A Basic Study to Prevent Non-Earnest Responses in Web Surveys by Arranging the Order of Open-ended Questions

Ikumi Yamazaki[1], Kenichi Hatanaka[1], Satoshi Nakamura[1] and Takanori Komatsu[1]

[1] Meiji University, Nakano 4-21-1, Nakano-ku, Tokyo, Japan
yama1225iku@gmail.com

Abstract. Although web surveys are convenient and allow easy collection of a large number of survey responses, there is a significant issue that some people give non-earnest responses such as "nothing in particular" in open-ended questions because they want to cut corners. We aim to realize a questionnaire system that could improve a quality of web surveys to gather better responses from respondents. In this study, we focused on the order of open-ended questions in a survey. Then, we experimented using crowdsourcing to examine the effect of the order of open-ended questions on the non-earnest response rate. As a result of the experiment, we found that there were fewer non-earnest responses when the open-ended questions were presented firstly than when the open-ended questions were presented at last.

Keywords: Web surveys, Non-earnest Responses, Order of questions.

1   Introduction

Web surveys are often used to collect responses on social research and services and to collect preliminary data for research purposes. Compared to paper-based surveys, web surveys can gather many answers, more easily. The number of people registering for crowdsourcing services is rapidly increasing because of easy participation and easy incomes. Researchers often use open-ended questions in surveys because they provide answers from various perspectives [1]. However, some respondents answer "nothing in particular" and "abcdefghij" in the open-ended questions and give non-earnest responses (abbreviated as NERs). Such NERs are influenced by the anonymous response format [2] and the ease of copying and pasting. These NERs in web surveys have become a problem.

Our laboratory asked 20,000 people to participate in web surveys and web-based experiments using *Yahoo! crowdsourcing* [3] in 2020. In these surveys, many people responded even though they were not the survey's target, and many answered "nothing in particular," "I don't know," and "asdfasdf" in the open-ended questions, where all respondents must write their opinions. Therefore, it takes time and effort to determine and remove NERs from the analysis and to register the persons in the block list. In addition, this process reduces the number of responses for analysis. To solve

this problem of NERs, we aim to clarify how to reduce NERs and realize a system that offers suggestions for improving the ability of a web survey to gather quality responses.

One of the reasons for NERs is that the respondents become bored with completing the task in the latter half of the survey. If they become bored, they may answer "nothing in particular" or "I don't know" in the open-ended questions. Schmidt et al. [4] found that the later the open-ended questions were asked, the lower the number of interpretable responses. However, these studies have not shown that changing the position of the open-ended questions improves the quality of the responses or affects the response text. We thought that if people faced the open-ended questions at an early stage of a survey, they might answer the questions more seriously because they would not feel bored. In addition, people who are not the survey's target would leave this survey quickly due to the presence of open-ended questions at an early stage. Then, researchers would be able to gather quality answers.

In this study, we investigate the effect of the position of open-ended questions on the number of NERs, aiming to realize a system that offers suggestions for improving a web survey to gather better quality responses. In particular, we hypothesize that people would give better answers to open-ended questions when they encounter them early in the process rather than when they encounter them later. This hypothesis is because people are expected to gradually become bored. We compare the number of NERs and the response content between a study that asks the respondents to answer the open-ended questions first and a general type of study in which the respondents answer the open-ended questions last.

The contributions of this paper are as follows.

- We conducted an experimental test on the order of open-ended questions and clarified that the number of NERs (Non-Earnest Responses) was lower when the open-ended questions were asked first than when they were asked last.
- We found that the instant respondents did not spend much time answering the open-ended question compared with the basic respondents.

2   Related Work

Various studies have been conducted on points to keep in mind when creating surveys. Regmi et al. [5] noted that there are many items to consider when planning an online survey, such as the simplicity of the questions, whether the survey is suitable for online implementation, and whether cultural and ethical considerations are taken into account. Tobias et al. [6] state that when designing a survey, it is necessary to consider not only the number and format of questions, but also how participants will respond to the survey design. Thus, it is believed that situation and response consideration are important when structuring the survey.

