

ARTICLES

Worth the Wait? Mandatory Breaks and Data Quality in Surveys of Online Panel Respondents

Seon Yup Lee, Ph.D.¹, Azdren Coma, Ph.D.²

¹ Kansas State University, ² University of California, Los Angeles

Keywords: Online panel, data quality, cognitive burden, web survey design, mandatory break

<https://doi.org/10.29115/SP-2026-0033>

Survey Practice

Vol. 20, 2026

A large proportion of web surveys rely on responses from online panels such as Prolific. Surveys with longer questionnaires tend to be cognitively burdensome, which risks data quality. To reduce the cognitive burden of respondents when completing longer questionnaires, this study examines an alternative proactive tool: the use of mandatory breaks. We assessed the effect of including a mandatory break using a randomized experiment embedded in a survey measuring climate policy attitudes. In a 71-question survey (with a mean completion time of approximately 11 minutes), respondents either completed all items without a break or were given a mandatory 60-second break midway through. Building on previous research (Peer et al. 2021), we analyze five markers of data quality (Internal Reliability, Straightlining, Self-reported Honesty, Attention Check, and Sociodemographic Data Inconsistency). We found no significant effect of a mandatory break on overall data quality in our survey configuration, while other survey configurations may produce different results. Future research can investigate whether mandatory breaks influence data quality across different break lengths and placements, different respondent characteristics, and in other survey conditions.

Introduction

Self-administered web surveys have become increasingly common (Couper 2017; DeLeeuw 2018; Groves 2011; Prandner et al. 2023). Among these surveys, a large proportion of social research today relies on online panels from crowdsourcing platforms such as Amazon's Mechanical Turk or Prolific. Online panels now account for a substantial share of data published in peer-reviewed empirical work in the social and behavioral sciences (Aguinis et al. 2021; Chandler et al. 2019; Paolacci and Chandler 2014). While these surveys offer promising results, concerns remain about their validity and reliability. This paper aims to improve response quality by examining the implementation of mandatory breaks within the online questionnaire.

Responses collected from online panels often raise data quality concerns, such as respondent inattentiveness, dishonesty, and survey fatigue (Chandler et al. 2020; Cornesse and Blom 2023). One major threat to data quality is a high respondent cognitive load (or cognitive burden). High cognitive load can lead to measurement errors, item non-response, or even break-offs (Dillman

et al. 2014; Ward et al. 2017). Survey length is a key contributing factor affecting respondents' cognitive burden (Emery et al. 2022). While shorter questionnaires are preferable, not all surveys can be shortened. This study aims to understand whether a mandatory break embedded within a survey can improve data quality, even in conditions where the survey is already at an ideal length of between 10 to 15 minutes (Revilla and Höhne 2020). To our knowledge, no prior experimental study has directly compared mandatory mid-survey breaks against no breaks on data quality outcomes.

To address concerns about low-quality data in online surveys, researchers have developed and implemented a range of assessment and monitoring methods to improve data quality, including attention-check questions and completion-time filtering. However, these approaches are reactive rather than proactive. A more proactive approach is to gamify surveys (Aubert et al. 2023), but not all surveys can be gamified.

For longer surveys, some researchers propose a “split questionnaire” design. Split questionnaire design can take the form of matrix sampling, in which different subgroups receive different sub-questionnaires, or delayed completion surveys, in which respondents complete one portion of the survey at a later time. However, these methods have pitfalls: matrix sampling requires imputation of missing data, and the delayed-completion method risks high drop-off and lower response rates (Adigüzel and Wedel 2008; Axenfeld et al. 2022; Peytchev and Peytcheva 2017; Raghunathan and Grizzle 1995). As an alternative approach to maintaining higher data quality, we explore introducing a mandatory break within the surveys. Unlike split-questionnaire designs, a mandatory break does not risk attrition due to delayed participation. Evidence from psychological literature suggests that introducing breaks into tasks can improve response quality. A short break within a task can encourage individuals to process more deliberately (Bilancini et al. 2024; Shirasuna et al. 2025). Meta-analytic evidence finds that short rest breaks produce small but statistically significant benefits for fatigue reduction and task performance, particularly for less cognitively demanding tasks (Albulescu et al. 2022). Our approach is designed to reduce cognitive burden and improve response quality with the goal of achieving similar benefits to the performance of respondents.

