



# Erratum to Peer et al. (2021) Data quality of platforms and panels for online behavioral research

Eyal Peer<sup>1</sup> · David Rothschild<sup>2</sup> · Andrew Gordon<sup>3</sup> · Ekaterina Damer<sup>3</sup>

Accepted: 16 June 2022 / Published online: 6 July 2022  
© The Psychonomic Society, Inc. 2022

## Abstract

This erratum reports on a technical error that was discovered in Study 2 of Peer et al. (2021). Because of this technical error, some specific findings on participants' proclivity for dishonesty reported in the paper have been found incorrect. We detail the error, which only affected female participants, and its impact on the findings and report on the reanalyzed findings accounting for the error. The new findings do not change the conclusions provided in the paper, and show again that participants from MTurk are more likely to engage in dishonest behavior than participants from Prolific or CloudResearch.

**Keywords** Online research · Data quality · Erratum

## The original finding

In Study 2 of Peer et al. (2021), we used an “imposter” question to measure participants' dishonest behavior between samples from MTurk, CloudResearch (CR) and Prolific. As detailed in the paper: “The ‘imposter’ question, which came at the very end of the study, asked participants whether they would like to be invited to a study in the future. We told participants that the future study investigates a specific subpopulation of people, and thus it offers higher pay than usual (‘up to 15 USD per hour’). Participants were told that the study is open to participants who are male/female (according to the gender of the participant, which they indicated at the beginning of the study) and are at a given age range, which was programmed to be from 5 to 9 years older than the age participants reported in the beginning of the study (e.g., for a person who said they were 30 years old, the age range for recruitment was 35–44). Participants could choose to say that they (1) fit the criteria

and wanted to take part, (2) fit the criteria but did not want to take part, (3) did not fit the criteria, or (4) other. Responses of 1 were coded as dishonestly claiming false eligibility.”

We reported the following results for this question in Study 2: “In the ‘imposter’ question, 60% of MTurk participants claimed false eligibility, compared to 55% on CR and 48% on Prolific,  $\chi^2(2) = 12.69$ ,  $p < .01$ . The specific difference between CR and Prolific was also significant,  $\chi^2(1) = 4.06$ ,  $p = 0.04$ , showing that Prolific participants were more honest than CR and MTurk.”

## The technical error

A researcher<sup>1</sup> who analyzed the data of Study 2 using the files published on the Open Science Framework (OSF) noticed that in the survey comments, some participants reported that the text of the question included the term “{Invalid Expression}” where the age range should have been. That researcher further identified that this error occurred only among females. This was because the imposter question was programmed using two items on the Qualtrics survey, one for males and one for females (to display for each participant according to their gender; “other” was joined with “female” for this question, and “prefer not to answer” was joined with “male”). We were able to reproduce the error using the survey files (which were also published on OSF).

<sup>1</sup> We thank Tory Taylor for bringing this issue to our attention.

Behavior Research Methods, <https://doi.org/10.3758/s13428-021-01694-3>

✉ Eyal Peer  
eyal.peer@mail.huji.ac.il

<sup>1</sup> School of Public Policy, The Hebrew University of Jerusalem, Jerusalem, Israel

<sup>2</sup> Microsoft Research, New York City, NY, USA

<sup>3</sup> Prolific Inc, Wareham, UK

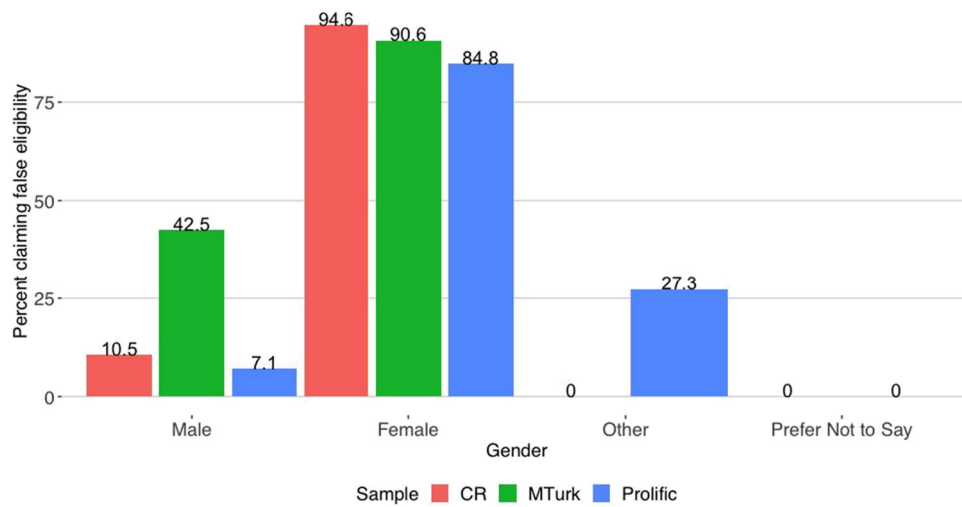


Fig. 1 Percent claiming false eligibility by gender and sample

### Reanalysis arrives at same conclusion

We thus examined whether the responses to the “imposter”

question differed between males and females, and discovered that they did. As can be seen in Fig. 1, most female participants claimed eligibility for the study, which is justifiable if

