KANTAR PUBLIC

Department for Digital, Culture, Media & Sport

DCMS Participation Survey 2021/22: Technical Note Quarter 1 (October-December 2021)

31 March 2022

Contents

1. Introduction

1.1 Background to the survey

In 2021, the Department for Digital, Culture, Media and Sport (DCMS) commissioned Kantar Public to design and deliver a new, nationally representative 'push-to-web' survey to assess adult participation in DCMS sectors across England. The new survey serves as a successor to the Taking Part Survey, which ran for 16 years as a continuous face to face survey.[1]

The scope of the survey is to deliver a nationally representative sample of adults (aged 16 years and over) in England. The data collection model for the Participation Survey is based on **ABOS** (**A**ddress-**B**ased **O**nline **S**urveying), a type of 'push-to-web' survey method. Respondents take part either online or by completing a paper questionnaire. In 2021/22 the sample consisted of approximately 33,000 interviews across two quarters of fieldwork (October-December 2021 and January-March 2022).

This technical note relates to Quarter 1 fieldwork, conducted between 1st October and 31st December 2021.

1.2 Survey objectives

The key objectives of this new survey are as follows:

- To inform and monitor government policy and programmes in DCMS and other governmental departments on adult engagement with the DCMS sectors. The survey will also gather information on demographics (e.g. age, gender, education).
- To assess the variation in engagement with cultural activities across DCMS sectors in England, and the differences in social-demographics such as location, age, education, and income.

---

[1] https://www.gov.uk/guidance/taking-part-survey

- To monitor the impact of previous and current restrictions due to the COVID-19 pandemic on cultural events/sites within its sectors, as well as feeding directly into the Spending Review Metrics, agreed centrally with the Treasury, to measure key departmental outcomes.

In preparation for the main survey launching in October 2021, Kantar Public undertook questionnaire development work and a pilot study to test various elements of the new design.[2]

1.3   Survey design

The basic ABOS design is simple: a **stratified random sample of addresses** is drawn from the Royal Mail's postcode address file and an invitation letter is sent to each one, containing username(s) and password(s) plus the URL of the survey website. Sampled individuals can log on using this information and complete the survey as they might any other web survey. Once the questionnaire is complete, the specific username and password cannot be used again, ensuring data confidentiality from others with access to this information.

It is usual for at least one reminder to be sent to each sampled address and it is also usual for an **alternative mode** (usually a paper questionnaire) to be offered to those who need it or would prefer it. It is typical for this alternative mode to be available only on request at first. However, after nonresponse to one or more web survey reminders, this alternative mode may be given more prominence.

Paper questionnaires ensure coverage of the offline population and are especially effective with sub-populations that respond to online surveys at lower-than-average levels. However, paper questionnaires have measurement limitations that constrain the design of the online questionnaire and also add considerably to overall cost. For the Participation Survey, **paper questionnaires are used in a limited and targeted way**, to *optimise* rather than maximise response.

---

[2] https://www.gov.uk/government/publications/participation-survey-methodology

2. Sampling

2.1 Sample design: addresses

The address sample design is intrinsically linked to the data collection design (see 'Details of the data collection model' below) and was designed to yield a respondent sample that is representative with respect to neighbourhood deprivation level, and age group within each of the 33 ITL2 regions[3] in England. This approach limits the role of weights in the production of unbiased survey estimates, narrowing confidence intervals compared with other designs.

The design also sought a minimum two-quarter respondent sample size of 900 for each ITL2 region. Although there were no specific targets per quarter, the sample selection process was designed to ensure that the respondent sample size per ITL2 region was approximately the same per quarter.

As a first step, a stratified master sample of 150,000 addresses in England was drawn from the Postcode Address File (PAF) 'small user' subframe. Before sampling, the PAF was disproportionately stratified by ITL2 region (33 strata) and, within region, proportionately stratified by neighbourhood deprivation level (5 strata). A total of 165 strata were constructed in this way. Furthermore, within each of the 165 strata, the PAF was sorted by (i) local authority, (ii) super output area, and finally (iii) by postcode. This ensured that the master sample of addresses was geographically representative within each stratum.

