how a person completes survey questions (see Figure 1). In step one (the intelligence phase), information is collected, processed, and examined to identify the problem; this equates to a respondent reading the survey question. In step 2 (the design phase), alternative decision choices are reviewed and considered based upon objectives and the context of the situation; this equates to a respondent evaluating the various response options for a given survey question. In step 3, (the choice phase), an alternative is chosen, or a response is given as the final decision. Simon's model is widely referred to as a "rational" decision-making process, suggesting that decision-making is consciously analytic, objective, and sequenced. While Simon's model is elegant and intuitive, humans often make non-rational decisions due to various emotions and constraints. Many of these non-rational response constraints, when viewed in the context of answering survey questions, reflect the influence of response biases.

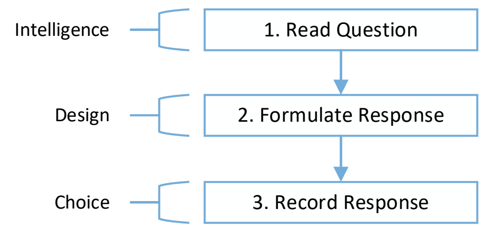


**Figure 1. Simon's rational decision-making model as applied to completing a survey question.**

In some response biases, the respondent has pre-decided their response when engaging in biases such as acquiescence and extreme responding as well as when engaging in various types of satisficing. In these contexts, respondents are more likely to be less engaged in the intelligence process, less engaged in response deliberation, and more quickly select this response. Such respondents will have overall faster response times and show lower levels of deliberation and reconsiderations than population baseline averages. Alternatively, other forms of response biases act to slow the intelligence process, response deliberation, and final response selection or generation. For instance, for a person wanting to provide a more socially desirable answer, they will more likely iterate between reading the question, evaluating possible responses, and selecting a final response. Such respondents will therefore have longer response times and show greater deliberation and answer switching than baseline averages. Thus, different response biases generate meaningful and predictable differences in how questions are processed, how responses are identified, and ultimately how selections are made. Consequently, participants with atypical answering behavior are more likely to be biased than those with typical answering behavior. It is therefore valuable for researchers conducting online surveys to know when and where a respondent enters a biased response.

## METHODOLOGY

We designed a two-part study to explore the effects of using respondent answering behavior to identify behavioral outliers and the influence of such removal on the explained variance in models. In the first part, we asked participants to self-report their personality dimensions of neuroticism and extraversion. The personality dimensions of neuroticism and extraversion have been shown to predict if someone will be deceptive in given scenarios (Conrads, Irlenbusch, Rilke, and Walkowitz, 2013; Michikyan, Subrahmanyam, and Dennis, 2014). In the second part of the study, participants are provided with a scenario to engage in deceptive behavior. We then examine whether removing behavioral outliers increases how well neuroticism and extraversion predict deceptive behavior.

### Survey Design

The survey consists of two parts. In the first part, participants provided answers to various demographics' questions and a 37-question Big Five Inventory questionnaire adapted from previous literature (Goldberg, 1990). These include 7 questions relating to Neuroticism, Extraversion and Openness and 8 questions relating to Conscientiousness and Agreeableness. Participants could then optionally complete the second part – a prescreening application for a future study with a larger payout.

In the second part, participants were told that experience with Excel and its mathematical functions, though not required, would be considered during the selection process. Importantly, all participants in the second part of the survey were asked to rate their skills on a non-existent Excel Plugin (i.e., StatView). The range for all the experience related questions were from 0 (e.g., beginner) to 10 (e.g., expert). As this plugin does not exist, any response greater than 0 was considered to be deceptive, with a greater number indicating a greater tendency to be deceptive.

### Participants

We recruited participants from Amazon Mechanical Turk. They were paid $0.50 for completing the demographic part of the study and another $0.50 for completing the second optional part of the study. Only participants who completed both parts of the study were considered in the analysis. 283 participants completed both parts of the study. Approximately 59% of the participants were male, and 57% of the participants were less than 35 years of age.

