Final methodological and implementation report

Findings from the EU-grant project "Improvement for the Norwegian AES 2022"

Anna-Lena Keute, Katharina Rossbach, Magnar Lillegård, Dag Gravem and Sara Grimstad

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

Symbols in tables	Symbol
Category not applicable Figures do not exist at this time, because the category was not in use when the figures were collected.	.
Not available Figures have not been entered into our databases or are too unreliable to be published.	..
Confidential Figures are not published to avoid identifying persons or companies.	:
Decimal punctuation mark	.

Preface

The Adult Education Survey (AES) is a European survey on adults' participation in different learning activities. Statistics Norway received EU-grants in the period 2021 – 2023 to finance the project "Improvement for the Norwegian AES 2022" (Grant Agreement Number 101051714 — 2021-NO-AES).[1] One aim of the project was to examine how to improve the user experience by doing expert reviews and cognitive and usability testing of the questionnaire. We also conducted a user journey analysis to get insight into the entire process of participation, from being contacted to answering or rejecting participation. Furthermore, the project aimed at improving the sampling and weighting design of the survey. This report presents the findings from this project.

The project team consisted of Anna-Lena Keute, Katharina Rossbach, Dag Gravem, Magnar Lillegård, Sara Grimstad. Elise Alstad programmed the questionnaire for the AES 2022 pilot survey.

Statistisk sentralbyrå, 07.09.2023

Ann-Kristin Brændvang

Funded by the European Union

Contents

1. Introduction

This report provides the final results of the project "Improvements for the Norwegian AES". The project was funded within the framework of the EU grants for the Adult Education Survey (AES) 2022 – Statistical infrastructure under IESS.

The project aims at improving the data quality in the Norwegian AES 2022 by focusing on two tasks, namely improving the user experience, and improving the sampling and weighting design. The user experience task encompasses two main sub-tasks: firstly, to test and improve new and existing AES questions. Secondly, to map the user (respondent) journey and to develop strategies for reaching and communicating with underrepresented groups in a mixed/multi-mode setup. Methods used include expert evaluation, focus groups, cognitive interviewing and usability testing. We selected the underrepresented groups based on earlier experiences which show that young adults and adults with low educational attainment (below ISCED level 3) are underrepresented in surveys.

The second task of the project is related to methodological work on AES sampling design and weighting scheme. The aim of this task is to improve the sampling design and to explore how to use register information on participation in formal education and training in the weighting procedure. By using register information on formal education and training as a weighting variable in the 18 to 24 age group, we aim at reducing bias in estimates. Using register information also reduces the overall response burden and we aim at improving the data quality and assuring that data comply with precision requirements.

The project period was September 2021–April 2023. This report gives an overview on the results of the grants project.

2. Planned tasks and implementation of the tasks

2.1. User experience

To get an extensive insight into user experience and challenges for the AES survey, as well as improving it, we originally planned to divide our activities into three phases. In the first phase, we conducted an expert review to map potential problems with questions. This was conducted in November 2021. The expert review helped us to know which questions to focus more on in the later user testing and already implement some improvements in question wordings, answer alternatives and help texts before conducting the subsequent user tests. In the second phase, we conducted cognitive and usability testing on actual users to get more knowledge about possible challenges with selected questions from the AES questionnaire. This was done in January 2022. In the third phase, in April/May 2022, we conducted a user journey mapping. User journey mapping not only gave us users' experiences of survey questions, but also how the respondents/users experience the whole process of participating: from receiving invitation letters and messages, to participating or not participating in the survey. It revealed pain points as well as which elements are perceived as positive. The fourth phase, which had not been planned but deemed necessary for this survey, consisted of additional user testing. This phase was important since several challenges in the section on non-formal education and training activities in the questionnaire persisted.

Table 2.1   Overview of user experience activities and when they were conducted

User experience activity	Time conducted
Phase 1	
Expert review	End of November 2021
Revision of AES questionnaire	December 2021
Phase 2	
Round 1: User testing on CATI and CAWI	January 2022
Revision of AES questionnaire	February–March 2022
Phase 3	
Pilot 1 – Target group: Age 25-34	25th April–3rd May 2022
Focus groups and exploratory interviews	2nd May–6th May 2022
Pilot 2 – Target group: ISCED below 3	16th May–24th May 2022
Focus groups and exploratory interviews	20th May–31st May 2022
Revision of AES questionnaire	End of May–Beginning of June 2022
Phase 4	
Round 2: User testing on CAWI	June 2022
Revision of AES questionnaire	July 2022

Source: Statistics Norway

2.2. Expert review

Main objective

The expert review was conducted by three experienced survey methodologists. The expert review is often conducted before testing and improving surveys in order to identify potential problems with the formulation of question, answer alternatives and the flow of the questionnaire. It is a fast and relatively inexpensive method compared with other methods (Tourangeau, Maitland, Steiger and Ting, 2020, p. 48). Moreover, it can detect more problems than other methods, and experts can often make good problem-solving proposals. We based our expert review on the cognitive process model (see e.g., Biemer & Lyberg 2003, p. 123 ff.) and the review coding system developed by Forsyth and Lessler (1996). Moreover, the expert review identified possible sources of mode effects based on the approach by Campanelli et al. (2013). Each survey methodologist independently reviewed a predefined group of questions (see details below). Findings were compared and discussed within the expert review team, and subsequently an expert review report was presented

to the subject matter experts at the Division for Education and Culture Statistics, and the Division for Social surveys.

The expert review covered the following topics: access to learning possibilities, ten questions about formal education, non-formal education and training and informal learning. These topics were selected for one or more of the following reasons:

1) Previous experience show that the respondents tend to find it difficult to answer the selected questions;
2) The selected questions had never been tested on the 18-24 age group;
3) The questions were new and hadn't been tested before;
4) There were possible measurement differences between Computer Assisted Telephone Interviewing (CATI) and Computer Assisted Web Interviewing (CAWI) (Based on Gravem and Berg, 2019)

CAWI is a surveying technique in which the participants receive questionnaires in a web browser or mobile application on mobile phone, PC, or tablet, and complete the survey themselves. CATI, on the other hand, is a telephone surveying technique. Participants are called by interviewers, and their answers are administered by the interviews.

Proposals for improvements were made regarding simplifying question wordings, using correct tense forms, splitting double-barrelled questions into two separate questions, and ensuring coherence between questions and response options.

Expert reviews can only give limited insight into potential challenges and problems for a specific survey, and they cannot replace user tests with persons from the survey's target groups. User tests are necessary for better understanding of the respondents' perspective and to be certain about which difficulties real test persons experience when they answer a survey.

Results
Access to learning possibilities
Concerning access to learning possibilities, the questions included technical terms (GUIDE_3), terms which were vague (GUIDE_1) or could be interpreted in different ways (SEEKINFO). There were also ambiguous question wordings in several questions within this topic. Moreover, the connection between previous questions and question GUIDESOURCE was not clear.

Formal education
For formal education, we found that in FEDMAINSTAT the term "main labour status" was a very technical term, as the question was phrased "What was your main labour status when you started with FEDNAME?". Hence, we looked at other Statistics Norway surveys that contain similar questions, for instance the Labour Force Survey (LFS), and rephrased the question to "When you started with FEDNAME, did you mainly regard yourself as…" We also found that FEDOUTCOME should be reworked into a sequence of yes/no questions, to avoid mode effects between CATI and CAWI.

Non-formal education and training
Regarding the topic on non-formal education and training, the questions about courses, private lessons and seminars refer to both work-related and leisure-related activities. E.g.: "Have you attended a course organised by an employer or in [your] leisure time the last 12 months?"). After the expert review, we proposed to split such questions into two individual questions by clearly separating work-related from leisure-related activities. Every question should focus on one domain only, because work-related courses and leisure-related non-formal education can differ vastly from the other.

Furthermore, we proposed to use more dynamic texts e.g., NFEACTxx_TYPE where we replaced the general term "activity" with the dynamic "<specific name of activity>" which is then named by the respondent in questions NFEACTxx_NAME.

For NFEACTxx_PAIDBY, the question was "Did your employer or prospective employer pay partially or fully for NFEACTxx_NAME", which was essentially a two-part question: if the employer or prospective employer pays for NFEACTxx, and if yes, whether they pay fully or partially. However, the original answer choices were "yes" or "no", which were a mismatch to the questions. We therefore changed the answer categories to: Yes, the employer paid fully; Yes, the employer paid partially; No, the employer did not pay; No, I did not have a job.