There have been many studies on open-ended questions. Reja et al. [1] compared the responses to online surveys in both close-ended and open-ended questions and found more missing data in the open-ended questions than in the questions with a close-ended format. Zhou et al. [7] experimented using a survey containing two types

of questions: a single open-ended question and an open-ended question asking the respondents to give reasons for the selected question. More than 75% of the respondents did not answer either open-ended question. Holland et al. [8] investigated how the level of interest in a survey topic affects the responses. They found that people with a high level of interest in the case had a high quality of responses, while people with no or low interest had more non-responses to open-ended questions. Thus, it is difficult to obtain many good-quality answers to open-ended questions.

There have been many studies on the order of responses. However, although it is said to be better to place questions that require time to answer, such as open-ended questions, in the latter half of the survey, placing them there may be disadvantageous. Galesic et al. [9] found that if a question was asked later, the respondents responded in a shorter time and with a shorter text. This study investigates the effect of placing open-ended questions, which are considered difficult to answer, early in the survey.

There have been various studies on the relationship between the survey's medium and the open-ended responses. Denscombe [10] conducted a survey containing four open-ended questions on the web and on paper. As a result, although there was no significant difference between the web and paper surveys in three of the four open-ended descriptions, the responses in the web survey tended to be slightly longer than those in the paper survey. Rada et al. [11] conducted an online and paper survey for the citizens of Andalusia. The results showed that the online survey had fewer unanswered questions and more detailed responses to the open-ended questions. Sara et al. [12] compared the content of open-ended questions in email and paper surveys. They found that emails received more socially undesirable responses than paper, with longer responses and more information disclosed. These studies compared web-based and paper-based surveys and found that web-based surveys collected better responses. However, the relationships between the position of open-ended questions in the web survey and the quality of the answers were not clarified.
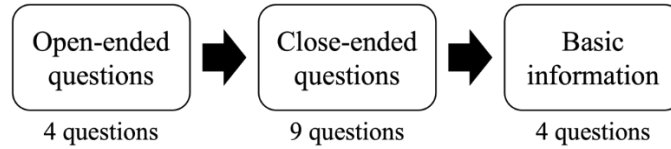
3   Experiments

3.1   Outline of the experiment

The objective of this experiment is to investigate whether the number of NERs is affected by the position of the open-ended question in the web survey.

This experiment divided participants into two groups. One group faced the open-ended questions first (hereafter referred to as Group-FIRST). The other faced the open-ended questions last (referred to as Group-LAST). The survey for Group-FIRST first asked participants to answer open-ended questions (hereafter referred to as the open-ended questions phase). Then, it asked them to answer close-ended questions (hereafter referred to as the close-ended questions phase) and finally demographic information such as gender, age, and profession (hereafter referred to as the basic information phase) in that order (see Fig. 1). The order was reversed for Group-LAST (see Fig. 2). The order of the questions in each phase was the same: four questions in

the open-ended questions phase, nine questions in the close-ended questions phase, and four questions in the basic information phase.

Fig. 1. Phase order for the group with the open-ended questions first (for Group-FIRST).
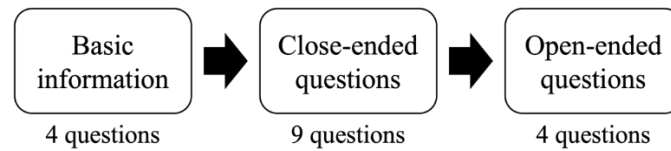
Fig. 2. Phase order for the group with the open-ended questions last (for Group-LAST).

The survey topic in this experiment was for people who had engaged in teleworking and working from home. The contents and order of the four questions in the open-ended questions phase were as follows. The phrase "even if it is a trivial matter" was added to encourage the respondents to answer each question.

- Q-a: What are the advantages of teleworking and working from home?
- Q-b: What are the disadvantages of teleworking and working from home?
- Q-c: What are the advantages of teleworking and working from home other than work (personal life, family, and so on)?
- Q-d: What are the disadvantages of teleworking and working from home other than work (personal life, family, and so on)?