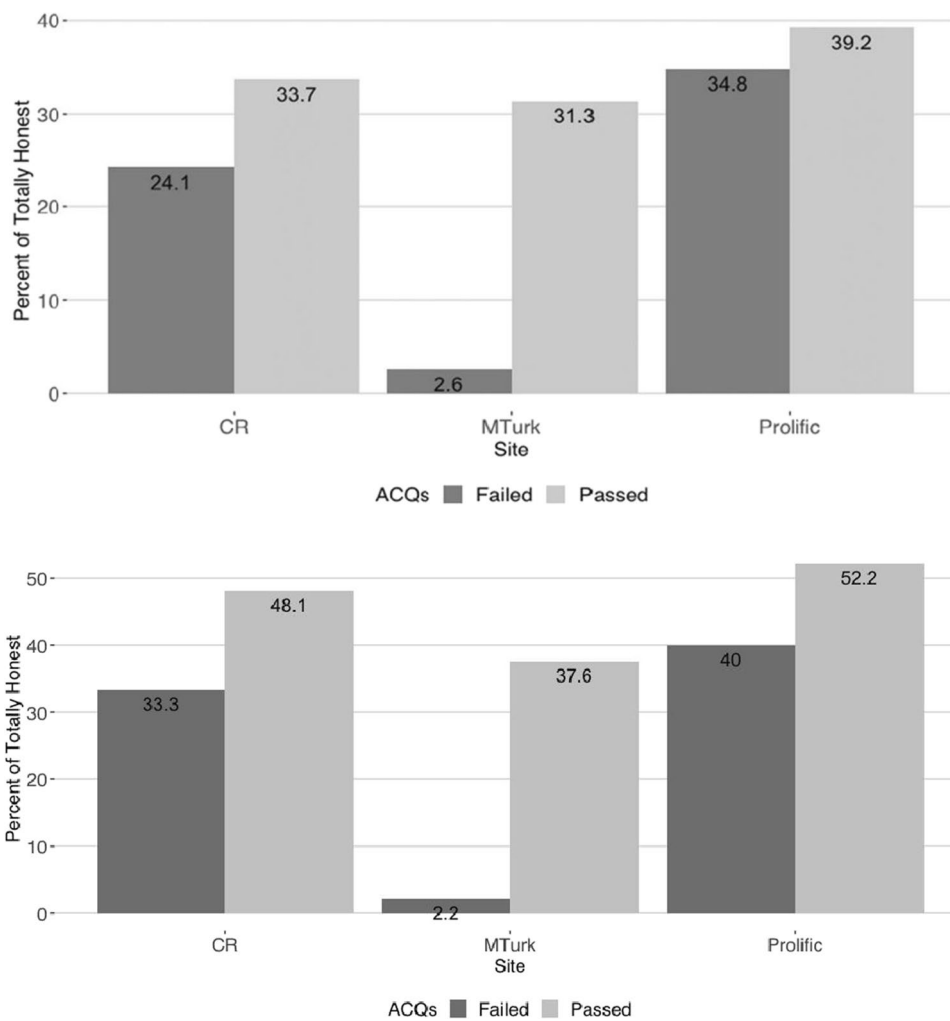


Fig. 2 Percent totally honest, with original Fig. 9 on top and updated Fig. 9 on bottom

indeed the age range did not appear for them. For males, in contrast, we found lower rates of “imposters” compared to those reported in the paper. However, the differences were in the same direction as reported in the paper, and larger. Among MTurk participants, 42.5% claimed false eligibility, compared to only 10.5% among CR and 7.1% among Prolific. These differences were statistically significant,  $\chi^2(2) = 119.33$ ,  $p < .001$ . The specific difference between CR and Prolific was not significant,  $\chi^2(1) = 1.64$ ,  $p = 0.20$ . Thus, we found that dishonesty rates were higher among MTurk participants compared to CR or Prolific participants, as we originally concluded in the paper as well.

Knowing this, we reanalyzed the results used for Fig. 9 in the original paper (in Fig. 2 of this erratum), which showed the rates of “fully honest” participants (those who did not claim false eligibility and also did not report in the matrix task any unsolvable problem as solved). Similar to the original findings, in this updated Fig. 9 including only male participants, we found that attention (passing both attention-check questions [ACQs]) also predicted honest behavior, as the percentage of participants who neither cheated on either of the two unsolvable problems nor claimed false eligibility for a future study was higher if they passed ACQs versus failing them (46% vs. 14%),  $\chi^2(1) = 70.5$ ,  $p < .001$ . The difference was most

pronounced among MTurk users, where only 2.2% of those who failed ACQs were fully honest.

## Conclusion

As reported in the original paper, we find that after reanalyzing the “imposter” question only among the participants that did not experience the technical error (males), the original conclusion, that higher rates of dishonesty are found on MTurk versus CR or Prolific, remains valid. The conclusion might only apply to the male participants in the study if females were to show a reverse pattern—which is that female participants on MTurk would cheat less than female participants on CR or Prolific. Because we did not find a difference between genders in the other measure of dishonest behavior (reporting unsolvable problems), we hold that that result is unlikely. Nevertheless, future studies may choose to replicate or explore this question further.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.