

Choosing Web Surveys: mode choices among Youth Cohort Study respondents

Heather Wardle & Chloe Robinson

Abstract

Web surveys are increasingly being offered as an alternative mode of completion within many large scale surveys, with the view that this will increase rates of co-operation. The Youth Cohort Study (YCS), has offered respondents the option of completing the questionnaire on-line as an alternative to the standard postal questionnaire at all four sweeps of the 11th cohort. Beginning in 2002, the proportion of respondents choosing to complete the questionnaire on-line has increased from 1% in sweep 1 to 21% in sweep 4. Analysis of who chooses to complete the web survey shows that respondents do not consistently use the same mode of completion at subsequent sweeps of the study. By sweep 4, 65% of respondents who completed the questionnaire on-line had not chosen this mode previously and were new “web responders”. The characteristics of these new web responders were substantially different from those who returned the postal questionnaire. Those who completed the questionnaire on-line were more likely to still be in full time education than those who completed the postal questionnaire. Likewise, web respondents had significantly higher levels of academic qualifications than postal respondents. In short, the profile of web-respondents evident at sweep 4 is that of respondents who are already known to be the most likely to co-operate in the study. Therefore, in the case of YCS, offering a web survey option as an alternate mode of completion, an increasingly popular practice among survey researchers, is unlikely to help address non-response bias.

Keywords

Web surveys, attrition, non-response, mode choices

1. Introduction

The Youth Cohort Studies (YCS) have been running since 1985 and are a series of surveys among young people aged 16 and upwards, which monitor their decisions and behaviour in making the transition from compulsory education to further or higher education, employment or another activity. The main aim of the research is to identify and explain the major factors influencing these transitions after school, such as levels of educational attainment, training opportunities and school experiences. The YCS is funded and managed by the Department for Education and Skills (DfES). In this paper findings are described from the 11th cohort of the series which ran for four sweeps between 2002 and 2005. The sample members for the first

sweep of the survey were drawn from a sampling frame of 35,000 pupils taken from the 2000/01 year 11 school registers.

This paper will discuss questionnaire mode choices among respondents to this 11th cohort. Respondents were offered the option of completing a paper questionnaire and returning it by post or electronically submitting an equivalent, on-line, version of the questionnaire. Attempts were also made to interview cohort members at each sweep who had not responded by a certain date through the method of computer assisted telephone interviewing (CATI). CATI respondents have been excluded from the analysis in this paper because their characteristics (as initial non-responders) differ to those respondents who proactively returned a postal questionnaire or completed a questionnaire on-line. Only two mode choices were advertised to respondents, postal or web, and we are primarily interested in who chose each method. As well as both the postal and internet modes being proactive mode choices for the respondents, there are also minimal mode effects between them because both are administered in a visual format and neither involve an interviewer.² This allows for a reasonable comparison of the two modes..The focus of this paper is not to compare responses between the two modes but to compare the actual respondents to the two modes.

This paper will present the proportion of respondents who chose to complete the questionnaire on-line at each sweep of the 11th cohort of YCS and explore whether web-responders consistently chose this mode at each sweep as an indication of how preferable the internet is as a means for responding to the survey. An on-line version of the survey offers a number of advantages including being able to automatically monitor questionnaire returns and reduced financial and environmental costs and therefore there is considerable benefit in knowing whether people do opt to respond on-line. This paper will also look at who chooses to complete the survey on-line and consider whether an on-line option as an alternative, potentially more flexible and manageable, mode encourages those who may otherwise have not taken part in the survey.

2. Mode choices by YCS11 respondents

The first sweep of the Youth Cohort Study (YCS), 11th cohort, began in 2002. This was the first year in which respondents were presented with two options of how to complete the questionnaire. The first was the standard postal questionnaire, used in previous YCS cohorts. The second was a specifically developed web questionnaire, which replicated the postal survey questions. All selected respondents were presented with both options and given a unique web survey password with which to access the web questionnaire. Mode of completion was left entirely to the discretion and preference of the respondent. Additional interviews were conducted using CATI with a subset of the sample who, approximately 20 days from initial contact, had not responded by either the postal or web methods. For purposes of comparison, these initial “non-responders” have been excluded from the figures presented in table 1, so to focus on proactive mode choices made by YCS11 respondents.