Our experiment system displayed one question per page to control the order of answers. When a participant answered a question, the system showed the next question.

We conducted this survey on *Yahoo! Crowdsourcing* [3] and Google Form with different responses to verify whether the order of the open-ended questions affected the quality of the responses.

3.2 Experimental procedure

We recruited 1,000 participants (500 males and 500 females) for this experiment. The experiment was conducted on *Yahoo! Crowdsourcing* [3], and the participants were divided into two groups by their gender to avoid gender variation. The start time of each request was 8:00 am.

First, the system showed the flow of investigations and the precautions. Then, the system asked participants to check the checkboxes of each description after reading that description. The precautions indicated that the experiment was intended for people who had teleworked or worked from home and that they should not press the

back button or the reload button on their browsers. After that, the system randomly divided the participants into two groups, Group-FIRST and Group-LAST, and displayed the link to the Google Form for each group. The order of the survey phases was as indicated in either Fig. 1 or Fig. 2. The system showed a progress bar to let the participants know how many questions remained to complete the survey. When a participant completed the survey, the system presented a code at the end. As a reward, the participants received 10 Paypay-points (a popular point program in Japan), which was equivalent to 10 Japanese yen (0.08 US cents) after returning to the crowdsourcing screen and selecting the correct code.

3.3   Results

We obtained 1,101 responses (515 males, and 586 females) except for illegal responses in this survey. The reason for over 1000 responses was possible that some experiment participants answered the survey but did not return to the crowdsourcing screen to enter the code.

In order to investigate the percentage of NERs in the survey, it is necessary to determine the criteria for NERs. Therefore, we asked two university students (evaluators) to classify the open-ended responses as earnest or not earnest. The classification standards for an NER are as follows.

- It was not accompanied by an answer to the question.
- Its meaning could not be captured by the answer itself.

Note that NERs such as "nothing in particular" and "I do not know" were removed in advance. After the classifications, we evaluated the degree of agreement between the classifications using the Kappa coefficient. The results showed that the Kappa coefficient was 0.623, indicating a high agreement. Examples of NERs are as follows.

- Q-a: A little easy. I don't have beneficial to an own business, but a side business is email and phone calls, so it does not matter too much.
- Q-b: I work in childcare, so telework is not a basic requirement.
- Q-c: I can work from home. I'm single, so no relationship.
- Q-d: Without distinction. Nothing really. Tiresome.

We annotated each response as earnest or not earnest based on the evaluators' classification. If their response classifications differed, we ourselves judged whether the response was earnest or not earnest.

Fig. 3 shows the percentage of NERs for each open-ended question. This figure clarifies that the number of NERs was lower in Group-FIRST in all questions. We tested the difference in the proportions for each question and found significant differences for three questions (Q-a and Q-d: $p < 0.01$, Q-c: $p < 0.05$). In addition, the percentage of NERs increased sharply in the fourth question in both groups.
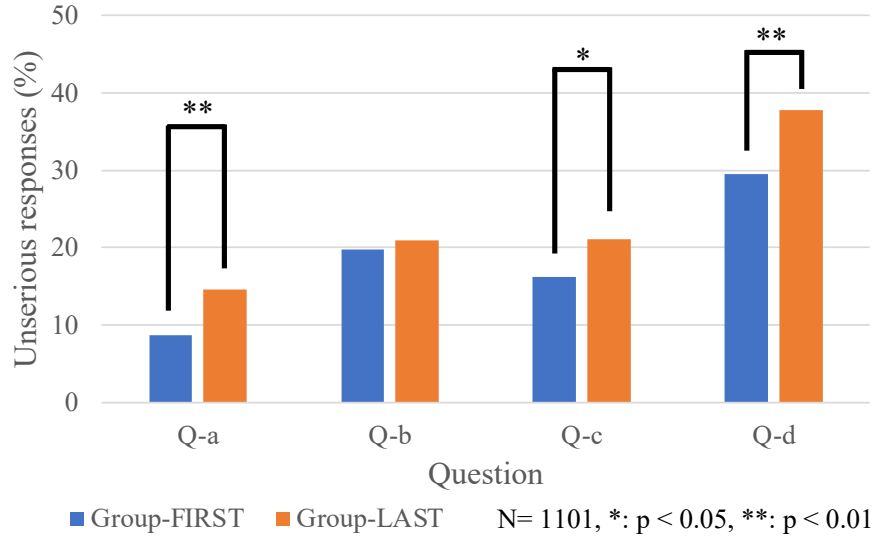
Fig. 3. Percentage of NERs in each question and group.

