

Using Paradata to Evaluate the Effect of Changes in the Small Screen Layout

Elise Alstad and Erdal Kilicdogan, Statistics Norway

1. Abstract

With Blaise 5 multimode capabilities, we can create dynamic and adjustable layouts for different screens—including layout for small screens—which is important because more respondents are using smart phones to respond to our surveys. Understanding the best design practices for small screens becomes essential as we aim to improve our surveys on mobile and to use a mobile-first approach when designing new surveys. A significant challenge when adapting web surveys initially designed for pen and paper or PC is to adjust questions in a group that are presented in a grid table to small screens. As grid tables are unsuitable for small screens, we usually change the presentation of the group questions to item-by-item questions on small screens. The Quality of Life Survey (QLS) is a web survey that uses grid tables to present question groups on PC, and between 2022 and 2023, we attempted to improve the small screen layout for these question groups. We transitioned from a pagination design, where each question is presented on a single page, to a scrolling design, where several questions are on the same page. Hoping to have reduced respondent burden and improved user experience by transitioning from pagination to scrolling, we will assess this hypothesis by comparing paradata indicators of QLS 2022 and 2023. We have focused on the following indicators in our analysis: (1) previous page actions, (2) response times, (3) breakoff rates, and (4) route errors. We find lower rates of previous page clicks, route errors, and breakoffs among mobile respondents in the 2023 QLS compared to the 2022 QLS. These results could indicate that the scrolling layout might be more suitable for small screens than the pagination layout. However, we see that we need to do controlled A/B testing of the layout to arrive at a solid conclusion, and we discuss how we aim to conduct these tests in the future.

2. Introduction

At Statistics Norway, we see that a majority of web responses are completed on small screens in various web surveys for individual persons (Pettersen & Engvik, 2022; Holmøy & Rossbach, 2021). Consequently, we aim to have a mobile-first approach when designing new surveys and to enhance our current web surveys for mobile use. Mobile phones may introduce other sources of survey error, which are not as prominent on PC. For instance, on small screens, response burden may increase if respondents have to scroll and zoom to see the questions, or if respondents have to handle small slider scales or radio items to select their responses. Therefore, it is important that the questionnaire and the layout are designed such that it is easy for the respondent to respond to the questions on small screens.

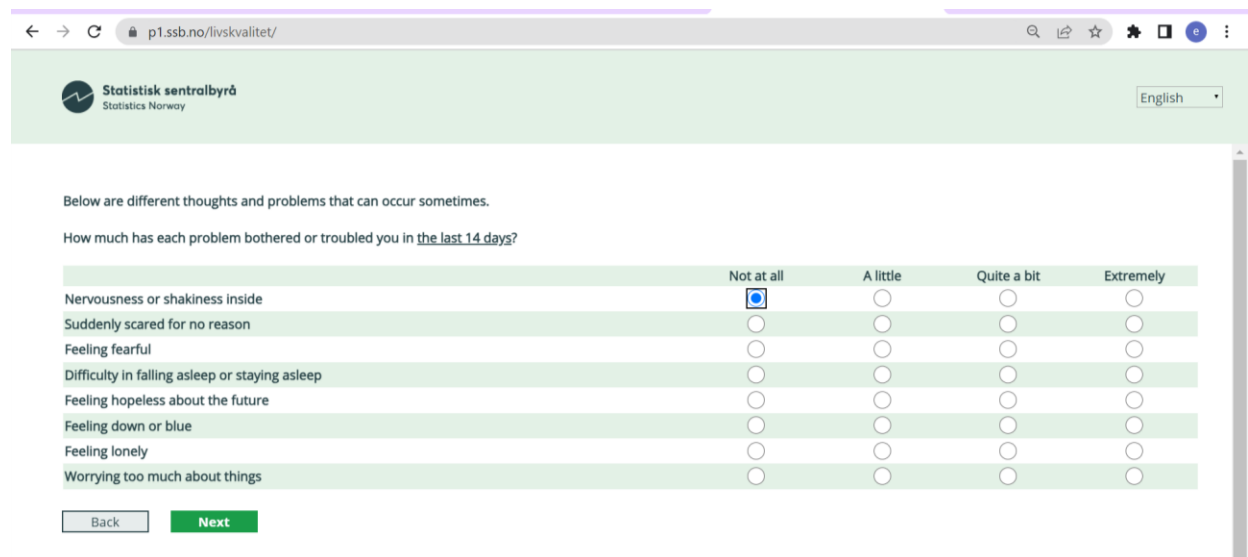
Our team has experienced that adapting our web surveys to small screens can be challenging if the web surveys use several large grid tables. Grid table is a format widely used to present a group of questions with the same answer categories, as it is an efficient use of space and text. But grids tables are unsuitable for small screens because they take up too much space, requiring the user to zoom and scroll. Additionally, the radio buttons in the grid tables are too small for small screens to comply with accessibility standards. It has been recommended to transform grid questions to item-by-item questions for mobile (Vehovar et al., 2022), but also on PC, to ensure comparability between mobile and PC respondents (Revilla et al., 2017). At Statistics Norway, we have, for most web surveys, adjusted the format for grid questions into item-by-item questions on small screens.

The Quality of Life Survey (QLS), which has been conducted yearly since 2020, uses several grid tables. Out of 175 question fields in the QLS survey questionnaire, 55 question fields are presented using grid tables in the PC layout. There are a total of 9 grid tables for 9 question groups in the PC layout of the

survey. These question fields are in groups, but they are presented as item-by-item questions and not in a grid table in the small screen layout, and we will therefore, for the sake of ease, refer to these specific questions of the QLS questionnaire as the “question groups” in this article. Between 2022 and 2023, we attempted to improve the small screen layout of the question groups by changing the presentation from a pagination design to a scrolling design. We utilized paradata from the Blaise audit trail to evaluate the effect of changing the layout.

Prior to the 2023 QLS, we assessed how we could improve the small screen layout of the question groups. Figure 1 shows a question group presented in a grid table in the large screen layout. The same question group in the small screen layout was presented in an item-by-item format with a pagination design and one item per page, as shown in Figures 2, 3, and 4. We suspected that the pagination layout might be unideal for the question groups because the introductory group question text was only visible on the page with the first question in the group. If respondents forget the group question text in the middle of the question group, it could be a substantial response burden to go back through previous pages to see the group question text. We therefore evaluated how we could improve the presentation of the question groups in the small screen layout.

Figure 1. Screenshot of Quality of Life Survey 2022 and 2023, Chrome Browser



We did a review of the literature to explore mobile alternatives for grid questions. A common mobile alternative to the grid table presentation is item-by-item in a scrolling layout, where all the group questions are presented on the same page. De Bruijne and Wijnant (2014) compared different alternative layouts for grid questions and saw a shorter completion time for a scrolling layout compared to a pagination layout. Similarly, Mavletova and Couper (2014) found that a scrolling layout seemed to be more suitable than pagination when assessing breakoff rates, reports of technical problems, and respondent rating of the questionnaire.

Additionally, we tested scrolling, pagination, and other layout alternatives within the team. Most of the team members were positive towards changing the small screen layout to a scrolling presentation, as they found it easier and more preferable to navigate by scrolling rather than having to navigate by using the “Next” and “Previous” page buttons. Thus, we decided to adjust the small screen layout of the question groups from pagination to a scrolling layout in the 2023 QLS (see Figures 5, 6, and 7). As emphasized by Cheung et al. (2016), the layout of a survey can substantially influence how respondents perceive and

respond to the questionnaire. To assess if the scrolling layout is more suitable for small screens than the pagination layout, we assess four indicators that aim to measure response burden using paradata from the Blaise audit trail; namely, we compare the level of (1) previous page actions, (2) response times, (3) breakoff rates, and (4) route errors received between 2022 and 2023 to see if adjusting the small screen layout could reduce response burden and improve user experience on mobile.

Figures 2–4 show the Quality of Life Survey 2022 in the iPhone 11 Safari browser.

Figure 2. View When Entering the Page for the First Question, Which Includes the Group Question Text

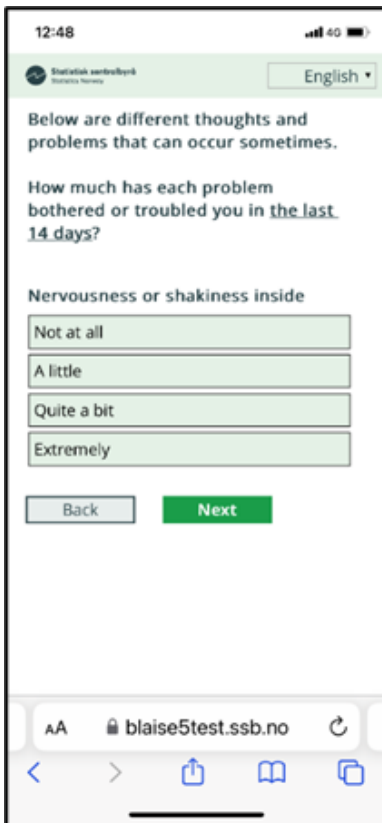


Figure 3. View of the Next Page

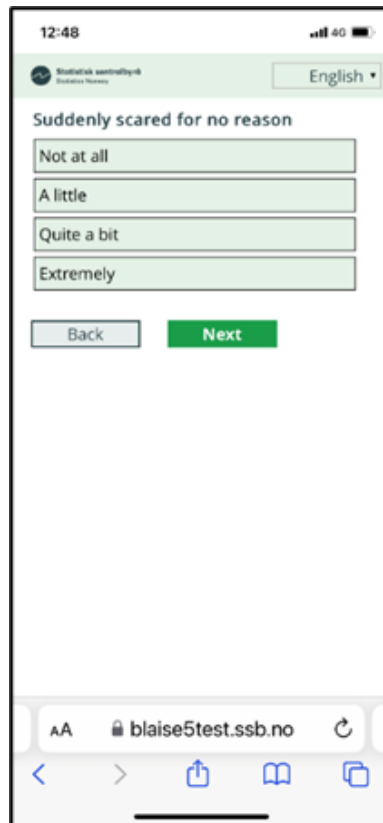


Figure 4. View of Third Group Question and the Route Error That Arises When the Respondent Attempts to Move to the Next Page without Responding to the Question



Figures 5–7 show the Quality of Life Survey 2023 in the iPhone 11 Safari browser.

Figure 5. Immediate View When Entering the Page

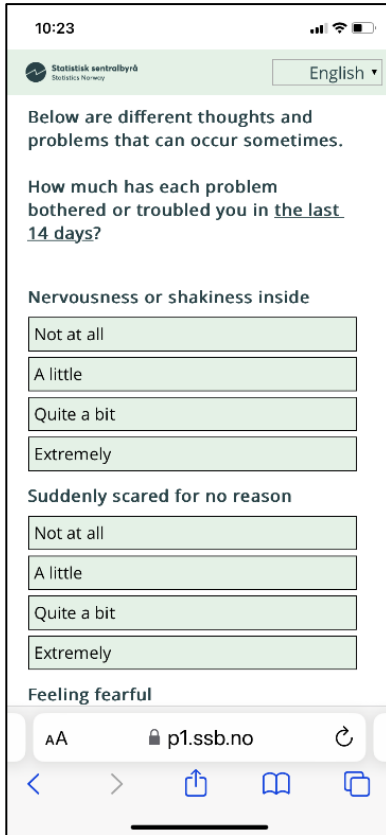


Figure 6. View in the Middle of the Page When Two Answers Are Selected

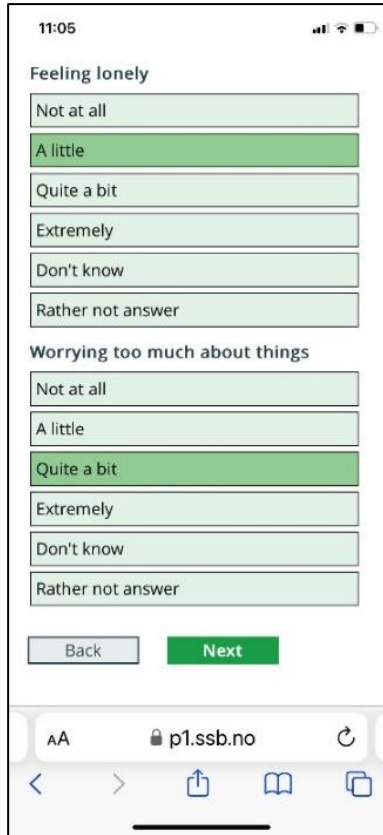
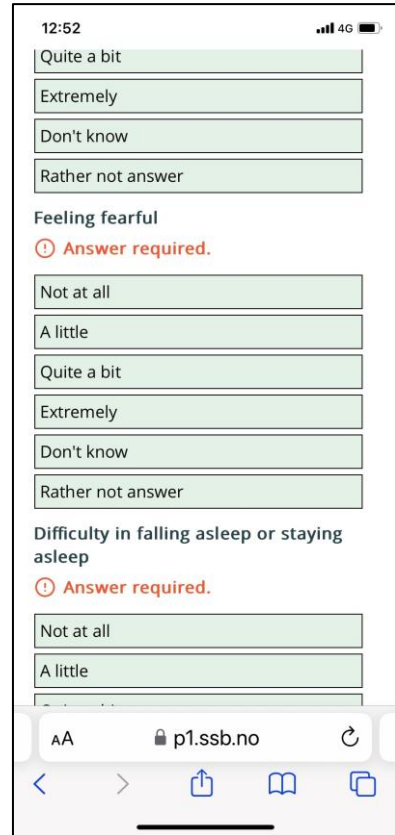


Figure 7. View in the middle of the Page When Two Unanswered Questions Trigger a Route Error



3. Paradata Indicators

Paradata from the Blaise audit trail can provide useful insights into how respondents move through the questionnaire and can be used to identify issues respondents experience. Paradata from the audit trail show the respondents' actions in the survey and the timestamp for each action. Thus, by using paradata, we can gain a greater understanding in areas of concern and learn how to improve survey questionnaires, potentially resulting in higher response rates, more accurate responses, and enhanced user experience. In this article, we use paradata to assess specific indicators that aim to measure response burden. We will look at (1) previous page actions, (2) response time, (3) breakoff rates, and (4) route errors. By assessing these indicators on activity within the question groups with the changed layout, we evaluate if changing from a pagination layout to a scrolling layout can influence how the respondent interacts with the survey and if it can help to reduce the response burden.

3.1 Previous Page

Previous page can be an especially useful indicator, as it can illustrate issues with the flow of the survey and identify problems with survey questions. Qualitative user tests at Statistics Norway have found that respondents go to the previous page for various reasons; namely, if the question is perceived as too

similar to the previous one, the respondent may go back again to the previous question to understand the difference between the current and the previous question. Additionally, respondents go to the previous page if they recognize, when presented with a question, that the previous question on another page was answered incorrectly. Importantly, respondents may click to the previous page because the current question is missing some information, and the respondent returns to the previous page to retrieve the missing information.