### Behavioral Data Tracking

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 and related behavioral data. This raw data was later processed through code developed by the team to obtain the metrics to screen for behavioral outliers. The metrics calculated from the raw data can conceptually fall into three categories: 1. Time, 2. Navigation, and 3. Behavior. Given these three conceptual categories of measures, we generated 3 metrics to identify participants whose responses were likely influenced by a response bias. These metrics are *Baselined response time* (a time-based measure), *Speed* (a navigation measure), and *Response switch* (a behavioral measure). Figure 2 summarizes the categorization of these metrics.

"Baselined response time" analyzes change in response times at an individual level – i.e., compared to how long it normally takes a person to respond. Conceptually, this metric aims to identify inconsistencies among responses. "Speed" is a navigation-based metric designed to identify participants who engage in fast mouse movements and is analyzed at a question level. "Response switch" captures the number of instances when a participant switches his/her response to a question before confirming the selection. We utilize participants' response behaviors to generate scores for these metrics for each response i.e., at question level. If a participants' metric score for any question qualifies as an outlier, their responses for related questions are discarded.

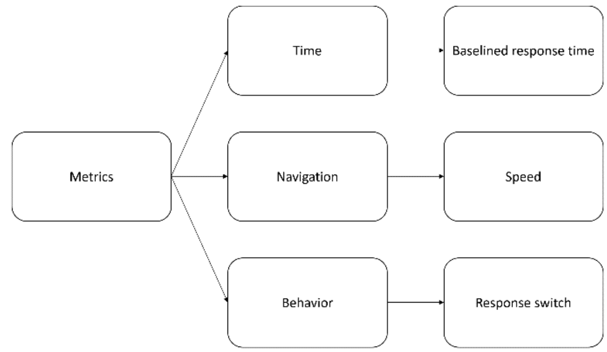


**Figure 2. Categorization of metrics**

## ANALYSIS AND RESULTS

We created two different models to predict if someone deceived using their self-reported responses to the Neuroticism and Extraversion scales individually. The equations for both these linear models are shown in Figure 3. We generated 6 models for each equation. First, we generated the linear model on the entire dataset (n=283). We then removed outliers based on survey completion times for the next model. Next, in separate models, we removed outliers based on the newly proposed metrics individually and combined. Table 1 summarizes the approaches followed in developing the models.

$$\text{Neuroticism Model: Deception}_i = \beta_0 + \beta_1 \text{ Neuroticism}_i + \varepsilon_i$$

$$\text{Extraversion Model: Deception}_i = \beta_0 + \beta_1 \text{ Extraversion}_i + \varepsilon_i$$

**Figure 3. Linear model equations for Neuroticism and Extraversion**

| Model | Name | Outlier Description |
|---|---|---|
| 1 | Entire dataset | No outliers removed |
| 2 | Baseline comparison: Outliers Completion Time | Completion time for entire survey is greater than three MAD away from the median completion time of the entire survey for all participants. |
| 3 | Outliers: Baselined response time | Time to answer the question is greater than three MAD away from median time. (The MAD and median time are calculated from the individual's response times on survey questions). |
| 4 | Outliers: Response switch | Two or more answer switches for a question |

| | | |
|---|---|---|
| 5 | Outliers: Speed | Mouse speed on a question is greater than three MAD away from the median speed for that question for all participants. |
| 6 | Combined Outlier Strategy: All metrics | All 3 behavioral measures used to identify outliers. |

**Table 1. Summary of Models**

Tables 2 and 3 summarize the results of the two linear equation models (Neuroticism & Extraversion) on the entire dataset and its various outlier treatments. Note that for all models mentioned in the tables, neuroticism/extraversion predict deception ($p<0.05$). As suggested in the table, the r-squared value increases when outliers are removed using the survey completion time metric (the baseline outlier removal strategy). However, greater improvements are observed using the metrics suggested in this study (i.e., Baselined response time in both models, Response switch and Speed for neuroticism model). The best performing model is obtained when all 3 outlier removal strategies are used together, resulting in a **93% increase** in the r-squared value for the neuroticism model and **10% increase** for the extraversion model.