However, a mandatory break may also be viewed as a task interruption, which might lower data quality. Previous research has found that task interruptions impair performance because they require additional effort to navigate the task (Baethge et al. 2015; Yuan and Zhong 2024). If a mandatory break is perceived as a stressor or an interruption, rather than reducing fatigue and encouraging deliberation, introducing a mandatory break may decrease data quality, rather than improve it.

Given the limited prior research on the implementation of mandatory breaks embedded in surveys, we hypothesize that the effect can be positive or negative. While prior research suggests that a break within a cognitively burdensome questionnaire may increase data quality, other research mentioned above suggests that a mandatory break may have a negative effect. Therefore, we propose two competing hypotheses about the effect of mandatory breaks in surveys on all respondents:

H1: Alleviating effect. A mandatory break improves data quality by refreshing cognitive load.

H2: Worsening effect. A mandatory break decreases data quality by disrupting respondents' engagement.

Methods

Our survey experiment examined the effect of a mandatory break on respondents in the United States, focusing on attitudes toward climate change policies. The questionnaire, comprising 71 questions, measured support for 22 climate policy items. The study employed a two-group experimental design. In the control group, participants completed the questionnaire without the mandatory break. In the experimental group, participants completed the first half of the questionnaire, including 11 key support item questions, before encountering a mandatory 60-second break. The climate policy items used standard closed-ended, Likert-type response scales—the dominant question format in attitudinal and public-opinion research—and the instrument fell within the 10-to-15-minute range that respondents and researchers typically regard as a reasonable survey length (Revilla and Höhne 2020).

The 22 climate policy items were administered as a parallel-structure battery, with respondents rating each policy on the same response scale. This battery format is widely adopted in attitudinal, public opinion, market research, and organizational surveys, and is the format in which concerns about respondent fatigue and straightlining are most often raised. This makes such instruments a natural test case for a mid-survey break intervention. In these respects, the questionnaire is representative of the moderate-length, closed-ended surveys that are routinely fielded on online panels, which supports the broader applicability of our findings beyond the specific topic studied here.

A total of 2,373 respondents were recruited via the Prolific platform in December of 2024. A convenience sample resembling U.S. demographics was created based on Prolific's Census sample matching feature. Respondents were compensated \$1.76 for a 10-minute survey. Additionally, half of the respondents assigned to the mandatory-break experimental condition received a \$0.17 bonus, bringing the average hourly participation rate to about \$10.02. This payment rate aligns with Prolific's fair pay guidelines.

Table 1. Descriptive Statistics

	Control (N = 1,054)	Experimental (N = 1,047)	Full Sample (N = 2,101)
Age (Range: 18 – 86)	41.55 (13.96)	41.23 (13.74)	41.39 (13.85)
Gender			
Man	47.53%	48.90%	48.22%
Woman	51.42%	50.24%	50.83%
Non-binary	0.95%	0.86%	0.90%
I use a different term	0.09%	0.00%	0.05%
Race and Ethnicity			
White (not Spanish, Hispanic, or Latino)	69.17%	70.20%	69.68%
Black or African American	15.28%	14.61%	14.95%
Spanish, Hispanic, or Latino	6.36%	4.87%	5.62%
Asian or Asian American	5.22%	5.73%	5.47%
American Indian or Alaska Native	0.47%	0.48%	0.48%
Other	0.57%	0.57%	0.57%
More than one race selected	2.94%	3.53%	3.24%

NOTE: Age is reported in terms of means (standard deviation in parentheses).

Gender and race/ethnicity are reported in terms of percentages.

A growing concern for surveys fielded on online panels is the possibility that responses are produced or assisted by large language models (LLMs) rather than genuine human participants (Veselovsky et al. 2025). We took several steps to mitigate this risk. At the platform level, Prolific restricts participation to an identity-verified pool and applies pre-screening and ongoing behavioral monitoring designed to detect automated accounts and fraudulent activity, and independent comparisons have generally rated their data quality favorably relative to other panels (Douglas et al. 2023; Peer et al. 2021). At the study level, beyond the data-quality measures that constitute our outcomes, we screened on completion time and excluded responses ($n = 260$) flagged as timed out, erroneous, or otherwise invalid, resulting in a study sample of 2,113. Because the survey relied on closed-ended policy items rather than open-text responses, it offered limited opportunity for the kind of LLM-assisted free-text answering that is most difficult to detect. We nonetheless encourage future researchers using this design to incorporate platform authenticity checks, which have become more widely available since our data were collected.