In the pagination layout, the group question text is shown only on the page with the first field, and the following group fields are missing important information. This may cause the respondent to return to the previous page, resulting in a higher number of previous page clicks for respondents in the pagination layout than in the scrolling layout. We can find the previous page actions in the Actions variable in the audit trail data. By counting the number of “PreviousPage” actions within the “Action” variable for each respondent, we can find the average number of times respondents clicked previous page within the question groups.

Hypothesis 1: Respondents have a higher frequency of previous page actions in the pagination layout (2022) than in the scrolling layout (2023).

3.2 Response Time

Response time is frequently used to assess respondent behavior and to infer results regarding the quality of the responses respondents provide. For instance, very short completion time is associated with careless responding (Leiner, 2019). On the other hand, a long response time may suggest respondents experience difficulties answering the questions; namely, Antoun et al. (2017) found that mobile respondents spent more time responding to questions, and at the same time, the answers they provided were less accurate. Thus, long response time can be related to a greater response burden and can indicate that respondents experience difficulties.

Time used to scroll versus navigating between pages could provide different completion times. Scrolling layout has been found to have shorter completion time compared to pagination layout (de Bruijne & Wijnant, 2014; Mavletova & Couper, 2014). A possible explanation for this is that it is quicker to scroll than to click to the previous or next page, and each page must also be loaded. Thus, we would expect to see lower completion time in 2023 with the scrolling layout. If we were to see higher response time for questions that were changed in the pagination layout, it may indicate that respondents experienced higher response burden.

Hypothesis 2: Response time is lower in the scrolling layout (2023) than in the pagination layout (2022).

We estimate the total response time for the question groups where the layout was changed using the TimeStamps in the paradata. Both rounds of the QLS had AutoEnter on the mobile buttons, such that the respondent is automatically and immediately moved to the next item after providing a response, either by moving to the next page or by auto focusing the view on the next question on the page.

3.3 Breakoff

Breakoff refers to when respondents initiate a survey but do not complete the entire questionnaire. A survey breakoff that occurs before completion or before a predetermined limit of the number of questions to be answered is often categorized as nonresponse. Researchers point out that respondent-related or survey-related aspects can cause increasing breakoff rates. For example, lack of motivation, higher task difficulty, and technical problems are some reasons which can increase the probability of breakoff (Steinbrecher et al., 2015).

Analyzing breakoff rate is an important assessment of survey quality and user experience. If respondents feel the response burden is very high or experience issues, they could decide to not complete the survey. On mobile, the respondents may become frustrated when they have to scroll and zoom to see the question text or if the question text is hard to read on the small screen (Cheung et al., 2016). Vehovar et al. (2022) compared the breakoff rates of different layouts for mobile grid alternatives and saw a higher breakoff rate in the pagination layout in comparison to a scrolling layout and other layout alternatives. We therefore expect that the pagination layouts will result in higher breakoff rates than the scrolling layouts. We consider breakoff rate an important indicator because a difference in breakoff rates may indicate that changing the layout can reduce the response burden. In our analysis, we assess breakoff rates by finding the number of respondents that break off the survey within the question groups with the changed layout.

Hypothesis 3: The breakoff rate in the pagination layout (2022) is higher than in the scrolling layout (2023).

3.4 Error Messages

One possible disadvantage with a scrolling design is that respondents may be more likely to miss a question when there are several questions on one page. For instance, de Bruijne and Wijnant (2014) found an indication of more item nonresponse in the scrolling layout than in the pagination layout; however, the difference was not significant. Mavletova and Couper (2014) did not find any difference in item nonresponse when comparing pagination and scrolling layouts.

In the QLS, response is required to move to the next page, and it is therefore not relevant to assess item nonresponse. Respondents who missed a question would receive a route error informing them to reply to the question before moving on to the next page (see Figure 7). Thus, if it was easier to miss a question in the scrolling layout, we can determine this with the frequency of route errors experienced. Moreover, in the scrolling layout, if respondents tried to move to the next page when they had missed a question, they would have had to scroll back up to reply to the question and then scroll down to move to the next page. This may have been a substantial response burden, especially in question groups with many fields. In the pagination layout, as each item is on a single page, feedback is more immediate if the respondent misses a question.

To examine if it was easier to miss a question in the scrolling layout, we assess the average number of route error messages received per respondent and the share of respondents who received at least one route error. From the audit trail, we find route errors using “Route” in the variable “ErrorKind.”

Hypothesis 4: There are more frequent experiences of route errors and a higher share of respondents receiving route errors at least once in the scrolling layout (2023) than in the pagination layout (2022).

4. Methodology

The Quality of Life (QLS) survey has a sample size of 40,000 people each year, and the sample is drawn to be representative of the Norwegian population aged 18 and above. Using paradata and defining complete response as responding to the last question in the survey, the response rate was 38% in 2022 and 45% in 2023. Between the two years, the questionnaire remained very much the same, except for the change in the small screen layout. However, the 2022 QLS and the 2023 QLS used different contact methods, which may have contributed to the increased response rate in 2023. In 2022, we sent the respondents the invitation to the survey and the link to the questionnaire via both email and SMS (Pettersen & Engvik, 2022). In 2023, we transitioned to using Altinn, the digital mail inbox where public agencies can communicate with individuals and businesses. In Altinn, individuals will, for example, receive information about their tax returns. Users are required to use two-factor authentication to log in to Altinn. In 2023, the invitation to the survey and the link to the questionnaire were sent in Altinn.

Respondents would get notifications about receiving an Altinn letter from Statistics Norway and reminders about the survey via both SMS and email, but the link to the survey was only sent in Altinn.

By comparing mobile respondents between the two years, this is essentially an observational study that will have limitations regarding self-selection effects. In particular, the contact channel where respondents receive a direct link to the questionnaire might influence the choice of device. For example, de Bruijne and Wijnant (2014) and Mavletova and Couper (2014) found that sending the survey invitation and the link to the questionnaire via SMS led to a higher share of respondents choosing mobile compared to sending them via email. Thus, changing the channel where respondents received the direct link to the survey could influence whether they choose to complete the survey on mobile or on PC. In the QLS 2022, when respondents received the link via SMS and email, 72% of completed surveys were done on mobile, while only 54% were done on mobile in 2023 (see Table 1). As the different contact channels may influence the respondents' choice of device, it may also influence the composition of the net sample on mobile. We see a difference in the composition of the mobile respondents between the two years, especially regarding age, as shown in Table 1. Respondent-related factors could have an impact on the measures we are analyzing; for instance, Zhang et al. (2014) found a relationship between speeding and straight-lining for people with lower education levels, and also for gender. Thus, the different mode of contact between the years could lead to different people choosing to use mobile to respond to the survey, resulting in two different mobile samples between the years. We attempt to reduce confounding effects by conducting exact matching on demographic variables, which we will describe further in the Data Preparation section. But the limitation of not conducting a randomized controlled experiment is something we recognize and will discuss further in the Limitations section.

Table 1. QLS 2022 and 2023 Descriptive Statistics of Sample

Year	2022 (Pagination Layout)			2023 (Scrolling Layout)			2022 vs 2023 (Pagination vs Scrolling)
Category	N	Completions (% of N)	Mobile Completions (% of Completions)	N	Completions (% of N)	Mobile Completions (% of Completions)	X2 (df)
Gender							13.434*** (df=1)
Male	20,385	36	65	20097	43	47	
Female	19,615	40	79	19903	47	64	
Age							179.710*** (df=4)
18-24	4,209	31	85	4,243	40	76	
25-44	13,694	34	81	13,760	43	68	
45-66	13,975	45	68	13,854	52	49	
67-79	5,814	42	60	5,902	45	37	
80+	2,307	19	56	2,241	20	32	
Education							20.006*** (df=3)
Elementary school or lower	9,028	24	80	8,810	30	66	
Upper secondary school	15,467	36	73	15,445	45	55	
University/ University college	13,724	51	69	1,3724	57	53	
Not stated	1,781	18	71	1,226	29	51	
Total	40,000	38	72	40,000	45	55	

*p < 0.05; **p < 0.01 ***p < 0.001.

5. Data Preparation

We removed sessions with contradicting screen size and layout, as well as respondents with missing information in LayoutSetName. We also excluded respondents that replied to the survey with two different LayoutSetNames. As the interest of the analysis is mobile devices, we will only examine respondents with a small screen layout, and therefore exclude PC respondents.

When assessing breakoff rates, we include all respondents who started the survey by responding to the first question. For the other indicators, we aim to assess activity on the question groups, and we therefore filter to include respondents who responded to at least one of the fields in the middle of the page. This is to ensure that the respondents included in the analysis responded to all the selected question groups we assess. These samples are the foundation of the matching process, which we describe later.

Additionally, we filtered on the middle questions in the question groups because of missing paradata. We were informed by the Blaise support team that missing paradata could occur because of the workings of the Safari browser. As the Safari browser does not always focus on items being clicked and as the audit trail relies on focus being triggered, actions from the Safari browser can be missing. We have not been able to effectively establish if a respondent is unaffected by missing paradata, as absence of activity is hard to determine. Because a substantial share of respondents uses the Safari browser and the missingness

does not apply to all observations using this browser, we did not want to exclude all Safari browsers from the analysis. Still, we noticed a pattern for missing paradata: activity on the first item was registered but not on the subsequent fields within a page. Therefore, we only included respondents who had responded to at least one of the fields in the middle of the page. As part of the sensitivity analysis, which we discuss in the Results and Data Analysis section, we also ran the analysis while excluding Safari browsers to see if this influenced the results.

As we mentioned in the Methodology section, this is an observational study, and the composition of the mobile samples might be different between the two years. We therefore conducted exact matching to adjust for possible confounding effects due to the different contact methods. The objective of the matching was to construct a subsample of the mobile respondents where the demographic composition would be identical between the two years. To do the exact matching, we stratified the sample of mobile respondents based on the categorical demographic variables gender, education, and age. The categories are shown in Table 4 in the Appendix. The stratification provided us with a total of 40 strata, yet we only used 36 of them because not all strata contained mobile respondents in both 2022 and 2023. In each stratum, we found which year had the smallest number of mobile respondents in the stratum and used that as the benchmark sample size for the stratum. Then for each stratum, we selected a random sample from the year with the highest number of mobile respondents in the stratum. For the breakoff rate analysis, the foundation for the matching process was respondents who answered the first question in the survey. For the other indicators, we selected respondents who responded to all the group questions with the changed layout. After the matching process for breakoff rates, we ended up with 9,640 respondents, using 82% of the mobile respondents from 2022 and 96% from 2023. The matching process for the other indicators left us with 8,371 respondents, using 83% of the mobile respondents from 2022 and 94% from 2023 (see Tables 4 and 5 in Appendix).

For each analysis, we excluded respondents with extreme values from the effective sample. To calculate response time, we used the TimeStamps in the paradata and calculated the time it took from one observed activity until the next activity for each respondent, and then summed up the total time of activities within the question groups for each respondent. We excluded activities that were completed 15 minutes or longer after the previous action, to exclude time spent on not answering the survey. Respondents with the top 5% longest response times were excluded. When counting the frequency of previous page clicks and route errors received per respondent, we removed the top 1% of respondents to avoid extreme values.

As we are interested in the group questions where the layout changed, the basis of the analysis is respondent activity within the 51 fields in the 8 question groups that had the layout changed from pagination to scrolling. Therefore, the results of the indicators are related to activity in the group questions with the changed layout, and not the whole survey.

5.1 Results and Data Analysis

Table 2. Paradata Indicators of Mobile Respondents

Year	2022 (Pagination Layout)			2023 (Scrolling Layout)			2022 vs 2023 (Pagination vs Scrolling)	
	Average	Std	N	Average	Std	N	Min - Max	p-value
Previous Page actions (count)	0.92	1.50	8,224	0.07	0.34	8,370	0 – 8	***
Route Errors (count)	1.40	1.97	8,223	0.21	0.65	8,369	0-11	***
Response time (seconds)	342	125	7,984	343	127	7,920	26-724	Not significant

*p < 0.05; **p < 0.01 ***p < 0.001, Mann-Whitney U-test.

In 2022, the average times respondents clicked previous page was 0.92 (see Table 1), so we can roughly say that respondents, on average, clicked to the previous page once, but not all respondents clicked to the previous page. The average number of previous pages per respondent within the question group dropped from 0.92 in 2022 to 0.07 in 2023, which supports Hypothesis 1. Naturally, with more pages in the pagination layout than in the scrolling layout, we would expect to see more previous page actions. But considering that having to click previous page increases the response burden, the decrease may suggest improved user experience in the scrolling layout. While we do not have data to see if respondents scrolled up and down on the scrolling layout, we use response time to indicate the total time it took respondents to navigate between the questions. Should the scrolling layout provide better navigation and possibly reduced response burden, we should also expect to see a reduced response time. But contrary to Hypothesis 2, response time did not decrease in 2023 with the scrolling layout. The response time in the question groups was essentially the same between the years. Even though respondents received fewer route errors and did fewer previous page clicks in 2023 than respondents in 2022, the response time was not shorter.

Table 3. Paradata Indicators of Mobile Respondents

Year	2022 (Pagination Layout)		2023 (Scrolling Layout)		2022 vs 2023 (Pagination vs Scrolling)
Indicator	Count (% of N)	N	Count (% of N)	N	p-value (df)
Experienced at least one Route Error	4,664 (56%)	8371	1,227 (15%)	8,371	*** X2 (df= 1)
Breakoff	471 (4.9%)	9640	357 (3.7%)	9,640	*** X2 (df= 1)

*p < 0.05; **p < 0.01 ***p < 0.001.

Interestingly, the average number of times respondents received route errors decreased from 1.40 in 2022 to 0.21 in 2023 (Table 2). Similarly, the share of respondents receiving at least one route error within the survey question groups was substantially reduced, from 56% of respondents having at least one route error in 2022 to 15% in 2023. The lower share of experienced route errors in the scrolling layout compared to the pagination layout contradicts Hypothesis 4. In fact, respondents were less likely to miss questions in the scrolling layout than in the pagination layout.

When analyzing breakoff rates within the questions with the changed layout, we see a similar trend as with the route error. Out of the respondents that started the survey, 4.9% broke off the survey in the question groups in 2022, while only 3.7% did the same in 2023, supporting Hypothesis 3. This can imply that respondents experienced higher task difficulty in the pagination layout than in the scrolling layout. Seeing that respondents experienced less route errors and were less likely to break off in 2023 than in 2022, it can indicate that respondents experienced lower response burden and higher willingness to complete the questions in the scrolling layout than with the pagination layout.

To test if the difference between years on the indicators we analyze would be due to changing the layout and not due to other factors, we completed some sensitivity analyses with different model inputs. Comparing PC respondents between 2022 and 2023, there is no significant difference between the years for any of the indicators assessed. Moreover, to check that our analysis was not affected by missing paradata, we ran the analysis while excluding sessions from the Safari browser, which is where the missingness occurs, and saw the same significant patterns as when including sessions from the Safari browser. Importantly, we assessed activity on all other questions where the layout was unchanged between the two years, and we saw the same significant patterns: more frequent previous page clicks and route errors received, lower breakoff rate in 2023 compared to 2022, and even significantly lower response time in 2023.

5.2 Limitations

According to our assumption that any changes between the years in the paradata indicators we analyzed would be due to the layout, we should not see the same pattern in the indicators on questions with no change in the layout. However, when we ran the sensitivity analysis, we still saw the same pattern: reduced frequency of previous page clicks, less experience of route error, and lower breakoff rate among mobile respondents in 2023 but on questions where the layout was the same as in 2022. Thus, the difference in the indicator's measures cannot be isolated to changes in the layout. This limitation is caused by our study design and may be exacerbated due to the different contact channels. We attempted to control for confounding effects by conducting exact matching on sociodemographic variables. Still, it is not possible to control for all possible confounding effects related to the choice of device.

Moreover, contact methods may influence the context and environment for when and how respondents reply to a survey; namely, distracting environments and multitasking while responding to a survey may reduce commitment and ability to concentrate on the questions (de Bruijne & Oudejans, 2015). There might be reason to suspect that the change from sending the survey link via SMS and email to using Altinn can influence when and how respondents reply to the survey. As Altinn is a place where respondents expect to receive important communication from public authorities, they might wait until they are in a familiar environment without distractions before they open the Altinn letter from Statistics Norway. In comparison, when receiving the questionnaire link directly in an SMS or an email, respondents might be more likely to open the survey in distracting environments or on the go. If more mobile respondents were in less distracting environments and in a different mindset when replying to the 2023 QLS, this can have a confounding effect on the indicators assessed here. Thus, we recognized from doing the sensitivity analysis that we would need to do randomized controlled experiments to determine the effect of changes in the layout.

6. Discussion

We hypothesized that the pagination layout in 2022 would have longer response times, as other research found that response time in a scrolling layout was comparatively shorter (de Bruijne & Wijnant, 2014; Mavletova & Couper, 2014). However, the response time in the question groups remained the same in 2022 and 2023. Considering that respondents received more route errors and returned more frequently to the previous page in the pagination layout, but the response time is the same, this suggests that even if these actions may have broken the survey flow, it did not have an impact on the response time. However, in the pagination design, we can clearly identify when respondents return to the group question text on the first page, while in the scrolling layout, we cannot see if respondents have to scroll to the top of the page to see the text, thus breaking the survey flow.

We saw that the frequency of previous page clicks and route errors per respondent decreased between 2022 to 2023 when changing the layout of the question groups from pagination to scrolling. Similarly, breakoff rates within the questions groups were lower in 2023 compared to 2022. Using route errors as an indicator for how easy it was to miss questions, we found that in 2023, fewer respondents experienced a route error at least once and the average number of time respondents received route errors decreased, which may indicate that the scrolling layout is not more prone to making respondents miss a question. Our results showing reduced frequency of route errors, previous page clicks, and lower breakoff rates from 2022 to 2023 may suggest that the scrolling layout is more suitable for small screens. However, as we have pointed out in the Limitations section, our study design is not a randomized controlled experiment, and we can therefore not be sure whether the improvements are due to the changes in the layout.

Still, we found this analysis useful to start utilizing paradata, and we have gained a more thorough understanding of how we should conduct assessments in the future. Mobile phones are increasingly becoming more popular, and we ought to understand how to improve user experience on small screens and ensure high data quality from the responses. We have started to think about how we can conduct A/B testing to assess small adjustments in the layout using paradata. Importantly, as there are various alternatives for presenting grid questions on mobile (see Vehovar et al., 2022), we hope to conduct future experiments with different layouts and improvements using the knowledge learned from this analysis. Regarding improving the layout further for a better user experience, we would like to assess how respondents experience AutoEnter in our surveys, as well as differentiating the presentation of multiple-choice buttons and single response answer buttons.

In terms of improving how we analyze paradata, we hope to develop more indicators to assess user experience and survey quality. This article focused mainly on the objective response burden of the survey, but an important consideration is to assess subjective user experience and the data quality of the responses. Indicators for data quality include straight-lining, careless responding through random responding, inconsistent answers, inaccurate responses, and response order effects. For instance, when assessing response order effects, Liu and Cernat (2018) found that the mean shifted in the direction to the responses on the left of the screen for mobile respondents for 9- and 11-point scales. We would therefore like to assess how questions with different scales and the number of questions in one group can be influenced differently when adjusting the layout.

