

# It's more than money: The true costs of spam in online survey research

Research Ethics  
1–18© The Author(s) 2025  
Article reuse guidelines:sagepub.com/journals-permissions  
DOI: 10.1177/17470161251357454  
journals.sagepub.com/home/rea

Shubham Agrawal<sup>1</sup> , Gwendolyn Paige Watson<sup>2</sup> ,  
Amy M. Schuster<sup>1</sup> , Shelia R. Cotten<sup>1</sup> , Michael L. Tidwell<sup>3</sup> ,  
Nathan Baker<sup>4</sup>  and Sicheng Wang<sup>5</sup> 

## Abstract

Online surveys are a popular method for collecting data in the social sciences. Despite its cost-effectiveness, concerns regarding the legitimacy of data from online surveys are increasing. One such concern is fraudulent responses or “spam” by malicious agents intentionally deceiving the survey process to gain monetary incentives or sway research results. The research costs of “spam”—their influence on research conclusions and their threat to scientific integrity—are not well understood. Here we show the differences in financial and research costs of spam using data from an online survey of transportation workers that was cleaned using a stringent battery of spam detection techniques that utilized commercially available features and a custom spam detection algorithm. We found that we would have wasted about 73% of our budget on incentivizing spammers if we had stopped data collection upon reaching the intended sample size. We also found significant differences in research conclusions related to the relationships between key organizational constructs, including affective commitment, job satisfaction, and turnover intention, between subsamples with and without spam. Our results demonstrate that researchers who are unaware of spam or do not adequately clean their data may spend substantially more monetary and human resources, as well as derive misleading conclusions. This study highlights the importance of survey researchers being cognizant of spam responses and employing robust spam detection techniques to ensure the scientific integrity of non-probability online survey research.

<sup>1</sup>Clemson University, SC, USA

<sup>2</sup>Auburn University, AL, USA

<sup>3</sup>University of Georgia, Athens, GA, USA

<sup>4</sup>Michigan State University, East Lansing, MI, USA

<sup>5</sup>University of South Carolina, Columbia, USA

## Corresponding author:

Shubham Agrawal, Department of Psychology, Glenn Department of Civil Engineering, Clemson University, 418 Brackett Hall, Clemson, SC 29634, USA.

Email: agrawa3@clemson.edu



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

## Keywords

web-based survey, bot screening, data quality, research integrity, internet research

## Introduction

Researchers are ethically responsible for conducting research and disseminating scientific findings that are supported by reliable and valid data and accurate interpretations of the results (American Sociological Association, 2018). The integrity of scientific research is in jeopardy given the proliferation of disinformation affecting data quality (Pozzar et al., 2020). Online survey research, where the distribution of survey instruments and collection of participant responses are conducted through the Internet, has grown as a dominant data collection approach for quantitative studies across the social sciences and various other disciplines (Chandler and Shapiro, 2016; Daikeler et al., 2020). Online surveys as a ubiquitous tool for data collection has introduced a potential challenge to maintaining the integrity of scientific research and increasing disinformation via fraudulent responses (Griffin et al., 2022; Levi et al., 2022).

Many online surveys rely on non-probability sampling methods, where members of the target population have an unequal and unknown probability of being sampled (Crano et al., 2015; Vehovar et al., 2016). A common non-probability approach used in online surveys is convenience sampling, where the sample consists of individuals who are available and willing to participate in the study (Crano et al., 2015). Convenience samples in online survey research are often obtained using social media postings, emails, and online panels (Vehovar et al., 2016). Such non-probability sampling-based online surveys offer benefits of cost-effectiveness, convenience, and the ability to engage hard-to-reach populations (Das et al., 2018; Dillman et al., 2014). However, they suffer from a potential lack of generalizability of results to the target population (Vehovar et al., 2016) and other methodological challenges such as low response rates (though response rates for surveys regardless of mode of administration have declined over the past two decades) and concerns about data quality (Daikeler et al., 2020; Manfreda et al., 2008). Traditionally, data quality concerns in online survey research have focused on data compromised by careless responders (Arthur et al., 2021; Meade and Craig, 2012; Nichols and Edlund, 2020). In recent years, concerns have increased as fraudulent responses and misinformation have been proliferated by devious agents in online surveys (Bybee et al., 2022; Levi et al., 2022). With researchers continuing to accumulate scientific knowledge through online surveys using the convenience sampling approach, it is imperative that researchers devote attention to understanding the impact of fraudulent survey responses to both research budgets and the integrity of science itself. This study presents an investigation of the financial costs (e.g. monetary participation incentives) and research costs (e.g. misleading research conclusions) of fraudulent responses in a large online survey utilizing a convenience sampling approach.

### *Fraud in online surveys*

Human fraud and automated bots are the two main types of fraud in online surveys (Simone et al., 2024). Human fraud refers to an invalid human response where a participant intentionally deceives (or “scams”) the survey process by providing phony answers, assuming fictitious identities, or attempting to take the survey multiple times using different fake profiles (Zhang et al., 2022). Automated bots (or “survey bots”) refer to computer-generated responses that are algorithmically programmed to automatically respond to surveys, often multiple times, with fake data (Kennedy et al., 2021; Storozuk et al., 2020). Automated bots are relatively easy to create and deploy, exacerbating the threat to the integrity

of online survey research (Dupuis et al., 2019). In this study, we refer to these fraudulent responses collectively as “spam” and the responders as “spammers.” Spammers attempt to deceive researchers through fabricated information or multiple completions to gain survey incentives (e.g. financial compensation) or to sway survey results (Griffin et al., 2022; Pozzar et al., 2020). Although research fraud is not unique to online surveys, as inauthentic participants have also been reported in qualitative research studies (Dougherty, 2021; Hoskins et al., 2025; Panicker et al., 2024), the rise of automated bots presents a much greater challenge in online survey research.

Fraudulent responses appear to be increasingly prevalent in online surveys, with several studies identifying most of their data as spam, especially those using non-probability sampling strategies (Bell and Gift, 2023; Buchanan and Scofield, 2018; Bybee et al., 2022; Griffin et al., 2022; Ruby et al., 2025). This increased prevalence results from technological advances, with modern survey bots able to circumvent many conventional data quality checks used in online surveys. For example, survey bots can effectively circumvent instructed response attention checks (e.g. “Please select ‘Strongly agree’ for this question”), which are commonly used to detect inattentive responses (Kennedy et al., 2021). Even complex CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) tests can be bypassed by bots that utilize advanced techniques like deep learning models (Zhang et al., 2022). Additionally, inexpensive access to virtual private networks has enabled spammers to falsify their IP location and respond to surveys that are otherwise restricted to certain regions (Pozzar et al., 2020). Some researchers analyze responses to open-ended text questions to identify automated spam. However, modern survey bots can generate convincing human-like text responses to open-ended questions by applying natural language processing techniques (Li et al., 2021). This issue will likely be exacerbated by the emergence of bots powered by artificial intelligence (AI) text generators, such as OpenAI’s ChatGPT, which are built on large language models capable of producing contextually relevant and consistent text responses similar to those provided by humans (Sawhney et al., 2025).