Table 1 shows the average length of answers for each open-ended question. In this table, the results include NERs. The table shows that the answer length was longer for Group-LAST. In addition, the answer length was the longest for both groups in the second question. Still, the answer length gradually decreased, and the fourth answer was the shortest.

In order to investigate whether the position of the open-ended questions affects the withdrawal rate of the survey, we calculated the withdrawal rate by dividing the number of surveys collected in Google Form by the number of people who accessed the survey system. The withdrawal rate for Group-FIRST was 23.7% and for Group-LAST it was 23.9%. This result shows that the attrition rate is almost the same for Group-FIRST and Group-LAST. However, this dropout rate is inaccurate because Google Form does not know how many respondents completed and left the survey. We will develop our survey system and conduct an additional experiment to clarify this withdrawal rate.

Table 1. The average length of answers (characters) in each question and group.

	Q-a	Q-b	Q-c	Q-d
Group-FIRST (N=566)	13.2	13.5	12.5	10.8
Group-LAST (N=535)	15.9	16.1	14.0	11.6

4   Additional Experiments Focused on Desertion Rate and Location

4.1   Outline of the experiment

This additional experiment aimed to clarify the disengagement point by the presentation timing of the open-ended question in the web survey. In order to clarify this, we implemented a new survey system that recorded participants' behaviors, such as the answering time for each question and which questions they could not answer in the survey.

In this experiment, we also divided participants into Group-FIRST and Group-LAST. The number of questions in each phase was the same as in section 3, and the questions in the open-ended questions phase were highly independent of those in the close-ended questions phase.

The survey topic in this additional experiment was for people who have a driver's license. The contents and order of the four questions asked in the open-ended questions phase were as follows.

- Q-e: Please tell us what you mainly drive for. If you do not usually drive, please tell us why you decided to get a driver's license.
- Q-f: Please describe the characteristics of the roads you mainly drive on. If you don't usually drive, please tell us what kind of roads are around your house.
- Q-g: If you are not good at driving, please tell us what aspects of driving are difficult for you or why you are not confident in driving. If you are confident in driving, please answer why you are so.
- Q-h: Please tell us what you pay attention to when you drive. If you do not usually drive, please answer what you were careful about when you got your driver's license. It can be something trivial.

Our experiment system displays one question per page to control the order of answers. When a participant answers a question, the system shows the next question.

4.2   Experimental procedure

We recruited 1,000 participants (500 males and 500 females) for this additional experiment, as in section 3. The requests were divided into those for males only and those for females only on *Yahoo! Crowdsourcing* [3].

In this additional experiment, the survey recorded the start time when a participant finished reading the explanation and proceeded to the next page. The order of the surveys was either Fig. 1 or Fig. 2. The system displayed the number of questions. After completing the survey, the system displayed a common code and ID. The participant was rewarded by money after returning to the crowdsourcing page, correctly selecting the code, and entering the ID.

4.3 Results

The number of participants who correctly entered their IDs was 979 (493 males and 486 females). We used them for the analysis of this experiment.

We also asked two university students (evaluators) to classify the open-ended responses as earnest or not earnest. The criteria for the classification of NERs were the same as in section 3. The Kappa coefficient of the agreement was 0.75, indicating that the classifications were almost identical. Examples of NERs were as follows.

- Q-e: Because it is necessary. Because it's better to have it.
- Q-f: Ordinary Roads. Public roads.
- Q-g: I want to drive at my own pace. I just drive with caution.
- Q-h: I don't usually drive. It was interesting.

We annotated each response as earnest or not earnest based on the evaluators' classification. If their response classifications differed, we ourselves judged whether the response was earnest or not earnest.

Fig.4 shows the percentage of NERs for each question. In all questions, the number of NERs in Group-FIRST was lower than in Group-LAST. We tested the difference in the proportions for each question and found a significant difference in Q-e and Q-h (Q-e: $p < 0.01$, Q-h: $p < 0.05$). In the present experiment in section 3, the percentage of NERs increased greatly for the last open-ended question (Q-d). However, in this experiment, the number of NERs did not increase for the last open-ended question (Q-h).
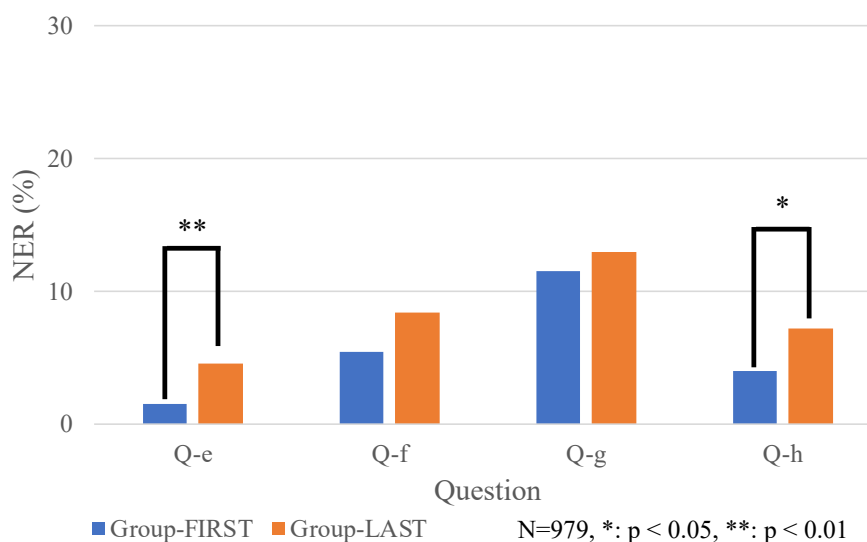
Fig. 4. Percentage of NERs in each question and group in the additional experience.

Table 2 shows the average length of answers, including the NERs, for each group and question. The answer length is longer in Group-LAST than in Group-FIRST in every question.

Table 2. The average length of answers in each question and group in the additional experience (characters).

	Q-e	Q-f	Q-g	Q-h
Group-FIRST (N=484)	12.7	10.6	18.5	15.4
Group-LAST (N=502)	14.0	11.6	20.1	16.8

The number of people who accessed the URL to this survey was 1,757 (883 for Group-FIRST and 874 for Group-LAST), and the number of participants who correctly entered their IDs was 979. This means that 778 people dropped out of this survey. Fig. 5 shows the withdrawal rate for each question. The figure shows that the withdrawal rate of Group-FIRST increased from Q1 to Q4. On the other hand, in Group-LAST, there were few dropouts, and most of the respondents answered all questions.
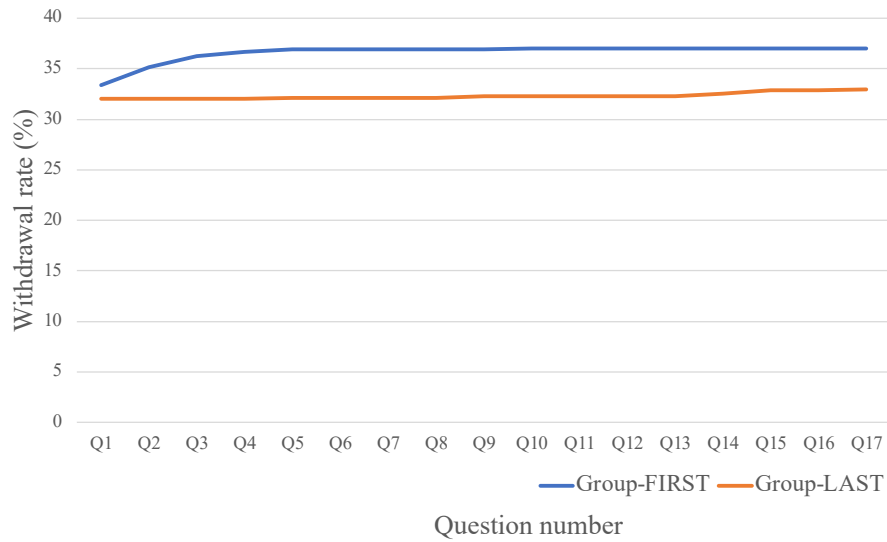
Fig. 5. Withdrawal rate per question in each question and group.

