

Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA

Participant-visible method context. Results, discussion, captions, and the target figure are excluded.

Participants We recruited adult participants ($N = 2729$) from the United States through MTurk ($n = 500$), CloudResearch ($n = 505$), Prolific ($n = 496$), Qualtrics ($n = 575$), and an undergraduate student sample overseen by our department, SONA ($n = 653$). Because participants self-selected into each of the five data collection platforms, assignment to the five groups was not random. For all paid platforms (i.e., MTurk, CloudResearch, Prolific, and Qualtrics) only participants who submitted the survey received payment. As such, the attrition rate from these platforms was 0%. SONA participants had a higher attrition rate with 17.92% of participants (117 participants) providing only a partial response. We removed participants who progressed through

PLOS ONE | <https://doi.org/10.1371/journal.pone.0279720> March 14, 2023

5 / 17

PLOS ONE

Data quality in online human-subjects research

less than 30% of the survey. A total of 555 participants recruited through SONA were included in our analyses. See Table 1 for a demographic breakdown of each of the five samples. Participants were paid \$0.96 on MTurk, CloudResearch, and Prolific. The payment rate was based on the US federal minimum wage (\$7.50 per hour). Payment information was not available from Qualtrics because Qualtrics does not allow researchers to determine participants' compensation. Note that the aforementioned dollar amounts refer to the compensation that participants actually received. These amounts are different from the costs for the researchers because each platform charges additional fees (i.e., a 20% fee on MTurk, a combination of a 20% fee plus additional charges on CloudResearch, a 30% fee on Prolific, and a flat rate based on survey length and additional restrictions on Qualtrics). Participants recruited through SONA received course credit for their participation. For all platforms that permitted the option (i.e., MTurk, CloudResearch, and Prolific), we recruited participants who had completed a minimum of 100 surveys in the past and had an approval rating of 95% or higher as is standard practice for online data collection in the social sciences [20]. It was not possible to include eligibility requirements based on either the number of past surveys taken or approval rate on Qualtrics or SONA. We did not use demographic quotas (i.e., quotas to ensure the same demographic distributions on all platforms) while recruiting participants.

Platforms Participants were recruited through five commonly used data collection platforms, MTurk, CloudResearch, Prolific, Qualtrics, and SONA. See S1 Appendix for information about how we determined which platforms to use in our study. As each platform operates slightly differently, we describe below how the platforms recruit and compensate participants. MTurk recruits participants for a given study from its own pool of potential workers. When a researcher posts a survey (also called a "hit") it appears on the dashboard of any worker who qualifies for the study. Workers can then opt to take the survey and are compensated with Amazon credit upon completion. CloudResearch includes multiple options for data collection. We included CloudResearch's interface for MTurk (the MTurk Toolkit) in the study. The MTurk Toolkit should not be confused with CloudResearch's other data collection options (i.e., Prime Panels and a managed research option) which we did not examine in our research. The MTurk Toolkit posts surveys to MTurk through CloudResearch's interface and applies various filters designed to only include respondents who will provide high-quality data. The data quality filters used in our study included CloudResearch's approved participants list, a duplicate IP address block, a suspicious geo location block, and a worker country location verification. Participants recruited through CloudResearch's MTurk Toolkit are paid with Amazon credit. Prolific is designed similarly to MTurk and uses its own pool of participants. Prolific's participants are paid cash for completing surveys. Qualtrics uses a variety of recruitment methods including directly emailing participants. Researchers connect with a representative from Qualtrics who then posts the survey on the researcher's behalf (for example, in our study we requested that Qualtrics recruit 500 adults from the US). Qualtrics then distributes the survey on other data collection platforms, however which platforms is not specified to the researcher. Participants recruited through Qualtrics are compensated with various rewards including cash, gift cards, air miles, and coupons for food. These rewards are determined based on the participant's preference. Our

undergraduate sample was recruited through SONA. SONA is a platform where organizations (in our study the University of Wisconsin–Madison Psychology Department) can create their own pool of participants and can be designed to meet the needs of a department or