### *Costs of survey spam*

There are two primary types of costs associated with spam responses in online survey research. First, the financial costs, which include the resources spent on collecting and incentivizing spam responses. For example, federal funding agencies in the U.S. allocate copious amounts of money to research efforts annually. In FY 2023, the National Science Foundation appropriated \$7.86 billion for research activities (NSF, 2024), while the National Institutes of Health spent \$34.9 billion on extramural research (Lauer, 2024). Researchers have an ethical responsibility to spend research funds responsibly on real participants; if they are not cognizant of spam responses in online surveys, funds will be wasted, and data will be compromised. Prior work has alluded to the financial costs associated with providing participation incentives to fraudulent respondents in online survey research (Levi et al., 2022). However, researchers have not explicitly examined the potential for research funds wasted on spam.

Second, and more importantly, are the research costs, which denote the influence of fraudulent responses on research conclusions and their threat to scientific integrity. Since fraudulent data do not represent the responses (e.g. attitudes, behavior, or opinions) of the target population in online survey research, inferences made from analyzing data contaminated with spam may be invalid and misleading. Griffin et al. (2022) noted that spam in online surveys may be even more harmful for research involving marginalized groups.

Spam responses can produce skewed results (Kennedy et al., 2021) and alter study findings (Bernerth et al., 2021; Chandler et al., 2020), thus contributing to the proliferation of disinformation. Even low rates of inaccurate responses can profoundly bias survey results (Arthur et al., 2021), suggesting that researchers who are not able to counter the more advanced forms of online survey spam risk jeopardizing the scientific integrity of their study results. Past studies have

reported the impacts of fraudulent responses by analyzing the differences in frequency of responses (Chandler and Shapiro, 2016), correlations among focal variables (Chmielewski and Kucker, 2020; Pratt-Chapman et al., 2021), reliability and validity indicators of constructs (Chmielewski and Kucker, 2020), and treatment effects and their confidence intervals (Kennedy et al., 2020). However, these studies either did not explicitly examine the impacts of spam on overall research conclusions or relied on specific participant recruitment platforms (e.g. MTurk) for data collection that may not be limited in their ability to recruit participants from certain population subgroups.

Hence, failing to adequately address spam responses in online survey research raises serious ethical concerns. For example, misallocation of research funds—particularly public funding—to incentivize spam responses reduces researchers' public accountability, and misallocation of staff time to manage and analyze fraudulent data takes away valuable expert resources that could otherwise be spent on meaningful research. Publishing findings based on compromised data, even if unintentional, can drive subsequent ill-informed research, leading to further waste of time, effort, and funding by other researchers. Moreover, unreliable and invalid results from compromised data can have dire consequences if used to inform public policy and can significantly deteriorate public trust in the scientific community.

### *The current study*

This study investigates the following research questions:

- **RQ1:** What are the “financial costs” associated with spam?
- **RQ2:** What are the “research costs” of spam, that is, the impact of spam on research inferences and conclusions drawn from the data?

We answer these questions using a large, convenience sampling-based online survey of U.S. transportation workers. The survey was affected by fraudulent activity and went through a stringent battery of spam detection techniques to identify fraudulent responses. We analyzed the financial costs of spam associated with this survey by estimating the amount of funds that would have been spent on incentivizing spammers if the researchers were unaware of them. We also examined the research costs of spam by empirically comparing the results and inferences for valid responses and spam responses in the survey.

## **Methods**

### *Online survey*

To analyze the costs associated with spam responses, we examined data from an online survey as part of a project funded by the National Science Foundation to investigate the potential workforce impacts of automated vehicles in three transportation industries: trucking, taxi services, and gig driving. Workers in seven occupational groups were recruited to participate in the survey from these industries, including truck drivers, trucking supervisors/dispatchers, trucking owners/managers, gig drivers (ride-hailing or app-based delivery service), taxi drivers, taxi supervisors/dispatchers, and taxi owners/managers. In line with the primary purpose of our study, the original target sample size was 350 for each of the seven industry groups based on the available research budget and a priori power analysis to achieve adequate power to detect medium-to-small effects. The online survey was not originally designed to investigate fraudulent responses; instead, our

experience with spam during data collection prompted the post-hoc analysis presented in this study. Details of the survey and a brief data collection timeline are provided below.

The 15–20-minute survey was administered online through Qualtrics, a web-based survey platform, in the U.S. between June 2022 and October 2022. The survey study was approved by the Institutional Review Board (IRB) at Clemson University and Michigan State University. Participants who completed the survey were redirected to a separate “incentive survey” where they were given an option to provide their name and email to receive a \$10 Amazon e-gift card as a participation incentive. At the beginning of the survey, participants were told that the survey includes questions about their current job and their perceptions of how fully automated vehicles may affect their jobs and industry. They were also informed that they would be compensated only if they answered questions to check respondent’s attention correctly. Given the limited research budget and hard-to-access samples, quota sampling (Battaglia, 2008) was implemented with a target sample size of 350 participants for each of the seven targeted worker groups.

During the data collection period, we recruited survey participants through four approaches. First, we contacted several industry organizations, worker unions, local transportation agencies, taxi companies, and a national gig-driving company to distribute the survey among their networks or employees. Second, we disseminated the survey through the research team’s relevant industry and professional networks. Third, we posted flyers and advertisements on social media platforms (e.g. Reddit, Facebook, LinkedIn). Finally, we placed flyers at truck stops (targeting truck drivers) and on some university campuses in the states of South Carolina, Michigan, and Georgia (targeting students who drove for gig mobility services). Those interested in participating could access the online survey through an anonymous link or by scanning a QR code. The heavy reliance on online convenience sampling strategies led to an unexpected influx of spam responses.

Two universities were involved in this project. The Qualtrics survey was first managed by Clemson University and then switched to Michigan State University. Clemson University collected 3,393 complete responses between June 2022 and July 2022. During the first 10 days of data collection, we identified an influx of spam responses (e.g. bursts of similar responses in a short period of time) and paused data collection to develop techniques for detecting spam responses (see “Spam prevention and detection techniques” section below). Given the extent of spam we experienced, we sought and received additional approval from the IRB to link data from the incentive survey to the main survey responses. The data were linked by matching IP addresses and aligning the start timestamps of the incentive survey with the end timestamps of the main survey. The survey was moved to Michigan State University’s Qualtrics platform as it offered more advanced spam detection features (e.g. RelevantID fraud score and invisible reCAPTCHA). We also updated the information for participants at the beginning of the survey to include that suspicious or low-quality survey responses would not be compensated. We resumed data collection and collected 9,685 completed responses between July 2022 and October 2022 through Michigan State University. Once data collection was complete, all survey responses from both universities were evaluated with a custom spam detection algorithm and a series of manual checks for consistency across the entire sample.

