National Disability Center
for Student Success

Survey Under Siege:
Exploring AI, Bot, and Human Interference in Online Surveys

Table of Contents

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

Authors

Ryan Mata, PhD

Stephanie W. Cawthon, PhD

Desirée Lama, MA

Funder Acknowledgement

This research was conducted by the National Disability Center for Student Success and was supported by funding from the Institute for Education Sciences (IES), U.S. Department of Education, under grant number R324C230008. The content and opinions expressed are those of the authors and do not necessarily represent the policy of or endorsement by the IES, the U.S. Department of Education, or the federal government.

Suggested Citation

National Disability Center for Student Success. (2025). Survey Under Siege: Exploring AI, Bot, and Human Interference in Online Surveys. Institute for Education Sciences, U.S. Department of Education. nationaldisabilitycenter.org

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

Objectives

This brief describes and explores the impact of fraudulent participation in the initial online distribution of the Campus Accessibility Measure. Our goal is to make transparent the challenges of online survey administration amid increasing use of artificial intelligence (AI) technology that can easily navigate and generate human-like responses to survey items, including a commonly used security integration, reCAPTCHA.

Suspicious data can now flow into surveys from both human and non-human respondents, warranting a multistep approach to data screening. These complicating factors of survey research have significant implications for those focused on community engaged research and online stakeholder relationship-building.

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

Steps to Understanding the Impact of Bots

There are four essential steps researchers can take to detect and address suspicious responses in online surveys. Drawing from our own experience, each step reflects a practical strategy for maintaining data quality in the face of bot, AI, and human interference.

Identify

Detect back-end indicators of suspected bot behavior in online surveys.

Evaluate

Assess the effectiveness of Google's "reCAPTCHA," a non-human screening extension to online programs which can be implemented to mitigate bot intrusions during survey data collection.

Analyze

Review the data cleaning process and the proportion of suspicious responses flagged using each identification strategy.

Discuss

Talk through the decision-making process for managing survey data and determining next steps for future surveys.

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

Surveys in Community Engaged Research

This brief is grounded in the perspective of community engaged research on disability in higher education. The National Disability Center for Student Success (Center) prioritizes engagement with audiences such as scholars, instructors, parents, and students to help build more than a network — it builds a thriving community with a shared vision, common goals, and focused purpose to ideate, execute, apply, and learn from this research. The Center also iteratively engages and builds partnerships with current and potential stakeholders.

Community Engagement is critical to achieving the first aim of the center:

> **"To provide a robust and comprehensive research foundation for future design of interventions to support disabled students in higher education."**

Community engagement involves both the leveraging and expanding of networks when conducting research. As a research center but also as a chance for advocacy and awareness raising, the survey discussed in the current study thus needed to reach beyond existing networks. The Center was also in its launch stages, so the survey was an introduction of its work to a larger, public audience. The center's effort to launch the survey in spring 2024 included a team of researchers, student fellows, and communications specialists.

For all its benefits, community engaged research also carries some risks, especially when relying on online survey recruitment and data collection tools susceptible to fraudulent responses, both human and non-human. Our efforts to open this survey to the community for participation was disrupted by an influx of suspected fraudulent data. This was evidenced by suspicious, patterned activity that became the basis of a systematic data-cleaning process and exploration of activity on the survey. The current analysis reconciles the needs of community engaged research with the challenges of maintaining data integrity within the context of rapidly advancing AI technology and imperfect security measures. By illustrating in detail our experience of fraud detection and data cleaning, this brief contributes to our understanding of data validity and reliability in the digital age.

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

Study Context

This analysis is a secondary, "post-mortem" quantitative exploration of suspected fraudulent or bot activity on the Campus Accessibility Measure online survey administered by the Center in the spring of 2024.

