

Pangborn Symposium 2025 Workshop: What Is Data Quality in Online Consumer Research? Why and How We Should Take It Seriously?

Handout for delegates: Examples of recommendations for improving data quality in online consumer research

Part 1. Table 1 from “Jaeger, S. R., & Cardello, A. V. (2022). Factors affecting data quality of online questionnaires: Issues and metrics for sensory and consumer research. *Food Quality and Preference*, 102, 104676.”

Table 1

List of factors affecting data quality in online consumer questionnaires and the available metrics of online data quality. The right column is the section number in this paper in which the factor or metric is discussed.

| Factors affecting data quality in online consumer surveys | Section |
|---|---------|
| Questionnaire design and administration | |
| Mode of administration and survey device | 2.1.1. |
| Questionnaire completion incentives | 2.1.2. |
| Questionnaire topic/respondent interest | 2.1.3. |
| Questionnaire length | 2.1.4. |
| Item difficulty | 2.1.5. |
| Item response modes and scales | 2.1.6. |
| <i>Respondent demographics and psychographics</i> | |
| Respondent identity: human vs robot | 2.2.1 |
| Age and gender | 2.2.2. |
| Respondent education and cognitive ability | 2.2.3. |
| Respondent familiarity/literacy with data collection devices | 2.2.4. |
| Respondent survey-taking experience | 2.2.5. |
| Respondent personality characteristics | 2.2.6. |
| <i>Distractions, carelessness and maladaptive attitudes and behaviors</i> | |
| Distractions in the questionnaire-taking environment | 2.3.1. |
| Respondent attitudes toward questionnaires and surveys | 2.3.2. |
| Careless responding | 2.3.3. |
| Satisficing behavior | |
| Metrics of online data quality | |
| <i>Ex-ante measures of data quality</i> | |
| Instructional manipulation checks and trap questions | 3.1.1 |
| Bogus, infrequency, nonsense, and inconsistency questions | 3.1.2. |
| Respondent engagement | 3.1.3. |
| <i>Post-hoc measures of data quality</i> | |
| Questionnaire completion rates | 3.2.1. |
| Questionnaire completion times | 3.2.2. |
| Item response times | 3.2.3. |
| Response accuracy | 3.2.4. |
| Response quality | 3.2.5. |

Part 2. Table 2 from “Jaeger, S. R., & Cardello, A. V. (2022). Factors affecting data quality of online questionnaires: Issues and metrics for sensory and consumer research. *Food Quality and Preference*, 102, 104676.”

Table 2

Examples of study information and data quality metrics that should be provided when reporting online questionnaire data in scientific journals.

1. Software platform used for questionnaire administration, database for participant selection, and any incentives to complete the questionnaire. If ISO certified vendor, so state.
 2. A statement regarding the questionnaire-taking experience of the participants, e.g., average number of questionnaires per month; if available
 3. Rates of participant completion, drop-outs, and post-hoc eliminations
 4. Proportion of skipped or missing responses or statement that software prevented respondents from proceeding without completion of all previous questionnaire items
 5. Task and/or survey completion times; if speeders eliminated, state criteria and percent eliminated
 6. Protocols used to identify and eliminate non-human (bot) responses; include percent eliminated, if available
 7. Post-hoc measures used to identify careless or inattentive participants, e.g., percent of straight-line or random responders
 8. Ex-ante measures used to assess data quality, e.g., trap questions, instructed manipulation checks, etc.
 9. Any measures used to assess participant engagement, topic interest or questionnaire enjoyment or satisfaction.
 10. Any proactive or embedded manipulations to improve attention or engagement
-

Part 3. Example data quality statement. Slide from Jaeger, S. R., Rasmussen, M. A., & Cardello, A. V. (2024). Factors affecting data quality of online questionnaires with consumers in sensory and consumer research: Metrics from the literature and empirical insights. *Science Talks*, 9, 100307. <https://doi.org/10.1016/j.sctalk.2024.100307>

1. Participants had self-registered with <name of research provider> (ISO-certified, 20252:2019) because they were interested in taking part in online research studies.
2. The final sample of 2114 people was reached by approaching 3229 people, of which 692 did not satisfy the eligibility criteria, and 312 began but did not complete the survey.
3. 31 people who had completed the survey were excluded *post-hoc* because they had completed the survey very quickly (operationalised as faster than one-third of the median response time), which is linked to careless responding (Goldammer, Annen, Stöckli, & Jonas, 2020).
4. 75 people who completed the survey and passed the “speeder criterion” were also removed *post-hoc* because they provided the same score (i.e., flatlining) (Schonlau & Toepoel, 2015) for all statements in the 10-item food neophobia scale (Pliner & Hobden, 1992).
5. A further 3 people were excluded because they had duplicate Internet Protocol (IP) address to a prior submission, which could suggest that the same person tried to complete the survey twice..
6. Finally, 2 people were removed because they provided random or nonsensical verbatim responses
7. There were no mindfulness exercise prior to starting the survey, and gamification was not used.
8. The median time to complete the required tasks was 9.5 (IQR=6.7) min.