### *Spam prevention and detection techniques*

The existing literature on survey spam detection suggests that a combination of common data quality checks and preventative measures serves as the best approach for detecting survey spam, as no single technique appears to capture all types of online survey fraud (Griffin et al., 2022; Lawlor et al., 2021; Simone et al., 2024; Storozuk et al., 2020; Zhang et al., 2022). To prevent low-quality and fraudulent responses, our research team implemented in-survey checks when the survey was

**Table 1.** Summary of survey spam prevention and detection techniques.

Type	Examples
<u>In-survey checks</u>	reCAPTCHA verification question; instructed response attention checks; common knowledge test; response consistency check; honeypot questions
<u>Custom spam detection algorithm</u>	
Response metadata	IP address location; short survey duration
Response survey data	Non-regular characters in text; suspicious patterns (e.g. repeated first and last names)
Metadata comparison	Quick bursts of responses with similar survey duration; multiple attempts from the same IP
Survey data comparison	Duplicated email address; duplicated text responses
Qualtrics security features	Invisible reCAPTCHA; RelevantID fraud score
<u>Manual checks</u>	Inspecting text responses (nonsensical or plagiarized); dubious email address; follow-up verification emails

administered. We also employed strategies to detect fraudulent responses, including a custom spam detection algorithm and manual checks. Using these techniques, we classified the completed survey responses as “valid” or “spam.” Table 1 presents a summary of the implemented methods. An overview of the spam detection process is provided in Supplemental Figure 1.

*Step 1: In-survey checks.* The survey contained several types of questions to identify inattentive, careless, or fraudulent responses (Abbey and Meloy, 2017; Simone et al., 2024). First, participants had to successfully respond to a reCAPTCHA verification question to continue the survey. Within the survey, participants had to correctly respond to two Likert-style instructed response attention checks (e.g. “Please choose ‘Strongly agree’”), one of the two randomly selected instructed response items (e.g. “Please type ‘99’ in the text field below”), and one of the two randomly selected common knowledge test questions (e.g. “Please enter the result of  $2 + 2$ ”). Randomization was added to reduce bots’ ability to memorize the in-survey check questions. A response consistency check was included where participants reported the industry sector of their primary occupation toward the beginning and end of the survey. The survey included two “honeypot” questions, which were visually hidden from human participants but were detectable by bots (Simone, 2019). The survey immediately ended for participants who failed any of these checks, and their responses were considered incomplete and removed from further analysis.

*Step 2: Custom spam detection algorithm.* Our research team developed a custom spam detection algorithm that classified each complete response as either “presumed-valid” (no apparent reason for suspicion of spam), “low-potential-spam” (indicating a mild level of suspicion), or “high-potential-spam” (indicating a strong basis for suspicion). The algorithm consisted of multiple filters that either checked metadata and survey data within individual responses or compared them between all survey responses. We included two Qualtrics fraud detection premium add-ons (Qualtrics, 2023): Google’s invisible reCAPTCHA technology (tracks user actions and behaviors) and Imperium’s RelevantID fraud score (comparison of user metadata to other online activity using digital fingerprinting) during the second phase of data collection. The details for filters used to flag “high-potential-spam” and “low-potential-spam” responses are presented in Supplemental Table 1.

Previous studies have used similar strategies to classify responses with an increasing level of suspicion of fraud (Goodrich et al., 2023; Miner et al., 2012). However, their filters were mostly limited to either IP address, survey duration, or suspicious responses to knowledge-based questions. All “low-potential-spam” responses were manually checked, as detailed in the next section. The remaining “high-potential-spam” responses were grouped as “spam.”

*Step 3: Manual checks.* As the final step for data that had not been removed during the first two steps, we performed a series of manual checks to classify “presumed-valid,” “low-potential-spam,” and certain “high-potential-spam” responses as either “valid,” “potential-spam,” or “spam.” A team member manually checked for suspicious patterns in responses to the three open-ended text response questions and email addresses across respondents. The survey included three text response items that inquired about participants’ day-to-day job responsibilities, factors they would consider if changing jobs, and their anticipated outcomes of autonomous vehicle adoption in the transportation industry. To assist the researcher in keeping track of suspiciously similar responses, we utilized a fuzzy string-matching algorithm implemented in the “fuzzywuzzy” Python package to flag similar text responses (Cohen, 2022). We also inspected for nonsensical (e.g. “ghsjehqb”) or inadequate (e.g. “There is no”) open-ended text responses, plagiarized text responses from the Internet, as well as suspicious email address patterns (e.g. “01email@mail.com” and “02email@mail.com”).

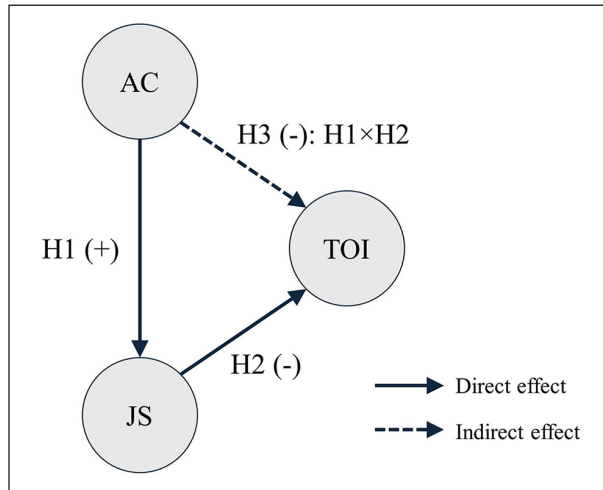
A verification email was sent to (1) all responses deemed “potential-spam” responses and (2) any “high-potential-spam” respondents who contacted the research team inquiring about their survey incentive. The verification email verified three pieces of information (i.e. age, educational attainment, industry), which were then compared to their original survey responses within a reasonable deviation. Small differences in age response (e.g. 1 year older than in the original data) were ignored, as were minor wording differences between industry and education responses. If the information was verified, the survey response was classified as “valid”; otherwise, it was classified as “spam.” These manual checks were time-intensive but crucial for enhancing confidence in the data.

### *Analysis of financial costs of spam*

To underscore the financial costs of online survey spam, we calculated the hypothetical amount of funds that would have erroneously been spent on spam responses. This calculation is based on the intended sample size of the study, assuming we had relied solely on conventional data quality checks (i.e. in-survey checks). Since our original intent was to incentivize the first 350 respondents in each of the seven industry groups that passed all typical survey methods to detect fraudulent responses, we stratified the data to capture (1) the first 350 responses in each group and (2) their final flag (valid or spam) after our battery of data quality checks. Stratifying the data in this way allowed us to calculate the cost of compensating respondents, which would have otherwise been valid based solely on traditional checks of fraudulent responses.

### *Analysis of research costs of spam*

To investigate if there were differences in research findings due to spam responses, we tested the hypothesized relationships between affective occupational commitment (AC; i.e. the emotional attachment one has toward their occupation), job satisfaction (JS; i.e. one’s overall evaluation of the favorability of their job), and turnover intention (TOI; i.e. the desire one has to voluntarily leave their industry). These constructs have been extensively studied in the



**Figure 1.** Hypothesized relationships between affective commitment (AC), job satisfaction (JS), and turnover intention (TOI).

organizational science literature, which provides us with the expected strength and direction of their relationships (Judge et al., 2017; Tett and Meyer, 1993; Xu et al., 2023). Past literature has found AC to be strongly correlated with JS as feeling emotional attachment and pride for one's occupation generates strong positive appraisals of their job ( $\rho=0.63$ ; Cooper-Hakim and Viswesvaran, 2005). JS and TOI have been found to have a strong negative relationship, as workers who are satisfied with their jobs are less likely to desire to leave ( $\rho=-0.55$ ; Özkan et al., 2020). Given these expected relationships, we hypothesized the following (see Figure 1):

- **H1:** Affective commitment will have a strong positive relationship with job satisfaction.
- **H2:** Job satisfaction will have a strong negative relationship with turnover intention.
- **H3:** Affective commitment will have a strong negative influence on turnover intention, and this relationship will be fully mediated by job satisfaction.

AC (Meyer et al., 1993), JS (Morgeson and Humphrey, 2006), and TOI (Meyer et al., 1993) were measured using established scales that were modified for the study context, as described in Table 2.

We tested these hypothesized relationships (see Figure 1) using structural equation modeling (SEM) on three subsamples independently: (1) valid responses only, (2) spam responses only, and (3) all completed responses (combined spam and valid responses). Before we estimated the SEM, we assessed the empirical equivalence of constructs between valid and spam groups by testing for measurement invariance of the three scales using confirmatory factor analysis (CFA; Kline, 2016).

The SEM and CFA model fits are assessed using the following fit indices: (1) Chi-square goodness-of-fit test, with  $p > 0.05$  suggests a good model fit, (2) Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), with  $CFI \geq 0.95$  and  $TLI \geq 0.95$  indicating a relatively good model fit, and (3) Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR), with  $RMSEA \leq 0.06$  and  $SRMR \leq 0.08$  indicate a relatively good model fit (Hu and Bentler, 1999).

**Table 2.** Overview of survey items.

Construct & items	Description
<u>Affective commitment (AC)</u> <sup>a</sup>	<i>To what extent do you agree or disagree with the following statements when thinking about the industry of your primary transportation job?</i>
AC1	Being {occupation} is important to my self-image.
AC2 [R]	I regret having entered the {industry} profession.
AC3	I am proud to be in the {industry} industry.
AC4 [R] [E]	I dislike being {occupation}.
AC5 [R]	I do not identify with the {industry} profession.
AC6	I am enthusiastic about the {industry} industry.
<u>Job satisfaction (JS)</u> <sup>a</sup>	<i>To what extent do you agree or disagree with the following statements when thinking about your primary transportation job?</i>
JS1	I like the kind of work I do.
JS2	I like my job better than the average worker does.
JS3	I find real enjoyment in my work.
JS4	I feel a great sense of personal satisfaction when I do my job well.
JS5	Considering everything, I am satisfied when I do my job well.
<u>Turnover intentions (TOI)</u> <sup>b</sup>	<i>To what extent do you agree or disagree with the following statements when thinking about your current primary transportation job and industry?</i>
TOI1	I frequently think about getting out of the {industry} industry.
TOI2	I am likely to explore other career opportunities in the transportation industry.
TOI3	I am likely to explore other career opportunities outside of the transportation industry.
TOI4	I am likely to leave the {industry} industry in the next year.

<sup>a</sup>Items were measured on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

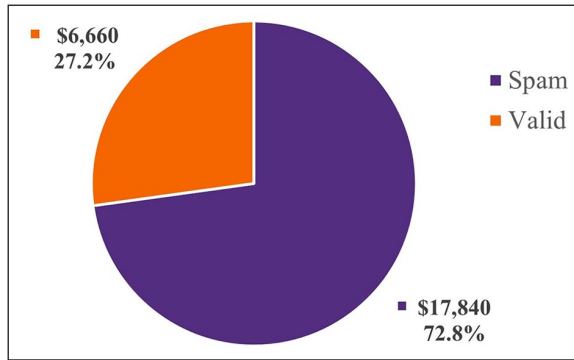
<sup>b</sup>Items were measured on a 4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree).

[R] indicates reverse coded items.

[E] indicates items that were excluded from the analysis due to missing data.

{ } indicates piped text from earlier responses.

Since the primary objective of this study is to compare the overall research inferences across subsamples, rather than directly comparing individual parameter estimates between subsamples, we used standardized parameter estimates to enhance the interpretability of the model results within each subsample. Typically, items with standardized factor loadings between 0.30 and 0.40 are considered to meet the minimum requirement, and loadings  $\geq 0.50$  are considered acceptable (Hair, 2010). AC4 was excluded from the analyses due to missing data for over half of the responses, as it was measured only for participants whose primary occupation is driving. We also assumed two correlated error terms: (i) AC2 and AC5, as they both are reverse-coded items, and (ii) TOI2 and TOI3, as they are created by decomposing the original item on the TOI scale “I am likely to explore other career opportunities” into “. . .opportunities in the transportation industry” (TOI2) and “. . .opportunities in a non-transportation industry” (TOI3). The CFA and SEM models were estimated using the maximum likelihood estimator implemented in the “lavaan” package (version 0.6-15) in the R programming language (Rosseel, 2012).



**Figure 2.** Hypothetical expenditure in our survey with conventional data quality checks.

## Results

We employed a stringent battery of spam detection techniques to classify completed responses as “valid” ( $n=1,774$ ) or “spam” ( $n=11,304$ ).

### *Financial costs of spam*

We offered a \$10 gift certificate as a participation incentive for each respondent, resulting in a budget of \$24,500. Out of our target sample ( $N=2,450$ ), 1,784 were spam responses and 666 were valid responses. Therefore, about 73% of our incentive budget (\$17,840) would have been wasted on survey spam if we had not implemented robust spam detection techniques, as illustrated in Figure 2.

In addition to the direct financial costs of participation incentives, there were several indirect costs related to personnel expenditure, including salaries and a lack of staff time to work on other project components. In this study, multiple graduate research assistants, a postdoctoral researcher, and an assistant professor spent substantial time and effort developing spam detection techniques and cleaning the data over several months. Like many academic researchers, we did not keep a detailed log of time spent on this data cleaning effort, which makes it difficult to estimate these costs precisely.

### *Research cost of spam*

Model results show that the configural invariance model did not have an acceptable fit based on the Chi-square test ( $\chi^2(144)=7107.6, p < 0.001$ ) as well as incremental and absolute model fit indices (CFI=0.901, TLI=0.875, RMSEA=0.087, SRMR=0.108; see Supplemental Table 2 for detailed results). Measurement invariance between the “valid” and “spam” groups was not supported at the configural level, indicating that the survey scales do not measure the same underlying focal constructs and should not be compared between the two groups.

Further, the results indicate that the estimated factor loading for TOI2 is less than the desired threshold of 0.50 in both subsamples (0.217 for valid and 0.426 for spam), which is not unexpected given that this item was not previously validated and is very similar to TOI3. The results also illustrate that while the estimated factor loadings for the two reverse-coded items (AC2 and AC3) are acceptable in the valid subsample (0.536 and 0.520), they are much lower than the desired

threshold in the spam subsample (0.196 and 0.200), indicating a poor factor structure. The SEM results show similar factor loading patterns.

Importantly, the SEM results indicate a good model fit for the valid sample ( $\chi^2(73)=507.7$ ,  $p < 0.001$ , CFI=0.950, TLI=0.938, RMSEA=0.060, SRMR=0.068), while they indicate a poor model fit for the spam sample ( $\chi^2(73)=6699.3$ ,  $p < 0.001$ , CFI=0.893, TLI=0.866, RMSEA=0.091, SRMR=0.125) and the combined sample ( $\chi^2(73)=7173.2$ ,  $p < 0.001$ , CFI=0.899, TLI=0.874, RMSEA=0.087, SRMR=0.119; see Supplemental Table 3 for detailed model results). Although the Chi-square goodness-of-fit test indicates that neither model fits the data perfectly, this test is highly sensitive to sample size and is often significant in large samples (Kline, 2016). The estimated model parameters for the hypothesized relationships between the valid, spam, and combined subsamples are illustrated in Figure 3. A comparison of results between the valid, spam, and combined samples is summarized in Table 3.

## Discussion

To our knowledge, this is the first study to: (1) explicitly assess the potential financial costs of spam, and (2) examine the research costs of spam by empirically comparing the results for valid and spam responses in online surveys using non-probability sampling. Our results show that there are substantial direct financial costs of survey spam—nearly three-fourths of our incentive budget would have been unintentionally misused to compensate spammers—and considerable indirect costs (e.g. personnel time expenditure, including tuition and salaries) associated with spam detection and data cleaning processes. However, the financial costs are only a part of the more significant issue with online survey spam.

Our analysis shows that the inclusion of spam would have compromised the scientific integrity of this research. First, the lack of configural invariance typically indicates differences in contextual/cultural factors or item interpretation (Kline, 2016). However, for the U.S. transportation workers sample, the most plausible cause for the lack of configural invariance is differences in response styles between valid and spam responses. This suggests that failing to exclude spam responses can compromise the factor structure and yield spurious relationships. Further, spam responses performed particularly poorly on reverse-coded items, as indicated by their low factor loadings. These findings suggest that spam respondents tend to select answers at one extreme, which is consistent with existing literature that has found a tendency for insincere respondents to choose positive answer choices (Kennedy et al., 2021). Further research is needed to determine if reverse-coded items can help identify spam responses without increasing false positives, while balancing issues like participant confusion (Venta et al., 2022) and lower internal consistency (Weijters and Baumgartner, 2012).

Second, as in our study and others (Griffin et al., 2022; Webb and Tangney, 2024; Xu et al., 2022), it is likely that spam responses in affected online survey studies outnumber valid responses due to automated spam. Researchers who are incognizant of spam in online surveys, particularly when using non-probability sampling methods, may draw false conclusions despite appropriate analysis. For example, our combined dataset would not have supported the hypothesized relationships due to poor model fit, unlike the valid subsample.

Furthermore, including spam responses would have produced inaccurate and potentially misleading results that could influence both theory and practice. For example, worker turnover is costly for organizations (Park and Shaw, 2013), stimulating longstanding interest in improving worker retention and identifying attrition factors (Lee et al., 2017; Shaw, 2011). However, if we had not detected and removed spam responses, we would have wrongly concluded that (1) job satisfaction has a significant but weak negative relationship with workers' turnover intention, and (2) affective commitment has a significant but weak negative indirect effect on turnover intention

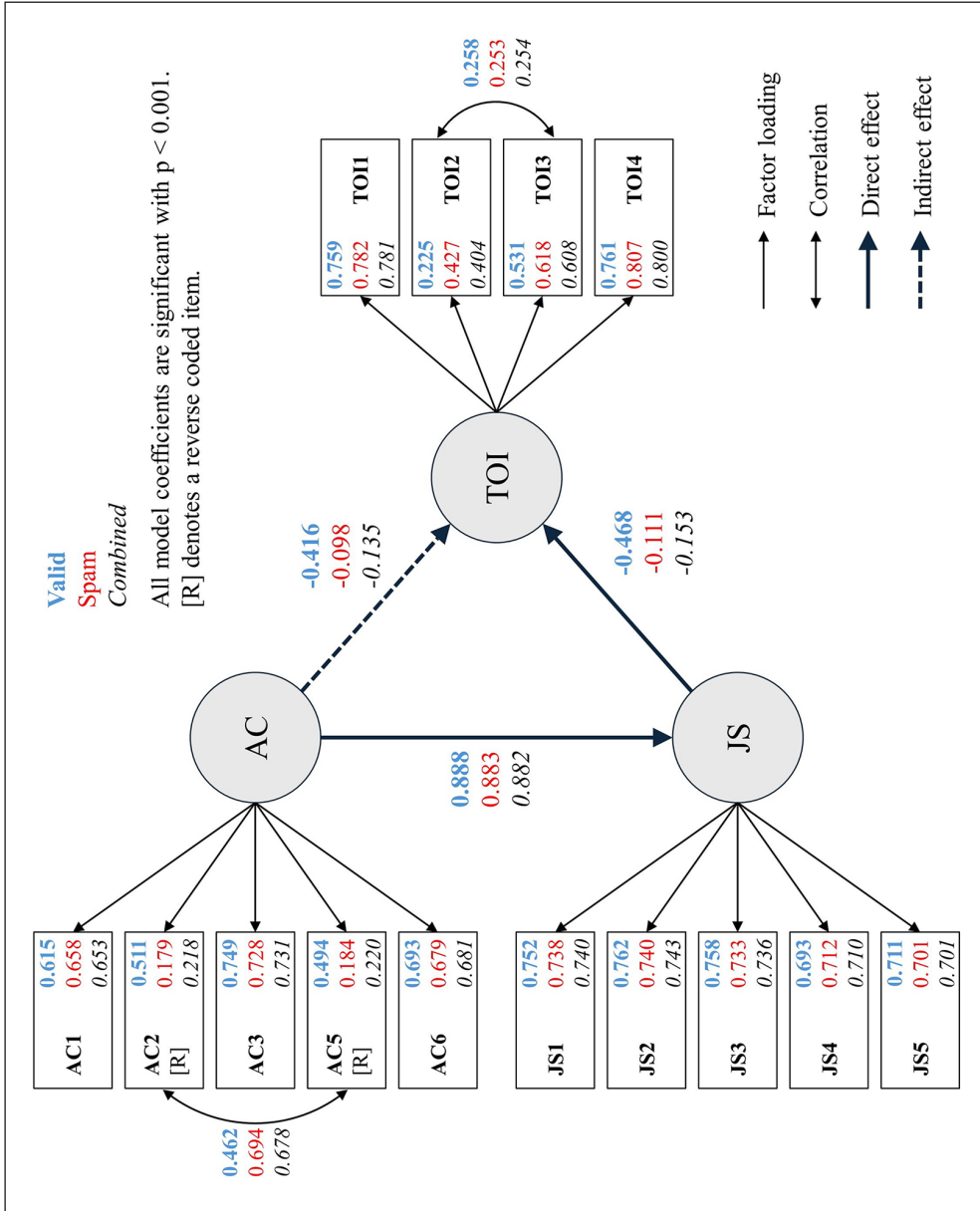


Figure 3. Model results for the hypothesized relationships across the three subsamples: Valid, Spam, and Combined.

**Table 3.** Summary of SEM results for the three subsamples.

Valid responses	Spam responses	All completed responses
<u>Model fit and respecification</u>		
Model fit is acceptable.	<u>Model fit is poor and not acceptable.</u>	<u>Model fit is poor and not acceptable.</u>
Model respecification is not required. However, TOI2 can be excluded.	Model respecification is required due to poor model fit. Respecify the factor structure and the relationships.	Model respecification is required due to poor model fit. Respecify the factor structure and the relationships.
<u>Factor structure<sup>a</sup></u>		
AC2 and AC5 have low factor loadings but are within the acceptable range.	Factor loadings of AC2 and AC5 are not within acceptable range.	Factor loadings of AC2 and AC5 are not within acceptable range.
Factor loading of TOI2 is not within acceptable range.	TOI2 has a low factor loading but is within the acceptable range.	TOI2 has a low factor loading but is within the acceptable range.
All other factor loadings are within the acceptable range.	All other factor loadings are within the acceptable range.	All other factor loadings are within the acceptable range.
<u>Covariances</u>		
Both covariances have the expected signs.	Both covariances have the expected signs.	Both covariances have the expected signs.
<u>Hypothesized relationships</u>		
H1 is supported.	H1 is supported.	H1 is supported.
H2 is supported.	H2 is partially supported. The relationship is weaker than expected.	H2 is partially supported. The relationship is weaker than expected.
H3 is supported.	H3 is partially supported. The relationship is weaker than expected.	H3 is partially supported. The relationship is weaker than expected.

<sup>a</sup>Factor loadings between 0.30 and 0.40 meet the minimum level of acceptance. Loadings  $\geq 0.50$  are acceptable.

through job satisfaction. Thus, including spam responses in the combined sample would underestimate the influence of affective commitment and job satisfaction on turnover intention in the transportation industry, leading researchers and practitioners to overlook them for transportation workers.

Though our study offers important implications into promoting the responsible use of research funds and preserving the scientific integrity of findings, it has limitations. Without the ground truth, we could not evaluate the spam detection techniques for false positives (valid response flagged as spam) and false negatives (spam response flagged as valid). Although such errors likely impact both research costs and financial costs, the practical constraints of non-probability sampling made this evaluation infeasible. Some studies have evaluated their spam detection strategies (Zhang et al., 2022); however, they did so by making strong assumptions that specific techniques (e.g. manually checking open-ended text responses) yield accurate results. More precise estimates of the human costs of identifying and resolving spam issues are also needed. In addition, we were

unable to benchmark our fraud rate in the context of the literature as they vary widely (1.2%–99%) across studies (Ballard et al., 2019; Bell et al., 2020; Kennedy et al., 2021; Pozzar et al., 2020; Thomas and Clifford, 2017), likely due to inconsistencies in how fraud was defined (e.g. different levels of rigor in spam detection indicators), recruitment strategies used (e.g. social media, crowdsourcing platforms, and opt-in panels), and other factors (e.g. whether the study was targeted by spammers or not). Next, our study was focused on specific industries with hard-to-recruit participants. More research examining a broader range of participants is needed. In addition, identifying the most effective spam detection techniques across recruitment and sampling strategies could help researchers with limited knowledge and expertise in this area. Finally, although some studies compared spam incidence across different recruitment sources (e.g. crowdsourcing vs social media; Antoun et al., 2016) and participation incentive amounts (Simone et al., 2024), we were unable to do so due to a lack of tracking data.

## Conclusion

The aim of this study is not to deter researchers from using online surveys with non-probability sampling methods but to highlight the importance of being cognizant of survey spammers and employing multiple survey spam prevention and detection techniques. A few studies have incorporated in-survey checks as part of their survey bot detection and prevention spam strategies (Goodrich et al., 2023; Moss et al., 2021), while some used in-survey checks—and them alone—to ensure their data does not contain fraudulent responses (Belliveau and Yakovenko, 2022; Kennedy et al., 2021). However, in-survey checks can potentially deter or even elicit noncompliant behavior from genuine participants, especially when an explanation for the check (e.g. “bot test”) is not included within the question text (Silber et al., 2022). Some studies have also suggested using statistical methods for data screening (Arthur et al., 2021; DeSimone et al., 2015; Dupuis et al., 2019); however, such techniques that rely on underlying data distributions might not work with a high proportion of spam responses in the data. Hence, as the more recent literature on survey spam detection suggests, a combination of common data quality checks and preventative measures serves as the best approach for detecting survey spam, as no single technique appears to capture all types of online survey fraud (Griffin et al., 2022; Lawlor et al., 2021; Simone et al., 2024; Storozuk et al., 2020; Zhang et al., 2022). Spam behavior has also been found to differ depending on the recruitment strategy (Kennedy et al., 2021), yet extant work on online survey spam has primarily been in data collected via opt-in panels or crowdsourced platforms such as Amazon Mechanical Turk (Chandler et al., 2020; Kennedy et al., 2021) rather than the increasingly widely used social media-based recruiting approach. Advertising online surveys through social media platforms may be more attractive and accessible to researchers because it provides a more cost-effective recruitment option compared to opt-in panels or crowdsourced platforms (Zindel, 2022), as researchers avoid paying platform fees. In conclusion, responsible science is reliant on valid data. Hence, it is imperative that researchers be cognizant of how spammers compromise online survey data, which can not only lead to unintentionally misused research funds but also lead to misleading or inaccurate results and conclusions, thereby reducing scientific integrity.


## ORCID iDs

Shubham Agrawal  <https://orcid.org/0000-0002-0990-9785>


Gwendolyn Paige Watson  <https://orcid.org/0000-0002-8838-4277>

Amy M. Schuster  <https://orcid.org/0000-0002-4472-0283>

Shelia R. Cotten  <https://orcid.org/0000-0002-8657-8262>

Michael L. Tidwell  <https://orcid.org/0000-0002-4398-1975>

Nathan Baker  <https://orcid.org/0000-0001-5860-7436>

Sicheng Wang  <https://orcid.org/0000-0001-6395-6235>

## Ethical considerations

In accordance with federal regulations 45 CFR 46.104(d), this study was qualified as Exempt under category 2 by the Office of Research Compliance at Clemson University (#IRB2020-183: “FW-HTF-RL: Preparing the Future Workforce for the Era of Automated Vehicles”; Approval date: July 6, 2022; PI: Shelia R. Cotten).

## Consent to participate

Respondents provided informed consent virtually at the beginning of the survey by choosing to continue with the survey. The survey informed consent page included the following statement: “To receive compensation, you must meet the eligibility criteria as determined by screening questions in the beginning of the survey, answer questions to check respondent’s attention correctly, and complete the entire survey. Any suspicious or low quality survey responses will not be compensated.”

## Author contributions

Shubham Agrawal: Conceptualization, Methodology, Investigation, Software, Formal analysis, Data curation, Visualization, Writing—original draft, Writing—review & editing. Gwendolyn Paige Watson: Methodology, Investigation, Data curation, Writing—original draft, Writing—review & editing. Amy M. Schuster: Investigation, Visualization, Writing—review & editing. Shelia R. Cotten: Conceptualization, Investigation, Funding acquisition, Writing—review & editing. Michael L. Tidwell: Investigation, Data curation, Writing—original draft. Nathan Baker: Investigation, Data curation, Writing—original draft. Sicheng Wang: Investigation, Writing—original draft.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Foundation under the grant “Preparing the Future Workforce for the Era of Automated VEHICLES (WEAVE)” [grant number #2041215]. Any opinions or findings expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Data availability statement

Data supporting the findings of this study are available upon request from the authors.

## Supplemental material

Supplemental material for this article is available online.

## References

Abbey JD and Meloy MG (2017) Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management* 53-56(1): 63–70.

- American Sociological Association (2018) Code of ethics. Available at: [https://www.asanet.org/wp-content/uploads/asa\\_code\\_of\\_ethics-june2018a.pdf](https://www.asanet.org/wp-content/uploads/asa_code_of_ethics-june2018a.pdf)
- Antoun C, Zhang C, Conrad FG, et al. (2016) Comparisons of online recruitment strategies for convenience samples: Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk. *Field Methods* 28(3): 231–246.
- Arthur W, Hagen E and George F (2021) The lazy or dishonest respondent: Detection and prevention. *Annual Review of Organizational Psychology and Organizational Behavior* 8(1): 105–137.
- Ballard AM, Cardwell T and Young AM (2019) Fraud detection protocol for web-based research among men who have sex with men: Development and descriptive evaluation. *JMIR Public Health and Surveillance* 5(1): e12344.
- Battaglia MP (2008) *Encyclopedia of Survey Research Methods*. Sage Publications, Inc.
- Bell AM and Gift T (2023) Fraud in online surveys: Evidence from a nonprobability, subpopulation sample. *Journal of Experimental Political Science* 10(1): 148–153.
- Bell CJ, Spruit JL and Kavanaugh KL (2020) Exposing the risks of social media recruitment in adolescents and young adults with cancer: #beware. *Journal of Adolescent and Young Adult Oncology* 9(5): 601–607.
- Belliveau J and Yakovenko I (2022) Evaluating and improving the quality of survey data from panel and crowd-sourced samples: A practical guide for psychological research. *Experimental and Clinical Psychopharmacology* 30(4): 400–408.
- Bernerth JB, Aguinis H and Taylor EC (2021) Detecting false identities: A solution to improve web-based surveys and research on leadership and health/well-being. *Journal of Occupational Health Psychology* 26(6): 564–581.
- Buchanan EM and Scofield JE (2018) Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods* 50(6): 2586–2596.
- Bybee S, Cloyes K, Ellington L, et al. (2022) Bots and nots: Safeguarding online survey research with under-represented and diverse populations. *Psychology and Sexuality* 13(4): 901–911.
- Chandler J and Shapiro D (2016) Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology* 12(1): 53–81.
- Chandler J, Sisso I and Shapiro D (2020) Participant carelessness and fraud: Consequences for clinical research and potential solutions. *Journal of Abnormal Psychology* 129: 49–55.
- Chmielewski M and Kucker SC (2020) An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science* 11(4): 464–473.
- Cohen A (2022) *fuzzywuzzy: Fuzzy string matching in python* (Version 0.18.0) [Python]. Available at: <https://github.com/seatgeek/fuzzywuzzy>
- Cooper-Hakim A and Viswesvaran C (2005) The construct of work commitment: Testing an integrative framework. *Psychological Bulletin* 131(2): 241–259.
- Crano WD, Brewer MB and Lac A (eds) (2015) *Principles and Methods of Social Research*, 3rd edn. Routledge.
- Daikeler J, Bošnjak M and Lozar Manfreda K (2020) Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology* 8(3): 513–539.
- Das M, Ester P and Kaczmirek L (2018) *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*. Routledge.
- DeSimone JA, Harms PD and DeSimone AJ (2015) Best practice recommendations for data screening. *Journal of Organizational Behavior* 36(2): 171–181.
- Dillman DA, Smyth JD and Christian LM (2014) *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons, Inc.
- Dougherty MV (2021) The use of confidentiality and anonymity protections as a cover for fraudulent field-work data. *Research Ethics* 17(4): 480–500.
- Dupuis M, Meier E and Cuneo F (2019) Detecting computer-generated random responding in questionnaire-based data: A comparison of seven indices. *Behavior Research Methods* 51(5): 2228–2237.
- Goodrich B, Fenton M, Penn J, et al. (2023) Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys. *Applied Economic Perspectives and Policy* 45(2): 762–784.