PLOS ONE | <https://doi.org/10.1371/journal.pone.0279720> March 14, 2023

6 / 17

PLOS ONE

Data quality in online human-subjects research

Table 1. Demographic information by platform. Measure

MTurk (N = 500)

CloudResearch (N = 505)

Prolific (N = 496)

Qualtrics (N = 575)

SONA (N = 555)

Age

38.75 (11.53)

41.99 (12.92)

37.23 (14.01)

64.34 (13.06)

18.53 (1.27)

Gender Female

36.20%

50.50%

67.54%

66.61%

62.34%

Male

63.40%

48.32%

30.44%

32.70%

29.01%

Another identity

0.00%

0.40%

1.41%

0.35%

0.36%

Prefer not to say

0.40%

0.79%

0.60%

0.17%

0.36%

No response

0.00%

0.00%

0.00%

0.17%

7.93%

Identify as Transgender Yes

7.00%

0.99%

2.42%

0.87%

0.72%

No

92.00%

98.22%

96.37%

98.43%

90.99%

I am unsure

0.00%

0.20%

0.20%

0.17%

0.36%

Prefer not to say

0.80%

0.59%

1.01%

0.52%

0.00%

No response

0.20%

0.00%

0.00%

0.00%

7.93%

American Indian or Alaskan Native

1.20%

0.99%

2.02%

0.70%

0.72%

Asian or Asian American

4.00%

9.31%

12.90%

0.70%

21.44%

Black or African American

15.00%

7.72%

9.88%

1.74%

3.78%

Latino, Hispanic, Chicano, or Puerto Rican

5.80%

4.95%

8.67%

2.43%

6.49%

Middle Eastern, Arab American, or North African

0.40%

0.79%

0.60%

0.35%

1.80%

Native Hawaiian or Pacific Islander

0.00%

0.59%

0.00%

0.00%

0.36%

White or European

75.40%

79.01%

72.38%

94.96%

67.93%

Another Identity

0.60%

1.19%

0.60%

0.35%

0.36%

Prefer not to say

1.00%

0.99%

0.40%

0.52%

0.36%

No response

0.00%

0.20%

0.00%

0.00%

8.11%

Ethnicity

Sexual Orientation Straight or Heterosexual

82.60%

89.70%

79.23%

92.70%

79.28%

Gay or Homosexual

1.80%

2.97%

3.23%

2.43%

1.98%

Bisexual

15%

4.55%

13.71%

2.78%

7.93%

Another identity

0.20%

1.19%

2.02%

0.52%

0.54%

Prefer not to say

0.40%

1.58%

1.81%

1.57%

2.16%

No response

0.00%

0.00%

0.00%

0.00%

8.11%

Family Income Less than \$10,000

2.00%

2.77%

6.25%

3.48%

1.44%

\$10,000 - \$19,999

4.60%

5.74%

7.26%

10.78%

0.90%

\$20,000 - \$29,999

9.00%

7.92%

9.07%

12.17%

2.16%

\$30,000 - \$39,999

11.00%

10.30%

10.48%

11.65%

3.96%

\$40,000 - \$49,999

15.40%

10.30%

11.69%

9.39%

4.68%

\$50,000 - \$59,999

19.00%

12.87%

9.07%

9.74%

5.23%

\$60,000 - \$69,999

7.60%

8.71%

8.27%

7.83%

4.68%

\$70,000 - \$79,999

8.80%

9.50%

9.27%

6.96%

6.13%

\$80,000 - \$89,999

7.00%

5.74%

4.44%

5.39%

3.42%

\$90,000 - \$99,999

7.40%

7.13%

5.44%

4.35%

4.68%

\$100,000 - \$149,999

6.40%

12.48%

10.48%
12.52%
18.38%
More than \$150,000

1.80%
5.94%
7.86%
5.74%
35.68%

No response

0.00%
0.59%
0.40%
0.00%
8.65%

Highest level of education completed (Continued)

PLOS ONE | <https://doi.org/10.1371/journal.pone.0279720> March 14, 2023

7 / 17

PLOS ONE

Data quality in online human-subjects research

Table 1. (Continued) Measure

MTurk (N = 500)

CloudResearch (N = 505)

Prolific (N = 496)

Qualtrics (N = 575)

SONA (N = 555)

Less than high school education

0.20%
0.59%
0.40%
0.70%
0.18%

High school graduate

5.00%

6.93%

12.90%

22.26%

48.29%

Some college

7.00%

17.43%

24.40%

21.04%

40.18%

2-year degree

5.40%

10.30%

10.69%

13.57%

0.72%

4-year degree

60.00%

46.53%

36.29%

25.91%

2.16%

Master's degree

21.80%

14.85%

11.09%

13.57%

0.18%

Doctorate or professional degree

0.60%

3.37%

4.03%

2.96%

0.18%

No response

0.00%

0.00%

0.20%

0.00%

8.11%

Strong Republican

21.60%

12.28%

6.85%

24.00%

5.23%

Weak Republican

7.60%

12.67%

7.66%

12.52%

13.51%

Independent Republican

2.20%

8.12%

5.44%

6.43%

7.39%

Independent Independent

3.60%

10.30%

11.69%

13.74%

13.51%

Independent Democrat

3.40%

10.50%

13.71%

6.61%

9.73%

Weak Democrat

15.60%

20.00%

19.35%

13.57%

22.70%

Strong Democrat

46.00%

25.94%

35.28%

23.13%

19.64%

No response

0.00%

0.20%

0.00%

0.00%

8.29%

Republican Party

28.80%

32.87%

19.35%

46.09%

22.88%

Libertarian Party

2.60%

6.34%

6.45%

2.78%

7.57%

Democratic Party

67.60%

56.24%

62.90%

47.83%

55.14%

Green Party

1.00%

3.56%

11.09

3.30%

5.95%

No response

0.00%

0.99%

0.20%

0.00%

8.47%

Political Affiliation (ANES)

Party Affiliation

Note. Table 1 includes mean and standard deviation for age by platform. For all other demographic measures, we report the percentage of participants who selected each response by platform. Participants could select more than one option for ethnicity. Some percentages may not add to 100% due to rounding.

<https://doi.org/10.1371/journal.pone.0279720.t001>

lab. Our SONA participant pool included students who sign up for studies in exchange for course credit. Participants on MTurk and CloudResearch were prevented from taking the survey on both platforms. It was not possible to implement similar preventions between other platforms because only MTurk and CloudResearch use the same worker IDs.

Materials Our study consisted of an online survey administered through Qualtrics' survey design software. Attention checks. We measured attention to the survey through five attention checks. In the first attention check we asked participants to identify the color that was mentioned in the consent form, a measure previously used by Douglas et al. [40]. In the second attention check participants were asked to leave a text-entry box empty. The third attention check occurred as part of the personality inventory to which we added an item asking participants to select "strongly agree" from five potential response options. In the fourth attention check participants reported if they have any type of colorblindness. This colorblindness check was adapted from Kan and Drummey [18]. The various response options included "red-blue colorblindness," which does not exist. Selecting this option indicated that participants were not paying

PLOS ONE | <https://doi.org/10.1371/journal.pone.0279720> March 14, 2023

8 / 17

PLOS ONE

Data quality in online human-subjects research

attention to the question. In the fifth attention check we asked participants to complete a simple arithmetic problem (what is 3×4 ?). Personality inventory. Participants' personality traits were assessed using the 50-item version of Costa and McCrae's [37] NEO-PI-R Domains obtained from the International Personality Item Pool (IPIP) [38]. Participants were asked to evaluate the extent to which they agreed with various statements about themselves on a 5-point Likert response scale ranging from "Strongly disagree" to "Strongly agree." Sample items are "I often feel blue" (neuroticism), "I feel comfortable around people" (extroversion), "I have a vivid imagination" (openness), "I believe that others have good intentions" (agreeableness) and "I pay attention to details" (conscientiousness). Sustainability. Participants were shown one of three 75-second videos about sustainability mid-way through the survey. Participants were later shown short descriptions of each of the three videos and were asked to choose the description of the video that matched what they had seen. Correct identification of the description indicated correct recall of the video. Participants were asked a series of follow-up questions including the extent to which they thought the government should invest in green energy. Participants were asked this same item again 21 questions later in the survey to assess the test-retest correlation of participant responses. Belief in conspiracy theories. We measured participants' belief in conspiracy theories using Brotherton et al.'s [41] Belief in Conspiracy Theories Scale. The scale includes 15 items each describing a different conspiracy theory (e.g., "Some UFO sightings and rumors are planned or staged in order to distract the public from real alien contact"). Participants responded to each of the items on a 5-point Likert scale ranging from "Definitely not true" to "Definitely true." Experience taking online surveys. Participants were asked the frequency with which they take online surveys. We asked participants how many surveys they think they had taken over the past seven days and over the past year. Survey meta-data. We also measured the amount of time each participant spent completing the survey. We categorized participants who took the survey in more than 3 minutes as having passed the speed indicator. The specific 3-minute cutoff was determined by having a research assistant take the survey as fast as possible. Three minutes was the fastest the survey could be completed while still reading each question. Additionally, because participants completed 106 survey items and watched one of three approximately 75-second videos, if participants spent 1 second per question the survey should take (at a minimum) 181 seconds to complete. Thus, the 3-minute cutoff was also consistent with Wood et al.'s [33] finding that response quality drops if participant spend only 1 second on each question. We also recorded participants' IP address, geolocation, worker ID, and the number of missing responses throughout the survey. Having a unique worker ID, IP address, and geolocation were also used to indicate high-quality data as these identifiers should be unique to each participant. Demographics. We included demographic measures for age, gender identity, identity as transgender, ethnicity, sexual orientation, family income, highest level of education completed, political affiliation, party affiliation, and political opinions. To measure political affiliation, we used the American National Election Studies' (ANES) Party Identification 7-Point Scale [42]. This measure involves a two-step procedure in which participants first are asked if they identify as a Democrat, Republican, or neither. If participants identify as a Democrat or Republican, they are then asked to specify how strongly they identify with that party. If participants identify as neither, they are prompted to select if they feel closer to the Democratic Party, Republican Party, or neither. Participants also reported their party affiliation from one of four political parties they would vote for in a hypothetical state election (Republican Party, Libertarian Party, Democratic Party, and Green Party). We included measures of political opinion

PLOS ONE | <https://doi.org/10.1371/journal.pone.0279720> March 14, 2023

9 / 17

PLOS ONE

Data quality in online human-subjects research

with respect to social and economic issues (e.g., "Please select the statement that most closely describes your political beliefs on social issues" with 5 response options ranging from "I am very conservative" to "I am very liberal"). End of survey questions. Participants were given the opportunity to provide feedback in an open-response text box. We also asked participants to check a box if they believed their data should be considered low-quality. Participants were told that responding to the survey while "distracted, respond[ing] to questions without reading them, or pick[ing] answers randomly" could lead to low-quality data. Selecting the box to indicate low-quality data did not affect their payment.

Procedure Prior to posting the survey on each of the platforms we reached out to representatives from CloudResearch, Prolific, and Qualtrics to ensure that we were using all possible means to guarantee high data quality. It was not possible to reach out to MTurk as they do not have a representative we could contact. Prior to

posting our study on SONA, the study was reviewed by our department's Introduction to Psychology Research Coordinator who oversees the department's SONA participant pool. We did not receive funding from any of the platforms. The survey was estimated to take 8 minutes to complete.

Ethics statement This study was approved by the Minimal Risk Research Board of the University of Wisconsin–Madison Institutional Review Board. A waiver of signed consent was approved by the IRB. Participants were required to read the consent form, agree to participate, but not sign the form prior to participating in the study.