Questionnaire Test Using Cognitive Approach: Lessons Learned

Redesign of Time Use Survey

Hyun-Jung Park[[1]](#footnote-2), Youngshil Park

Statistics Korea, Statistics Research Institute

|  |
| --- |
| **Abstract:** Statistics Korea (KOSTAT) established the Questionnaire Design Lab (QDL) in May 2013. In the testing of survey questionnaires, Statistics Korea used expert review and a field pilot test before the QDL was set up. Although the evaluation and design of questionnaires is one of the most important aspects of the national statistical office’s survey and data production work, Statistics Korea could not consider the Cognitive Aspects of Survey Methodology until recently. In the mean time, because of the complexity and diversity surrounding the survey environment, mixed mode surveys have been employed more in the field of survey methodology. Mixed mode surveys need to match the characteristics of survey instruments and methods. The main purpose of the evaluation of survey questionnaires is to review where and how the questionnaires can be improved with respect to the response burden and data quality. The questionnaire redesign was applied to three example Statistics Korea questionnaires over the last two years. These were experimented using two methods, cognitive interviewing and eye-tracking. The experiment results can provide evidence for question quality in terms of the problem with structure, content and terms used in questionnaires. Statistics Korea conducts about 42 questionnaire-based surveys in Korea and plans to review these questionnaires one by one through systematic and standardized procedures using cognitive interviewing or eye-tracking. It is a challenging task but a good opportunity to improve the quality of official statistics in Korea.*Keywords: questionnaire, data quality, cognitive interviewing, eye-tracking*  |

1. **Introduction**

Statistics Korea (KOSTAT) has done extensive work in the areas of data collection, editing, and imputation since the mid-2000s. However, questionnaires have largely been overlooked even though they are an essential element in minimizing measurement error for an accurate data production. Most survey researchers undermined the importance of questionnaire design and evaluation. Statistics Korea has mostly used expert review and field pilot tests for the evaluation of questionnaires. Recognizing the importance of the questionnaire design based on the Cognitive Aspects of Survey Methodology, Statistics Korea has been proactively researching on a few major household surveys since 2011. Since the establishment of the Questionnaire Design Lab (QDL) in May 2013, the questionnaires for three national official statistics surveys have been redesigned, including the Population and Housing Census, Economically Active Population Survey and Time Use Survey (TUS). These questionnaires were tested using two methods, cognitive interviewing and eye-tracking beyond the traditional pre-test approach. The experiment results provide evidence for question quality, in terms of problems related to structure, content and terms used in questionnaires.

 This paper will look at the results of a recently tested Time Use Survey (TUS) based on cognitive interviews. First, we give a short background on the cognitive approach to the experiment in section 2 and present test experience of TUS in section 3. Then in section 4, we present analysis of experiment results. In section 5, we draw our conclusions with recommendations and plans for future research.

1. **Application of Cognitive Approach**

Mixed mode surveys have been employed more recently with the complexity and diversity surrounding the survey environment. Mixed mode surveys need to match the characteristics of survey instruments (paper, computer, tablet pc, etc.), considering whether the questionnaires will be filled out by the respondent alone or by the interviewer. The main purpose of the questionnaire evaluation that takes a cognitive approach is reviewing and identifying where and how the questionnaire can be improved with respect to response burden and data quality.

**2.1 Questionnaire Design Lab**

Since the establishment of QDL at Statistics Korea, studies have been conducted about the questionnaire response process in data production. Using the cognitive model[[2]](#footnote-3), the response process can be understood more extensively. The Lab includes two cognitive interview rooms, a usability test room equipped with an eye-tracking system, a spacious focus group interview room, and an analyzing room. The main task of QDL is experimentation and research for the consultation and redesign of questionnaires of national official statistics.

**2.2 Testing methodology**

For the proper use of the QDL, literature review and case studies have been conducted in the field of research concerning appropriateness of two main methodologies used for questionnaire evaluation.

 First, cognitive interviewing can be conducted in testing draft questionnaires with a small sample of paid volunteers. This is a qualitative method for detecting a wide range of potential sources of errors[[3]](#footnote-4) by rationalizing how respondents understood or interpreted questions and how they reach to their answers in accordance to the stages of the cognitive model, mainly remembering these two aspects as a qualitative method to estimate measurement error. The advantage of this method is the information it provides on the existence of problems in questions and its potential sources as well as solutions (Brancato et al., 2006).