¹² Kaminska, O., Bautista, R., Serrano, E. *Best Combination of Modes* The American Association for Public Opinion Research (AAPOR) 62th Annual Conference, 2007

Table 1 shows the proportion of respondents in each sweep between 2002 and 2005 who chose each method of completion.

Data collection mode	Sweep 1 (2002)	Sweep 2 (2003)	Sweep 3 (2004)	Sweep 4 (2005)
	%	%	%	%
Web survey	1	9	14	21
Postal survey	99	91	86	79
<i>Base</i>	<i>13719</i>	<i>5493</i>	<i>3729</i>	<i>3687</i>

Base: All postal and web responders at each sweep

The proportion of respondents who chose to complete the questionnaire on-line rose from 1% in 2002 to 21% by 2005. This represents a significant increase in the percentage of respondents choosing to complete the questionnaires on-line within each sweep. Part of this increase, particularly between sweep 1 and sweep 2, can be attributed to practical improvements in the study administration. For example, the internet option in sweep 2 was advertised in a pre-notification letter to cohort members and also contact was made with cohort members by e-mail after the pre-notification letter which contained a link to the web survey allowing cohort members to access the questionnaire more easily. E-mail addresses were systematically reviewed at sweeps 2 to 4, for example, any spaces were removed and words such as 'hotmail' were checked for spelling errors. Furthermore, for sweep 2 the following URL was purchased www.pathways2003.com, as it was deemed that the URL used previously was too long and not very memorable. The increase in respondents choosing to complete the questionnaire on-line between sweep 1 and sweep 2 was viewed by researchers to be largely related to these improvements.

These improvements were replicated for all subsequent sweeps of the study. As such, not all of the observed increase in proportion of respondents choosing the web survey as a mode of completion can be attributable to these administrative improvements. Furthermore, analysis of cohort members who chose to complete the web survey in sweep 2 shows that cohort members do not consistently use the same mode of completion for subsequent phases, but rather use different modes in different sweeps.

Table 2. Choices in mode of completion in Sweeps 2, 3 and 4, by sex

Data collection mode	Young Men %	Young Women %	Total %
Web survey respondent all sweeps	5	3	3
Mainly web survey respondent (chose web survey in 2 out of 3 sweeps)	7	7	7
Mainly postal survey respondent (chose postal questionnaire in 2 out of 3 sweeps)	11	15	14
Postal respondents all sweeps	77	75	76
Base	629	1376	2005

Base: All postal and web responders at sweeps 2, 3 and 4 (excludes CATI responders at any wave)

Table 2 shows that only 3% of cohort members who participated in sweeps 2, 3 and 4 chose to complete the questionnaire on-line in all sweeps.³ A further 7% of cohort members chose to complete the web survey in two sweeps and the paper questionnaire in one sweep, whilst an additional 14% of respondents completed the questionnaire by returning the paper booklet for two sweeps but also completed the questionnaire on-line for a least one sweep. The net result of these varying choices was that by sweep 4, 65% of cohort members who chose to complete the questionnaire on-line at this data collection phase had not chosen this mode of completion in earlier sweeps. These respondents were therefore “new” web respondents who were now choosing to mix their mode of completion compared with previous phases.

3. Web response and non-response profile

Profile of web respondents

As all respondents were offered the option to complete the questionnaire on-line, and increasing numbers of respondents chose to do so, even if they had not used this method in previous sweeps, insight into the profile of cohort members who chose this mode of completion can be gained.

Isolating co-operating cohort members from sweep 4, table 3 compares some key statistics between those respondents who completed the questionnaire on-line and those who completed the postal questionnaire.

³ This analysis focuses on those respondents who completed either the paper questionnaire or the web survey at all sweeps. Therefore any respondents for whom information was collected by the CATI questionnaire are excluded.