5 Discussion

5.1 Comparing two groups based on several features

In the two experiments, the number of NERs was lower in Group-FIRST for all questions. This suggests that the position of the open-ended questions affects the

quality of the responses. This result might have occurred because those in Group-FIRST did not feel bored answering the open-ended questions, but those in Group-LAST did feel bored answering these questions.

For both experiments, the trend differed only for the fourth open-ended question. Here, in the experiment in section 3, all four open-ended questions were similar in that they asked about advantages and disadvantages. So, it is possible that the fourth question caused boredom. On the other hand, in the additional experiment in section 4, the fourth question was not similar to the other three questions. So, it is possible that the difficulty of the questions was low. These findings suggest that asking similar open-ended questions may increase the number of NERs. In particular, open-ended questions that ask for the reason for the selected answer may increase the number of NERs.

In both experiments, the average length of answers was shorter for Group-FIRST. This may have been because the participants in Group-LAST answered the close-ended questions before the open-ended questions phase in detail, which made it easier for them to answer the open-ended questions, leading to an increase in the length of answers. On the other hand, although the participants in Group-FIRST did not feel bored, they answered the open-ended questions phase without answering the close-ended questions, which may have made it difficult for them to answer in detail.

To sum up, Group-FIRST had a lower percentage of NERs but a shorter length of answers, while Group-LAST had a higher number of NERs but a longer length of answers. This suggests a trade-off relation between the percentage of NERs and the answer length, depending on the position of the open-ended questions.

In the additional experiment, when analyzing at which point the respondents dropped out of the survey (Fig. 5), the results show that in Group-FIRST, a certain number of participants left the survey at the open-ended questions from Q1 to Q4. In Group-LAST, almost no participants left the survey in the middle. This suggests that some people left the survey because they thought it was troublesome due to the open-ended questions coming first. In addition, since the percentage of NERs was also small, it is highly likely that we could exclude from the survey people trying to answer it frivolously or who were trying to answer it even though they were not the target of the survey.

5.2   Comparison by survey start time

The survey start time was adjusted to 8:00 am, although the dates answered were different by the gender. Therefore, we compared the percentages of NERs in the open-ended questions of those who accessed the website within 10 minutes of the start of the survey (hereafter referred to as instant respondents) and in the open-ended questions of those who accessed the website more than 10 minutes after the beginning of the survey (hereafter referred to as basic respondents) (see Fig. 6). The numbers for each group and the results are shown in Table 3.
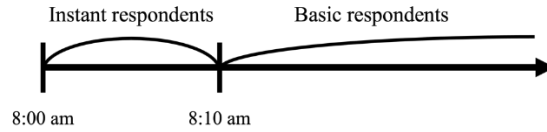
Fig. 6. Response time and the classification name of the respondent.

The table shows that NERs were lower for all questions in Group-FIRST and basic respondents. We conducted a chi-square test and found a significant difference ($p < 0.05$).

The comparison of the average length of answers is shown in Table 4. The results show that basic respondents gave longer responses in both groups than instant respondents.