NFEFIELD1 consisted of too many and too long answer alternatives and contained some very vague terms. Similarly, NFESKILLSMAIN1 consisted of many answer categories, which can be challenging in both CAWI and CATI. On NFECERT1 the response categories did not match the question phrasing, making it poorly suited for CATI in particular. Therefore, we split the question into two: "Did the activity lead to a certificate?" Yes/No (NFECERT1a) If yes, a follow-up question was asked: "Was the certificate required by your employer or a professional body for the execution of your current or planned activity as employer or employee?" Yes/no.

Informal learning
For questions about informal learning, the instruction INF and the question INFFAMILY were vaguely worded. We proposed to remove the term "organised" from the sentence: "Disregarding the education and [organised] learning activities we have previously asked about: During the past 12 months, have you deliberately tried to learn anything on a particular topic or area, or are you currently doing it, through one of the ways mentioned below?"

In addition, we moved some information from the question wording to the interviewer/respondent instruction below the question, including the term "random learning": "During the past 12 months, have you learned something from family members, friends or colleagues?" Do not include random learning, formal education, or non-formal education we have asked about earlier."

Furthermore, the term "printed media" is quite technical and not very concrete. Hence, we proposed to replace it with the Norwegian term "fagbøker", which means textbooks (INFMATERIAL). INFDEVICE contains the term "electronic devices", which has the same problem. We proposed to replace it with more concrete terms, such as "mobile phones, pc and similar". The term "learning centre" is also a vague term in INFLIBRARIES. Finally, INFPURP was unclear about the type of activities it referred to. Therefore, we suggested asking this question for each activity separately. Yet we decided against the suggested revision in the final version, because asking more questions was time-consuming, and EUROSTAT does not require information to such details. By having only one question about INFPURP, we also managed to fit all questions about informal learning on one page for CAWI.

Recommendations for Eurostat
Several of the above-mentioned problems can be attributed to the original source materials provided by EUROSTAT. Hence, it is also advisable to adjust the English version which is used by all countries' statistical offices as a base for their own translation and adaptation. This applies to, for instance, the term "electronic devices" which is a broad term and can be interpreted in different ways also in English. More concrete examples should be incorporated in the questionnaire source material, and it should be stressed that national statistical institutes are not obliged to follow the model questionnaire phrasings verbatim.

Another example where the source questionnaire should be improved, relates to our proposal for the question "Did your employer or prospective employer pay partially or fully for NFEACTxx_NAME?". The answer categories should also be extended to four categories: Yes, the

employer paid fully; Yes, the employer paid partially; No, the employer did not pay; No, I did not have a job.

The expert review also uncovered that AES source materials were poorly adapted to mixed-mode data collection. Many questions presuppose a visual mode, e.g., NFEREASON and NFEPAIDBY with "mark all that apply" answer formats, whereas several countries have conducted the survey fully or partially using CATI. We recommend reworking the source materials into a form that is applicable to any mode in use. This can help to reduce measurement differences between modes for each participating country, and measurement differences between countries that are attributable to modes (Cooperation on Multi-mode Data Collection- Mixed Mode Designs for Social Surveys- MIMOD"(2018)).

2.3. Redesign and pretesting

Main objective
The main objective of conducting cognitive interviews was to test how selected questions were perceived by actual respondents. We focused specifically on whether the flow/structure of the survey was satisfactory and challenges with the order of the questions. Problems in the process of answering questions can lead to low data quality, but also dropout during the survey, which is a concern especially for CAWI. We aimed to identify difficult terminology as well as vague question phrasings. As AES has a 12-month reference period, problems related to memory and information retrieval were also of interest. Cognitive tests were performed in CATI and CAWI modes to shed light on possible mode differences.

For the cognitive interviews, we recruited 12 persons of whom five had an educational attainment below ISCED-level 3, which means below upper-secondary education. Persons with low education is a group that is harder to recruit for surveys, more prone to drop out, and also one of the target groups for user journey mapping. AES was conducted as a mixed mode survey in Norway in 2016, meaning that respondents could either self-administer the survey by answering the questions in CAWI mode, or being called by an interviewer and answering as a telephone interview, CATI. As described in the MIMOD report "Cooperation on Multi-mode Data Collection- Mixed Mode Designs for Social Surveys- MIMOD" (2018), a major concern is potential mode related bias for AES, more specifically an education bias increase for CAWI in the net sample. In fact, Statistics Norway detected noticeable net sample differences between CATI and CAWI respondents regarding education: the education levels were higher for respondents answering in CAWI than for CATI respondents in AES 2016.

Experiences from AES data collection in 2016 and earlier had shown that questions regarding non-formal education were more challenging than the rest of the questions in this survey. Hence, a special emphasis was put on those questions, and they have always been included in expert reviews, user tests and focus groups.

Results from the cognitive interviews indicated that more testing was needed to evaluate the effects of the proposed changes discussed in the "Results" section below. For this second round of cognitive interviews, we recruited people aged 18 to 24, an age group that has not previously been part of the target population in AES. The second round of user testing was not planned in the original project description and will therefore not be described in detail in the "Results" section below.

Results

The cognitive tests revealed several weaknesses of the survey questionnaire. The questions about access to learning possibilities were often interpreted differently by the test persons and some of them were not sure what we meant. Hence, we included more examples of learning possibilities.

Table 2.2   Overview of test persons and their characteristics

Test person (TP)	Age group	Formal education last 12 months	Non-formal activities last 12 months	CAWI/CATI	Gender
TP1	25-34	Yes	No	CATI	Female
TP2	25-34	No	Yes	CATI	Male
TP3	25-34	No	No	CAWI-mobile	Female
TP4	18-24	Yes	No	CATI	Female
TP5	25-34	Yes	Yes	CATI	Male
TP6	18-24	Yes	Yes	CATI	Male
TP7	25-34	No	Yes	CAWI-pc	Male
TP8	25-34	Yes	Yes	CAWI-mobile	Female
TP9	18-24	Yes	Yes	CAWI-pc	Female
TP10	25-34	No	Yes	CAWI-pc	Female
TP11	Over 34	No	No	CAWI-pc	Male
TP12	Over 34	No	Yes	CAWI-mobile	Male

Source: Statistics Norway

Formal education

For formal education, we tested two different versions of the questions FEDREASON and FEDOUTCOME. In both these questions, one or more response options might be chosen by respondents because several alternatives can apply. In version A, we presented the questions as single multiple-choice questions with checkboxes, whereas in version B, each response option was presented as a yes/no question. For the CATI tests, version A was treated as an open question where the answer alternatives were not read out loud. Our findings suggest that the yes/no structure works better. One reason is that sometimes CAWI respondents do not see the instruction that several answer categories can be chosen. This could lead to different answers in CAWI and CATI modes, but also differences between countries conducting AES in various ways and modes (Appendix A of Gravem and Berg (2019), Murgia et al. (2019)).

Informal learning

For questions regarding informal learning, test persons interpreted some of the terms in various ways. For instance, the term "other learning centers" was interpreted in different ways, whereas "learning by chance" was not understandable for everyone. Moreover, there was sometimes uncertainty regarding what should be included and not included in question INFFAMILY "During the last 12 months, have you learnt something from family members, friends or colleagues?" One respondent was unsure about whether boyfriend/girlfriend should be considered. Another respondent wanted to know if we meant "to learn from someone" or "to teach someone".[2] One test person also thought that the questions about informal learning were job-related only.

Interestingly, findings for the NFENBHOURS variable were contrary to what has been previously reported by the MIMOD grant project (Gravem and Berg, 2019). Here, the MIMOD study uses data from the 2016 Norwegian AES. The respondents could choose to report numbers of days on which they received training. This was followed by a question on average hours of training per day. Based on this, a total number of training hours was calculated, and respondents were presented with a control question: "We have calculated that you received a total of x hours of training. Does this seem

---

[2] In Norwegian, the same verb is used in both senses, the difference being marked using prepositions. The distinction can be difficult for non-native speakers.

right, or do you think a different total would be more correct?" In the 2016 Norwegian AES, 96% of these respondents confirmed that the calculated number was correct.