This master sample of addresses was then augmented by data supplier CACI. For each address in the master sample, CACI added the expected number of resident adults in each ten-year age band. Although this auxiliary data will have been imperfect, Kantar's investigations have shown that it is highly effective at identifying households that are mostly young or mostly old. Once this data was attached, the

---

[3] International Territorial Level (ITL) is a geocode standard for referencing the subdivisions of the United Kingdom for statistical purposes, used by the Office for National Statistics (ONS). Since 1 January 2021, the ONS has encouraged the use of ITL as a replacement to Nomenclature of Territorial Units for Statistics (NUTS), with lookups between NUTS and ITL maintained and published until 2023.

master sample was additionally stratified by expected household age structure based on the CACI data: (i) all aged 35 or younger (15% of the total); (ii) all aged 65 or older (21% of the total); (iii) all other addresses (64% of the total).

The conditional sampling probability in each stratum was varied to compensate for (expected) residual variation in response rate that could not be 'designed out', given the constraints of budget and timescale. The underlying assumptions for this procedure will be updated quarter by quarter as more evidence is obtained about how the Participation Survey works in practice, as opposed to ABOS studies in general.

Kantar drew a stratified random sample of 106,223 addresses from the master sample of 150,000 and systematically allocated them with equal probability to quarters 1 and 2. Kantar then systematically distributed the quarter-specific samples to two equal-sized 'replicates', each containing c.26,555 addresses with the same profile. The first replicate was expected to be issued six weeks before the second replicate, to ensure that data collection was spread throughout the three-month period allocated to each quarter.

These replicates were further subdivided into three differently-sized 'batches', the first comprising 80% of the addresses allocated to the replicate, and the second and third batches comprising 10% each. This process of sample subdivision into differently-sized batches was intended to help manage fieldwork, given that the Participation Survey was new, with an unknown response level.

All 26,555 addresses allocated to quarter 1, replicate 1 were issued. However, the response rate was considerably higher than had been anticipated, so only a systematically sampled fraction (7,170/21,245) of the first (80%) batch of replicate 2 needed to be issued in order to complete quarter 1. In total, 33,725 addresses were issued for quarter 1.

Figure 1 shows the quarter 1 (issued) sample structure with respect to the major strata.

Figure 1: Initial address issue					
	Area deprivation quintile group				
Expected household age structure	Most deprived	2nd	3rd	4th	Least deprived
All <=35	1,743	1,487	1,081	810	647
Other	4,702	4,649	4,317	3,691	3,196
All >=65	1,278	1,359	1,605	1,814	1,346

2.2   Sample design: individuals within sampled addresses

All resident adults aged 16+ were invited to complete the survey. In this way, the Participation Survey avoided the complexity and risk of selection error associated with remote random sampling within households.

However, for practical reasons, the number of logins provided in the invitation letter was limited. The number of logins was varied between two and four, with this total adjusted in reminder letters to reflect household data provided by prior respondent(s). Addresses that CACI data predicted contained only one adult were allocated two logins; addresses predicted to contain two adults were allocated three logins; and other addresses were allocated four logins. The mean number of logins per address was 2.8. Paper questionnaires were available to those who are offline, not confident online, or unwilling to complete the survey this way.

2.3   Details of the data collection model

Figure 2 summarises the data collection design within each stratum, showing the number of mailings and type of each mailing: push-to-web (W) or mailing with paper questionnaires (P). For example, 'WWP' means two push-to-web mailings and a

third mailing with paper questionnaires included alongside the web survey login information. In general, there was a two-week gap between mailings.

Figure 2: Data collection design by stratum					
	Area deprivation quintile group				
Expected household age structure	Most deprived	2nd most deprived	3rd most deprived	4th most deprived	Least deprived
All <=35	WWPW	WWWW	WWWW	WWW	WWW
Other	WWPW	WWW	WWW	WWW	WWW
All >=65	WWPW	WWPW	WWP	WWP	WWP

3. Fieldwork

3.1 Contact procedures

All selected addresses were sent an initial invitation letter containing the following information:

- A brief description of the survey

- The URL of survey website (used to access the online script)

- A QR code that can be scanned to access the online survey

- Log-in details for the required number of household members

- An explanation that participants will receive a £10 shopping voucher

- Information about how to contact Kantar Public in case of any queries

The reverse of the letter featured responses to a series of Frequently Asked Questions.

All non-responding households were sent two reminder letters, at the end of the second and fourth weeks of fieldwork. A targeted third reminder letter was sent to households for which, based on Kantar Public's ABOS field data from previous studies, this was deemed likely to have the most significant impact (mainly deprived areas and addresses with a younger household structure). The information contained in the reminder letters was similar to the invitation letters, with slightly modified messaging to reflect each reminder stage.