 Second, eye-tracking is another good method that identifies issues related to visual and verbal aspects of survey instruments. However, this has been rarely utilized in survey methodological research because of the special instrument required for the experiment. With the advance of technology, today’s eye-tracking equipment no longer requires an awkward head-mounted helmet and is easy to use. With the combination of near-infrared beams that reflect off the retina and digital cameras that track the head position, the eye-tracker makes it possible to record eye movements with an adequate precision on many practical applications (Galesic et al., 2008). Statistics Korea’s QDL is set up with a Tobii X120 mobile device, which may be used with diverse computer devices, such as desktop computers, tablet PCs and smart phones.

1. **Cognitive Interview: Time Use Survey**

Statistics Korea conducted its first Time Use Survey (TUS) in 1999 and has been conducting it every 5 years since then. The QDL carried out the first cognitive experiment of the TUS in Feb. 2014. The main purpose of this study is to identify root causes of problems in the response process and to develop a questionnaire that is more accurate and easier for understanding prior to the upcoming field test. The TUS questionnaire is composed of 19 questions and a ‘time diary’, in which respondents provide a sequential list of their activities of the previous two days. The ‘time diary’ also includes duration of each activity and related information such concurrent activities, location and people with whom these activities were conducted. This survey is recognized as a complex one that gives respondents a considerable amount of burden.

**3.1 Method**

If enumerators were unable to conduct interviews face to face with respondents due to their absence from home, the questionnaire was self-administered by respondents. Since instructions on how to complete a ‘time dairy’ will affect the fidelity of the content, it is another important factor to review in this study. Therefore, we decided to experiment on the groups by controlling for the amount of instructions provided on the ‘time diary.’ Namely, we assumed that the group given with short instructions will find it more difficult to fill out the diary than the group provided with long instructions. Through this experiment, the results will be compared to differences of response time and the number of activities recorded for the day.

* + 1. **Experimental design**

In the first experiment, we recruited staff from Statistics Korea to assess the accuracy of respondent’s self-completed data and to measure the completion time. Prior to the second experiment, participants were recruited through Statistics Korea’s homepage, an advertisement in a weekly local newsletter, and by word of mouth. Each participant attended one session that lasted about one hour. Twenty selected respondents for the second experiment were randomly assigned to two groups. Table 1 shows the distribution of respondents’ characteristics over experiment conditions.

Table 1. Distribution of respondents

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Gender | Age group | Education | Household member | Work |
|  | Men | Women | 20-39 | 40-59 | College | University+ | 3-4 | 5+ | Yes | No |
| First (staff) | 3 |  5 | 3 |  4 | 1 |  7 |  8 | 0 | 8 |  0 |
| Second | 4 | 16 | 2 | 16 | 5 | 15 | 16 | 4 | 7 | 13 |
| ∙Treatment | 2 |  8 | 1 |  8 | 3 |  7 |  9 | 1 | 3 |  7 |
| ∙Control | 2 |  8 | 1 |  8 | 2 |  8 |  7 | 3 | 4 |  6 |

\* Four single person households included

* + 1. **Interview protocol**

The interview protocol for cognitive probing consists of five elements. First, we have the original form of questionnaire ready. Secondly, we prepare a list of potential cognitive issues to be investigated in the questionnaire, and thirdly we select appropriate probing questions to ask for each item. Then based on the results collected from probing, potential sources for error in the course of the cognitive test on each question item are identified, analyzed and summarized. Lastly, suggested revisions are provided based on experiment results. Scripted scenarios for probing questions are prepared in advance to minimize potential errors by interviewers, who are also trained through role play exercise

 Probing methods used for different question items in this study are as follows:

* Paraphrase: status on separated family and reasons
* Comprehension and paraphrase: members living in the same household, reference date for the ‘time diary’
* Response difficulty: status on separated family and reasons, household income, working hours
* Response accuracy: total floor area of the house, household income, working hours
* Reference to explanation/classification: status on separated family and reasons, unpaid family workers, and the ‘time diary’
	+ 1. **Interview process**

As shown in Figure 1, the interview process can be composed of four steps. The second experiment was designed by forming treatment (short instruction) and control (long instruction) groups to examine differences of results which were dependent upon the amount of information provided. Namely, interviewers spent a short time (around 1 minute) explaining completion instructions to the treatment group before participants responded to the questionnaire whereas the control group received more detailed explanation (less than 5 minutes) by interviewers.

