

Taking Part Child Boost

Technical Note



Contents

1. Introduction	3
2. Sample Design	4
3. Questionnaire	6
4. Fieldwork	9
5. Data Processing and Outputs	13
6. Appendix	20

1. Introduction

Department for Culture, Media and Sport together with the Department for Education commissioned a boost of children aged 11-15 at secondary schools as part of the Taking Part Survey in order to capture detail about the children's participation in sport activities both in their free time and at school and about sport provision at their schools. The survey was carried out from April 2014 to March 2015. 1840 11-15 year olds were interviewed at their homes by specially trained interviewers using Computer Assisted Personal Interviewing (CAPI). The sample was drawn from the National Pupil Database and issued on a monthly basis, with the first sample issued in April 2014 and the final sample issued in March 2015.

This note outlines the technical aspects of the 2014/15 Taking Part Child Boost Survey. It is structured as follows:

- Chapter two provides a detailed description of the sample design;
- Chapter three provides an overview of the child boost questionnaire;
- Chapter four covers fieldwork – this includes all fieldwork and management procedures and a summary of fieldwork performance;
- Chapter five covers data processing and outputs, including weighting and design factors.

2. Sample Design

In order to expand the sample of 11 to 15 year-old individuals responding to the Taking Part survey, a boost sample was drawn from the National Pupil Database (NPD). Respondents from this boost sample (that includes children in state-school secondary education) were combined with respondents from the Taking Part main child sample (that includes children in all types of secondary education).

This chapter outlines our approach to sampling the boost sample of 11 to 15 year olds from the NPD.

2.1 Selection of the boost sample

The sampling frame for selecting the boost sample comprised records of eligible individuals (i.e. individuals of 11 to 15 years of age) within the 724 Taking Part main sample primary sampling units (PSUs)¹ identified within the NPD Autumn 2014 Census. Additional auxiliary information from the NPD Spring 2013 Census was appended to the records in the sampling frame in order to facilitate the stratification of the frame. The sampling frame was sorted by:

- PSU
- The child's gender (recorded at the Autumn 2014 Census)
- The child's age at the month when the PSU they fall under will be sampled (computed based on information recorded at the Autumn 2014 Census)
- The child's ethnic background (recorded at the Spring 2013 Census)
- The child's main language (recorded at the Autumn 2014 Census)
- Whether the child receives free school meals (recorded at the Spring 2013 Census)
- Whether the child receives special education needs provision (recorded at the Spring 2013 Census)

Using this sorted database, a sample was selected systematically, employing a probability proportionate to size (PPS) design², 70% of which would be issued as the boost sample while the remainder (30%) would be kept as a reserve.

The total probability of a child being sampled from the boost sampling frame equals the product of the PSU selection probability³ and the 'inclusion' probability of being drawn into the boost sample from the sampling frame⁴:

¹ See: 2014/15 Taking Part Survey Technical Report; Chapter 2.

² The inverse of the PSU selection probability was set as the measurement of size.

$$\Pr_{(\text{child-boost})} = \Pr_{(\text{PSU-selection})} * \Pr_{(\text{child-inclusion} \mid \text{PSU selected})}$$

³ See: 2014/15 Taking Part Survey Technical Report; Chapter 2.

⁴ The average inclusion probability was 0.185. However, this varied between different strata, ranging from a minimum of 0.004 to a maximum of 0.332.

3. Questionnaire

3.1 Questionnaire overview

The child boost questionnaire comprised mainly of sport section, similar to that from the main Taking Part 11-15 questionnaire, as well as household and demographics sections. It was designed to capture detail about the child's participation in sporting activities at school and in their free time and about sports provision at their school in the 12 months prior to the interview.

The child boost questionnaire was split roughly evenly between the child, who answered sports questions, and their parent or guardian, who completed the household and demographics sections.

3.2 Cognitive testing

In preparation for the child boost survey, a number of new questions on sports provision and the provision of specialist sports coaching at secondary schools were tested cognitively. Cognitive interviews were conducted either at the respondent's home or at the TNS BMRB office. TNS BMRB research team tested respondents' comprehension and probed for the thought processes adopted by respondents when constructing their answers. Respondents were friends or relatives of TNS staff, recruited following a request for volunteers aged 11-15 years who currently attend secondary school.