As well as the online survey, respondents were given the option to complete a paper questionnaire, which consisted of an abridged version of the online survey. Each letter informed respondents that they could request a paper questionnaire by contacting Kantar Public using the email address or freephone telephone number provided.

In addition, some addresses received up to two paper questionnaires with the second reminder letter. This targeted approach was, again, based on historical data Kantar Public has collected through other studies, which suggests that provision of paper questionnaires to all addresses can actually displace online responses in some areas. Paper questionnaires were pro-actively provided to (i) sampled addresses in the most deprived quintile group, and (ii) sampled addresses where it was expected that every resident would be aged 65 or older (based on CACI data).

3.2   Fieldwork performance

In total, 17,127 respondents completed the survey during quarter 1 – 14,394 via the online survey and 2,733 by returning a paper questionnaire. Following data quality checks (see Chapter 4 for details), 850 respondents were removed, leaving 16,227 respondents in the final dataset.

This constitutes a 48% conversion rate, a 31% household-level response rate, and an individual-level response rate of 28%[4].

For the online survey, the average completion time was 29 minutes.

---

[4] Response rates were calculated via the standard ABOS method. An estimated 8% of 'small user' PAF addresses in England are assumed to be non-residential (derived from interviewer administered surveys). The average number of adults aged 16+ per residential household, based on the Labour Force Survey, is 1.89. Thus, the response rate formula: Household RR = number of responding households / (number of issued addresses*0.92); Individual RR = number of responses / (number of issued addresses*0.92*1.89).

4. Data processing

4.1 Data management

Due to the different structures of the online and paper questionnaires, data management was handled separately for each mode. Online questionnaire data was collected via the web script and, as such, was much more easily accessible. By contrast, paper questionnaires were scanned and converted into an accessible format.

For the final outputs, both sets of interview data were converted into IBM SPSS Statistics, with the online questionnaire structure as a base. The paper questionnaire data was converted to the same structure as the online data so that data from both sources could be combined into a single SPSS file.

4.2 Quality checking

Initial checks were carried out to ensure that paper questionnaire data had been correctly scanned and converted to the online questionnaire data structure. For questions common to both questionnaires, the SPSS output was compared to check for any notable differences in distribution and data setup.

Once any structural issues had been corrected, further quality checks were carried out to identify and remove any invalid interviews. The specific checks were as follows:

1. **Selecting complete interviews**: Any test serials in the dataset (used by researchers prior to survey launch) were removed. Cases were also removed if the respondent did not answer the fraud declaration statement (online: QFraud; paper: Q88).

2. **Duplicate serials check**: If any individual serial had been returned in the data multiple times, responses were examined to determine whether this was due to the same person completing multiple times or due to a processing error. If

they were found to be valid interviews, a new unique serial number was created, and the data was included in the data file. If the interview was deemed to be a 'true' duplicate, the more complete or earlier interview was retained.

3. **Duplicate emails check**: If multiple interviews used the same contact email address, responses were examined to determine if they were the same person or multiple people using the same email. If the interviews were found to be from the same person, only the most recent interview was retained. In these cases, online completes were prioritised over paper completes due to the higher data quality.

4. **Interview quality checks**: A set of checks on the data were undertaken to check that the questionnaire was completed in good faith and to a reasonable quality. Several parameters were used:

    a. Interview length (online check only)

    b. Number of people in household reported in interview(s) vs number of total interviews from household.

    c. Whether key questions[5] have valid answers.

    d. Whether respondents have habitually selected the same response to all items in a grid question (commonly known as 'flatlining').

    e. How many multi-response questions were answered with only one option ticked.

Following the removal of invalid cases, 16,227 valid cases were left in the final dataset.

---

[5] CARTS1, CARTS2, CARTS3, CARTS4, CLIBRARY1, CHERVIS12, CDIGHER12, CMUSVIS1, CSPOLIVE, CDOMT12, CEVEAW & CINTUSE

4.3   Data checks and edits

Upon completion of the general quality checks described above, more detailed data checks were carried out to ensure that the right questions had been answered according to questionnaire routing. This is generally all correct for all online completes, as routing is programmed into the scripting software, but for paper completes, data edits were required.

There were two main types of data edit, both affecting the paper questionnaire data:
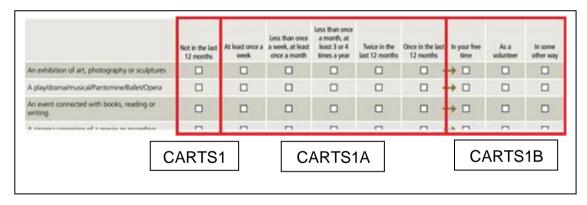
1. **Single-response questions edits**: If a paper questionnaire respondent had mistakenly answered a question that they weren't supposed to, their response in the data was changed to "-3: Not Applicable". If a paper questionnaire respondent had neglected to answer a question that they should have, they were assigned a response in the data of "-4: Not answered but should have (paper)".