| Treatment | Dataset Size | $R^2$ | % $R^2$ increase |
|---|---|---|---|
| Original dataset | 283 | 0.1382 | -- |
| Baseline outlier comparison (Completion Time) | 266 | 0.1498 | 8 |
| Outliers (Baselined response time) | 184 | 0.2184 | 58 |
| Outliers (Response switch) | 249 | 0.1696 | 23 |
| Outliers (Speed) | 211 | 0.1617 | 17 |
| Combined Outliers (All metrics) | **151** | **0.2666** | **93** |

**Table 2. Table summarizing results for Neuroticism model**

### DISCUSSION

In this paper, we demonstrate that using time, navigation and behavior outlier removal techniques can potentially be used to remove "low quality data" and improve model performance. We only collected data from participants who cleared the attention check question in their corresponding surveys. Thus, we demonstrate that removing data from participants who fail attention questions alone is not adequate, and that the suggested metrics are useful in identifying and eliminating additional outliers. A common trend across results obtained in both linear models is the use of the "Baselined response time" metric as an important tool to identify outliers. This result is consistent with several studies that use time-based metrics to improve data quality (Christian, Parsons, and Dillman, 2009).

| Treatment | Dataset Size | $R^2$ | % $R^2$ increase |
|---|---|---|---|
| Original dataset | 283 | 0.2152 | -- |
| Baseline outlier comparison (Completion Time) | 266 | 0.2242 | 4 |
| Outliers (Baselined response time) | 169 | 0.2356 | 9 |
| Outliers (Response switch) | 247 | 0.2205 | 2 |
| Outliers (Speed) | 206 | 0.2118 | -2 |
| Combined Outliers (All metrics) | **145** | **0.2369** | **10** |

**Table 3. Table summarizing results for Extraversion model**

Another observable trend across results in both models is the loss of data due to the developed metrics. For datasets created with individual metrics, we observe that most data are lost using the Baselined response time metric. Around 69% (184 of 266) of data remains for the neuroticism model while 63.5% (169 of 266) remains for the extraversion model. One explanation for this observation could be due to the definition of the Baselined response time metric which classifies a participant as an outlier if the time taken to generate a response is greater than three MAD away from the median time taken by the participant to generate responses on other similar questions. We analyzed the effect of changing the number of MAD allowed and its impact on model performance and data size and found that increasing tolerance for response generation time (by increasing the number of MAD) is not necessarily good for model performance and there is an optimal number for allowed MAD that allows for minimal data loss while maximizing model performance.

#### Implications for Research and Practice

There are three important benefits over existing approaches for dealing with response biases. First, our methods can be used to identify if response biases are likely to be present in a study (i.e., if the Baselined response time, Response switch score, and Speed are not significantly outside baselines, a response bias in not likely to be present). Second, our technique provides novel insight into understanding how response biases influence relationships that are often difficult or impossible to obtain through other measures and approaches. Third, and most importantly, the statistical metrics used to capture response biases and eliminate low quality data helps to account for various types of response biases in predictive statistical models, thus improving the explanatory power of the relationship between a survey construct and predictor variable.

## LIMITATIONS

In this paper, we developed three behavioral metrics which help identify individuals who likely provided biased responses. While the metrics are developed to identify responses which are biased, it is plausible that other factors could impact participant behavior and the metrics incorrectly flag certain responses as outliers. We also do not account for differences in magnitude of bias before excluding potentially biased responses. Additionally, while our developed metrics help in improving data quality, it is unclear for now as to which combination of metrics gives the best performance. We also acknowledge that the metrics presented in this paper are not exhaustive, and that metrics relating to more complicated statistical modeling (Tijdens, 2014) could also help improve data quality.