Responses between the experimental and control groups were compared to determine whether the mandatory break produced a statistically significant difference between those who had a break and those who did not. During this break, a screen displaying a natural scenery image appears (see Appendix). A countdown timer counted down from 60 seconds.

[Table 1](#) indicates the demographic composition of the full analytic sample. Starting from the 2,113 respondents retained after study-level screening, we further excluded 12 cases with item non-response on the main variables of interest. As a result, a total of 2,101 responses were analyzed in this

study. Examining key sociodemographic characteristics, our sample includes 1,013 respondents identified as male (48.22%) and 1,068 as female (50.83%). Twenty respondents identified as either non-binary or a different term (0.95%). In terms of race, 69.68% identified as white, 14.95% Black, 5.47% Asian, 5.62% Latino, 3.24% multi-racial, 0.48% Native American, and 0.57% as other. Overall, the sample demographics resemble those of the U.S. population, according to the 2020 U.S. Census (U.S. Census Bureau 2021a; 2021b).

Dependent Variable

Previous studies have employed multiple measures to assess the quality of responses from online panels. Similarly, we build on Peer et al. (2021), who compared the data quality of different online panel respondents across *four* criteria: attention, comprehension, honesty, and reliability. Additionally, we examine repetitive response patterns (or “straightlining”) and the accuracy of sociodemographic data, as described above. In this study, we assess the data quality of responses based on the following five criteria:

1. **Internal Reliability:** Based on (1) averaged 11-item measures, (2) standard deviation across the 11 items, and (3) Cronbach’s alpha of response consistency between experimental groups.
2. **Straightlining:** Measurement of whether respondents excessively repeat their answers (e.g., same scale point) in a row.
3. **Self-reported Honesty:** Measured via a 0–100 slider, asking, “Did you participate properly in this survey, reading carefully, and giving your actual opinion?”
4. **Attention Check:** Based on whether the respondents passed or failed the attention check that was shown after the break.
5. **Sociodemographic Data Inconsistency:** Based on the comparison of the demographic data (such as age and gender) collected at the end of the questionnaire and the demographic data provided by Prolific.

Results

Internal Reliability

The internal reliability is inferred based on three measures: the averages of each of the 11 policy items *after* the mandatory break, the standard deviations, and Cronbach’s alpha of those items. [Table 2](#) shows the differences in data quality between the experimental and control groups. The

Table 2. Data Quality Measures Between Control and Experimental Groups

	Control (N = 1,054)	Experimental (N = 1,047)
Internal Reliability		
11-Item Average (1-5)	3.90 (0.68)	3.92 (0.63)
11-Item Standard Deviation	0.91 (0.33)	0.92 (0.33)
11-Item Cronbach's Alpha	0.86	0.83
Straightlining (1-11)	5.75 (2.61)	5.57 (2.46)
Self-Reported Honesty (0-100)	98.36 (5.89)	98.09 (6.97)
Failed Attention Check Rate (0-100%)	4.65%	5.35%
Sociodemographic Data Inconsistency Rate (0-100%)	4.74%	5.44%

NOTE: Standard Deviations in parentheses.

11-item averages were 3.90 and 3.92 (on a 5-point scale) for the control and experimental groups, respectively. A t-test comparison showed no statistically significant differences ($p = 0.51$).

We also examined the standard deviation of each participant's responses. Individual-specific standard deviation tells us the extent to which each participant's responses are "spread" across the 11 items. The lack of spread can indicate that the individual responded with less variability, which may be a potential indicator of inattentiveness. The standard deviations of the 11-item responses were 0.91 and 0.92 for the control and experimental groups, respectively. No significant differences were found based on the t-test ($p = 0.41$).

Third, we compared Cronbach's alpha coefficients of the 11-item responses between the control and experimental groups. Cronbach's alpha is a statistic that denotes the internal consistency of those 11 items. The alphas were 0.86 and 0.83, for control and experimental conditions. Based on the Feldt test, the 95% confidence intervals of the two groups' coefficients overlap, indicating no statistically significant difference. Based on three measures, we found no statistically significant difference in how respondents "behave" in terms of their responses, regardless of whether they were given a mandatory break. These results fail to provide support for Hypotheses 1 and 2; we find no evidence that the mandatory break affects overall data quality.

Straightlining

For 11-item questions after the break, the "straightlining" measure ranges from 1 (no repetition) to 11 (entirely consistent).¹ Results from the t-test show that there is no statistically significant difference ($p = 0.11$) in the frequency of repeated same responses, regardless of control ($M = 5.75$, $SD = 2.61$) or experimental group ($M = 5.57$, $SD = 2.46$).