Figure 1. Interview process

 The results show statistically significant differences in time spent explaining about the questionnaire (p-value: 0.000) and reading instructions on the ‘time diary’ (p-value: 0.013). The cognitive interview was composed of 21 probing questions and was conducted by interviewers who asked one to three questions in a group on selected items in order to listen to thoughts and opinions of participants. Using a standard probing scenario, the probing time lasted about 26 minutes, similarly across the three groups (See Table 3).

1. **Analysis Results**

**4.1**

**Analysis of test time**

It took a total experiment time of 66.4 minutes on average (See Table 2). The age group 40 to 50 took more time[[4]](#footnote-5) (4.6 minutes) than the age group 20 to 30. Employed and non-employed groups spent a similar amount of time on the experiment, but the employed group completed the ‘time diary’ quicker than the non-employed group while the employed group took a longer time probing than the non-employed group. There appears to be a trade-off between the two tasks, which could have resulted from respondents’ fatigue.

Table 2. Experiment time by characteristics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | n | Totaltime | Explaining about questionnaire(interviewer) | Completingquestions | Reading instructions on ‘time diary’(respondent) | Writing‘time diary’ | Probing |
| Total | 28 | 66.4 | 2.1 | 11.9 | 1.3 | 24.9 | 26.2 |
| Men | 7 | 74.9 | 1.8 | 13.3 | 1.6 | 23.9 | 34.3 |
| Women | 21 | 63.6 | 2.3 | 11.4 | 1.2 | 25.2 | 23.5 |
| 20-39 | 9 | 63.3 | 1.6 | 11.7 | 1.1 | 21.0 | 27.8 |
| 40-59  | 19 | 67.9 | 2.4 | 11.9 | 1.4 | 26.7 | 25.5 |
| Employed | 17 | 66.3 | 2.0 | 11.9 | 1.4 | 23.6 | 27.4 |
| Non-employed | 11 | 66.6 | 2.4 | 11.7 | 1.1 | 26.9 | 24.4 |

 Because the first experiment group (staff) and treatment group in the second experiment completed the questionnaire using the self-completion method without interviewer’s intervention, the control group in the second experiment, which had interviewers reading instructions to them in detail, spent 3.1 minutes longer and relatively less time in reading instructions on the ‘time diary’ for themselves.

Table 3. Experiment time by group

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | n | Totaltime | Explaining about questionnaire(interviewer) | Completingquestions | Reading instructions on‘time diary’ (respondent) | Writing‘time diary’ | Probing |
| First(staff) | 8 | 56.6 | 1.1 | 9.8 | 1.3 | 18.8 | 25.6 |
| Second | 20 | 70.4 | 2.6 | 12.7 | 1.3 | 27.3 | 26.5 |
|  | Treatment | 10 | 67.1 | 1.0 | 12.2 | 1.9 | 25.3 | 26.7 |
|  | Control | 10 | 73.8 | 4.1 | 13.2 | 0.7 | 29.6 | 26.2 |

* 1. **Probing results**

To code the problems that were found through cognitive interviewing, we adopted a coding system to reflect the cognitive model (Willis et al., 1999). Then the problems identified were classified using this coding system[[5]](#footnote-6).

* + 1. **Comprehension**

It is very important to understand the terminology and wording in the encoding process.

 **▪ Comprehension of Terms**

On the ‘household income’ item, the main issues are to investigate the degree of response difficulty and reference to term descriptions. The probing questions[[6]](#footnote-7) were prepared based on the issues identified. About half of the participants (13/28) could understand questions without difficulty but several participants found it difficult to compute bonuses including annual incentives and to distinguish pre-tax from post-tax income. Some participants did not understand the term ‘transferred income’ so an example as added to the description (See Table 4).

Table 4 . Summary of probing results on ‘household income’ (example)

|  |  |
| --- | --- |
| 1. Original question | Last year, how much was the monthly gross household income before the tax deduction? |
| 2. Experiment results | * About half of the participants understood the question
* Some did not understand ‘transferred income’ in income description
 |
| 3. Suggested revision | * Specify ‘transferred income’ with examples

(Ex. pension, premium, living expenses form other household, etc.) |

 **▪ Usage of Term Descriptions**

Some questions provided information such as explanations or examples on terms to get a more accurate answer and to reduce response burden on respondents. We reviewed how efficiently term descriptions were being used in the actual response process. When it comes to the usage of descriptions, more than half of the participants answered to have used references on ‘head of household’, ‘household income’ and ‘employment status’, but rarely so on ‘unpaid family workers’ (See Table 5). It is likely that respondents did not read the additional information on ‘unpaid family workers’ since regularly paid workers could easily respond to the question without the reference.

Table 5. Reference to term descriptions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Head of household | Household income | Unpaid family workers | Employment status |
| Reference  | 23/28 | 23/28 | 8/28 | 17/21\* |

\* If only employed.

**▪ Respondents’ Rating**

Some probing questions evaluated the degree of difficulty and confidence related to questions and term descriptions. The difficulty was measured on a 10-point scale (1=easy, 10=difficult). Respondents evaluated how easily they could understand four items (See Table 6). The assessment results on ‘satisfaction on time use in the ‘time diary’’ (2.75) was much lower than the average (5.5, mid-point), which shows that this question was easily understood. The score on the ‘status on separated family and reasons’ was the highest (3.82) across four items reviewed. The reason is that it asked not only the general understanding of the question itself but also the reasons in a sequential question separately. It could be inferred that participants felt more response burden from logical questions in order. The results of the paraphrasing question show that participants understood questions well.

Table 6. Level of difficulty on questions and descriptions

|  |  |  |
| --- | --- | --- |
| Item | Probing questions | Score |
| Status on separated family and reasons | · How hard was this question to answer? | 3.82 |
| Average monthly household income | · How hard was this description to understand? | 3.24 |
| Unpaid family worker | · How hard was this description to understand? | 2.57 |
| Satisfaction on time use in ‘time diary’ | · How hard was ‘the day when you wrote the time diary’ to understand? | 2.75 |

\* Rated on a 10-point scale (1= definitely easy, 10 = definitely difficult)

* + 1. **Retrieval and computation**

The retrieval process often requires recalling information to compute answers. A good example is the biographical questions that rely heavily on the retrieval process (Eisenhower et al., 1991). Based on probing[[7]](#footnote-8) results of questions related to ‘working hour’, more than half of respondents (15/21) marked “no problem”, but some said that they have been confused by terms such as ‘main job’ and ‘second job’.

Table 7 shows the results of respondent’s confidence level on retrieval items. The scores on three items range from 7.44 to 8.81 on a 10-point scale (1=least confident, 10=most confident). Even for an open-ended question, ‘working hours’ scored the highest since the majority of respondents were regular workers. In the ‘total floor area of the house’ item, problems were identified such as “don’t know” and “unit of conversion.” The ‘household income’ item scored higher than expected for a close-ended question even though some participants faced a computational problem.

Table 7. Level of confidence on retrieval items

|  |  |  |  |
| --- | --- | --- | --- |
| Item | Probing questions(confidence rating) | Score | Remarks |
| Total floor area of the house | How certain are you of your answer? | 7.44 | Retrieval and unit conversion |
| Household income | How certain are you of your answer? | 8.11 | Response category use |
| Working hours | How certain are you of your answer? | 8.81 | Most respondents were regular worker |

\* Rated on a 10-point scale (1=definitely not certain, 10= definitely certain)

* + 1. **Response category**

Several question items were tested to identify cognitive problems on response categories. Table 8 shows how some of the response categories were revised based on probing results.

Table 8. Probing results of response categories

|  |  |  |
| --- | --- | --- |
| Item  | Probing questions | Remarks |
| Reasons for time shortage | How easy or hard was it to find your answer on the list? | One response category deleted  |
| Reasons for feeling tiredness | How easy or hard was it to find your answer on the list? | Some response categories modified  |
| Where in ‘time diary’ | Vignette classifications | Some categories modified:Walking → Moving on foot Other places → Other places (Ex. stores) |
| With whom in ‘time diary’ | Vignette classifications | Some categories modified:Other children → 10 years or older children |

* 1. **Additional analysis**

The analysis results on the number of activities recorded by characteristics show statistically significant differences on gender and employment status (Table 9).