Table 3. Percentage of NERs by survey start time (%).

	Q-e	Q-f	Q-g	Q-h
Group-FIRST and instant respondents (N=256)	2.3	5.5	12.1	5.1
Group-FIRST and basic respondents (N=221)	0.5	5.4	10.9	2.7
Group-LAST and instant respondents (N=239)	3.8	8.0	13.4	6.7
Group-LAST and basic respondents (N=263)	5.3	8.4	12.6	7.6

Table 4. The average length of answers by survey start time (characters).

	Q-e	Q-f	Q-g	Q-h	Total
Group-FIRST and instant respondents (N=256)	12.2	9.9	16.6	14.3	53.0
Group-FIRST and basic respondents (N=221)	13.3	11.5	20.6	16.8	62.2
Group-LAST and instant respondents (N=239)	14.2	11.2	18.4	15.8	59.6
Group-LAST and basic respondents (N=263)	13.8	12.1	21.7	17.6	65.2

Tables 5, 6, and Fig. 7 show the response time of each open-ended question in each group and respondent type. The two tables show that basic respondents took longer to complete all open-ended questions than instant respondents in both groups. Fig. 7 also shows that the completion time of the survey tended to be longer for basic respondents than for instant respondents, regardless of whether they were in Group-FIRST or Group-LAST.

These results suggest that open-ended surveys of basic respondents are valid for analysis, and Group-FIRST is especially valid. Suppose that the order of the questions does not affect the content of the responses. In that case, it can be expected that dynamically changing the survey, for example, by presenting the free text first for respondents who accessed the site after 10 minutes, would be effective. The number of NERs was highest for Group-LAST and basic respondents. This result suggests that Group-LAST and basic respondents may be more likely to give not earnest or

inaccurate responses. In addition, those who completed the survey in a short period of time might focus on getting rewarded and trying to finish it quickly.

Table 5. Open-ended questions response time for Group-FIRST (seconds).

	Introduction +Q1 (Q-e)	Q2 (Q-f)	Q3 (Q-g)	Q4 (Q-h)
Instant respondents (N=256)	68.0	38.9	47.7	38.6
Basic respondents (N=221)	69.0	44.9	58.9	47.2

Table 6. Open-ended questions response time for Group-LAST (seconds).

	Q14 (Q-e)	Q15 (Q-f)	Q16 (Q-g)	Q17 (Q-h)
Instant respondents (N=239)	35.9	41.6	50.5	39.5
Basic respondents (N=263)	40.6	43.1	61.6	48.4
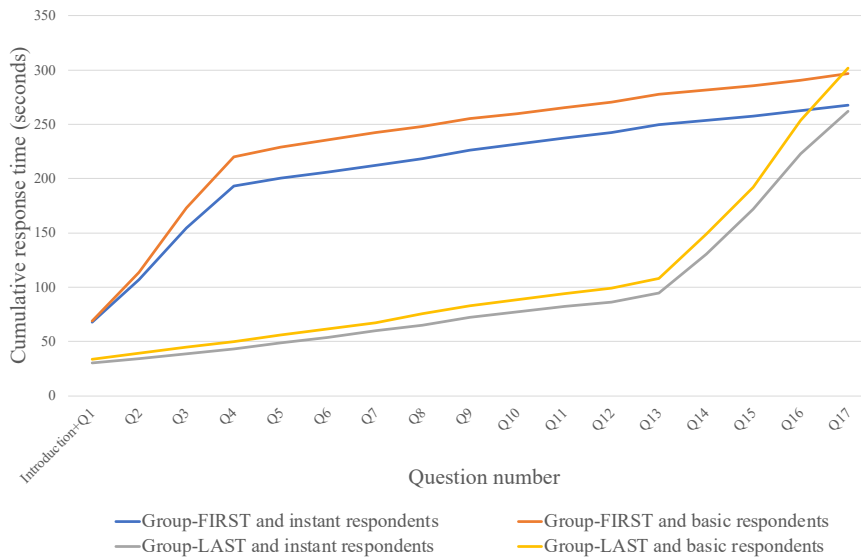
Fig. 7. Cumulative response time for each time period when the survey was accessed.