¹ The entire sample had a mean of 5.66 and a standard deviation of 2.54, with a positively skewed distribution (skewness coefficient = 0.267).

In addition to a t-test, a Poisson regression was also analyzed. Poisson regression is a generalized linear model that predicts the expected “count” of the outcome, rather than normally distributed or binary outcomes (Hoffmann 2016, 136). Considering that the straightlining measure is in a discrete “count” form, a Poisson regression is considered appropriate. The results show that the experimental group is associated with a 3.10% decrease in the expected number of repeated answers. However, the coefficient is not statistically significant ($p = 0.09$). In other words, based on both t-test and Poisson regression results, no evidence is found to support that a break can decrease the frequency of careless, repetitive responses. Therefore, the findings from the straightlining measure also show no support for either hypothesis.

Self-Reported Honesty

The experimental group has a mean self-reported honesty score of 98.09 (on a scale from 0 to 100) and a standard deviation of 6.97. For the control group, the honesty score had a mean of 98.36 and a standard deviation of 5.89. No statistical significance was observed in the difference between those who had and did not have the mandatory break ($p = 0.35$), supporting neither Hypothesis 1 nor 2.

Attention Check

The attention check was shown to the respondent after policy item #14, which was three questions after the mandatory break was given to the experimental group (after policy item #11). Results show that those who received the break (the experimental group) had a higher rate of failing the attention check (5.35%) than the control group (4.65%). However, the proportion test shows that there is no statistically significant difference in the failure rates between the two groups ($p = 0.46$), providing no support for either Hypothesis 1 or 2.

Sociodemographic Data Inconsistency

We assessed response accuracy by comparing the experimental and control groups, focusing on the inconsistency in self-reported sociodemographic data between participants’ Prolific profiles and their responses in our survey. The three sociodemographic variables that we examined that had the most complete data supplied by Prolific were the respondents’ gender, age, and race and ethnicity. A discrepancy in any of these three variables for each respondent was recoded into a dichotomous variable.

Minor differences emerged between the experimental and control groups in the accuracy of their self-reported demographic data. 5.44% of the experimental group showed inconsistency between Prolific and survey responses, whereas 4.74% of the control group showed inconsistency. The

proportions test shows that there is no statistically significant difference between the experimental and control groups ($p = 0.47$), thus providing no support for either hypothesis.

All five measures of data quality show that there is no statistically significant difference between the control group and the experimental group. At the full sample level, the mandatory breaks did not affect the data quality among the respondents.

Conclusion

This study sought to determine whether a mandatory break could enhance data quality in web surveys administered to online panels. Building on previous research that used four measures of data quality (Peer et al. 2021), we used five measures to assess the effectiveness of this intervention. Overall, our results show that the mandatory break provided no consistent benefit (or loss) to data quality within the context and configuration of our survey.

For practitioners designing surveys of comparable length and content, the practical implication is straightforward: adding a one-minute mandatory break is unlikely to be a worthwhile use of resources, since it raises costs and respondent burden without improving any of the data-quality indicators we examined. Effort is better directed toward established practices to increase data quality, such as by keeping the questionnaire concise, by not including the mandatory break, and by using the bonus money to increase the pay rate. Other conditions and configurations of an experiment might yield different results (e.g., different survey subjects, varying survey lengths, different break lengths, different target populations).

In addition, we speculate that demographic factors may moderate the effect of mandatory breaks. Prior work suggests that gender variations exist in task quality outcomes under various conditions that alter the effort or attention a task demands, including time pressure (Bilancini et al. 2024), noise (Abbasi et al. 2022), and psychosocial stress such as induced competition (Cahlíková et al. 2020). If mandatory breaks similarly reshape the demands of the survey response, their effects may likewise differ across respondents. Including gender, demographic characteristics (e.g., race and ethnicity, age, educational attainment) may therefore influence how respondents respond to mandatory breaks. A systematic examination of such moderators was beyond the scope of the present study, but it is a promising direction that we invite future researchers to pursue.

Although our overall findings show null or negative results, future studies can test whether offering optional breaks, varying break duration, or informing participants in advance may alter our findings. Previous literature shows that the effect of introducing breaks in tasks can depend on other factors, such as

the break timing, duration, frequency, or break content (for review, Albulescu et al. 2022), and survey design more broadly requires a tailored rather than one-size-fits-all approach (Dillman et al. 2014).