Table 9. Completion time and number of activities in ‘time diary’ by characteristics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | n | Completion time | T | p-value | No. of Activities | T | p-value |
| Total | 28 | 24.9 |  |  | 31.1 |  |  |
| Men | 7 | 23.9 | 0.112 | 0.911 | 24.1 | -4.28 | 0.000 |
| Women | 21 | 25.2 | 33.4 |
| 20-39 | 9 | 21.0 | 1.047 | 0.305 | 29.2 | -0.82 | 0.42 |
| 40-59 | 19 | 26.7 | 32.0 |
| Employed | 17 | 23.6 | -0.86 | 0.932 | 27.1 | -3.85 | 0.002 |
| Non-employed | 11 | 26.9 | 38.2 |

 In this experiment, we reviewed the reliability of responses in the ‘time diary’ through the number of activities recorded and efficiency in completion time of the ‘time diary’. The results are as follows:

* Will completion time and the number of activities recorded in the ‘time diary’ be different across experiment groups? (Both are not statistically significant.)
* Will completion time and the number of activities recorded in the ‘time diary’ be different depending on the time pressure that respondents usually feel? (Both are not statistically significant)
* Will completion time and the number of activities recorded in the ‘time diary’ be different between reader and skimmer groups[[8]](#footnote-9)? (Significant and not significant, respectively)

The results on the number of activities show no statistical significance in comparison to other groups (See Table 10). Completion time for the ‘time diary’ also matched the number of activities in the experiment and time pressure groups, but presented statistically significant results between reader and skimmer groups (p-value = 0.035). Namely the readers used more time to complete the ‘time dairy’ than the skimmers, but it is debatable whether the former group is less efficient than the latter group.

Table 10. Completion time and number of activities in ‘time diary’ by groups

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | n | Completion time | T(F) | p-value | No. of Activities | T(F) | p-value |
| Staff | 8 | 18.77 | 1.76 |  | 26.12 | 2.30 | 0.121 |
| Treatment | 10 | 25.28 | 0.194 | 33.60 |
| Control | 10 |  25.98\* |  | 32.50 |
| Pressure | 16 | 25.29 | 1.21 | 0.239 | 31.94 | 0.64 | 0.527 |
| No pressure | 12 | 21.10 | 29.92 |
| Reader | 8 | 26.99 | 2.30 | 0.035 | 31.25 | 0.42 | 0.678 |
| Skimmer | 10 | 18.70 | 29.50 |

\* The data for the participant with longest completion time was excluded.

1. **Conclusion**

Through this experiment, we could examine the cognitive response process of the Time Use Survey (TUS) and predict the average response time and identify problems in the questionnaire. On average, it may estimate about 64 minutes[[9]](#footnote-10) to complete the questionnaire in cases where respondents were asked to write a ‘time diary’ for two day. The average time for 21 probing questions which assessed respondents’ comprehension of terms and descriptions, retrieval and computation of information, and understanding of response categories was 26 minutes.

 In conclusion, the cognitive interviews based on probing results revealed interesting insight into questionnaire design of the TUS. Respondents found it less difficult to understand questions or terms descriptions on the comprehension stage and were more confident about their responses on the retrieval and computation stage. On the response stage, through probing on the easiness and vignette classifications on the validity of response categories some response categories were modified or deleted.

The study shows extensively that respondents had difficulties understanding terms and where to write the answers. It was clearly shown that respondents did not read instructions thoroughly on the ‘time diary’ as expected. Therefore, instructions in the questionnaire need to be improved and presented more user-friendly by using visuals, as an example.

This experiment shows that efficiency and integrity of the survey are two key influencing factors to completion time and the number of activities in the ‘time diary’. In comparison to other groups, the results related to the ‘time diary’ show statistically significant differences in completion time of only reader/skimmer groups and the number of activities recorded on the reference day by gender and employment status. Although the readers spent more time on the ‘time dairy’ than the skimmers, it is debatable whether the former is less efficient than the latter. It is an issue to be studied further in the future.

