

# Bright Futures: UK survey

## User manual

---

### Contents

‘Bright Futures’: A Comparative Study of Internal and International Mobility of Chinese Higher Education Students .....	2
Overview of the project .....	2
Scope of this data collection .....	2
Sampling design .....	2
Population of interest .....	2
Sample design .....	2
Stratification and selection of universities .....	3
Sampling of students within selected universities.....	3
Survey instruments .....	4
Fieldwork.....	5
Response rate at student level and sample size .....	5
Survey weights .....	5
Quality issues .....	6
Data processing.....	6
Variable naming conventions .....	6
Variable labels.....	7
Coding of open answers.....	7
Adaptive appearance questions .....	7
Types of missing values.....	7
Codebook .....	8
References .....	9
Project team.....	9

# **‘Bright Futures’: A Comparative Study of Internal and International Mobility of Chinese Higher Education Students**

## **Overview of the project**

Young people moving away from home to seek 'bright futures' through higher education are a major force in the urbanization of China and the internationalization of global higher education. Chinese students constitute the largest single group of international students in the richer OECD countries of the world, making up 20 percent of the total student migration to these countries. Yet systematic research on a representative sample of these student migrants is lacking, and theoretical frameworks for migration more generally may not always apply to students moving for higher education. Bright Futures is a pioneering study that investigates key dimensions of this educational mobility through large-scale, representative survey research in China, the UK and Germany. We explore this phenomenon in two related aspects: the migration of students from the People's Republic of China to the UK (this data collection) and Germany for higher education, and internal migration for studies within China. This research design enables an unusual set of comparisons, between those who stay and those who migrate, both within China and beyond its borders. We also compare Chinese students in the UK and Germany with domestic students in the two countries. Through such comparisons we are able to address a number of theoretical questions such as selectivity in educational migrations, aspirations beyond returns, the impact of transnationalization of higher education on individual orientations and life-course expectations, and the link between migration and the wellbeing of the highly educated.

Bright Futures is a collaborative project, involving researchers from University of Essex, University of Edinburgh, UNED, University of Bielefeld and Tsinghua University. The research was funded by the Economic and Social Research Council (UK), German Research Foundation (Germany) and the National Natural Science Foundation (China).

## **Scope of this data collection**

This data collection comprises of the UK survey data for Chinese international students and UK home students. Data of the other survey countries will be made available at a later date.

## **Sampling design**

### **Population of interest**

The population of interest are taught Chinese students studying in UK universities. This includes students in Undergraduate (UG) and Postgraduate (PGT) programmes. Not part of the target population are Postgraduate research students. UK home students studying at the same universities as the Chinese students serve as a comparison group.

### **Sample design**

The sample design is a two-stage stratified sampling, with universities as Primary Sampling Units (PSUs) and the secondary unit being the (Chinese students) at the universities selected for the sample.

We sampled 7,914 Chinese students, anticipating around 80% non-response. We sampled the same number of UK home students.

## Stratification and selection of universities

The sampling frame was constructed based on the most up-to-date data from the Higher Education Statistics Agency (HESA) at the time, pertaining to academic year 2013/2014. At university-level, the sample is stratified according to two factors that are significantly associated with the number of Chinese students enrolled in UK universities (Cebolla-Boado, Hu, Soysal, 2018). These factors are university prestige (ranking) and the number of Chinese students enrolled at the university.

The 115 universities in the sampling frame were sorted according to their position in the Guardian University ranking<sup>1</sup>, and then divided into quintiles: Five groups of universities, hosting an approximately similar number of Chinese students corresponding to roughly 20% or around 16,785 of the total Chinese student population in 2013/14 of 83,918.

Within each quintile the universities were then sorted by the size of their Chinese student population to create two sub-strata: one with universities that host more Chinese students (hosting a combined 50% of Chinese students in this quintile), and one with universities hosting the other half of Chinese students in this quintile (with each university hosting relatively fewer Chinese students). Of the resulting ten strata nine contain between 4 and 16 universities, while one stratum contains 44 universities due to the very small numbers of Chinese students hosted by many institutions in that stratum.

In each stratum between one and four universities were selected. If a university declined to participate, another university in the stratum was selected as replacement. With few exceptions due to practical constraints the selection probability is proportionate to the population of Chinese students. We contacted 47 universities in total in order to achieve the sample of 20 participating universities.

## Sampling of students within selected universities

Depending on the size of the Chinese student population in a sampled university, we asked universities to sample all eligible Chinese students or to take a random sample. For the Chinese sample, students with Chinese nationality, who were enrolled in a full (minimum 3-year) Undergraduate degree or in a taught (minimum 1-year) Postgraduate degree were eligible. This means students in a foundation year were not eligible. Students on 'linked programmes' (where part of the degree is delivered by a Chinese university, and part of the degree is delivered by a UK university, e.g. 1+3 or 2+2) were included. Students on exchange programmes were not eligible.

For the UK home students comparison sample the eligibility criteria were similar to that of Chinese students: UK nationals on a full (minimum 3-year) Undergraduate degree or on a (minimum 1-year) taught Postgraduate degree were eligible. In addition, the Home student sample was restricted to full-time students because Chinese international students can only study full-time. The Home student sample serves as a comparison group to the Chinese sample. Therefore, when sampling Home students, the ratio of UG to PGT Chinese students at a given university was applied. This means that PGT Home students, compared to UG Home students, are on purpose overrepresented.

---

<sup>1</sup> Other rankings were tested and delivered similar results.

Please note that the term ‘home students’ is to be understood as the opposite of ‘international’ students; it is not necessarily the same as ‘home student’ status in terms of tuition fees.

## Survey instruments

For the overall project with surveys in the UK, Germany and China, different questionnaires were developed, each adapted to respondent group and survey country.

The starting point were the questionnaires for the main population of interest: Chinese international students (in the UK and Germany). For the comparison groups (Chinese home students in China, and home students in the UK and Germany), instruments that were as comparable as possible were developed.

The questionnaire follows a temporal structure, from students’ past background, prior-migration and academic experiences/decisions, current well-being/experiences through to future life-course orientations. More specifically, the following areas are covered:

Sociodemographic characteristics and course details

Family background (parental education, occupation, household income, siblings)

Prior education (academic achievement and educational migration)

Motivations for study abroad and decision-making process

Personality traits and values (e.g. risk-taking attitude)

Study experience in current course

Health and wellbeing

Future life course aspirations

Cosmopolitan vs national orientations

To address the key research questions in the project, a large number of survey questions were developed by the Bright Futures research team. Decision-making related questions were based on the factors identified in the annual report of the Education International Cooperation Group (EIC) , adapted and further expanded for the project purposes (<http://www.eic.org.cn/special/report/> ; accessed March 2016). A number of survey experiments were used to produce answers that otherwise face a high risk of suffering from social desirability biases. These include questions on racism and xenophobia (list experiments); the respondents’ position in the axis of ‘cosmopolitan-national’ orientation (donation game). The newly developed questions were pre-tested with cognitive interviews, to ensure students understand the question as intended. This was followed by a small-scale pilot survey of the whole questionnaire in one university in December 2016.

Where available, suitable questions from other social surveys, such as the Chinese General Social Survey, China Family Panel Studies, Chinese College Student Survey, European Social Survey, European Social Survey and the World Values Survey, were used.

The questionnaire for Chinese international students takes around 20 minutes to complete. The questionnaire for UK home students is shorter as some aspects are not applicable to domestic students (e.g. migration decision). The questionnaire for Chinese students was in Chinese, for UK home students it was in English. As recommended for translations into languages not belonging to the same linguistic family, we opted for the use of multiple translators (Harkness et al, 2010). Three

professional translators provided independent translations from English into Chinese. Where the questions were extracted from existing Chinese surveys, the original wording in the Chinese surveys was used. A synthesized version of the translations was then verified against the original English version by another translator.

## Fieldwork

The survey fieldwork started in April 2017 and ended in April 2018. This means the fieldwork runs over academic years 2016/17 and 2017/18. The majority of the responses are from the academic year 2017/18.

Universities received detailed instructions on how to randomly sample students from their register. With few exceptions, students were invited to take part in the survey by their university via email. We provided email templates in Chinese and English for the survey invitation and with a set of personalised links to the survey which ensures each student can submit the questionnaire only once. We asked universities to send two reminders to increase response, which most of them did.

We offered Amazon vouchers as incentives for responding. Initially, the longer questionnaire for Chinese students was incentivised with £10, the shorter UK home student questionnaire with £5. In light of a much shorter than anticipated survey duration we reduced the amount for international students to £8 pounds. It is possible that this affected the response rate negatively, though it is hard to ascertain because the fieldwork timing and other factors on the side of the university also play a role.

## Response rate at student level and sample size

Response rates at the student-level differed substantially across universities. They range between 2% and 19% (mean 11.3%) for Chinese and 5% and 22% (mean 14.1%) for UK home students. Table 1 displays the achieved sample sizes for the two groups.

*Table 1 Achieved sample sizes*

Group	Chinese students (UK)	Home students (UK)	Total
Full response	13050	1581	2886
Partial response	141	97	238
<b>Total</b>	<b>1446</b>	<b>1678</b>	<b>3124</b>

## Survey weights

The dataset includes two survey weight variables, the sample design weight *sweight2* and the analysis weight *cweight2*. Analysis weights need to be applied in analysis in order to make the results representative of the Chinese student population in the UK.

The sample design weight *sweight2* adjusts for coverage error, i.e., differences between the sample frame (which is based on HESA data for 2013/14) and the actual population (HESA data for 2016/17,

which is used as reference population)<sup>2</sup>. The willingness to respond to a survey may vary according to the characteristics of the students. The analysis weight, *cweight2*, helps to adjust for this differential survey non-response. It calibrates the data to characteristics of the reference population using the following information: Stratum, level of study (UG versus PGT), gender, age, subject area.<sup>3</sup> The analysis weight incorporates the sample design weight.

Note that British respondents are calibrated to the characteristics of the Chinese student population as it is meant to serve as a comparison group only. The survey data on British students, with or without using survey weights, is not suited to describe the British student population in the UK.

## Quality issues

In a small number of cases, the sample information provided by participating universities suggest that the sampling criteria were not entirely followed and the sample included some ineligible students (e.g. postgraduate research students). Once we had noticed this, we added an additional response category “postgraduate research student” to the questionnaire so that these students could be screened out. As far as possible, ineligible respondents have been removed from the data.

## Data processing

### Variable naming conventions

The variable names consist of a main part and where necessary one or two suffixes. The first part describes the question that the variable pertains to. The suffix(es) are separated from the main part by an underscore (\_). The first suffix uses lower case letters and describes the respondent group (c for Chinese; h for UK home students). The second suffix consists of upper-case letters and describes in which survey country/countries the question was asked.

The suffixes distinguish variables where the underlying question differs either in terms of question wording, or in terms of response options. Questions where both question wording and response options were the same across all groups (and survey countries) have no suffix. Equally, for questions that were only asked of one group (e.g., *alevels1*, the number of A-levels with grade A+, was only asked of UK home students) and there is no equivalent question for other groups, no suffix is used.

For example, father’s and mother’s highest level of education was asked using standard questions for the UK and China for UK home students and Chinese students, respectively. The variable names indicate this with the suffix: *fedu\_c* is father’s education for Chinese students; *fedu\_hUK* is father’s education for UK home students.

Similarly, if a question varied slightly between survey countries (e.g., the question wording is the same, but the countries listed as response options differed), this is captured by using a country

---

<sup>2</sup> At the time when survey weights were calculated, 2017/18 HESA data were not yet available. Furthermore, distinguishing between the two years of the fieldwork as reference would have resulted in inefficiencies in the calculation of the weights. Therefore, it was decided to ignore potential differences between the two years and use 2016/17 as reference year.

<sup>3</sup> Subject area groups university subject into five areas: Business and economics; science, engineering, computer science and maths; humanities, social science (excl. economics) and education; arts and architecture; other (which includes in particular medicine and law).

suffix. For example, the grading system for UG degrees differs across survey countries, therefore expected final grade for (Chinese and UK home) UG students in the UK has the suffix UK: *expgradeug\_UK*.