Table 3. Key statistics by mode of completion at sweep 4

Key statistics	Postal survey	Web survey
	%	%
Proportion with 5 or more GCSES, grades A*- C	76	90
Proportion reporting at sweep 1 all/most peers planning on going into higher education	72	80
Proportion who had been expelled or suspended	3	1
Proportion who ever truanted	27	20
Base	2915	772

Base: All postal and web responders at sweep 4 (excludes CATI responders at sweep 4 only)

A significantly higher proportion of respondents who chose to complete the web survey in sweep 4 had 5 or more GSCE's, grades A* - C than those who completed the postal questionnaire, 90% and 76% respectively. In addition to achieving higher levels of educational attainment, those completing the web survey were less likely to have been expelled or truanted whilst in compulsory education than their postal respondent counterparts, though some element of co-linearity between these factors and educational attainment may be expected. Multivariate logistic regression has been used to examine the independent factors associated with choosing to complete YCS11 questionnaire on-line in sweep 4. The regression technique adjusts for several explanatory variables simultaneously. Key variables of interest were entered, including a number of socio-demographic factors such as sex, National Statistics Socio-Economic Classification (NS-SEC) of the cohort member, educational attainment, and education/employment/training status. Only variables that were significant in the final model have been presented in table 4.

Table 4 Estimated odds ratios for completing the YCS11 questionnaire on-line in sweep

	N	Odds ratio	95% C.I. ^a
Ethnic group (p>0.05)			
White	5434	1	
Black/Black British	98	0.22	(0.07, 0.70)
Asian/Asian British	401	0.74	(0.53, 1.03)
Other	220	0.98	(0.65, 1.46)
Not answered	27	0.67	(0.16, 2.93)
Whether living with parents (p=0.000)			
Lives with both parents	2423	1	
Lives with one parent	630	1.14	(0.81, 1.61)
Does not live with either parent	2509	1.16	(0.93, 1.44)
Not answered	618	2.47	(1.99, 3.06)
NS-SEC of cohort member (p=0.000)			
Managerial/professional	463	1	
Intermediate	722	1.03	(0.67, 1.59)
Routine/semi-routine	2336	0.74	(0.50, 1.08)
Not answered	2659	1.46	(0.99, 2.15)
Highest level of educational attainment (p=0.000)			
2 or more A/AS levels	3522	1	
1-1.5 A/AS levels	393	0.70	(0.48, 1.02)
5 or more GCSES A*- C	915	0.69	(0.51, 0.92)
Less than 5+ GCSES A*- C	1350	0.44	(0.32, 0.61)
Whether in employment, education or training			
Full time education	3495	1	
Employment with training	771	0.46	(0.30, 0.68)
Employment without training	1048	0.50	(0.36, 0.70)
Other education and training	185	0.43	(0.22, 0.84)
Government Supported Training	292	0.44	(0.24, 0.81)
Not in Education, Employment or Training (NEET)	389	0.50	(0.32, 0.76)

The odds of completing the questionnaire on-line were significantly associated with educational attainment, with the odds becoming lower as level of maximum educational attainment decreased. For example, odds of completing the questionnaire on-line were some 0.44 times lower among those whose maximum educational attainment was less than 5 or more GCSEs, grades A*-C than those who had achieved 2 or more A/AS levels by sweep 4. Likewise, odds of completing the questionnaire on-line were also lower among those respondents who were in employment (either with or without training) than those who were in full time education and were lowest among those who reported being in other education or training (0.43). Other variables were significant in the final model, including NS-SEC status of the respondent, where the odds of completing the questionnaire on-line were lower among those in routine and semi-routine occupations (0.74) and whether the cohort member lived with any of their parents or not. Ethnic group was also significant, with the odds of Black or

Black British cohort members choosing to complete the questionnaire on line being 0.22 times lower than White cohort members. Gender, experience of truancy and/or suspension or expulsion were included in preliminary analysis, but these variables were not significant in the final model.