Though cognitive interviewing generally only identifies problems and does not guarantee statistical significance of the results (Willis, 2005), it is a cost effective method for collecting quality information with a small sample of population compared to a large scale of a field pretest. The experiment results can provide insights to question quality in terms of problems with structure, content and terms used in the questionnaire. Statistics Korea covers about 42 questionnaire-based surveys for the production of national statistics data. Statistics Korea plans to review these questionnaires one by one through systematic and standardized procedures using a cognitive interviewing and eye-tracking. It is a challenging task but a good opportunity to improve the quality of official statistics in Korea.

**References**

[1] Brancato, G., Macchina, S., Murgia, M., Signore, M., Simeoni, G., Blanke, K., Körner, T., Nimmergut, A., Lima, P., Paulino, R., Hoffmeyer-Zlotnik, J.H.P., (2006). Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System, European Statistical System.

[2] Biemer, P. P. and Lyberg, L. E. (2003), Introduction to Survey Quality, Hoboken, NJ: John Wiley & Sons.

[3] DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey researc*h (pp. 177-195). San Francisco: Jossey-Bass.

[4] Eisenhower, D., Mathiowetz, N.A., and Morganstein, D. (1991), Recall Error: Sources and Bias Reduction technique. In P. Biemer, R. M. Groves, L. E. Lyberg, N.A. Mathiowetz, and S. Sudman(Eds.), *Measurement errors in survey*(pp.127-144). New York: John Wiley & Sons.

[5] Galesic, M., Couper, M., Tourangeau, R. and Conrad, F. (2008) Eye-tracking data : New insights on response order effects and other signs of cognitive shortcuts in survey responding. Public Opinion Quarterly, 72(5).

[6] Tourangeau, R. (1984). Cognitive science and survey methods, In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), Cognitive aspects of survey methodology: Building a bridge between disciplines (pp. 73-100). Washington, DC.: National Academy Press.

[7] Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2004). Survey methodology. Hoboken, NJ: John Wiley & Sons.

[8] Willis, G. B. (1999). Cognitive Interviewing: A “How to” Guide. Research Triangle Institute. <http://appliedresearch.cancer.gov/archive/cognitive/interview.pdf>.

[9] Willis, G. B. (2005). Cognitive Interviewing: A Tool for Improving Questionnaire design, Sage Publication: Thousand Oaks, California.

1. With special thanks to Jinwoo Bae and HyeJin Kim. [↑](#footnote-ref-2)
2. The response process can be formulated differently, although it can be done following similar stages as suggested (Tourangeau, 1984; Eisenhower et al., 1991; Biemer and Lyberg, 2003; Groves et al., 2004); encoding, comprehension, retrieval, judgment and reporting (Brancato et al,. 2006). [↑](#footnote-ref-3)
3. There are potential sources of specification error, measurement error related to respondents, and non response error as well as mode effect (DeMaio and Rothgeb, 1996; Biemer and Lyberg, 2003; Brancato et al., 2006). [↑](#footnote-ref-4)
4. The longest writing time was 59.4 minutes by a non-employed woman in age group 40, followed by 38.5 minutes by an employed man in age group 40, while the shortest writing time was 8.4 minutes by an employed man in age group 30. [↑](#footnote-ref-5)
5. Developed by Willis and colleagues through cognitive interviews of NCHS (National Center Health Service) and NORC (National Opinion Research Center) (Willis, 2005). [↑](#footnote-ref-6)
6. “Did you have any difficulty in responding to this question? If so, what is it?”, “Did you read the description under the question?” [↑](#footnote-ref-7)
7. “Did you have any difficulty answering? How did you get to your answer? Describe the response process in your own words.” [↑](#footnote-ref-8)
8. The skimmer group was classified by reading time for instruction on the ‘time diary’ under 100 seconds from the first experiment (staff) and the second experiment (treatment group), vice versa for the reader group with a threshold of 135 seconds. [↑](#footnote-ref-9)
9. It includes reading time for instructions (2 min.) and the completion time for questions (12 min.) and time diary (25 min. per day). [↑](#footnote-ref-10)