As this data collection only comprises the UK survey, the country suffix “UK” is somewhat redundant, however, it is retained so that the data can be combined with the data from the remaining survey countries which will be deposited at a later date.

*Table 2 Suffix denoting respondent group*

<b>Respondent group</b>	<b>suffix</b>
Chinese students	c
(UK) home students	h
<b>Survey country</b>	
Question asked in UK only	UK
Question asked in some other but not all survey countries	Other country suffixes as appropriate, e.g. UKCN

## Variable labels

The variable labels describe the question content and (in the case of multiple choice questions) also the response option. For questions that appeared with slightly different wording in the questionnaire for different groups, e.g., for Chinese students in the UK and Germany, we use placeholders. For example, *arriv\_i* is labelled “year of arrival in COUNTRY OF STUDY” instead of “year of arrival in the UK” because the equivalent question was asked of international students in Germany as well.

## Coding of open answers

Many closed questions had a response option “other, please specify” with a text box provided to write in text. These open responses were re-coded into another existing category as far as possible. The open text variables are not part of the dataset.

## Adaptive appearance questions

For some questions, if the respondent did not respond, the question would be shown a second time (either in the same format, or in a simplified version). The variables relating to the second appearance have the same name as the variable relating to the initial question, with an “x” at the end of the name (but before any suffixes). For example *foccp* is Father’s occupation, *foccp<sub>x</sub>* is the ‘2<sup>nd</sup> appearance’ version of this. The variable label indicates this by starting with “2<sup>nd</sup>”.

## Types of missing values

The dataset differentiates between a number of missing value types for item nonresponse and other missing values (see Table 3Table 1).

Most questions did not offer a don’t know or refusal option. Therefore, one cannot know for which reason a respondent did not respond to a question that he or she did see. This type of item non-response is coded with -99.

For the few questions where explicit ‘don’t know’ or ‘refuse’ options were offered codes -97 and -98 are used.

*Table 3 Types of missing values*

Code	Type	Note
<i>Item nonresponse</i>		
-95	<b>not filled in</b> (most likely meaning 0). Variables <i>nysister, nesister, nybrother, nebrother</i>	In the early part of the fieldwork, the question about number of siblings had textboxes to write in. Most respondents seemed to leave boxes empty to signify “0”, but this can’t be distinguished with certainty from skipping the question as a refusal.
-96	<b>not applicable</b> Variables <i>foccpgp, moccpgp</i>	
-97	<b>don’t know</b>	Only for questions that had an explicit don’t know option (see questionnaires)
-98	<b>refused</b>	Only for questions that had an explicit refusal option (see questionnaires)
-99	<b>skipped</b>	The respondent has seen this question but did not answer.
<i>Other missingness</i>		
.	<b>filtered/system missing</b>	Question did not apply to respondent or respondent broke off prior to question
.a	<b>not determinable/ not codeable</b>	For example, open response too vague to be coded
.b	<b>implausible</b>	For example, value for BMI or family income too low to be feasible

## Codebook

The codebook shows the labels assigned to the values for each variable. For missing value codes that are the same across all variables please refer to Table 3.

A variable note indicates if, compared to the question in the questionnaire, response categories have been collapsed in order to reduce statistical disclosure risk. Derived variables are also explained.



## References

Héctor Cebolla-Boado, Yang Hu & Yasemin Nuhoğlu Soysal (2018) Why study abroad? Sorting of Chinese students across British universities, *British Journal of Sociology of Education*, 39:3, 365-380, DOI: 10.1080/01425692.2017.1349649

Janet A. Harkness (ed) et al., *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Hoboken N.J.: Wiley, 2010

## Project team

Principal investigator	Prof. Yasemin Soysal, University of Essex
Co-investigators	Dr. Hector Cebolla Boado, UNED Dr. Sophia Woodman, University of Edinburgh
Senior Researcher	Dr. Dorothee Schneider, University of Essex
Project website	<a href="http://brightfutures-project.com/">http://brightfutures-project.com/</a>