As web-based data collection expands, improving instrument design remains a central objective. Rather than implementing mandatory breaks indiscriminately, researchers should test alternative break strategies across conditions while remaining attentive to group-level sensitivities. We suspect mandatory breaks may still prove beneficial in certain survey contexts.

.....

Corresponding author contact information

Seon Yup Lee

seonyupl@k-state.edu

828 Mid-Campus Drive, Manhattan, KS 66506

Submitted: May 28, 2026 EDT. Accepted: July 16, 2026 EDT. Published: August 18, 2026 EDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-ND-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc-nd/4.0> and legal code at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode> for more information.

REFERENCES

- Abbasi, Ali Mohammad, Ebrahim Darvishi, Matilde A. Rodrigues, and Kourosh Sayehmiri. 2022. "Gender Differences in Cognitive Performance and Psychophysiological Responses during Noise Exposure and Different Workloads." *Applied Acoustics* 189: 108602. <https://doi.org/10.1016/j.apacoust.2021.108602>.
- Adigüzel, Feray, and Michel Wedel. 2008. "Split Questionnaire Design for Massive Surveys." *Journal of Marketing Research* 45 (5): 608–17. <https://doi.org/10.1509/jmkr.45.5.608>.
- Aguinis, Herman, Isabel Villamor, and Ravi S. Ramani. 2021. "MTurk Research: Review and Recommendations." *Journal of Management* 47 (4): 823–37. <https://doi.org/10.1177/0149206320969787>.
- Albulescu, Patricia, Irina Macsinga, Andrei Rusu, Coralia Sulea, Alexandra Bodnar, and Bogdan Tudor Tulbure. 2022. "'Give Me a Break!' A Systematic Review and Meta-Analysis on the Efficacy of Micro-Breaks for Increasing Well-Being and Performance." *PLOS ONE* 17 (8): e0272460. <https://doi.org/10.1371/journal.pone.0272460>.
- Aubert, Alice H., Andreas Scheidegger, and Sara Schmid. 2023. "Gamified Online Surveys: Assessing Experience with Self-Determination Theory." *PLOS ONE* 18 (10): e0292096. <https://doi.org/10.1371/journal.pone.0292096>.
- Axenfeld, Julian B., Annelies G. Blom, Christian Bruch, and Christof Wolf. 2022. "Split Questionnaire Designs for Online Surveys: The Impact of Module Construction on Imputation Quality." *Journal of Survey Statistics and Methodology* 10 (5): 1236–62. <https://doi.org/10.1093/jssam/smab055>.
- Baethge, A., T. Rigotti, and R. A. Roe. 2015. "Just More of the Same, or Different? An Integrative Theoretical Framework for the Study of Cumulative Interruptions at Work." *European Journal of Work and Organizational Psychology* 24 (2): 308–23. <https://doi.org/10.1080/1359432X.2014.897943>.
- Bilancini, Ennio, Leonardo Boncinelli, and Tatiana Celadin. 2024. "Manipulating Response Times in the Cognitive Reflection Test: Time Delay Boosts Deliberation, Time Pressure Hinders It." *Journal of Behavioral and Experimental Economics* 112: 102273. <https://doi.org/10.1016/j.jsocec.2024.102273>.
- Cahlíková, Jana, Lubomír Cingl, and Ian Levey. 2020. "How Stress Affects Performance and Competitiveness Across Gender." *Management Science* 66 (8): 3295–310. <https://doi.org/10.1287/mnsc.2019.3400>.
- Chandler, Jesse, Cheskie Rosenzweig, Aaron J. Moss, Jonathan Robinson, and Leib Litman. 2019. "Online Panels in Social Science Research: Expanding Sampling Methods Beyond Mechanical Turk." *Behavior Research Methods* 51 (5): 2022–38. <https://doi.org/10.3758/s13428-019-01273-7>.
- Chandler, Jesse, Itay Sisso, and Danielle Shapiro. 2020. "Participant Carelessness and Fraud: Consequences for Clinical Research and Potential Solutions." *Journal of Abnormal Psychology* 129 (1): 49–55. <https://doi.org/10.1037/abn0000479>.
- Cornesse, Carina, and Annelies G. Blom. 2023. "Response Quality in Nonprobability and Probability-Based Online Panels." *Sociological Methods & Research* 52 (2): 879–908. <https://doi.org/10.1177/0049124120914940>.
- Couper, Mick P. 2017. "New Developments in Survey Data Collection." *Annual Review of Sociology* 43 (1): 121–45. <https://doi.org/10.1146/annurev-soc-060116-053613>.