During the cognitive testing, 6 in-depth interviews were completed amongst the following profile of respondents:

Gender	Male	4
	Female	2
Age	11-12 years	3
	13-15 years	3

Interviews lasted between 15 and 20 minutes, and each respondent received £10 cash as an incentive for participation.

Following the cognitive testing, a number of new questions were added to the child boost and main Taking Part 11-15 questionnaires. The full report can be found in Appendix H.

3.3 Content of the questionnaire

The following sections were covered in the child boost questionnaire:

Household

This section of the questionnaire asked about number of adults and children living in the household and collected details about them, including names, sex, ages, marital status and relationship to the respondent. Furthermore, this section included question establishing who the Household Reference Person was.

School and school year

These questions collected information on which school the child went to and which school year the child was in at the time of interview. Alternatively, if the child did not attend school (for example they were in receipt of home education or had not yet started school) then this information was also collected at this point. The child was also asked whether they were on school holidays in 7 days prior to the interview.

Sport

This was the main section of the questionnaire, aiming to establish the child's level of sport participation and began with a question asking which sports the child had done in the last 4 weeks. This question was administered using a sport prompt pack consisting of a comprehensive list of sports. This was followed by questions asking which of the sports they had done in school lessons in last four weeks and in the last seven days, and similar questions asking about the sports done during school lessons. Then the respondent was asked about the number of days in the last week that they had spent participating in these sports for a minimum of 30 minutes in their own time.

Competitive sport

The sports participation section progressed by asking a couple of questions about competitive sport. These questions collected data on the types of sporting activities organised by the school and not organised by the school that the child took part in in the past 12 months.

Sport provision at school

Following questions about competitive sport, the child was asked which sports were offered by their school, both during school lessons and as part of activities outside of school hours. This section also collected information on sports their school did not currently offer, which the respondent would like to do. This was followed by a question asking whether the child had

received specialist sports coaching from external coaches or specialist sports teachers at school, from coaches or teachers, other than their regular PE or games teachers.

Well-being

This was a short section, consisting only of 1 question asking the respondent how happy they were on scale of 1 to 10.

NPD link

This was the last question asked of the child respondent. It asked for permission for the child's answers in the survey to be linked to their NPD data. If they said yes, it was also asked of the parent or guardian to ensure that we had their consent to create the link. If the child refused to give their permission, parent was not asked for their permission.

Demographics

This section was completed by the parent or guardian and collected detailed demographic information about the child respondent and the household, including social classification and income of the adult respondent and the HRP, household tenure and vehicle ownership questions.

Re-contact questions

The questionnaire concluded with a question establishing whether the adult respondent would be happy to be re-contacted by TNS BMRB in the future for similar research, followed by questions collecting their contact details.

4. Fieldwork

4.1 Fieldwork management

The fieldwork for the Taking Part Child Boost survey was managed on a monthly basis. Assignments were generally distributed evenly throughout the year, and were issued on a monthly basis, together with the main Taking Part assignments.

4.2 Briefings

Child boost briefings were incorporated in the main Taking Part briefings for new interviewers. Interviewers who had been briefed and had been working on Taking Part before the child boost survey commenced, were sent a memo and updated interviewer instructions containing information about the child boost survey and fieldwork procedures.

4.3 Fieldwork procedures and documents

4.3.1 Advance Letter and Leaflet

Parents of all selected children were sent a letter and a Taking Part Child Boost respondent leaflet in advance of an interviewer calling at the address. The letter packs were sent out by interviewers two or three days before they started their assignment.

The letters and the leaflets gave information about the survey, stressed its importance and its confidentiality, explained how respondents were selected, and informed about a respondent incentive for taking part. They also informed that an interviewer would call round in the next couple of weeks and contained contact details of TNS BMRB and TNS Field for respondents to use if they required more information about the survey, to make an appointment for an interviewer to call, or to opt out. The letters were despatched on DCMS headed paper in On Her Majesty's Service envelopes and signed by the project manager at DCMS to authenticate the survey.