## FUTURE RESEARCH

In this paper we identify participants who provide biased responses through time, navigation and behavior-based metrics. While the approach provided encouraging results, there are limitations that exist which could be addressed in future studies. As mentioned earlier, there are other kinds of metrics that haven't been analyzed yet (metrics derived from mouse-movement, participant behavior like acquiescence bias etc.). These metrics, and their dependence on the type of questions asked in the survey raise several potential research questions that require further study. Future studies could also test the efficacy of the proposed methods and metrics in different contexts.

## CONCLUSION

In this paper, we address the growing issue of poor data quality in online surveys by identifying and excluding participants who provide biased responses. We developed metrics based on time, navigation and behavior to identify participants belonging on either end of the response time spectrum. To test the efficacy of this approach, we conducted a survey where participants self-reported their personality dimensions on neuroticism and extroversion – two personality traits shown to be correlated with propensity to be deceptive in the past. They were then presented with a scenario to exhibit deceptive behavior. We generated predictive models that estimated the degree of deception based on scores obtained from neuroticism and extraversion constructs. We found that model performance improves when outliers identified using our metrics are removed. We posit that researchers can utilize these methods to not only improve model performance but also improve understanding of relationships between constructs measured through surveys.

## REFERENCES


1. Barge, S., and Gehlbach, H. (2012). Using the Theory of Satisficing to Evaluate the Quality of Survey Data, Research in Higher Education (53), pp. 182–200. (https://doi.org/10.1007/s11162-011-9251-2).
2. Christian, L., Parsons, N., and Dillman, D. (2009). Design Scalar Questions for Web Surveys, Sociological Methods and Research - SOCIOL METHOD RES (37), pp. 393–425. (https://doi.org/10.1177/0049124108330004).
3. Chung, J., and Monroe, G. S. (2003). Exploring Social Desirability Bias, Journal of Business Ethics (44:4), pp. 291–302. (https://doi.org/10.1023/A:1023648703356).
4. Conrads, J., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2013). Lying and Team Incentives, Journal of Economic Psychology (34), pp. 1–7. (https://doi.org/10.1016/j.joep.2012.10.011).
5. Goldberg, L. R. (1990). An Alternative Description of Personality: The Big-Five Factor Structure., Journal of Personality and Social Psychology (59:6), American Psychological Association, p. 1216.
6. Gove, W. R., and Geerken, M. R. (1977). Response Bias in Surveys of Mental Health: An Empirical Investigation, American Journal of Sociology (82:6), University of Chicago Press, pp. 1289–1317.
7. Jenkins, J. L., Valacich, J. S., and Williams, P. A. (2017). Human-Computer Interaction Movement Indicators of Response Biases in Online Surveys, in 38th International Conference on Information Systems: Transforming Society with Digital Innovation, ICIS 2017.
8. Michikyan, M., Subrahmanyam, K., and Dennis, J. (2014). Can You Tell Who I Am? Neuroticism, Extraversion, and Online Self-Presentation among Young Adults, Computers in Human Behavior (33), Elsevier, pp. 179–183.
9. Navarro-Gonzalez, D., Vigil-Colet, A., Ferrando, P. J., and Lorenzo-Seva, U. (2019). Psychological Test Toolbox: A New Tool to Compute Factor Analysis Controlling Response Bias, Journal of Statistical Software (91:1), pp. 1–21.
10. Paulhus, D. L. (1991). Measurement and Control of Response Bias., Measures of Personality and Social Psychological Attitudes., Measures of Social Psychological Attitudes, Vol. 1., San Diego, CA, US: Academic Press, pp. 17–59.
11. Sheeran, P. (2002). Intention—Behavior Relations: A Conceptual and Empirical Review, European Review of Social Psychology (12:1), Taylor & Francis, pp. 1–36.
12. Simon, H. A. (1976). Administrative Behavior; a Study of Decision-Making Processes in Administrative Organization-3, New York, NY (USA) Free Press.
13. Tijdens, K. (2014). Dropout Rates and Response Times of an Occupation Search Tree in a Web Survey, Journal of Official Statistics (30:1), Sciendo, p