Example data quality statement

Part 4. Web Appendix C from “Berry, C., & Burton, S. (2024). Response satisficing across online data sources: Effects of satisficing on data quality and policy-relevant results. *Journal of Public Policy & Marketing*, 43 (4), 318-328.

Measures of Response Satisficing, Multitasking, Effort, and Attention Checks

Satisficing

- For questions in this survey, as long as I provided an answer, any response would be fine.
- I was not willing to waste much time in responding to questions in this survey.
- In responding to this survey, I selected answers that could be chosen most easily.
- In answering questions in this survey, I was okay with any of the response alternatives.
- In choosing between response options in this survey, I stopped at the first one that seemed reasonable.
- If I was satisfied with one of my responses, I did not spend any time seeking out better response options.

Multitasking

- While answering the questions in this survey, I was multitasking.
- When responding to this survey I was also engaged in other activities.

Effort

- For this survey, I made a strong effort to provide accurate answers to the questions asked.
- In completing this survey, I did my best to follow the instructions I was given.
- I did my best to be careful in responding to the questions in this survey.

Attention Check Measures

- The Sun rotates around the Earth
- I have never heard of Facebook
- Obama was the first President of the U.S.
- The product shown in the advertisement was a cereal product

Notes: All measures were assessed using seven-point scales with endpoints of “strongly disagree” and “strongly agree.” For the first three attention check measures, a response of “strongly disagree” (coded as a ‘1’) indicates that the participant was attentive to the question. For the fourth item, which was specific to the experiment), a response of “strongly agree” (coded as ‘7’) indicates attention. The first three items were recoded such that higher values indicate greater attention, and an average of the four items was used as an index of attention check quality.

For those interested in future use of the survey satisficing measure, reliability for more parsimonious versions remains acceptable ($\alpha > .80$) for three and four-item versions. For example, for the above satisficing items, 3-item scales consisting of items 1, 3, and 4; items 1, 3, and 5; items 3, 5, and 6; items 1, 2, and 3; and items 1, 2, and 5 all result in coefficient alphas $> .80$.

The satisficing and related items were included at the end of the survey and were preceded with the following instruction: “For this next set of questions, it is **CRITICAL** to our research that you answer each of these questions **HONESTLY** and **TRUTHFULLY** as possible. Your answers will in no way affect your compensation or approval ratings of your work or any future opportunities to respond to online surveys. However, these questions **ARE EXTREMELY IMPORTANT** to our research, and we very much appreciate your honest responses.”

Part 5. Table 5 from “Berry, C., & Burton, S. (2024). Response satisficing across online data sources: Effects of satisficing on data quality and policy-relevant results. *Journal of Public Policy & Marketing*, 43 (4), 318-328.

Recommendations for Sample and Quality Check Disclosures to Include for Journal Submissions

| Information Type | Specifics to Include |
|---|--|
| Source of the sample | <ul style="list-style-type: none"> • Report the sample source (vendor) and/or platform used • Include the requested sample size |
| Sample composition | <ul style="list-style-type: none"> • Provide any screening criteria specified for the sample source (e.g., % approval rate; minimum age or other sample restrictions; approved respondent list [CloudResearch]) • Include criteria used for sample composition (e.g., census matching) |
| Recruiting procedures and quality checks performed by vendor | <ul style="list-style-type: none"> • Reference available, online information about sources of crowdsourced participants who are recruited by the vendor and the checks and procedures performed by the vendor to enhance response quality |
| Data quality checks performed by the researcher | <ul style="list-style-type: none"> • Report specific types of screeners and checks used (e.g., instructional manipulation checks, attention checks, response satisficing scale) • Provide the specific items used • Include the criteria used for excluding respondents • Some considerations for exclusion criteria: If there are multiple checks used, what are the specific criteria across checks? If an interval scale is used for a fictitious attention check statement, does only the most extreme disagree level lead to the exclusion of the respondent? |
| Participant exclusion report | <ul style="list-style-type: none"> • Report the number of participants / % of total sample who were excluded based on the quality checks |
| Researcher analyses | <ul style="list-style-type: none"> • Report results of primary analyses with and without participant exclusion (see Tables 3 and 4) • In instances in which the % excluded is very low, just confirming that results are consistent with and without participant exclusions should be sufficient • In instances in which the excluded % is higher, supplementary analyses (provided in the form of an appendix or web appendix may be warranted |

Note: While basic information such as the source of the sample and sample criteria could be reported in the manuscript text, to minimize manuscript length, the more specific information should be reported in a web appendix.