6   Conclusion

This study examined the relationship between NERs and the order of questions in open-ended questions in a web survey. We experimented by dividing participants into two groups, Group-FIRST and Group-LAST, based on the hypothesis that people would give better answers to open-ended questions when they encountered them early in the process compared to when they encountered them late in the process.

As a result of the experiment, we found a trade-off relationship between the number of NERs and the length of answers. In Group-FIRST, the number of NERs decreased. However, the length of answers also decreased, while in Group-LAST the number of NERs increased, but the length of answers increased too. In addition, it was suggested that placing the open-ended questions at the beginning would cause a certain number of people to leave the survey. Furthermore, we analyzed the percentage of NERs of those who accessed the website. We found that the NERs, the length of answers, and the response time were greatly affected by whether the respondent answered the survey within or after 10 minutes.

Before conducting the survey this time, the procedure and precautions were presented. However, the length of the explanatory text and the checkboxes that prevented people from skipping over the text may have significantly affected the response rate for the first question. In future experiments, we plan to use a system that starts the survey immediately after accessing the site, without setting up a page of instructions and precautions. Also, in the additional experiment, the response time of the first question was not accurate because it was obtained from the time when the website was accessed, not when the question was presented. We plan to experiment again to get the exact time. In addition, since it may not be appropriate to categorize the responses into only two types: earnest response and NERs, it is necessary to examine the method of classifying the responses.

Acknowledgement

References

1. Reja, U., Manfreda, L. K., Hlebec, V., and Vehovar, V.: "Open-ended vs. Close-ended Questions in Web Questionnaires." Adv Methodol Stats, 19(1), 159-177 (2003).
2. Dickinson, D. L., and McEvoy, D. M.: "Further from the Truth: The Impact of In-Person, Online, and mTurk on Dishonest Behavior." Journal of Experimental and Behavioral Economics, 90(4), 101649 (2021).
3. *Yahoo! Crowdsourcing*, https://crowdsourcing.yahoo.co.jp/, last accessed 2023/02/09.
4. Schmidt, K., Gummer, T., and Roßmann, J.: "Effects of Respondent and Survey Characteristics on the Response Quality of an Open-Ended Attitude Question in Web Surveys." Methods, Data, Analyses, 14(1), 3-34 (2020).

5. Regmi, P. R., Waithaka, E., Paudyal, A., Simkhada, P., and van Teijlingen, E.: "Guide to the design and application of online questionnaire surveys." Nepal Journal of Epidemiology, 6(4), 640-644 (2017).

6. Tobias, G., and Joss, R.: "Explaining Interview Duration in Web Surveys: A Multilevel Approach." Social Science Computer Review, 33(2), 217-234 (2015).

7. Zhou, R., Wang, X., Zhang, L., and Guo, H.: "Who tends to answer open-ended questions in an e-service survey? The contribution of closed-ended answers." Behaviour & Information Technology, 36(12), 1274-1284 (2017).

8. Holland, J. L., and Christian, L. M.: "The Influence of Topic Interest and Interactive Probing on Responses to Open-Ended Questions in Web Surveys." Social Science Computer Review, 27(2), 196-212 (2009).

9. Galesic, M., and Bošnjak, M.: "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." Public Opinion Quarterly, 73(2), 349-360 (2009).

10. Denscombe, M.: "The length of responses to open-ended questions: a comparison of online and paper questionnaires in terms of a mode effect." Social Science Computer Review, 26(3), 359-368 (2008).

11. Rada, V. D. de, and Domínguez-Álvarez, J. A.: "Response Quality of Self-Administered Questionnaires: A Comparison Between Paper and Web Questionnaires." Social Science Computer Review, 32(2), 256-269 (2014).

12. Sara, K., and Lee, S. S.: "Response Effects in the Electronic Survey." Public Opinion Quarterly, 50(3), 402-413 (1986).