However, in AES cognitive tests, 4 participants got the same control question, but only one judged the calculated number of hours to be correct. This apparent discrepancy is based on a small number of observations and could be by chance. However, it could also be that user test participants invested more time and effort into answering and correcting their answers, whereas in a survey, respondents often tend to spend the minimum amount of time deemed necessary – also known as satisficing (Biemer & Lyberg, p. 124). This could imply that the calculated control only has a small effect on improving the data quality for this variable.
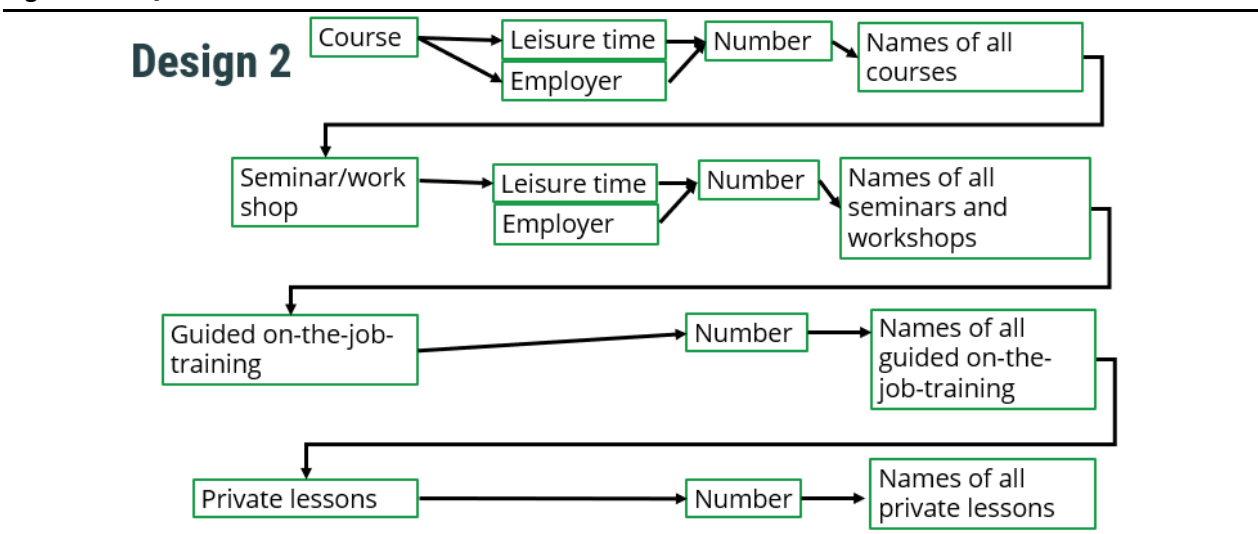
Non-formal education and training
For the questions on non-formal education and training, we tested two different question flows which are shown in the figures below.

Figure 2.1   Question Flow 1
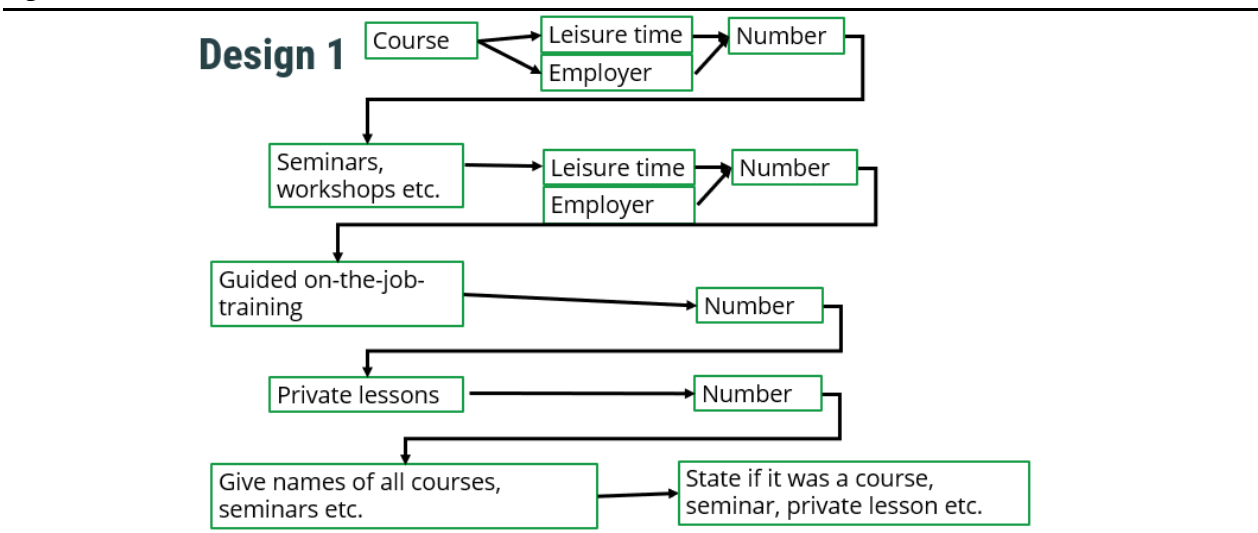
Source: Statistics Norway

Figure 2.2   Question Flow 2

Source: Statistics Norway

We decided to test two different designs because experience from AES 2016 indicated that these questions were difficult to answer. The aim was to find an appropriate structure for non-formal activities which makes it easier for participants to respond to the questions and thereby improve response quality. Design 1 is the original design, which Statistics Norway used in 2016, whereas design 2 is new. Design 2 was developed in the hope that it would be easier for participants to name the activities after reporting the number of activities. In design 1, the test persons could name up to ten activities, regardless of type. In design 2, the test persons could name up to 5 activities of each type, i.e., a total of 20 possible named activities.

Irrespective of the design, test persons had problems naming the activities. They often did not remember the name of the activity, or the activity did not have a specific name. Another problem is double reporting, for instance that the same activity was reported both as a course and as a seminar. Moreover, the Norwegian wording "Hvor mange ganger har du tatt privattimer de siste 12 månedene?" (How many times have you taken private lessons in the last 12 months) for questions on private lessons and guided-on-the-job training can lead to over-reporting as has been the case for guided on-the-job training.

Regarding private lessons, none of the participants answered that they had taken any private lessons the last 12 months; thus, the question about number of private lessons was never tested. Overreporting occurred because the Norwegian term "ganger" (times) is an ambiguous term and has been interpreted as how many occasions one had training of the same kind. For instance, one test person reported having received guided-on-the job training 12 times and afterwards had difficulties answering the subsequent questions. Had it been sufficiently clear that this should have been reported as one activity, the test person would likely have had fewer issues.

Since the question wording for private lessons and guided-on-the-job training are quite similar, we suspected that there would be overreporting for private lessons as well. Hence, we changed the question to "How many different types of paid private lessons have you taken the last 12 months?". Yet, through user testing in June (otherwise outside of the scope of this report) we found out that 2 out of the 6 test persons who answered yes, overreported both. Therefore, we included an instruction with an example in the final version: "If you have [e.g.] taken five driving lessons, count this as one type of paid private lessons".

In the pilot survey prior to the user journey focus groups and individual interviews, we included an overview of all the non-formal learning activities the respondent had reported. This gave the respondents a chance to correct the number of courses, seminars, and other non-formal educational activities.

Test persons who had no job or no job-related learning activities, felt that the survey contained too many job-related questions. This might have led to the assumption that leisure activities should not be included, as several test persons reported that they did not think of leisure activities when answering the questions. Based on this feedback from test persons, we decided to include examples for the different activities regarding non-formal learning, e.g., for private lessons we included "For example private music, sports or driving lessons". Another approach is to include a filter for those who answer that the activity was not job-related. Follow-up questions on whether the non-job-related activities took place during working hours, or were paid for by the employer will often seem illogical to respondents.

2.4. Focus groups

Main objective
User journey mappings are conducted to get insights into how users experience a product or service, and in this project, how respondents experience their participation in the Adult Education Survey. An important term regarding user experience mapping is the "touchpoint". As defined by Kalbach (2020), touchpoints are points where interaction between individuals and an organisation can (but do not necessarily) take place, for instance through a phone call, e-mail interaction or web sites (Kalbach, 2020, Chapter 2: Fundamentals of Mapping Experiences).

Mapping the user journey for respondents of a survey is a relatively new approach in Statistics Norway, although it has to some extent been used for the Time Use Survey. Internationally, the approach has been pioneered by the UK's ONS (Wilson & Dickinson, 2021.)

Discussing users' experiences offers useful perspectives on whether each touchpoint is actually encountered, and what challenges or problems users face. It also highlights users' possible misconceptions and how touchpoints are perceived: if users have positive or negative emotions or are indifferent. Moreover, discussing users' experiences can offer a better understanding of how far the goals of an organisation are achieved and how to improve the experience of touchpoints to increase the organisation's goals.

A major goal for conducting this user journey mapping was to make the survey experience more salient and thereby increase representativity of two specific demographic groups. The first group consists of persons aged 25 to 34. The second group consists of those who are registered with an education below ISCED3 (EU Grants: Application form (SMP ESS): V1.0 – 15 .04.2021, p. 5).

Before conducting the user journey mapping, we determined possible touchpoints between respondents and Statistics Norway. The process was as follows: a survey methodologist mapped specific touchpoints and then presented the touchpoints to the Division of Education and Culture Statistics and the Division for Social surveys in a meeting; the other team members then shared useful insights on how relevant those touchpoints were and proposed other relevant touchpoints. This helped us get a more accurate picture of all relevant experiences for users and remove aspects which are of no importance to users. After this process, the following possible touchpoints were identified:

- Information letters (digital)
- Email notifications of information letter
- SMS notifications of information letter
- Email reminders
- SMS reminders
- Survey information page on Statistics Norway's website
- Support service (phone or email)
- CAWI questionnaire
- Interviewer interaction (appointments, refusal etc.)
- CATI interview
- "Thank you" notification SMS after the interview

After having detected all possible touchpoints with Statistics Norway, the next step was to find out if respondents come in contact with those touchpoints or not, which ones are more important, and how to improve them for respondents in the target groups. For Statistics Norway, it is important that users taking part in a survey have a positive survey experience overall, as well as understand and answer the questions the best they can.

The data collection process

To investigate actual user journey experiences, we drew a sample of 400 people, and invited them to participate in a pilot survey using a slightly shortened version of the actual AES questionnaire. 200 were people aged 25-34 and the other 200 were people with an educational attainment below ISCED 3. After the data collection, we contacted some of them, both respondents and non-respondents, and invited them to focus groups to discuss their user experience. The advantage with approaching the sample after they have completed their user journey, is that their experience is as realistic as possible. Had we informed them about the focus group before they had completed their user journey, the participants could be affected by this awareness, which might influence their user journey experience.

We limited the pilot survey field period to one week and two days and recruited focus group participants immediately afterwards. This was to ensure that the user journey would be fresh in the participants' mind.

We sent out invitation letters, reminders and the so-called mode letter digitally via the Altinn portal.[3] Respondents received an e-mail and a text message from Altinn notifying them that Statistics Norway have sent them a letter in Altinn, and that they should log into Altinn to read the letter. All these pieces of communication from us were sent within one week and a half. They all contained a personalised link to the online questionnaire. The mode letter included information that interviewers would soon make phone calls to conduct a telephone interview, but that they could still choose to complete the survey online. CATI interviewing started the same day the mode letter was sent out. Figures 3 and 4 illustrate the day-by-day response figures for the two groups and the timing of SMS and letter dispatches.

Figure 2.3   Day-by-day response from target group 1, Age group 25-34
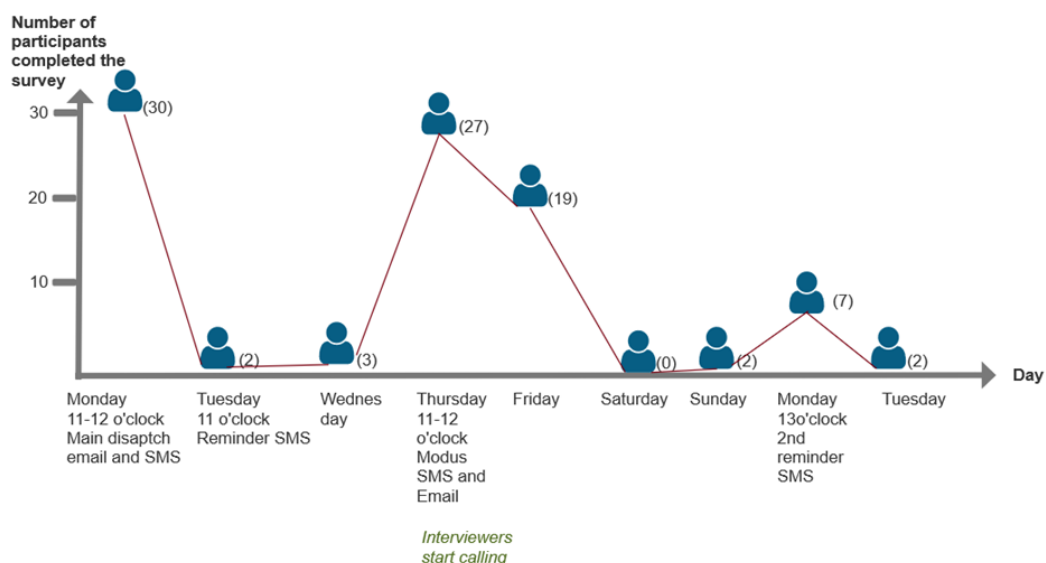
Source: Statistics Norway

---

[3] Altinn is an internet portal for digital dialogue between businesses, private individuals and public agencies in Norway. It is developed and managed by several government institutions who co-operate it. When a new message from a government institution is sent to your Altinn inbox, you are notified of this through an SMS and/or e-mail sent from Altinn (not Statistics Norway). Most citizens are familiar with Altinn through yearly tax returns.

Figure 2.4   Day-by-day response from target group 2, educational attainment level below ISCED 3
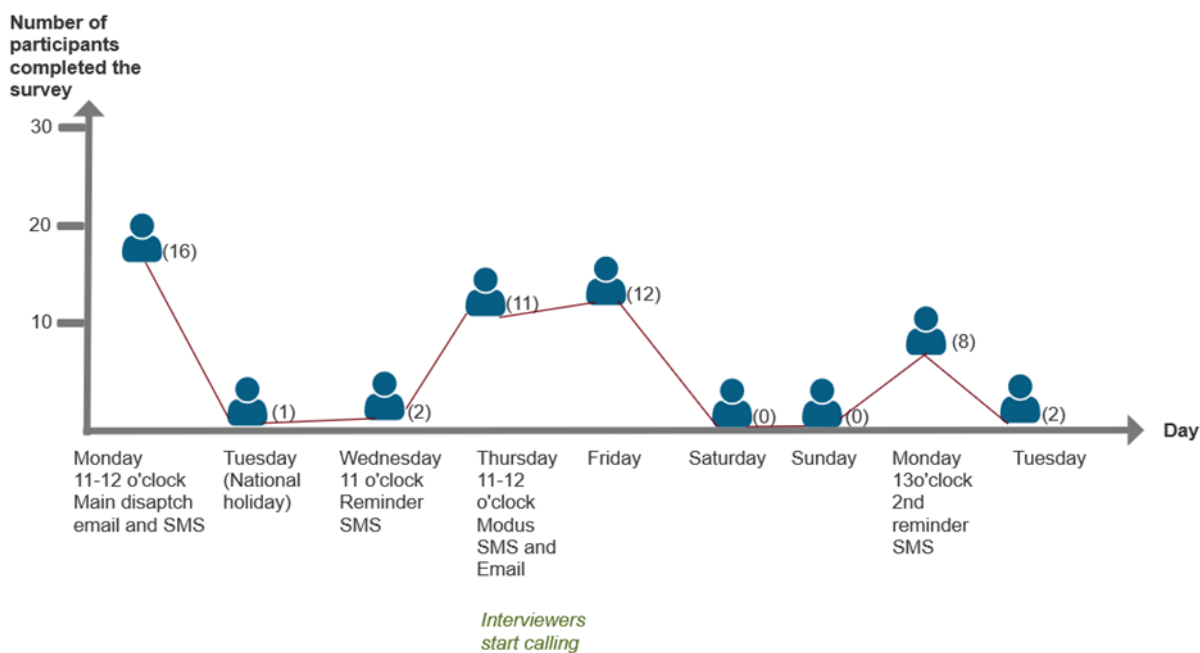
Source: Statistics Norway

In short, the response rate of the low education group was nearly half of that of the 25-34 age group. This is not surprising. Research by Dillman and Messer (2011, p. 445) have similar findings of mail and web responses: those who answer online are younger and have higher educational attainment compared with those who needed to be followed up by mail.

Our experience with recruiting to focus groups after the respondents had participated in a realistic user journey, was very positive. Although individual interviews sometimes replaced focus groups (see below), we received useful insights into how respondents experienced their participation, or non-participation, in AES. To conclude, we recommend using pilot surveys as the basis for recruitment for future user journey mappings. However, better strategies need to be developed for recruiting pilot survey non-respondents (see below).

Tables 3 and 4 show characteristics of the participants in the focus groups and exploratory interviews for target group 1 and 2.

Table 2.3   Overview of participants in exploratory interviews and focus groups – target group 1

Participant (P)	Status	Formal education last 12 months	Non-formal activities last 12 months	CAWI/CATI	Gender	Focus group (F) or exploratory interview (E)
P1	Completed	No	Yes	CAWI	Female	F
P2	Completed	Yes	Yes	CAWI	Male	F
P3	Started	No	Yes	CAWI	Male	E
P4	Completed	No	No	CATI	Male	F
P5	Completed	No	Yes	CATI	Male	F
P6	Completed	No	Yes	CAWI	Male	E

Source: Statistics Norway

Table 2.4   Overview of participants in exploratory interviews and focus groups – target group 2

Participant (P)	Age group	Formal education last 12 months	Non-formal activities last 12 months	CAWI/CATI	Gender	Focus group (F) or exploratory interview (E)
P1	25-34	No	Yes	CATI	Male	E
P2	25-34	No	No	CATI	Female	E
P3	25-34	No	Yes	CAWI	Male	E
P4	25-34	No	Yes	CAWI	Female	F
P5	25-34	No	No	CATI	Male	E
P6	Over 34	No	Yes	CATI	Male	E
P7	Over 34	No	No (back and forth)	CAWI (started)	Male	F
P8	Over 34			No answer	Male	E

Source: Statistics Norway

Results

The participants' user journey experiences were discussed in focus groups and exploratory interviews, helping us to get useful insights into which methods of communication are preferred (Altinn letter, e-mail notifying them of the letter from Statistics Norway in Altinn, telephone, SMS). We learned that all our participants were concerned about spam and phishing. Yet, they had different opinions on which methods of communication are secure, and which are not.

Several participants stated that they only trusted the survey was real because an interviewer called them, and some others stated that they would not take part in the survey if they had to use a web questionnaire. It became apparent in the low education group that offering CATI is essential. The reasons were various: some participants did not like or feel the need to take the initiative in answering the survey, while others had difficulties reading.

Additional information about the survey, which was included in the Altinn invitation letter in the form of a link to a Statistics Norway web page, had not been accessed or used by any of the participants. Neither had anyone sought help or information from the Statistics Norway support service. The e-mail address and telephone number for this service, which answers or forwards any user/respondent query, was also provided in the invitation letter.

Another major finding was that the survey name was associated with something else than we intended. We used the name "Undersøkelsen om voksnes læring" (Survey about Adults' learning) but several participants mistook this for the term "voksenopplæring" which refers narrowly to courses for adults to gain a certificate for primary school, high school or Norwegian language. This in turn made them believe the survey was not relevant for them, as they had not participated in "voksenopplæring". Therefore, we proposed several alternative names for the survey and discussed with test persons about what kind of questions they would expect from a survey with the names in question in the user testing in June.

Other findings from the focus groups and exploratory interviews align with those from the user tests in January and June: naming learning activities is a challenge. Another source of annoyance mentioned in focus groups and interviews for target group ISCED below 3, was the question regarding educational background of parents. The respondents wondered why we need this information.

At the end of the data collection period, we sent out a "thank you" text message to those who completed the pilot survey. This was perceived as a nice gesture by the participants but did not have any impact on their overall experience of participation in the survey. Therefore, we decided not to send out such messages in the actual survey.

Regarding which languages participants speak or understand (LANGUSED), it became obvious in the interviews and focus groups in target group 2 that the participants had different interpretations of what to count, and what not. Some participants counted only languages they were fluent in, and one of them stated that he only counted those he also lists in his CV. Another participant counted the languages he had learnt at school. Despite the instruction that said "Scandinavian languages should be included", several participants did not do that. Therefore, we decided to replace "Scandinavian languages" with the specific languages "Danish and Swedish". In addition, we replaced a sentence in the instruction. In the questionnaire used for the user journey mapping, the instruction read "It is not necessary to have complete mastery of the language". In the final version, we replaced that with "Also include languages which you understand just a little of."

On obstacles to learning, several participants in target group 1 stated that this topic is very important and relevant. Including and focusing on topics that can be salient for respondents who have participated in few or no learning activities can thus have a positive impact on user journey experience.

Non-respondents
With one exception, all the user journey mapping was performed on survey respondents. Non-respondents were considerably harder to recruit. We nevertheless were able to conduct one exploratory interview with a person in the ISCED below 3 sample, who did not take part in the survey. The interview with this non-respondent had a different focus than the interviews with respondents. In that interview, we focused on exploring why the interviewee chose not to take part in the survey and what we could do to motivate him as well as others to participate in our surveys. He revealed that the reasons for not responding Is that he doesn't like to take part in surveys in general, and that he doesn't think it's something worth spending time on.

The moderator suggested several measurements and checked with the interviewee whether they would increase his motivation to take part in the survey. The first suggestion was to offer a gift card or lottery. The participant stated that this was not important for him and hence would have had no effect. The second proposal from the moderator was to change the information in the invitation letter, for instance to not use register information, or to emphasise the importance of making this the information available to the society. Here, the participant proposed to simplify the invitation letter.

When asked whether he'd prefer to receive a private e-mail or SMS with a direct link to the survey, as opposed to receiving an e-mail or SMS only referring to the Altinn inbox, the participant stated he was not sure. Another idea was to change the information in the SMS or the Altinn letter, for instance to inform how many days the participant would be able to answer. Again, the respondent was also not sure if that would have had an impact on motivation.

Regarding the proposal to extend the deadline of participation, the participant answered with a clear yes that this would have motivated him to respond. Finally, the moderator asked if the participant trusted Statistics Norway's privacy policy or not. The participant contemplated and then said yes because it is a public institution, and that he had less trust in private institutions. He concluded that it is important to have more personalised letters and messages.

Since we only managed to recruit one non-respondent for an exploratory interview, it is debatable whether his answers are representative of other non-respondents. We therefore advise future interviews with non-respondents, both for this survey and for other surveys.

Mode assignment and communication strategies
Another finding from the respondent user journey is related to different answering modes. We asked whether the focus group participants preferred to answer the survey online or over the

telephone. For the first group aged 25-34, people generally said they had few problems with answering the questionnaire online. Although they were more inclined to answer online, some emphasised that it was nice with a choice. For the second group with ISCED below level 3, however, the majority said they preferred to answer the survey as a telephone interview rather than online. Lack of initiative to take the time to answer self-administered, as well as challenges with reading and understanding information, were some reasons pointed out by the focus group participants for not answering online. Based on these insights, we decided to opt for an adaptive mixed-mode design in AES 2022, using CATI first for people aged 25-34 and with ISCED level below 3, and web first for the rest of the sample. We hoped that this would increase response rates for this group and thereby contribute to less bias and better data quality. Results from AES 2022 data collection indicate that this was the case.

We sent out reminders via SMS and Altinn throughout the data collection period and asked the focus group participants what they thought about that. People generally regarded reminders as something that reminded them to participate. They found them neither motivating nor demotivating as such. The participants said they understood that it was necessary for Statistics Norway to send reminders to get people to participate in the survey. We also wanted to know how the participants experienced receiving the mode letter the same day we started with telephone interviewing. Some said that it was good with a short time interval between the mode letter and the telephone call, because then they had not forgotten that Statistics Norway was going to call. However, in AES 2022, we decided to wait a few days before we started with telephone interviewing, because we wanted people to answer online as much as possible first.

3. Sampling and weighting

3.1. Mapping of register information

Main objective

The Adult Education Survey (AES) maps adults' participation in formal education and non-formal education. In 2022, AES will also include young adults aged 18-24. Regulation (EU) No 2019/1700 of the European Parliament and of the Council establishes a common framework for European statistics based on individual data collected through sample surveys, and this also includes AES. The regulation stipulates a precision requirement for the indicator for participation in formal education in the last 12 months for this age group.

However, results from surveys are not the only source of information on young adults' participation in education. The National Education Database (NUDB) contains information on participation at the individual level, and with the help of such register information, we want to improve the data quality in AES. Register information can be used in weighting to obtain more precise indicators. The use of register information in weighting also means that the sample size for the age group 18-24 years can be reduced and will thus lead to a reduced response burden.

As people aged 18-24 were not in the target population in AES 2016, we use data from another survey, the Learning Conditions Monitor (LCM), as the LCM also maps participation in formal education in the last 12 months and includes this age group.

The main objective of this task is to explore how register information on participation in formal education and training can be used in the weighting procedure to reduce bias in estimates. We explore this by comparing results from LCM with register information, and mapping how register information about participation in education can be used in the weighting scheme for AES.

Results

For the number of formal education participants according to the register, we use data from the NUDB. Data on pupils or students who are enrolled in or who have completed education are collected from the county municipalities' administrative computer system for admission to upper secondary education (VIGO) for figures for upper secondary education. For higher education, data are obtained from the Database for Statistics on Higher Education (DBH). DBH is also the source of data on education at a higher vocational level (post-secondary vocational school).

In Statistics Norway's official education statistics, a pupil or student is considered to have participated in formal education if he or she is enrolled in at least one education per October 1. For completed education, we have tested two different time ranges to see if they might influence the calibration and standard errors. Statistics Norway's official education statistics cover completed education during a school year, which lasts from October 1 to September 30 the year after. As LCM interviews are conducted in the period January - March, and the reference period for participation in education last 12 months thus becomes either January, February or March last year through January, February or March the current year, we want to see if we capture more participants in the registers than with the usual time frame of a school year. We test both the correspondence between LCM and current/completed education separately, and a combination of information from both register files, as described below:

Completed education type 1, four register files:
- Completed education in the period 01.01.2016 – 30.03.2017
- Completed education in the period 01.01.2017 – 30.03.2018
- Completed education in the period 01.01.2018 – 30.03.2019
- Completed education in the period 01.01.2019 – 30.03.2020

Completed education type 2, four register files:
- Completed education in the period 01.10.16 – 30.09.17
- Completed education in the period 01.10.17 – 30.09.18
- Completed education in the period 01.10.18 – 30.09.19
- Completed education in the period 01.10.19 – 30.09.20

Ongoing education, four register files:
- Registered in education on the 01.10.16
- Registered in education on the 01.10.17
- Registered in education on the 01.10.18
- Registered in education on the 01.10.19

Combined completed and ongoing education, four register files:
- A combination of files on completed education type 1 (completed education in the period January 2016 – March 2017) and ongoing education as of 1.10.2016
- A combination of files on completed education type (completed education in the period January 2017 – March 2018) and ongoing education as of 1.10.2017
- A combination of files on completed education type 1 (completed education in the period January 2018 – March 2019) and ongoing education as of 1.10.2018
- A combination of files on completed education type 1 (completed education in the period January 2019 – 2020) and ongoing education as of 1.10.2019

Learning Condition Monitor (LCM) data

The target population for LCM is persons aged 15-66 living in Norway. The survey focuses on participation in education and training, and data is collected through telephone interviews. The LCM is an annual supplementary survey to the Labour Force Survey (LFS) in the first quarter.[4] For the years examined, LFS has a response rate of approximately 85 per cent, which is largely transferable to LCM.

LCM has been conducted since 2008. To get a more comprehensive picture of the correspondence between LCM and register information, we have chosen to use the last four available years:

- Learning Conditions Monitor 2017
- Learning Conditions Monitor 2018
- Learning Conditions Monitor 2019
- Learning Conditions Monitor 2020

In LCM, an interviewee is considered to have participated in formal education if he or she answer yes to either of the following questions:

- Have you gone to school, studied or been an apprentice during the last 4 weeks? (Utd105a)
- Have you gone to school, studied or been an apprentice during the last 12 months? (A1)

---

[4] From 2022, the LCM is conducted every second year.

The correspondence between register files and LCM sample
For each person in LCM for the year t, we check whether he or she is in the following register files:

- Completed education type 1
- Completed education type 2
- Ongoing education

Completed and ongoing education
LCM 2017 is thus checked against files on completed education in the period January 2016 -March 2017, completed education in the period 1.10.2016-30.09.2017, ongoing education as of 1.10.2016 and the combination of completed education type 1 (completed education in the period January 2016 – March 2017) and ongoing education as of 1.10.2016. We do the same for the other LCM files and their corresponding register files in the matching time period. See the correspondence between LCM and the four versions of the education register data in table 3.1.

Table 3.1   Comparison of formal education in LCM with the occurrence in education registers. Years 2017, 2018, 2019 and 2020 combined

Formal education in LCM	Completed education type 1		Completed education type 2		Ongoing education		Combination completed and ongoing education	
	Yes	No	Yes	No	Yes	No	Yes	No
Yes	2 655	535	2 376	814	2 589	601	2 974	216
No	465	1 250	249	1 466	287	1 428	561	1 154

Source: Statistics Norway

We see that there is a large degree of agreement between LCM and register information for all four register files we look at. However, there are persons who have stated that they participated in formal education during the last 12 months, whose information we do not find in our registers, and vice versa. There can be several reasons to it. For example, current education is counted per October 1, but the register on ongoing education does not capture educations that have started at a later time in the school year.

Also, when we look at results for each individual year, there is a large degree of agreement between the LCM results and register information, as shown in table 3.2 – table 3.5.

Table 3.2   Comparison of formal education in LCM 2017 with the occurrence in education registers

Formal education in LCM 2018	Completed education type 1		Completed education type 2		Ongoing education		Combination completed and ongoing education	
	Yes	No	Yes	No	Yes	No	Yes	No
Yes	636	124	574	186	625	135	711	49
No	123	282	61	344	63	342	137	268

Source: Statistics Norway

Table 3.3   Comparison of formal education in LCM 2018 with the occurrence in education registers

Formal education in LCM 2018	Completed education type 1		Completed education type 2		Ongoing education		Combination completed and ongoing education	
	Yes	No	Yes	No	Yes	No	Yes	No
Yes	605	144	547	202	599	150	684	65
No	107	298	67	338	70	335	133	272

Source: Statistics Norway

Table 3.4   Comparison of formal education in LCM 2019 with the occurrence in education registers

Formal education in LCM 2018	Completed education type 1		Completed education type 2		Ongoing education		Combination completed and ongoing education	
	Yes	No	Yes	No	Yes	No	Yes	No
Yes	659	123	605	177	648	134	740	42
No	141	354	80	415	86	409	164	331

Source: Statistics Norway

Table 3.5   Comparison of formal education in LCM 2020 with the occurrence in education registers

Formal education in LCM 2018	Completed education type 1		Completed education type 2		Ongoing education		Combination completed and ongoing education	
	Yes	No	Yes	No	Yes	No	Yes	No
Yes	755	144	650	249	717	182	839	60
No	94	316	41	369	68	342	127	283

Source: Statistics Norway

Estimation by calibration against register education

We now want to estimate the total number of persons who have participated in formal education in the last 12 months in the age group 18–24 years based on LCM data. The easiest way to calculate this is to take the proportion (p) who have answered "Yes" to formal education in LCM, and then multiply it by the population size N. In this way, each person in the response group (net sample) is given the same weight N/n, where n is the size of the net sample. For such a method to work well, the net sample must be representative for the population we are to examine, i.e. the distribution of important variables is approximately the same in the sample as in the population. This is usually not the case. Those who do not respond to the survey will typically differ demographically from those who respond. Thus, some groups are overrepresented in the response group, while others are underrepresented. If the proportion p is not the same in the two groups, a population estimate calculated with the method described above will give a biased picture of the number we are looking for.

The sample weights must therefore be improved so that the groups which are least represented in the survey are given a larger weight than those that are well represented, and this can be achieved by calibration of the weights. The basis of calibration is that there is a correlation between the variable of interest, in this case participation in formal education in the last 12 months in LCM, and one or more register variables. The idea is that weights are constructed in such a way that the population totals of relevant register variables are correctly estimated, and that they provide a good estimate of the variable of interest. Here we will illustrate this by first calibrating against the register variable gender, which has a relatively small correlation with formal education in LCM, and then against gender + registered education, where registered education has a relatively large correlation with formal education in LCM, as shown below.

Calibration of weights can reduce both the bias and the standard error / uncertainty of the estimate. In table 3.6 – table 3.9, we see that including registered education in addition to gender reduces both the estimate for the number of people in education and its standard error. The reduction in the estimate indicates that the bias may have been reduced, i.e., that we have come closer to the truth, but this is not something we can say with certainty. The reduction in the standard error, however, is exclusively positive, and the size of this reduction can be used as a measure of which register education we should calibrate against.

The fact that calibration reduces the standard error given that the sample size does not change, also means that we, by calibrating, can reduce the sample size without increasing the standard error. From table 3.6 – table 3.9, we see a reduction in standard errors of at least 20 per cent. We know that the standard error is inversely proportional to the square root of the sample size. Hence, in order to achieve the same reduction in standard error without calibration, we must increase the

sample size by around 50 per cent, e.g. from 2000 to 3000, $(1/\sqrt{3000} \approx 0.8 \times 1/\sqrt{2000})$. In other words, using calibration means we can reduce the sample size by approximately 30 per cent.

Table 3.6   Estimate and standard error of the proportion under education in LCM, age group 18–24 years, using sample weights calibrated against various register variables. Year 2017

	Calibration against gender		Calibration against gender and register variable	
Register variable	Estimate	Standard error	Estimate	Standard error
Completed education type 1	65.25	1.21	63.16	1.21
Completed education type 2	65.25	1.21	62.92	1.19
Ongoing education	65.25	1.21	63.23	1.10
Combination of completed and ongoing education	65.25	1.21	63.61	1.07

Source: Statistics Norway

Table 3.7   Estimate and standard error of the proportion under education in LCM, age group 18–24 years, using sample weights calibrated against various register variables. Year 2018

	Calibration against gender		Calibration against gender and register variable	
Register variable	Estimate	Standard error	Estimate	Standard error
Completed education type 1	65.05	1.40	64.95	1.19
Completed education type 2	65.05	1.40	63.36	1.21
Ongoing education	65.05	1.40	63.95	1.13
Combination of completed and ongoing education	65.05	1.40	64.98	1.11

Source: Statistics Norway

Table 3.8   Estimate and standard error of the proportion under education in LCM, age group 18–24 years, using sample weights calibrated against various register variables. Year 2019

	Calibration against gender		Calibration against gender and register variable	
Register variable	Estimate	Standard error	Estimate	Standard error
Completed education type 1	61.33	1.35	60.82	1.13
Completed education type 2	61.33	1.35	59.52	1.12
Ongoing education	61.33	1.35	60.68	1.05
Combination of completed and ongoing education	61.33	1.35	61.42	1.02

Source: Statistics Norway

Table 3.9   Estimate and standard error of the proportion under education in LCM, age group 18–24 years, using sample weights calibrated against various register variables. Year 2020

	Calibration against gender		Calibration against gender and register variable	
Register variable	Estimate	Standard error	Estimate	Standard error
Completed education type 1	68.83	1.27	67.23	1.06
Completed education type 2	68.83	1.27	67.43	1.08
Ongoing education	68.83	1.27	67.11	1.06
Combination of completed and ongoing education	68.83	1.27	67.13	0.97

Source: Statistics Norway

Conclusion

We have investigated the connection between the information on education in LCM 2017–LCM 2020 and different types of register files. Overall, we find that there is a large degree of correspondence between reported participation in formal education in LCM and the register information on formal education that we have used. Regardless of whether we use information about completed education, current education, or a combination of these in the weighting, the standard error is considerably reduced compared to when we use only gender in the weighting.

When we compare the results across years, we find that the use of a combination of completed and ongoing education gives the lowest standard error for all the years we have examined. To use only

information about ongoing education per October 1 each year also gives very good results. The use of ongoing education also gives lower standard errors than when using information about completed education alone.

Combining information about ongoing and completed education provides the most comprehensive picture of participation in formal education in the last 12 months. However, the use of two different types of information about participation in education, means that the use of register information in the weighting becomes more complex. In addition, information about two different outcomes is mixed. Using only information about ongoing education, on the other hand, will simplify the use of register information in the weighting and ensure that the weighting system is not unnecessarily complicated. In addition, the use of only ongoing education ensures that we do not mix information about different outcomes.

Another argument for using only information about ongoing education is that the AES has a different data collection period than the LCM. While the AES2022 field period was November - January, the LCM is conducted in the first quarter of each year. Originally it was planned that the data collection period should be November – March and creating special files for completed education would thus have to cover a much longer period and should be adapted to the actual interview time to get the correct information. The starting point for data collection for AES is also quite close to the counting time for ongoing education, October 1.

We conclude that we use information about ongoing education per October 1 in the weighting for respondents aged 18–24 years in AES2022.

3.2. Revision of the sampling design

Main objective
The main objective of this task is to revise the sampling design for AES. It is necessary to revise the sampling design for the survey because the target population was changed from persons aged 25–64 years in AES 2016 to 18–69 years in AES 2022. Furthermore, the sampling scheme needs to be revised to take into account precision requirements in the regulation No 2019/1700 of the European Parliament and of the Council.

Results
The regulation No 2019/1700 of the European Parliament and of the Council specifies two precision requirements for AES:

- Participation rate in formal education and training (age 18–24)
- Participation rate in non-formal education and training (age 25–69).

The regulation specifies parameters a and b that shall be used in order to calculate the number of observations required to meet the precision requirements. This is calculated using the formula $a\sqrt{N} + b$, where N is the country population in the required aged group residing in private households, in million persons and rounded to 3 decimal digits. Table 3.10 shows the parameters used and the number of respondents (n) required using the formula.

Table 3.10   Parameters and calculated number of respondents to meet precision requirements

	N	a	b	n
Participation rate in formal education and training (age 18-24)	0.475	200	1 500	1 638
Participation rate in non-formal education and training (age 25-69)	3.086	400	2 000	2 703

Source: Statistics Norway

To calculate the sample size required to obtain results in accordance with the precision requirements, we assumed a response rate of 55 per cent and draw a sample of 3 000 persons in age group 18–24 years and 5 000 persons aged 25–69 years.

With these assumptions, the required sample size is 3 000 individuals for the age group 18-24 years. However, we have shown above that the use of register information in weighting makes it possible to reduce the sample size by one third. Therefore, the required sample size is 2 000 for this age group given the use of register information in weighting. For the age group 25 – 69 years, the required sample size is 5 000 individuals.

For the age group 18-24 years, we draw a simple random sample. For 25-69 years, we draw a stratified sample using age, gender and highest educational attainment in stratification. This is in line with the sampling approach used in AES 2016.

About age, target population and sampling frame
The sample was drawn in October 2022, and the start of the field phase was November 2022. The target population in AES 2022 is, as previously mentioned, residents aged 18–69 years, where age is defined as age at the time of the first interview. This means that the target population is actually undefined, as interviews only exist for the sample and not for the population outside the sample. Hence, the coverage error is also undefined. We can, however, try to construct a sampling frame in such a way that the sample is mostly in the target population, i.e., in the age group 18–69 years by the time of interview. In the previous AES, the sampling frame was defined as those who were in the group 25–64 years at the beginning of the data collection period, which was then October. This approach guaranteed that no one ended up outside the target population because they had not reached the age of 25. However, because the collection period at that time was six months, an age bias was introduced due to the fact that the persons in the sample were on average three months older at the time of interview than when the data collection started. If the interviews were randomly distributed over the entire collection period of six months, we would expect that 25 per cent had reached their next birthday before the interview.

In AES 2022, the data collection period is reduced to three months - from November 2022 to January 2023.[5] This alone will reduce the difference between age in the sampling frame and age at interview. To further reduce the difference, we now propose to define the sampling frame as those who are in the age group 18–69 years on 31 December 2022, which we call "register age". If the interviews with persons with register age x years, $x = 18, \ldots, 69$, are uniformly distributed over the data collection period, a simple probability calculation will show that 5.6 per cent of that group are expected to be $x - 1$ years at the time of the interview, while 1.4 per cent are expected to be $x + 1$ years. It is of course possible to reduce both these numbers by choosing time of interview time based on the date of birthday.

3.3. Revision of weighting scheme

Main objective
The main objective of this task is to revise the existing weighting scheme for AES. As already discussed, AES 2016 did not include age group 18-24. Previously, we have shown that using register information on participation in formal education and training reduces bias in estimates. The new weighting scheme therefore must be revised to include the age group 18-24.

---

[5] Originally, the field period was planned for November 2022 through March 2023, but this five-month period was additionally reduced due to lack of interviewer resources. The project team decided that shortening of the field period was defensible given the volume of CAWI interviews and because the user journey had insights had helped in allocating resources to improve the representativeness of the net sample.

Moreover, for the age group 25-69, we revise the existing weighting scheme by exploring how the use of additional weighting variables reduces bias in estimates.

Results
Age group 18–24 years
As shown in part 2.3, we find that the use of register information reduces bias in estimates for the participation in formal education and training considerably compared to using gender only as a weighting variable. Moreover, we have explored the use of different types of register information in weighting. Because we decided to use information on participation in education and training by October 1 each year, we developed a new weighting scheme for this age group, using both gender and register information on participation in education and training as weighting variables. The new weighting scheme is developed in R, using the R-package ReGenesees.

Age group 25-69 years
Previous rounds of AES included individuals aged 25 – 64 years, and Statistics Norway used the tool CLAN in the weighting procedure. Data were weighted and calibrated against population totals according to two dimensions. Dimension one is a combination of sex*age*educational attainment. Age consists of three age groups (25–34, 35–54 and 55–64 years) and educational attainment is divided into low, medium and high educational attainment. The second dimension is region (7 regions).

First, initial weights were calculated based on the number of individuals in the net sample compared to the known number of individuals within the different strata. These weights were then adjusted through calibration using known population totals for number of individuals in the 7 regions and number of individuals according to the dimension sex*age*educational attainment.

According to the guidelines for AES 2022, 'demographic characteristics such as age, sex and region should be used in the weighting. Due to its relevance in the context of AES, educational attainment level is another prioritised characteristic (European Commission, p. 220). In addition, main activity status and occupation are recommended as additional variables.

As part of the project, we investigated how the use of additional weighting variables can improve the quality of AES2022-data.