- DeLeeuw, Edith D. 2018. "Mixed-Mode: Past, Present, and Future." *Survey Research Methods* 12 (2): 75–89. <https://doi.org/10.18148/srm/2018.v12i2.7402>.
- Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4th edition. Wiley. <https://doi.org/10.1002/9781394260645>.
- Douglas, Benjamin D., Patrick J. Ewell, and Markus Brauer. 2023. "Data Quality in Online Human-Subjects Research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA." *PLOS ONE* 18 (3): e0279720. <https://doi.org/10.1371/journal.pone.0279720>.
- Emery, T., S. L. F. Cabaço, L. Fadel, et al. 2022. "Breakoffs in an Hour Long Online Survey." *Survey Practice*. <https://doi.org/10.31235/osf.io/ja8k4>.
- Groves, Robert M. 2011. "Three Eras of Survey Research." *Public Opinion Quarterly* 75 (5): 861–71. <https://doi.org/10.1093/poq/nfr057>.
- Hoffmann, John P. 2016. *Regression Models for Categorical, Count, and Related Variables: An Applied Approach*. University of California Press. <https://doi.org/10.1525/9780520965492>.
- Paolacci, Gabriele, and Jesse Chandler. 2014. "Inside the Turk: Understanding Mechanical Turk as a Participant Pool." *Current Directions in Psychological Science* 23 (3): 184–88. <https://doi.org/10.1177/0963721414531598>.
- Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. 2021. "Data Quality of Platforms and Panels for Online Behavioral Research." *Behavior Research Methods* 54 (4): 1643–62. <https://doi.org/10.3758/s13428-021-01694-3>.
- Peytchev, Andy, and Emilia Peytcheva. 2017. "Reduction of Measurement Error Due to Survey Length: Evaluation of the Split Questionnaire Design Approach." *Survey Research Methods* 11 (4): 361–68. <https://doi.org/10.18148/srm/2017.v11i4.7145>.
- Prandner, Dimitri, Alexander Seymer, and Martin Weichbold. 2023. "The New Normal of Online Survey Research: General Population Studies Are Going Online." In *Researching the "New Normal" Social World*. Routledge. <https://doi.org/10.4324/9781003386087-3>.
- Raghunathan, Trivellore E., and James E. Grizzle. 1995. "A Split Questionnaire Survey Design." *Journal of the American Statistical Association* 90 (429): 54–63. <https://doi.org/10.1080/01621459.1995.10476488>.
- Revilla, M., and J. K. Höhne. 2020. "How Long Do Respondents Think Online Surveys Should Be? New Evidence from Two Online Panels in Germany." *International Journal of Market Research* 62 (5): 538–45. <https://doi.org/10.1177/1470785320943049>.
- Shirasuna, Masaru, Rina Kagawa, and Hidehito Honda. 2025. "Pause before Action: Waiting Short Time as a Simple and Resource-Rational Boost." *Scientific Reports* 15 (1): 4362. <https://doi.org/10.1038/s41598-025-87119-z>.
- U.S. Census Bureau. 2021a. "Median Age by Sex." In *Decennial Census*. <https://data.census.gov/table/DECENNIALDHC2020.P13>.
- U.S. Census Bureau. 2021b. "Race for the Population 18 Years and Over." In *Decennial Census*. <https://data.census.gov/table/DECENNIALPL2020.P3>.
- Veselovsky, V., M. H. Ribeiro, P. J. Cozzolino, A. Gordon, D. Rothschild, and R. West. 2025. "Prevalence and Prevention of Large Language Model Use in Crowd Work." *Communications of the ACM*. <https://doi.org/10.1145/3685527>.
- Ward, M. K., Adam W. Meade, Christopher M. Allred, Gabriel Pappalardo, and J. William Stoughton. 2017. "Careless Response and Attrition as Sources of Bias in Online Survey Assessments of Personality Traits and Performance." *Computers in Human Behavior* 76: 417–30. <https://doi.org/10.1016/j.chb.2017.06.032>.

Yuan, X., and L. Zhong. 2024. “Effects of Multitasking and Task Interruptions on Task Performance and Cognitive Load: Considering the Moderating Role of Individual Resilience.” *Current Psychology* 43 (28): 23892–902. <https://doi.org/10.1007/s12144-024-06094-2>.

Appendix



Image 1. Image Used in Mandatory Break Page