An initial influx of suspected fraudulent responses occurred following the start of the recruiting campaign on social media platform X (formerly Twitter), resulting in increased security measures implemented by the researchers. Center researchers used a "CAPTCHA" (Completely Automated Public Turing Test to tell Computers and Humans Apart) tool capable of discerning non-human responses and traffic on websites. Qualtrics allows for survey integration with Google's free-use CAPTCHA tool, "reCAPTCHA." Qualtrics offers both Versions 2 and 3 of reCAPTCHA. While Version 2 relies on user interaction with logic puzzles with audio and visual components and can boot suspicious respondents from the survey, Version 3 allows all respondents to progress through the items and collects user metadata. It then reports a score from zero to one based on the probability that the user is human.

The decision to implement Version 3 (the passive model) over Version 2 came after some consideration with accessibility issues posed by the requirements of Version 2's logic puzzle interactions (Hollier et al., 2021). Thus, the tradeoff was to implement Version 3 for greater accessibility for disabled respondents in exchange for having potentially more fraudulent responses to screen through. To address this, the researchers implemented screening criteria beyond reCAPTCHA to flag potential fraudulent responses such as:

- **Flagging suspicious email addresses.**

- **Analyzing open-ended responses for gibberish, irrelevant, or Latin responses (which were commonly used by bots).**

- **Identification of "rapid repeat" or "ballot stuffing" instances where multiple fraudulent responses are recorded in a matter of minutes from similar locations.**

- **Monitoring consistency of answers across different survey portions.**

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

Data Cleaning and Analysis

The data cleaning and analysis components of the current study are combined processes of flagging, describing, and visualizing aspects of fraudulent submissions to the survey. These stages are outlined as follows:

Analyzing reCAPTCHA scores, including making comparisons between groups of flagged and non-flagged responses using descriptive statistics and unequal variance t-tests to determine the efficacy of reCAPTCHA as a diagnostic tool.

Employing multiple identification strategies to flag remaining suspected fraudulent responses.

We conducted quantitative analyses on the total data set to determine the proportion of fraudulent responses. Further, we examined the impact of our flagging criteria on the overall number of flagged responses to understand which strategies were most useful in detecting fraudulent data.

This study uses data (n = 2,286) from a 38-item survey that was designed and administered online via Qualtrics in the spring (March-April) of 2024 by the Center. The survey was 65 items with a mix (insert numbers) of select all-that-apply, matrix (Likert), and open-ended questions National Companion Report. The survey took approximately 8 minutes to complete and was piloted with a sample of 500 undergraduates in fall 2023 and five cognitive lab interviews in winter 2023/2024 as part of the measure development process. Eligible students were entered into a raffle with a chance to win a $50 online gift card.

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

The online survey yielded the following types of data:

Survey items: This is the primary source of data focusing on demographics, perceived accessibility, and open-ended responses about individual experiences of disability and life as a college student.

Survey metadata: Another type of data useful in the bot-detection process is some of the various metadata collected by Qualtrics and enabled as viewable by researchers as a "Survey Option." This metadata includes reCAPTCHA scores, email addresses, and geolocation data.

Both types of data were exportable to a collated spreadsheet in Microsoft Excel for preliminary steps of data cleaning. For descriptive statistics, group comparisons, and data visualizations, the researchers used R Studio (Version 2023.12.1+402).

Two-Stage Review Process

Test emails for bounce back

An initial group of 50 participants were emailed to see how many messages "bounced back," indicating the use of a fake email address. Of the 50 participants selected, 46 email addresses sent bounce-back messages to the researchers.

Response checks

A. Open response questions for use of gibberish/fake/Latin language **(299 flagged responses).**

B. "Rapid repeat" or "ballot stuffing" attempts **(994 flagged responses).**

C. Inconsistencies in responses between survey sections **(528 flagged responses).**

Some responses were flagged under multiple criteria. Overall, this combination of strategies produced more effective culling of unwanted responses than any single strategy alone, although checking for ballot stuffing instances screened out most of the suspicious responses.

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

Results

This study presents one case of a national research center's struggle with maintaining data integrity during an online survey recruitment campaign. This process involved a reconciliation of the convenience and accessibility of online sampling with the needs of community engaged research and the potential risks posed by fraudulent, bot, and/or AI-driven responses.