Variables used in weighting scheme
When testing different weighting approaches, we used the following variables:

- Gender (G)
  - Male
  - Female
- Age (A)
  - 25–34 years
  - 35–54 years
  - 55–69 years
- Education (E)
  - Low educational attainment (ISCED 0-2) or unknown education
  - Medium educational attainment (ISCED 3- 4)
  - High educational attainment (ISCED 5- 8)
- Labour status (L)
  - Employed
  - Unemployed
- Region (R), 6 categories

Notation

If the calibration model is written as G + A + E, it means that we calibrate the weights such that the eight population totals for gender (2), age (3) and education group (3) are correct. The model G + (A × E) means that we calibrate the weights so that the two population totals for gender and the nine population totals for cross classified age and education are correct.

The initial/design weights are set equal to 1. We also tried initial weights equal to post-stratified weights based on division into the nine strata A × E, which is the method used in the present weighting scheme, which had no effect on the results.

Table 3.11 shows that especially calibration against education and employment reduces both the estimate and its standard error. The calibration against only marginals generates as good an effect as calibration against cross-classification of variables. A model that is simple, gives good results and at the same time preserves important population totals is G + A + E + L + R. However, because AES guidelines require that the weighting procedure should respect the cross classified distribution of gender (2 groups), age (3 groups) and educational attainment level (3 groups), this simple calibration model is not an option. Therefore, we end up with the calibration model R + L + (G × A × E).

Table 3.11   Estimated proportion, with standard error (SE), age group 25–69 years, who have participated in informal training according to the Learning Condition Monitor (LCM), for different calibration models

Calibration model[1]	LCM 2020		LCM 2019		LCM 2018		LCM 2017	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
None (population total only)	44.41	0.49	42.68	0.49	42.72	0.49	41.32	0.49
G	44.43	0.49	42.69	0.49	42.73	0.49	41.32	0.49
G + A	44.36	0.49	42.49	0.49	42.76	0.48	41.41	0.49
G + E	43.08	0.47	41.38	0.47	41.46	0.47	39.98	0.47
G + L	41.88	0.46	40.34	0.46	40.90	0.46	39.54	0.46
G + A + E	43.07	0.47	41.26	0.47	41.50	0.47	40.07	0.47
G + A + L	41.92	0.46	40.34	0.46	40.99	0.46	39.67	0.46
G + E + L	41.27	0.45	39.66	0.45	40.24	0.45	38.76	0.45
G + A + E + L	41.28	0.45	39.67	0.45	40.29	0.45	38.85	0.46
G + A + E + L + R	41.32	0.45	39.77	0.45	40.32	0.45	38.89	0.46
G + (A × E)	43.16	0.47	41.29	0.47	41.56	0.47	40.08	0.47
G + (A × L)	41.97	0.46	40.38	0.46	41.05	0.46	39.73	0.46
G + (E × L)	41.29	0.45	39.77	0.45	40.30	0.45	38.89	0.45
G + A + (E × L)	41.30	0.45	39.77	0.45	40.34	0.45	38.97	0.45
G + A + (E × L) + R	41.34	0.45	39.86	0.45	40.37	0.45	39.01	0.46
R + L + (G × A × E)	41.37	0.46	39.77	0.46	40.39	0.45	38.92	0.46
Today's method	41.27	0.47	39.60	0.47	40.26	0.47	38.82	0.47

[1] Gender (G), Age (A), L (Labour status), E (Education), R (Region)
Source: Statistics Norway

Note that "today's method" refers to a method using post stratified initial weights and the calibration model R + (G × A × E).

Weighting and sampling frame

The weighting should respect the distribution of the sampling frame. To avoid using two different age definitions in the sample and population, we use register age, i.e. age per 31 December 2022, in all places where age is included. In practice, this means that a person with register age 24 years, but who is 25 years at the time of the interview, will be weighted against the 18–24 years population - with the calibration model belonging to this. Note that if we instead had weighted as if the person were in the group 25–69 years, the calibrated sampling weight would not have reflected the person's sampling probability, as he/she was sampled from the group 18–24 years and this group is overrepresented in the sample.

4. Summary self-evaluation

The aim of this project was to improve the data quality in the Norwegian AES 2022 by improving the user experience and by improving the sampling and weighting design.

The tasks related to the user experience consisted of three sub-activities: expert review, redesign and pretesting and focus groups. The activities not only contributed to identifying difficult questions in the questionnaire. A main advantage of our approach is that we mapped the user experience in a broader way. Previously, we focused only on the questionnaire and other elements such as invitation letter isolatedly, whereas the user journey captures the entire experience of participating in the survey. We gained important insight that helped us improve the implementation of AES2022, such as making the survey topic more salient for non-participants in education and optimizing mode combinations and resource utilization by using CATI first for participants with an education below ISCED3.

The project also gave us important insights into how to design user journeys in the future for AES and other surveys, as well as how to conduct user journey analyses for surveys to create better user experience.

We also identified potential improvements in the AES manual and example questions where we recommend that action is taken to improve them and to optimize them for mixed-mode data collection. This is transferable to other surveys of the European statistical system. The documentation of the various surveys should be reviewed, assessing the mixed mode suitability of individual survey questions and question formats. This could help improve user experience and data quality of statistics both on national and on European level.

With regard to the sampling and weighting, the main outcome is that we were able to identify register information that can be used to reduce bias in the estimates. By using register information on participation in education, we get more precise results for age group 18-24. At the same time, our findings show that using register information makes it possible to reduce the sample size for this group by 30 per cent, thereby contributing to a lower response burden.

Our findings also show that results for age group 25-69 can be improved by revising the weighting scheme by taking into account employment status. These findings can be transferred to other surveys on national and European level.

Regarding the implementation of the project, some tasks were delayed, while others were implemented according to the schedule. Moreover, the team decided to change methodology where it was found to be acceptable and necessary for the progress of the project. For example, the team decided to conduct another round of user testing to further improve the quality of the questionnaire, which delayed the finalisation of the questionnaire. However, the final questionnaire was delivered for programming before July 2022. Since user testing was conducted, large parts of the questionnaire had already been programmed.

Overall, the team is satisfied with the progress and delays in the project did not affect the schedule for the implementation of AES 2022 in Norway. The team is confident that work related to the user experience and the weighting/sampling design have given important insights that improved the quality in the Norwegian AES 2022, and that findings from this project are of high relevance to other surveys both on national and on European level.

References

Biemer. P.P. & Lyberg. L.E. (2003) *Introduction to survey quality*. Hoboken: Wiley.

Campanelli P,. Nicolaas G., Jäckle A., Lynn P., Hope S., Blake M., & Gray M. (2011). *A Classification of Question Characteristics relevant to Measurement (Error) and Consequently Important for Mixed Mode Questionnaire Design*. paper presented at the Royal Statistical Society. October 11. London. UK.

EU Grants: Application form (SMP ESS): V1.0 – 15 .04.2021

European Commission (2022). 2022 *AES-manual. Methodological guidelines for the adult education survey*. Version 1.2.

Gravem. D. F. & Berg, N. (2019). *Recommendations for key questionnaire elements. questions and question types in mixed mode settings. MIMOD WP4 deliverable 3.*

Murgia. M. et al. (2019). *Mixed-mode strategies for social surveys: how to best combine data collection modes. MIMOD WP1 - Deliverable 2*.

Kalbach. J. (2020). *Mapping experiences*. O'Reilly Media.

Lessler. J. T., & Forsyth. B. H. (1996). *A coding system for appraising questionnaires*.

Tourangeau. R., Conrad. F., & Couper. M. P. (2013). *The science of Web surveys*. Oxford University Press.

Tourangeau. R., Maitland. A., Steiger. D. & Ting. Y. (2020). A Framework for Making Decisions About Question Evaluation Methods. In P. Beatty., D. Collins., L. Kaye., J. Padilla., G. Willis & A. Wilmot. (Ed.) *Advances in Questionnaire Design. Development. Evaluation and Testing* (pp.47-73) John Wiley & Sons. Ltd.

Wilson. L. & Dickinson. E. (2021). *Respondent Centred Surveys: Stop. Listen and Then Design*. Sage.

Appendix A: Revised timetable

Tabell A1   Revised timetable

Milestone	Delivery date	Delivery date (actual)
Expert review completed	30 November 2021	30 November 2021
Focus groups completed	31 May 2022	31 May 2022
Cognitive and usability tests completed	31 January 2022	31 January 2022
Respondent journey mapping done	31 May 2022	31 May 2022
Questionnaire for programming submitted	31 May 2022	08 July 2022
Mapping of the use of register information completed	31 December 2021	14 January 2022
Weighting scheme revised	28 February 2022	06 April 2022
Sampling design revised	31 May 2022	31 May 2022

Source: Statistics Norway