2. **Multiple response question edits**: If a paper questionnaire respondent had mistakenly answered a question that they weren't supposed to, their response was set to "-3: Not Applicable". If a paper questionnaire respondent had neglected to answer a question that they should have, they were assigned a response in the data of "-4: Not answered but should have (paper)". Where the respondent had selected both valid answers and an exclusive code such as "None of these", any valid codes were retained and the exclusive code response was set to "0".

Other, more specific data edits were also made, as described below:

1. **Additional edits to library question**: The question CLIBRARY1 was formatted differently in the online script and paper questionnaire. In the online script it was set up as one multiple-response question, while in the paper questionnaire it consisted of two separate questions (Q15 and Q21). During data checking, it was found that many paper questionnaire respondents followed the instructions to move on from Q15 and Q21 without ticking the "No" response. To account for this, the following data edits were made:

a. If CFRELIB12 was not answered and CNLIWHYA was answered, set CLIBRARY1_001 was set to 0.

b. If CFRELIDIG was not answered and CNLIWHYAD was answered, CLIBRARY1_002 was set to 0.

c. CLIBRARY1_003 was set to 0 for all paper questionnaire respondents.

2. **Additional edits to grid questions**: Due to the way the paper questionnaire was set up, additional edits were needed for the following linked grid questions: CARTS1/CARTS1A/CARTS1B, CARTS2/CARTS2A/CARTS2B, CARTS3/CARTS3A/CARTS3B, CARTS4/CARTS4A/CARTS4B, CHERVIS12/CFREHER12/CVOLHER, CDIGHER12/CFREHERDIG/CREPAY5.

The figure below shows an example for the CARTS1 section in the paper questionnaire.

Marking the option "Not in the last 12 months" on the paper questionnaire was equivalent to the code "0: Have not done this" at CARTS1 in the online script. As such, leaving this option blank in the questionnaire would result in CARTS1 being given a default value of "1" in the final dataset. In cases where a paper questionnaire respondent had neglected to select any of the options in a given row, CARTS1 was recoded from "1" to "0".

4.4   Coding

Post-interview coding was undertaken by members of the Kantar coding department. The coding department coded verbatim responses, recorded for 'other specify' questions.

For example, if a respondent selected "Other" at CARTS1 and wrote text that said they went to some type of live music event, in the data they would be back-coded as having attended a "a live music event" at CARTS1_006.

For the sets CASRT1/CARTS1A/CARTS1B, CASRT2/CARTS2A/CARTS2B and CHERVIS12/CFREHER12/CVOLHER data edits were made to move responses coded to "Other" to the correct response code, if the answer could be back coded to an existing response code.

4.5   Data outputs

Once the checks were complete a final SPSS data file was created that only contained valid interviews and edited data. From this dataset, a set of data tables were produced.

4.5   Weighting

A three-step weighting process was used to compensate for differences in both sampling probability and response probability:

1. An address design weight was created equal to one divided by the sampling probability; this also served as the individual-level design weight because all resident adults could respond.

2. The expected number of responses per address was modelled as a function of data available at the neighbourhood and address levels. The step two weight was equal to one divided by the predicted number of responses.

3. The product of the first two steps was used as the input for the final step to calibrate the sample. The responding sample was calibrated to the August-

October 2021 Labour Force Survey (LFS) with respect to (i) gender by age, (ii) educational level by age, (iii) ethnic group, (iv) housing tenure, (v) region, (vi) employment status by age, (vii) household size, and (viii) internet use by age.

An equivalent weight was also produced for the (majority) subset of respondents who completed the survey by web. This weight was needed because a few items were included in the web questionnaire but not the paper questionnaire.

It should be noted that the weighting only corrects for observed bias (for the set of variables included in the weighting matrix) and there is a risk of unobserved bias. Furthermore, the raking algorithm used for the weighting only ensures that the sample margins match the population margins. There is no guarantee that the weights will correct for bias in the relationships between the variables.

The final weight variables in the dataset are:

- 'Finalweight' – to be used when running estimates for online and paper completes combined.
- 'Finalweightweb' – to be used when running estimates for online completes only.