4.3.2 Contact sheets

Similarly to the main Taking Part survey, interviewers were issued with a paper Address Contact Sheet (ACS) for each sampled respondent, which allowed them to carry out the different tasks that make up Taking Part Child Boost assignment and to record and manage their own calling strategies for each address.

An example of the ACS is included in Appendix C.

4.3.3 Fieldwork procedures

A minimum of eight calls were made at each address before regarding it as a non-contact. Interviewers were required to make calls on different days of the week and at different times of day: at least two of the calls had to be made on a weekday evening (after 7.00 p.m.) and at least one call at a weekend (10.00 a.m. – 9.00 p.m.). Having made contact, the interviewer needed to obtain written permission from a parent or guardian of the selected child to be able to attempt to complete the interview with the child. The adult was shown the Parental Permission Card (see Appendix D) indicating what the interviewer would be asking the child, and asked to sign the “parental/guardian permission” section of the address contact sheet. All children who completed an interview were incentivised with a £5 high street gift card.

Movers were not followed up in the child boost survey. If it was established that the named respondent had moved, they were treated as deadwood. Similarly, when respondent had turned sixteen by the time interviewer made contact with them, they were no longer eligible for the study.

4.4 Fieldwork figures

Table 4.1 shows the fieldwork outcomes for the child boost survey.

The final response rate was **72.3%**, with a contact rate of 91.7% and co-operation rate of 78.9%. The conversion rate was 61%.

Table 4.1 Child Boost Fieldwork Outcomes

Outcome		Outcome grouping		% of total issues	% of non-deadwood
Derelict/demolished	1	Deadwood	472	15.6%	-
Vacant/empty housing	16				
Named respondent has died	1				
Named respondent has moved from England	10				

Other ineligible	74				
Child now age 16 or over	125				
Respondent has moved	245				
No contact with anyone at address	163	Non-contact	211	7%	8.3%
No contact with parent	11				
No contact with named respondent	37				
Information about household refused	15	Refusals	349	11.6%	13.7%
Office refusal	64				
Parent refused permission to interview	114				
Refusal by named respondent	120				
Proxy refusal	36				
Contact made but no appointment made	23	Other unproductive	144	4.8%	5.7%
Broken appointment	42				
Respondent away throughout survey period	3				
Respondent ill at home during survey period	-				
Selected person physically or mentally	20				

unable to participate					
Selected person has inadequate English	-				
Other unproductive	56				
Full interview	1,840	Interview	1,840	61%	72.3%
TOTAL			3,016		

4.5 Interview length

The mean interview length was 27 minutes 56 seconds (median 25 minutes 53 seconds).

The interview length has been calculated after capping the lower and upper extreme values – the lower 0.2% and the upper 0.4% were capped. Extreme lower (including negative) and upper values are likely to have arisen from interviews being split into two or more sessions, since the computation is not date-sensitive (e.g. if an interview was concluded on a subsequent day but earlier in the day, the difference between relative start and end times could be negative, or unexpectedly small).

5. Data Processing and Outputs

5.1 Coding

The Taking Part child boost questionnaire had a number of full and partial open-ended questions.

Partial open-ended questions have response lists with an 'other specify' option. For the partial-opened questions, the coders were provided with the code frames used in the questionnaire as a starting point, which were the same as code frames in the main 11-15 Taking Part survey. The Coding team checked whether any of the verbatim responses could actually be coded in one of the pre-coded response options (this exercise is commonly known as back coding).

The coding of open-ended questions was carried out using a web-based package called Ascribe by an experienced team of coders. Five per cent of open-ended answers were checked by senior coders. New coders had 100% of their work checked until the required standard was reached and thereafter their work was systematically spot-checked.

The coding team also coded socio-economic data for this survey to produce Standard Occupational Classification (SOC2010) and National Statistics Socio-economic Classification (NS-SEC) categorisation, from a series of standard questions asked about Household Reference Person, which were designed for NS-SEC and SOC categorisation.

TNS BMRB researchers kept in close contact with the coding team throughout fieldwork to ensure that coding was carried out at regular intervals. At least every quarter of the survey year the coding was accessed by the TNS BMRB research team to check the quality of the coders' work in terms of what had been back-coded to each answer category, and to see what sort of answers had been left in "Other".

A list of all of the code frames used on open-ended and partially open-ended questions in the child boost survey can be found in Appendix G.

5.2 SPSS outputs

The main delivery was a SPSS file which contained all data from the Taking Part child boost survey, as well as relevant data from the main Taking Part 11-15 child and adult surveys. The dataset contained 1840 child boost cases and 650 main Taking Part cases. Interviews completed on each survey can be identified by filtering the dataset using the variable "samplotype".

The dataset included all variables from the child boost questionnaire, along with a number of derived variables and area variables. Details are provided in Appendix E (questionnaire) and Appendix F (list of all additional variables).

5.3 Weighting

This chapter outlines our approach to weighting the combined main and boost samples of 11 to 15 year-old respondents.

The combined main and boost responding sample of 11 to 15 year old individuals contains a total of 2,490 cases, which for the purposes of weighting were grouped into:

- **Fresh state-school cases.** This group includes 2,147 respondents in state-school secondary education responding to the Taking Part survey for the first time. Within this group, 307 cases came from the main sample and 1,840 cases came from the boost sample. In theory, cases in this group have a dual probability of being selected in the sample, as the main and the boost sampling frames are overlapping, both including 11 to 15 year-old children at state schools.
- **Fresh non-state-school cases.** This group includes 14 respondents in non-state-school secondary education responding to the Taking Part survey for the first time. These respondents could only have been sampled from the main sample frame (the Postcode Address File), as the NPD does not include children at non-state schools.
- **Repeat cases.** This group includes 329 respondents (out of which 316 are in state-school secondary education) who were sampled in a previous round of the Taking Part survey and re-interviewed.

The weighting process involves two phases:

1. The application of *pre-weights* that account for responding cases' differential probabilities of being sampled and of responding to the Taking Part survey; and

The application of *post-stratification weights*, which align the marginal distribution of key demographic variables (i.e. gender, age, and government office region) in the combined sample to their distribution within the actual population of 11 to 15 year old individuals.

The following paragraphs describe how the pre-weights and post-stratification weights were calculated.

5.3.1 Calculating the pre-weights

Pre-weights were specific to each of the three groups of cases identified in the combined sample (see paragraph 5.3) and their calculation involves three *base-weight components*:

- i. **The main sample base-weight BW_{main}** comprises three components; one that accounts for the probability of a household being selected in the main sample⁵, one that accounts for the probability of a particular child being selected within a specific household⁶, and finally one that accounts for the estimated probability of obtaining at least one interview from the household, given known information about the neighbourhood⁷:

$$BW_{main} = 1 / (Pr_{(address-main)} * Pr_{(child-main)} * Pr_{(response-main)})$$

- ii. **The boost sample base-weight BW_{boost}** comprises two components, accounting for the total probability of a child being sampled from the boost sampling frame (see paragraph 2.1) and for the probability of obtaining an interview given information on the NPD sample frame:

$$BW_{boost} = 1 / (Pr_{(child-boost)} * Pr_{(response-boost)})$$

To estimate the probability of obtaining an interview $Pr_{(response-boost)}$ a logistic regression model was fitted to the boost sample cases that were issued during the fieldwork process. Constructing the model involved assessing the role of several variables available within the boost sampling frame⁸ as possible predictors of whether an interview was obtained⁹. The final model included as predictors (a) whether English is the child's first language, (b) whether the child is receiving free school meals, and (c) a segmentation of the PSU where the child's household is situated. This segmentation of PSUs was implemented by means of the CHAID algorithm and was designed to cluster PSUs into groups of maximally different response probabilities¹⁰. Table 5.1 presents some key parameters of the response probability model.

Table 5.1: Probability of obtaining an interview from the issued boost sample – Logistic regression odds ratios and their 95% confidence intervals (C.I.s)

Variable (recorded in boost sampling frame NPD)	Odds ratio	Lower 95% C.I. for odds ratio	Upper 95% C.I. for odds ratio
Constant	0.749		
Whether English is the child's first language	0.828	0.681	1.005

⁵ The household sampling probability is equal to the address sampling probability assigned to the child's co-resident adult respondent multiplied by the probability of a certain dwelling unit being selected at the address, if more than one dwelling unit was found at the address (see 2014/15 Taking Part Survey Technical Report; Chapter 2).

⁶ The probability of a child being selected is equal to one divided by the number of eligible children living at the selected household.

⁷ The address-level response propensity is assumed equal to the modelled address-level response propensity of the co-resident adult respondent, computed based on the most recent twelve month issued sample for which fieldwork was complete (see 2014/15 Taking Part Survey Technical Report; Chapter 6).

⁸ Variables examined as candidate predictors of response were: gender, whether English is the child's first language, ethnic background, whether the child is receiving free school meals, whether the child receives Special Education Needs provision, and the PSU where the child's household is situated.

⁹ Candidate variables were assessed based on the likelihood ratio criterion.

¹⁰ The segmentation was dictated by the fact that logistic regression coefficients could not be calculated for all PSUs separately. PSUs and the response probability predicted based on a previous version of the model were used as input variables for the CHAID algorithm.

- Yes [vs. No]			
Whether the child is receiving free school meals - Yes [vs. No]	1.354	1.130	1.622
PSU segment 1 [vs. Reference PSU segment]	1.609	1.189	2.177
PSU segment 2 [vs. Reference PSU segment]	1.312	0.971	1.772
PSU segment 3 [vs. Reference PSU segment]	1.177	0.871	1.588
PSU segment 4 [vs. Reference PSU segment]	0.952	0.705	1.285
PSU segment 5 [vs. Reference PSU segment]	0.920	0.682	1.241
PSU segment 6 [vs. Reference PSU segment]	1.303	0.965	1.758
PSU segment 7 [vs. Reference PSU segment]	1.616	1.194	2.187
PSU segment 8 [vs. Reference PSU segment]	1.226	0.908	1.654
PSU segment 9 [vs. Reference PSU segment]	1.432	1.059	1.937

- iii. **The repeat sample base-weight BW_{repeat}** is equal to the population weight that was assigned to a repeat case for its previous interview. We note that the process for constructing the latest population weight is identical to the process of constructing the main sample base-weight BW_{main} (see paragraph 5.3.1.1) but also includes a post-stratification component. This post-stratification component was designed to align the marginal distribution of demographic variables (i.e. gender, age, and government office region) in the latest child Taking Part survey data to their distribution within the population of 5 to 15 year olds.

Using these three base-weight components, the group-specific pre-weights were calculated as follows:

- (a) The **fresh state-school pre-weight ($PW_{fresh-state}$)** was designed to account for this group's dual probability of being selected from the main sample frame or from the boost sample frame and then responding to the survey¹¹:

$$PW_{fresh-state} = (BW_{main} * BW_{boost}) / (BW_{main} + BW_{boost} - 1)$$

- (b) The **fresh non-state-school pre-weight ($PW_{fresh-non-state}$)** accounts for the probability of being selected from the main sample frame and then responding to the survey:

$$PW_{fresh-non-state} = BW_{main}$$

¹¹ We note that in order to calculate BW_{boost} for the cases coming from the main sample, the cases were assigned the average boost response probability for their PSU. This is because as for these cases one of the predictors used in the model (i.e. whether the child is receiving free school meals) was not available, thus not allowing the estimation of response probabilities based on the model coefficients presented in Table 1.

- (c) Finally, the **repeat pre-weight (PW_{repeat})** accounts for the probability of being sampled from the main sample and then providing a repeat interview to the Taking Part survey:

$$PW_{repeat} = BW_{repeat}$$

The distributions of the calculated pre-weights $PW_{fresh-state}$, $PW_{fresh-non-state}$, and PW_{repeat} were 'trimmed' in order to suppress extremely small or extremely large values thus limiting the pre-weight related design effect on the estimates produced based on the combined Taking Part dataset¹². The distribution of PW_{repeat} was then scaled so that the average PW_{repeat} would be equal to a combined $PW_{fresh-state}$ and $PW_{fresh-non-state}$ average. This ensured that the cases where PW_{repeat} was applied (and which should represent the target population of 11 to 15 year-old individuals in state and non-state secondary education) would not be systematically over or under-weighted in comparison to the cases where $PW_{fresh-state}$ or $PW_{fresh-non-state}$ were applied (which *only collectively* represent the target population).

5.3.2 Calculating the post-stratification weights

The combined sample was post-stratified by means of a RIM weighting algorithm. The algorithm (using the pre-weights to derive an initial distribution) aligned the marginal distribution of gender by age and of government office region in the combined sample to their distribution within the actual population of 11 to 15 year old individuals, as defined by the 2013 Mid-year Population Estimates published by the Office for National Statistics (Table 5.2). The weighting factors yielded by the RIM weighting algorithm were 'trimmed' to suppress extremely small or extremely large values¹³ and a second post-stratification was implemented using a Stata script. The design effect based on the final weights is estimated at 1.15¹⁴.

Table 5.2: Distribution of gender by age and government office region in the population of 11 to 15 year olds: Office for National Statistics, 2013 Mid-year Population Estimates

Subgroup	Population count
Male - 11 to 13 years old	904,544
Male - 14 to 15 years old	642,566
Female - 11 to 13 years old	863,475
Female - 14 to 15 years old	611,476
North East – 11 to 15 years old	139,880
North West – 11 to 15 years old	400,394
Yorkshire and The Humber – 11 to 15 years old	299,477
East Midlands – 11 to 15 years old	258,649
West Midlands – 11 to 15 years old	332,658
East – 11 to 15 years old	339,280

¹² The trimming exercise suppressed the maximum $PW_{fresh-state}$, $PW_{fresh-non-state}$, and PW_{repeat} weighting factor to five times the $PW_{fresh-state}$, $PW_{fresh-non-state}$, and PW_{repeat} medians, respectively. It also fixed the minimum $PW_{fresh-state}$, $PW_{fresh-non-state}$, and PW_{repeat} weighting factor to the $PW_{fresh-state}$, $PW_{fresh-non-state}$, and PW_{repeat} medians divided by five.

¹³ The maximum weighting factor was fixed to five times the median weighting factor, while the minimum weighting factor was fixed to the median weighting factor divided by five.

¹⁴ Design effect = $1 + (\text{Standard deviation of weighting factors} / \text{average of weighting factors})^2 = 1 + (478.142 / 1213.679)^2 = 1.15$.

London – 11 to 15 years old	455,079
South East – 11 to 15 years old	506,089
South West – 11 to 15 years old	290,555
Total	3,022,061

5.4 Design Factors

Significance tests assume that the achieved sample is a simple random sample from the survey population. The design factor takes into account the actual complexity of the sample design, reflecting the compromises necessary for real world survey practice. A standard error calculated using simple random sample assumptions should be multiplied by the design factor to obtain a robust standard error that properly reflects the sample design.

For Taking Part, the design is affected by clustering, weighting and stratification. Stratification usually helps to narrow the margin of error around estimates, while clustering and weighting increase the margin of error around estimates.

For the Taking Part Survey, a series of design factors are generated for the different sectors that the survey covers (arts, heritage, libraries, museums, galleries and archives, sport). The main reason different design effects are used for different sectors is due to the differential impact of the sample clustering. For child boost survey, which covers only sport sector, a similar approach to design factors was taken.

Table 5.3 details the typical design factor – and design effect - for sport sector in the child boost survey. The design effect is equal to the square of the design factor and shows how much bigger the sample would have to be to match the precision of a simple random sample.

Table 5.3: Typical design effect and design factor

Sector	Representative variable	Design effect	Design factor
Sport - Whether done at least one sports activity in last 4 weeks	c11anysport	1.284	1.133

Table 5.4 details the typical design factors for a number of key sub-groups.

Table 5.4 Child boost design factors by sub-group

	All	Limiting disability	No disability	BME	White	Male	Female
Sport - Whether done	1.133	0.932	1.194	1.373	1.073	1.044	1.133

at least one sports activity in last 4 weeks							
--	--	--	--	--	--	--	--

6. Appendix

A	Advance Letter
B	Respondent leaflet
C	Address Contact Sheet
D	Parental Permission Card
E	Questionnaire
F	Dataset Variables
G	Codeframe document
H	Cognitive testing report