Although the researchers implemented Google reCAPTCHA (Version 3) its value as a screening tool needed to be explored due to the range of observed reCAPTCHA scores among responses that had otherwise been flagged as suspicious. Visual inspection of reCAPTCHA score densities shows a high level of discrimination and very few "grey area" scores around the 0.5 threshold **(See Figure 2)**.

The mean and median reCAPTCHA scores for flagged responses were 0.41 and 0.3, respectively, falling below Google's recommended threshold of 0.5 for evaluating reCAPTCHA scores and indicating probable non-human responses. Meanwhile, for responses that were not flagged in the data cleaning process, the mean and median reCAPTCHA scores were 0.62 and 0.8, respectively **(See Figure 1 and Figure 2)**.

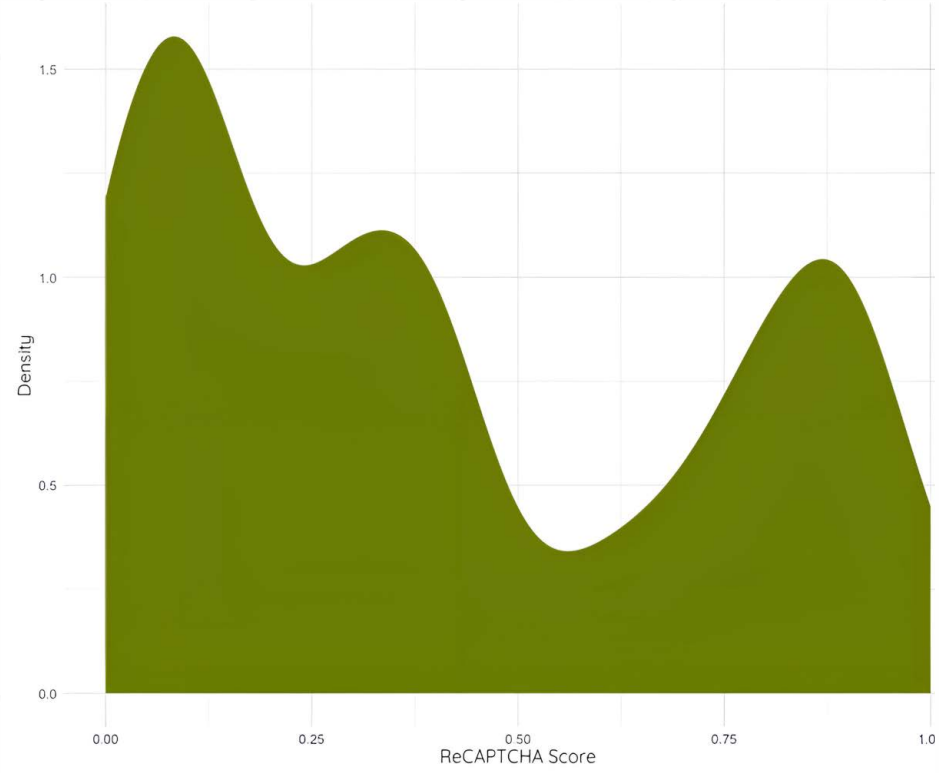
The researchers flagged a total of 1,622 responses as potentially fraudulent, nearly 71% of the total sample, after application of the screening criteria.

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

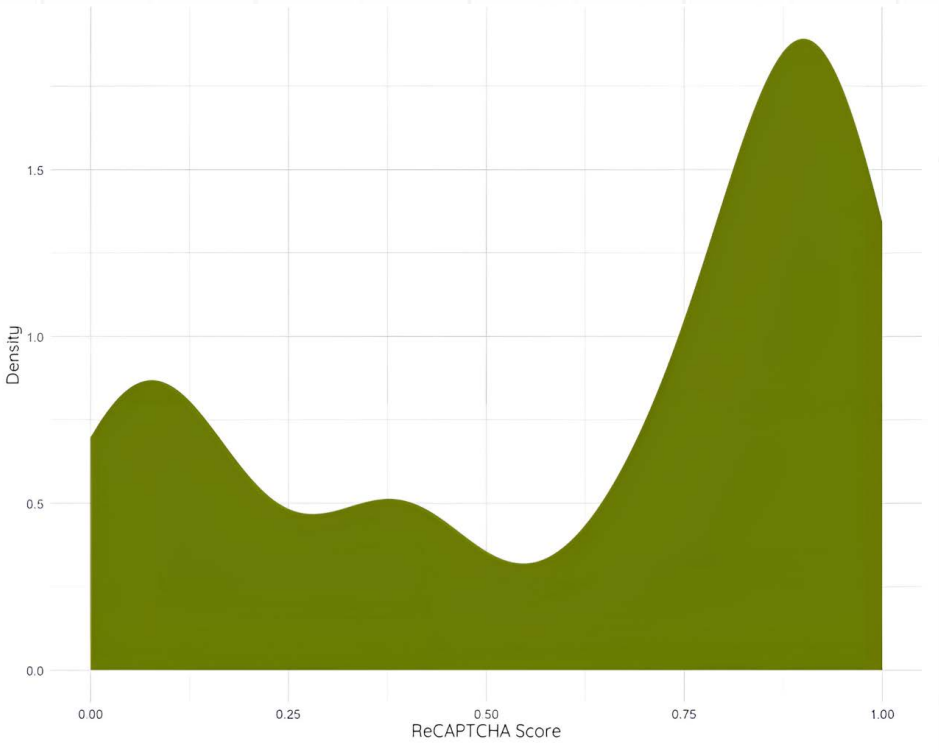
Figure 1

ReCAPTCHA Score Distributions of Flagged and Non-Flagged Responses

ReCAPTCHA Scores - Flagged Responses Only



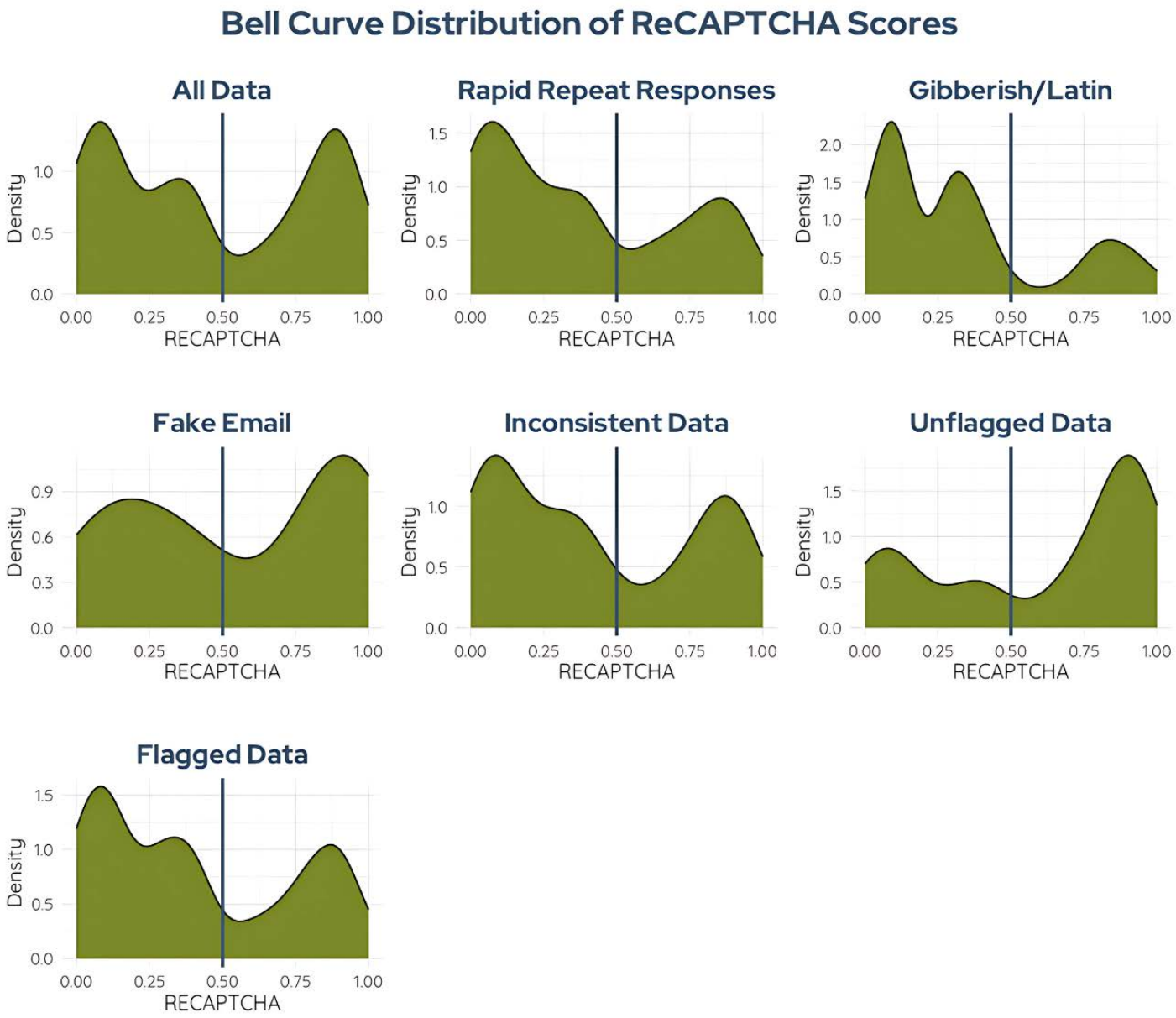
ReCAPTCHA Scores - Non-Flagged Responses Only



© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

Figure 2

ReCAPTCHA Score Distributions of Flagged and Non-Flagged Subsamples with Recommended 0.5-Score Threshold



Mean differences between flagged and non-flagged responses were significant after unequal-variances t-tests, $t(1109.5) = 12.92$, $p < .001$, 95% CI [0.179, 0.405]. However, the potential to reduce false negatives coupled with the possibility of human fraudulent responses (that can pass the reCAPTCHA) led to researchers endorsing a multi-step screening approach.

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

Conclusion

This brief presents some of the challenges and solutions related to maintaining survey data integrity while leveraging technology tools and social media platforms for recruitment, data collection, network-building, and community engagement. It illustrates a unique challenge of maintaining community-facing research opportunities within a research center's overall communications and stakeholder relationship plan.

Key contributions begin with the researchers' example of screening/flagging strategies by implementing a multi-step approach to examining both quantitative and qualitative survey data. These strategies can be taken by survey researchers as practical recommendations with an overall theme of implementing robust security measures and adaptive strategies that help balance survey accessibility with data integrity. As shown by the current study, such a balance is necessary given the shortcomings observed in reCAPTCHA's ability to flag (or not flag) suspicious respondents to the survey, echoing previous research (Bonett et al., 2024).

Addressing the challenge of fraudulent, AI-driven, and/or bot responses is something researchers must embrace. In the digital age, the question should not be, "Will my survey yield fake responses?" but rather, "What can we do to systematically account for suspicious data?" Given that in our field, online recruitment and data collection are increasingly common, this simple exploration of one research center's survey can inform future researchers on best practices for data integrity.

References

1. Bonett, S., Lin, W., Sexton Topper, P., Wolfe, J., Golinkoff, J., Deshpande, A., Villarruel, A., & Bauermeister, J. (2024). Assessing and Improving Data Integrity in Web-Based Surveys: Comparison of Fraud Detection Systems in a COVID-19 Study. JMIR formative research, 8, e47091. Assessing and Improving Data Integrity in Web-Based Surveys

2. Hollier, S. Sajka, J., White, J., & Cooper, M. (2021, December 16). Inaccessibility of CAPTCHA: Alternatives to Visual Turing Tests on the Web. W3C. Turing Tests

© 2025 National Disability Center for Student Success | nationaldisabilitycenter.org

National Disability Center
for Student Success

nationaldisabilitycenter.org

Suggested Citation

National Disability Center for Student Success. (2025). Survey Under Siege: Exploring AI, Bot, and Human Interference in Online Surveys. Institute for Education Sciences, U.S. Department of Education. nationaldisabilitycenter.org