Analysis presented in tables 3 and 4 show that there are some key differences between cohort members who chose to complete the questionnaire on-line and those who did not. On-line responders are more likely to have a higher level of academic achievement, still be in full time education at the age of 19-20 and, when asked in sweep 1, more likely to have reported that most peers were continuing to full time education. These differences may not in themselves be surprising, easier access to the internet for those in full time education, for example, may be an important influencing factor. However, it is important to assess how these findings might affect overall survey precision, especially as data shows completing the questionnaires on-line was an increasingly popular choice among cohort members.

Gaining high response rates is considered an important quality marker of a study. However, in parallel to this, researchers must also consider the possibility that bias can be introduced into a study if the profile of those who do not respond to the study is vastly different to those who do. Cohort studies are effected by initial non-response, resulting from who refused to co-operate at the very first sweep, and also non-response between sweeps due to attrition in the sample. Earlier work on non-response to YCS has suggested that those who are less likely to co-operate are those who have lower levels of academic achievement, those who truanted at school, men, those of black ethnic origin, the unemployed, those looking after a family and those whose father is unemployed.⁴ The potential for bias to be introduced in the sample due to attrition is considered more fully below.

Attrition among cohort members

Cohort studies can be subject to bias due to attrition caused by cohort members “dropping” out in between data collection phases as these people may have different characteristics to those who remain in the sample. Analysis of YCS11 cohort members who responded at sweep 1 but had dropped out by sweep 4 shows that, like other cohort studies, this potential for bias caused by attrition between phases was evident. Table 5 shows that there are very real differences between YCS11 cohort members who participated in all four sweeps (by any mode) and those who had only partial co-operated in some sweeps and by sweep 4 were non-responders. Significantly more fully co-operating cohort members had 5 or more GCSES, grades A*- C than those who had dropped out of the sample by sweep 4; 75% compared with 59% respectively. Likewise, fully co-operating cohort members were also more likely to report that all or most of their friends planned on continuing into higher education (76%), whilst partially co-operating cohort members were somewhat less likely to report this (69%). Significantly more cohort members who did not participate in all four sweeps were more likely to report ever being suspended or expelled, or playing truant than their fully co-operating counter parts.

⁴ Taylor, S and Lynn, P (1996) England and Wales Youth Cohort Study (YCS). The effect of time between contacts, questionnaire length, personalisation and other factors on response to the YCS. DfEE Research Series.

Key statistics	Cohort members who responded at all four sweeps (fully co-operating)	Cohort members who did not respond to all four sweeps (partially co-operating)
	%	%
Proportion with 5 or more GCSES,	75	59
Proportion reporting at sweep 1 whether most/all friends were planning on going into higher education	76	69
Proportion who had been expelled or	4	8
Proportion who ever truanted	29	37
Sex:		
Male	40	48
Female	60	52
Base	6180	10527

Base: All respondents at sweep one using any mode

At sweep 1, 45% of the original cohort were male and 55% were female. Population estimates for 2002, showed that approximately 51% of the English population aged 16-17 were male and 49% female. Therefore, at sweep 1, women were already overrepresented and men underrepresented in the sample population. As table 5 shows, by sweep 4 (where respondents were aged 19-20), 60% of cohort members were female and 40% were male, indicating that attrition among cohort members was most acute among males. The net result was that the sex profile of fully co-operating respondents was further removed from the sex profile of the population as a whole. (Population estimates show that in 2005, 52% of 19-20 year olds were male and 48% female). This, along with differences in educational attainment and behavioural differences such as expulsion from school, indicate areas of potential bias introduced into YCS through attrition between sweeps.⁵

4. Discussion

Web surveys are increasingly being offered as an alternative mode of completion within many large scale surveys, often with the view that this will increase rates of co-operation. A web-based mode of completion was introduced into the YCS11 as it was felt that internet completion would appeal to the younger age group from whom information was being collected. Furthermore, it was also felt that mode effects would be minimal between postal and internet completion as both involve a visual channel of communication and neither involves an interviewer. The purpose of this paper was not to consider mode effects between each mode, but rather to assess any systematic differences between those respondents who chose to complete the questionnaire on-line and those completing the postal questionnaire. The increasing proportion of cohort members choosing to complete the questionnaire on-line across all four sweeps certainly shows that this was becoming an increasingly popular method

⁵ Final YCS data are weighted to address non-response.

of completion. By sweep 4 in 2005, one fifth of proactive, co-operating respondents chose on-line completion and it was thought that “increased access to the internet, particularly for young people aged 19/20 could have contributed to this considerable increase in the internet response”.⁶

Consideration of the profile of those who chose on-line completion in sweep 4 supports this assertion. Cohort members remaining in full time education, with a higher level of academic achievement, were most likely to complete the questionnaire on-line and possibly likely to have greater access to the internet through their respective academic institutions. It was thought by YCS researchers that offering flexibility through different mode options and incorporating an on-line option would appeal to the age group represented by the YCS11 cohort and have potential importance in maintaining the goodwill of cohort members. The fact that cohort members frequently swapped their mode of completion between different sweeps suggests that cohort members utilised a mode of completion at each sweep that was convenient for them at that time. By sweep 4, 65% of those who completed the questionnaire on-line had not do so in previous sweeps and were “new web responders”.

Offering the internet as a mode of completion gave cohort members flexibility which they increasingly utilised. However, analysis shows that the on-line option appealed to a particular subset of the cohort sample. Those who chose the on-line option had significantly higher levels of educational attainment, were less likely to be from routine/semi routine occupations, more likely to be still be in full time education at the age of 19/20 and reported significantly fewer behavioural problems, such as expulsion or suspension when in compulsory education. Although, notably there were no significant differences evident in the choices made between male and female respondents. When these characteristics are compared with those who by sweep 4 had become “non respondents”, the differences between the two groups are acute. The profile of those who “dropped” out between cohort sweeps showed that significantly greater numbers were male, those with a lower level of educational attainment and those who reported great behavioural problems when in compulsory education. Therefore, whilst offering the internet as mode of completion may be useful for respondents as it gives them greater flexibility and allows researchers to utilise a number of different contact methods with cohort members (for example, via e-mail), this analysis shows that those who are most likely to choose the on-line questionnaire are those people who are already known to be most likely to co-operate in all phases of the study. This reinforces the viewpoint that whilst, in the case of YCS, offering a web survey option may have greater appeal to some cohort members and some cost saving efficiencies, overall it is unlikely to help address non-response bias and any evaluation of this methodology should take this into account.

References

For ONS mid-population estimates see:

<http://www.statistics.gov.uk/statbase/Expodata/Spreadsheets/D9390.xls>

Finch, S (1996) England and Wales Youth Cohort Study: An investigation of the feasibility of using different data collection methods for the YCS, DfEE Research Series, 38, 1996.

⁶ Phillips M, Johnson M, Robinson C, Tipping S (2005) Youth Cohort Study: Survey of 18 and 19 year olds (Cohort 11 Sweeps 3 and 4) Technical Report. Prepared for Department for Education and Skills. National Centre for Social Research 2005.

Jarvis L, Exley S, Tipping S (2006) *Youth Cohort Study: Survey of 17 Year Olds (Cohort 11 Sweep 2)*

Technical Report Prepared for Department for Education and Skills. National Centre for Social Research 2005.

Knight I, "Measuring Change in Labour Force Panel Studies" *Social Survey Methodology Bulletin*, 51, 2003.

Phillips M, Johnson M, Robinson C, Tipping S (2005) *Youth Cohort Study: Survey of 18 and 19 year olds (Cohort 11 Sweeps 3 and 4) Technical Report*. Prepared for Department for Education and Skills. National Centre for Social Research 2005.

Taylor, S and Lynn, P (1996) *England and Wales Youth Cohort Study (YCS): The effect of time between contacts, questionnaire length, personalisation and other factors on response to the YCS*, DfEE Research Series.

About the Authors

Heather Wardle and Chloe Robinson are both Senior Researchers with the National Centre for Social Research, 35 Northampton Sq, London, EC1V0AX. The YCS study was funded by Department for Education and Skills (DfES). The views expressed in this paper are the authors' and do not necessarily reflect those of DfES.