- Griffin M, Martino RJ, LoSchiavo C, et al. (2022) Ensuring survey research data integrity in the era of internet bots. *Quality & Quantity* 56(4): 2841–2852.
- Hair JF (ed.) (2010) *Multivariate Data Analysis*, 7th edn. Pearson Prentice Hall.
- Hoskins K, Sebetka K, Thomas A, et al. (2025) Pragmatic and ethical challenges in navigating inauthentic participation in remote qualitative research. *Research Ethics*. Epub ahead of print 5 February 2025. DOI: 10.1177/17470161251316286
- Hu L and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling A Multidisciplinary Journal* 6(1): 1–55.
- Judge TA, Weiss HM, Kammeyer-Mueller JD, et al. (2017) Job attitudes, job satisfaction, and job affect: A century of continuity and of change. *Journal of Applied Psychology* 102(3): 356–374.
- Kennedy C, Hatley N, Lau A, et al. (2021) Strategies for detecting insincere respondents in online polling. *Public Opinion Quarterly* 85(4): 1050–1075.
- Kennedy R, Clifford S, Burleigh T, et al. (2020) The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods* 8(4): 614–629.
- Kline RB (ed.) (2016) *Principles and Practice of Structural Equation Modeling*, 4th edn. Guilford Press.
- Lauer M (2024) FY 2023 By the Numbers: Extramural Grant Investments in Research – NIH Extramural Nexus. Available at: <https://nexus.od.nih.gov/all/2024/02/21/fy-2023-by-the-numbers-extramural-grant-investments-in-research/>
- Lawlor J, Thomas C, Guhin AT, et al. (2021) Suspicious and fraudulent online survey participation: Introducing the REAL framework. *Methodological Innovations* 14(3): 20597991211050467.
- Lee TW, Hom PW, Eberly MB, et al. (2017) On the next decade of Research in voluntary employee turnover. *Academy of Management Perspectives* 31(3): 201–221.
- Levi R, Ridberg R, Akers M, et al. (2022) Survey fraud and the integrity of web-based survey research. *American Journal of Health Promotion* 36(1): 18–20.
- Li S, Liu H, Dong T, et al. (2021) Hidden backdoors in human-centric language models. In: *Proceedings of the 2021 ACM SIGSAC conference on computer and communications security*, pp.3123–3140. DOI: 10.1145/3460120.3484576
- Manfreda KL, Bosnjak M, Berzelak J, et al. (2008) Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research* 50(1): 79–104.
- Meade AW and Craig SB (2012) Identifying careless responses in survey data. *Psychological Methods* 17(3): 437–455.
- Meyer JP, Allen NJ and Smith CA (1993) Commitment to organizations and occupations: Extension and test of a three-component conceptualization. *Journal of Applied Psychology* 78: 538–551.
- Miner MH, Bockting WO, Romine RS, et al. (2012) Conducting internet research with the transgender population: Reaching broad samples and collecting valid data. *Social Science Computer Review* 30(2): 202–211.
- Morgeson FP and Humphrey SE (2006) The Work Design Questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of Applied Psychology* 91(6): 1321–1339.
- Moss AJ, Rosenzweig C, Jaffe SN, et al. (2021) *Bots or inattentive humans? Identifying sources of low-quality data in online platforms*. PsyArXiv. DOI: 10.31234/osf.io/wr8ds
- Nichols AL and Edlund JE (2020) Why don't we care more about carelessness? Understanding the causes and consequences of careless participants. *International Journal of Social Research Methodology* 23(6): 625–638.
- NSF (2024) National Science Foundation: FY 2023 performance and financial highlights. Available at: <https://www.nsf.gov/pubs/2024/nsf24003/nsf24003.pdf>
- Özkan AH, Elci M, Erdilek Karabay M, et al. (2020) Antecedents of turnover intention: A meta-analysis study in the United States. *E+M Ekonomie a Management* 23(1): 93–110.
- Panicker A, Nurain N, Ibrahim Z, et al. (2024) Understanding fraudulence in online qualitative studies: From the researcher's perspective. In: *Proceedings of the 2024 CHI conference on human factors in computing systems*, pp.1–17. DOI: 10.1145/3613904.3642732
- Park T-Y and Shaw JD (2013) Turnover rates and organizational performance: A meta-analysis. *Journal of Applied Psychology* 98(2): 268–309.

- Pozzar R, Hammer MJ, Underhill-Blazey M, et al. (2020) Threats of bots and other bad actors to data quality following research participant recruitment through social media: Cross-sectional questionnaire. *Journal of Medical Internet Research* 22(10): e23021.
- Pratt-Chapman M, Moses J and Arem H (2021) Strategies for the identification and prevention of survey fraud: Data analysis of a web-based survey. *JMIR Cancer* 7(3): e30730.
- Qualtrics (2023) Fraud detection. Available at: <https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/>
- Rosseel Y (2012) Lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48(2): 1–36. DOI: 10.18637/jss.v048.i02
- Ruby E, Ramlawi S, Bowie AC, et al. (2025) Identifying fraudulent responses in a study exploring delivery options for pregnancies impacted by gestational diabetes: Lessons learned from a web-based survey. *Journal of Medical Internet Research* 27(1): e58450.
- Sawhney G, Bijlani A and DeSimone J (2025) *Are Existing Bot-Detection Techniques Sufficient: An Exploration With Real Bots*. OSF.
- Shaw JD (2011) Turnover rates and organizational performance: Review, critique, and research agenda. *Organizational Psychology Review* 1(3): 187–213.
- Silber H, Roßmann J and Gummer T (2022) The issue of noncompliance in attention check questions: False positives in instructed response items. *Field Methods* 34(4): 346–360.
- Simone M (2019) *How to battle the bots wrecking your online study*. Behavioral Scientist. Available at: <https://behavioralscientist.org/how-to-battle-the-bots-wrecking-your-online-study/>
- Simone M, Cascalheira CJ and Pierce BG (2024) A quasi-experimental study examining the efficacy of multimodal bot screening tools and recommendations to preserve data integrity in online psychological research. *American Psychologist* 79(7): 956–969.
- Storozuk A, Ashley M, Delage V, et al. (2020) Got bots? Practical recommendations to protect online survey data from bot attacks. *Quantitative Methods for Psychology* 16: 472–481.
- Tett RP and Meyer JP (1993) Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. *Personnel Psychology* 46(2): 259–293.
- Thomas KA and Clifford S (2017) Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior* 77: 184–197.
- Vehovar V, Toepoel V and Steinmetz S (2016) *The Sage Handbook of Survey Methodology*. Sage Publications Ltd.
- Venta A, Bailey CA, Walker J, et al. (2022) Reverse-coded items do not work in Spanish: Data from four samples using established measures. *Frontiers in Psychology*. Epub ahead of print 23 June 2022. DOI: 10.3389/fpsyg.2022.828037
- Webb MA and Tangney JP (2024) Too good to be true: Bots and bad data from Mechanical Turk. *Perspectives on Psychological Science* 19: 887–890.
- Weijters B and Baumgartner H (2012) Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research* 49(5): 737–747.
- Xu X, Zhao P, Hayes R, et al. (2023) Revisit the causal inference between organizational commitment and job satisfaction: A meta-analysis disentangling its sources of inconsistencies. *Journal of Applied Psychology* 108: 1244–1261.
- Xu Y, Pace S, Kim J, et al. (2022) Threats to online surveys: Recognizing, detecting, and preventing survey bots. *Social Work Research* 46(4): 343–350.
- Zhang Z, Zhu S, Mink J, et al. (2022) Beyond bot detection: Combating fraudulent online survey takers. In: *Proceedings of the ACM web conference 2022*, pp.699–709. DOI: 10.1145/3485447.3512230
- Zindel Z (2022) Social media recruitment in online survey research: A systematic literature review. *Methods, Data, Analyses* 17(2): 207–248.