7. References

- Antoun, C., Couper, M. P., and Conrad, F. G. (2017). "Effects of mobile versus PC web on survey response quality: A crossover experiment in a probability web panel", *Public Opinion Quarterly*, 81(S1), pp.280-306.
- Cheung, G., Piskorowski, A., Wood, L. and Peng, H. (2016). "Using Survey Paradata", Presented at the 17th International Blaise Users Group Conference, The Hague, Amsterdam
- de Bruijne, M., and Oudejans, M. (2015). "Online surveys and the burden of mobile responding", *Survey Measurements: Techniques, Data Quality and Sources of Error*, pp.130-145.
- de Bruijne, M., and Wijnant, A. (2014). "Improving response rates and questionnaire design for mobile web surveys", *Public Opinion Quarterly*, 78(4), pp.951-962.
- Holmøy, A. and Rossbach, K. (2021) "IKT i husholdningene 2020" (Notater 2021/4). Statistisk sentralbyrå. <https://www.ssb.no/teknologi-og-innovasjon/artikler-og-publikasjoner/attachment/445604?ts=177a4b71958>
- Leiner, D. J. (2019). "Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys", *Survey Research Methods*, 13(3), pp.229-248.
- Liu, M., and Cernat, A. (2018). "Item-by-item versus matrix questions: A web survey experiment", *Social Science Computer Review*, 36(6), pp.690-706.
- Mavletova, A., and Couper, M. P. (2014). "Mobile web survey design: scrolling versus paging, SMS versus e-mail invitations", *Journal of Survey Statistics and Methodology*, 2(4), pp.498-518.
- Pettersen, A and Engvik, M. (2022) "Livskvalitetsundersøkelsen 2022" (2022/35). Statistisk sentralbyrå. <https://www.ssb.no/sosiale-forhold-og-kriminalitet/levekraer/artikler/livskvalitetsundersokelsen-2022.dokumentasjonsnotat>
- Revilla, M., Toninelli, D., and Ochoa, C. (2017). "An experiment comparing grids and item-by-item formats in web surveys completed through PCs and smartphones", *Telematics and Informatics*, 34(1), pp.30-42.
- Steinbrecher, M., Roßmann, J., and Blumenstiel, J. E. (2015). "Why do respondents break off web surveys and does it matter? Results from four follow-up surveys", *International Journal of Public Opinion Research*, 27(2), pp.289-302. <https://doi.org/10.1093/ijpor/edu025>
- Vehovar, V., Couper, M. P., and Čehovin, G. (2022). "Alternative layouts for grid questions in PC and mobile web surveys: An experimental evaluation using response quality indicators and survey estimates", *Social Science Computer Review*. <https://doi.org/10.1177/08944393221132644>
- Zhang, C., and Conrad, F. (2014). "Speeding in web surveys: The tendency to answer very fast and its association with straightlining", *Survey Research Methods*, 8(2), pp.127-135. <https://doi.org/10.18148/srm/2014.v8i2.5453>

8. Appendix

Table 4. QLS 2022 and 2023 Sample Matching for Previous Page Actions, Route Error, and Response Time

Year	2022 (Pagination layout)	2023 (Scrolling layout)	2022/2023 Matched samples
Category	N (% of Total)	N (% of Total)	N (% of Total)
Gender			
Male	4,329 (43%)	3,586 (40%)	3,133 (40%)
Female	5,703 (57%)	5,311 (60%)	5,058 (60%)
Age			
18-24	1,082 (11%)	1,221 (14%)	1,056 (13%)
25-44	3,564 (36%)	3,767 (42%)	3,410 (41%)
45-66	3,958 (39%)	3,078 (35%)	3,078 (37%)
67-79	1,232 (12%)	722 (8%)	720 (9%)
80+	196 (2%)	109 (1%)	107 (1%)
Education			
Elementary school or lower	1,674 (17%)	1,659 (19%)	1,450 (17%)
Upper secondary school	3,720 (37%)	3,386 (38%)	3,086 (37%)
University/University college	4,426 (44%)	3,680 (41%)	3,680 (44%)
Not Stated	212 (2%)	172 (2%)	155 (2%)
Total	10,032	8,897	8,371

Table 5. QLS 2022 and 2023 Sample Matching for Breakoff Rates

Year	2022 (Pagination layout)	2023 (Scrolling layout)	2022/2023 Matched samples
Category	N (% of Total)	N (% of Total)	N (% of Total)
Gender			
Male	5,183 (44%)	4,111 (41%)	3,916 (41%)
Female	6,584 (56%)	5,907 (59%)	5,724 (59%)
Age			
18-24	1,383 (12%)	1,465 (15%)	1,355 (14%)
25-44	4,339 (37%)	4,317 (43%)	4,052 (42%)
45-66	4,446 (38%)	3,334 (33%)	3,334 (35%)
67-79	1,364 (12%)	780 (8%)	778 (8%)
80+	235 (2%)	122 (1%)	121 (1%)
Education			
Elementary school or lower	2,193 (19%)	2,028 (20%)	1,905 (20%)
Upper secondary school	4,293 (36%)	3,761 (36%)	3,516 (36%)
University/University college	4,961 (42%)	4,007 (42%)	4,003 (42%)
Not Stated	320 (3%)	222 (2%)	216 (2%)
Total	11,767	10,018	9,640