Student Income and Expenditure Survey 2011/12

User Guide

Authors: Jess Bailey, Senior Research Data Manager at NatCen Social Research
Date: November 2014
Prepared for: UK Data Service (UKDS)
Contents

1 Background .............................................................................................................3
2 Survey Design ........................................................................................................4
3 Documentation .......................................................................................................5
4 Using the data .........................................................................................................6
  4.1 Variables on the files............................................................................................6
  4.2 Standard classification lists ................................................................................7
  4.3 Weighting variables ............................................................................................7
  4.4 Missing values conventions ...............................................................................9
5 SIES 2011/12 Report .............................................................................................10
1 Background

The data file contains data from the Student Income and Expenditure Survey (SIES) 2011/12; the survey was jointly commissioned by Department for Business, Innovation and Skills (BIS) and the Welsh Government (WG) and was conducted by NatCen Social Research (NatCen) and the Institute for Employment Studies (IES).

SIES is carried out at approximately three year intervals. However, the methodology used in the 2011/12 survey is very different from the previous surveys, the last of which was run in 2007/08. In 2011/12 a 25 minute web/telephone survey was used, whereas in previous waves of SIES the methodology consisted of a 60 minute face-to-face interview.

The 2007/08 data is also available from the UK Data Service (UKDS). The methodology for this latter data was kept as similar as possible to the previous wave carried out in 2004-2005 (not currently available from the UKDS). A research report for 2004/05 is available from the ‘Digital Education Resource Archive’ hosted by the Institute of Education.¹

The main aims of the Student Income and Expenditure 2011/12 survey were to:

- provide detailed information on the income, expenditure and debt levels of higher education (HE) students in England and Wales
- allow for analysis on larger and more memorable spending captured in the main questionnaire, as well as day-to-day spending recorded in the seven-day spending diary
- provide a baseline for assessing the impact of changes in student finance introduced in September 2012 for those starting HE in the 2012/13 academic year

Fieldwork was conducted between February 2012 and June 2012, broadly corresponding to the Spring and Summer terms in the academic year.

¹ [http://dera.ioe.ac.uk/view/organisations/NATCEN.html](http://dera.ioe.ac.uk/view/organisations/NATCEN.html) (last accessed 28/11/2014)
2 Survey Design

The survey aimed to cover England and Wales (with separate samples of English and Welsh-domiciled students) and to develop a representative sample of those students using random probability sampling.

For the Student Income and Expenditure Survey 2011/12, a number of institutions in England and Wales were selected using stratification and with probability proportional to their weighted size. Of the selected institutions 54 HEIs (including the Open University) and 31 FECs in England and 8 HEIs and 3 FECs in Wales agreed to take part and provided a sample of students. Full-time and part-time students were identified and sampled from each institution taking part. Students in their second year and above were drawn directly from Higher Education Statistics Agency (HESA) records, and students in their first year were randomly sampled by the institutions with guidance from IES. Further details are provided in the technical report.

In 2011/12, a 25 minute web/telephone survey was used, and all students were initially invited to complete an online questionnaire. They received up to four emails and one text message to remind them to complete the main questionnaire, and if they did not fill out the questionnaire online – or only partially completed – they were contacted by a telephone interviewer and given the opportunity to complete the survey over the phone. On completion of the main survey questionnaire, students were asked to fill out a seven day online spending diary, and were provided with a voucher on completion. The spending diary was only available online.

In total 2,986 full-time and 927 part-time students of English domicile responded to the survey and 914 full-time and 180 part-time students of Welsh domicile responded to the survey. This represents an overall response rate of 36 per cent.
3 Documentation

The documentation has been organised into two sections, besides this user guide:

- Interview (contains the household questionnaire)
- Data (contains the list of variables and derived variable specification).
4 Using the data

The 2011/12 data consists of one file including data from both the main survey questionnaire, and the spending diary:

<table>
<thead>
<tr>
<th>SIES 11/12 Questionnaire Archiving</th>
<th>5,007 records</th>
<th>data for all responding students who completed the main survey questionnaire.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIES 11/12 Spending Diary Archiving</td>
<td>2,625 records</td>
<td>data for all responding students who completed a seven day online spending diary on completion of the main survey questionnaire</td>
</tr>
</tbody>
</table>

4.1 Variables on the files

Each of the data files contains questionnaire variables (excluding variables used for administrative purposes) and derived variables. The variables included in the individual file are detailed in the “List of Variables” document in the data section of the documentation. This document is the best place to look at in order to plan your analysis. It includes:

- Full list of all variables included in data file (557 in total)
- Source of each variable (e.g. Household questionnaire, Spending diary)
- Major categories of variables (e.g. Course details, Background, Fees, Higher Education-related income, Student choices, Other income, Commercial credit, and Expenditure)

Once you have decided which variables to include in your analysis, you can look up details of the question wording in the household questionnaire or spending diary, using the interview section documentation, or use the “Derived Variables Specification” document in the data section of the documentation for derived variables.
4.2 Standard classification lists

The variables NSClass1-NSClass3, NSClasst, and NSClast3 are based on the National Statistics Socio-Economic Classification (NS-SEC). Full details of NS-SEC are in the Office for National Statistics guide 'The National Statistics Socio-Economic Classification User Manual'.

4.3 Weighting variables

The SIES dataset contains two weighting variables: a questionnaire weight and a diary weight and the choice of weighting variable to use depend on the variables analysed.

The SIES weights were calculated from the estimated probability of being both selected and responding to the survey. For the questionnaire, this overall probability was calculated as the product of two main components:

- the probability that the institution/student was selected
- the probability of taking part in the main interview.

The weights were calculated as the inverse of the overall probability. Extreme weights were trimmed at each stage so as to reduce their impact. Finally, the weights were adjusted so that the characteristics of the weighted sample matched that of the student population (as recorded by HESA) in terms of age and sex.

As with the 2007/08 survey the overall aim was to generate a weighted sample that matched the population of students as closely as possible, whilst at the same time generating weights that were not so variable that the standard errors of survey estimates were unnecessarily inflated.

As shown in the table above the diary response rate was low, with about half of questionnaire respondents also completing a spending diary. Significant discrepancies were also found in the profiles of respondents to the main questionnaire and those completing a diary. A separate non-response weight was therefore computed for analysis of diary data.

---

2 www.ons.gov.uk (last accessed 23/05/2014)
The questionnaire weight (\texttt{wt\_final}) should be used for analysis of only questionnaire variables. Any analysis that includes diary variables should be weighted using the diary weight (\texttt{wt\_d\_final}) which adjusts for both questionnaire and diary non-response. The “List of Variables” document and “Derived Variables Specification” document can be used to identify which expenditure variables are diary variables, and also which derived variables include information collected using the diary.

More detailed information on the weighting strategy and how the weights were calculated can be found in the technical report.

4.4 Complex sample design variables

The complex sample design of the survey outlined above and discussed in more detail in the technical report should be accounted for in any analysis by specifying the clustering variable (\texttt{PSU}) and the stratification variable (\texttt{STRATA}).

This can easily be done by using the relevant commands in the analysis software, for example by specifying a design plan in SPSS (using the CSPLAN command available in the CS module) or by using the svyset command in Stata. This ensures the correct treatment of complex sampling weights in the analysis and accurate calculation of standard errors.
4.5 Missing values conventions

-1 Item not applicable: Used to signify that a particular variable did not apply to a given respondent usually because of internal routing. For example, men in women only questions.

-8 Don’t know, Can’t say.

-9 No answer/ Refused

Derived Variables

The SIES questionnaire includes a number of question ‘sets’ which build to provide a figure for each element of student income and expenditure. Missing values could occur in any one of the questions that make up the set, and would lead to a missing value for the derived variable and any other derived variables higher up the scale. Prior to treatment of missing values, the cumulative size of the missing data (>20%) across the income data was relatively large, potentially introducing bias and reducing the power of the statistical analysis. For income data, it was decided therefore to give missing data an imputed value in order to retain all the cases for analysis, and to make full use of the data that students did provide. Imputed values were either a zero value or a median recipient value (based on the median value of a similar group of recipients). Non-zero values were used when there was sufficient additional data to be able to estimate a likely response value. For expenditure, savings and debt, the cumulative size of missing data was much smaller (<10% across the entire section) and missing values were treated as missing i.e. if the sets have over a certain amount of components missing then the value for the derived variable was set to missing. There are a few exceptions to this where zero values were imputed instead of being excluded:

- Total spending on children (XCHILD)
- Career development loan debt amount (XCDLDEBT)
- Access to finance (XACCDEBT)
- Commercial credit (XCREDIT)
- Used savings for income (XSAVEINC)
5 SIES 2011/12 Report

Further information about the Student Income and Expenditure Survey 2011/12 is provided in the Technical Report deposited with this data set, and in the Department for Business, Innovation and Skills research report on www.gov.uk: