The mathematical backgrounds of undergraduates: Interim report

By

Jeremy Hodgen[i], Michael Adkins[i] and Anthony Tomei[ii]
[i]University of Nottingham, [ii]King's College London

January 2017

Contact:
Professor Jeremy Hodgen
School of Education
University of Nottingham
Dearing Building (Room C32)
Wollaton Road
Nottingham, NG8 1BB
Email: jeremy.hodgen@nottingham.ac.uk

Please cite as: Hodgen, J., Adkins, M. & Tomei, A. (2017). *The mathematical backgrounds of undergraduates: Interim report.* Nottingham: University of Nottingham.

Copies of this interim report are available for download from:
www.nottingham.ac.uk/research/mathematical-backgrounds

Acknowledgement: This study is funded by the Core Maths Support Programme. The authors are grateful for their support.

Introduction

This interim report describes the initial findings of an investigation into the mathematical backgrounds of students from England entering UK universities. Specifically, we report our findings broken down into HESA's 19 high-level subject groups, together with findings from two of those subject groups at a more fine-grained level. The study builds on our earlier work in this area and we hope that the findings will contribute to the ongoing debate on post-16 Mathematics in England.[1]

Our full report, due to be available in Spring 2017, will report on the mathematical backgrounds of students in all 140 Principal Subjects together with an analysis of the distribution across HEIs. This will form the basis of a resource which will enable staff in higher education to examine the mathematical backgrounds of students taking courses in their disciplines, and to see how their institutions compare with others.

The report draws on a dataset formed from the combination of two large national databases: the National Pupil Database (NPD) and the Higher Education Statistical Agency (HESA) database. The dataset tracks 253,557 students who took their GCSEs at age 16 in 2008, took a level 3 qualification at some point between 2009 and 2011 and entered university at some point between 2010 and 2012. The data cover all HE subjects.

Our methodology is described in Appendix 1. The dataset is subject to some limitations:

- Our data extract includes grades for GCSE and A-level/AS-level Mathematics and a variable indicating each student's principal Level 3 route (A-level, BTEC, IB etc). We do not have records of students' actual BTEC qualifications. As a result, one limitation is that we do not have details of any Level 3 Mathematics taken by students as part of BTEC or vocational qualifications. In some cases, notably those wishing to study Engineering, these qualifications do contain some Level 3 Mathematics, including in some cases calculus.
- Our focus is on English-domiciled students who were aged 16 in 2007/8. As a result, our analysis does not include students from Wales, Scotland or Northern Ireland, international students, EU students, mature students and other older students who were aged 16 prior to 2007/8.
- For students on combined degrees, we follow the HESA practice of apportioning students in proportion to the contribution of each separate subject.[2] This 'full-time equivalent' approach avoids double-counting these students. As a result, our estimates will differ from actual headcount figures.
- Our dataset records GCSE Mathematics grades at age 16, but does not record GCSE retakes between the ages of 16 and 19. Additionally, the NPD at that stage did not record iGCSE records, so these data are missing.

[1] Hodgen, J., McAlinden, M., & Tomei, A. (2014). *Mathematical transitions: a report on the mathematical and statistical needs of students undertaking undergraduate studies in various disciplines*. York: Higher Education Academy (HEA); Hodgen, J., Pepper, D., Sturman, L., & Ruddock, G. (2010). *Is the UK an outlier? An international comparison of upper secondary mathematics education*. London: The Nuffield Foundation.

[2] For more detail on this 'full time equivalent' type approach, see: https://www.hesa.ac.uk/index.php?option=com_content&view=article&id=2880&limit=&start=#subject2d

The mathematical backgrounds of undergraduates: initial results

Table 1 gives the mathematical backgrounds of students, broken down by Subject Group. (HESA and UCAS use the Joint Academic Coding System (JACS) to categorise university subjects for statistical and planning purposes.[3] University subjects are aggregated into 140 Principal Subjects that are then further aggregated into 19 high-level Subject Groups.) We have divided the Subject Groups into three categories which we have called High, Medium and Low demand subjects, based on an estimation of the mathematical demands of most of the subjects in those Groups.

- **High demand**. Subjects for which a level 3 mathematics qualification, usually A or AS level, is in most cases essential.
  Subject Groups: Engineering and Technology, Mathematical sciences, Physical sciences.
- **Medium demand.** Subjects which need mathematics, but where the demands are usually not as high as for the high demand subjects. In many cases the primary need is for statistics and the ability to analyse and interpret data.
  Subject Groups: Agriculture and related subjects, Architecture, building and planning, Biological sciences, Business and administrative studies, Computer science, Education, Medicine and dentistry, Social studies, Subjects allied to medicine, Veterinary science.
- **Low demand**. Subjects with generally low mathematical demands.
  Subject Groups: Combined/general subjects, Creative arts and design, Historical and Philosophical studies, Languages, Law, Mass communications and documentation.

These are of course approximate categories. There are many exceptions within them and there is ample room for debate about where individual subjects belong. The detailed data contained in our final report will make possible more fine grained analysis, based on detailed knowledge of the mathematical demands of different subjects.

- *For High demand subjects*, 64% of students had a post 16 mathematics qualification. Somewhat surprisingly, as many as 40% of engineering and technology students and 47% of physical sciences students do not have a mathematics qualification beyond GCSE.
- *For Medium demand subjects,* 24% of students had a post 16 mathematics qualification. By way of illustration, in Tables 2 and 3, we show the breakdown of the biological sciences and business and administrative studies groupings into principal subjects. We note that for the two largest subjects in the subject groupings, psychology and business, only 15% of students have a mathematics qualification beyond GCSE.
- *For Low demand subjects,* 12% of students had a post 16 mathematics qualification

Overall, the proportion of students (10%) in our dataset who did not achieve a C grade at GCSE Mathematics at age 16 is surprisingly high. It is important, however, to note that this figure does not include GCSE Mathematics retakes. We estimate the proportion of students entering university without at least a C grade in GCSE Mathematics to be somewhere between 5% and 7.5%.[4]

[3] Our analysis uses the JACS categories that were in place between 2010/11 and 2012/13: https://www.hesa.ac.uk/jacs3

[4] Between 2008 and 2011, approximately 3% of the entire age cohort gained a C grade in GCSE Mathematics between the ages of 16 and 19. See Wolf, A. (2011). Review of Vocational Education: The Wolf Report. London: DfE.
As a result, those retaking GCSE were likely to be skewed towards students intending to take university degrees. Additionally, proportion of students whose iGCSE Mathematics grades are missing from the NPD is likely to be skewed towards higher attaining students from independent schools.

Subject Grouping	Total Number of Students	Post-16 Mathematics (AS, A2, IB)	Maximum Pre-University Mathematics Qualification: All GCSE	A*	A	B	C	D-G	Total GCSE	Total with GCSE B or C
High Mathematical Demand										
Engineering & technology	13711	60%	40%	0%	5%	14%	15%	6%	5503	3887
Physical Sciences	13385	53%	47%	2%	12%	18%	11%	2%	6321	3781
Mathematical sciences	6081	97%	3%	0%	0%	1%	1%	0%	165	101
Total	*33177*	*64%*	*36%*	*1%*	*7%*	*13%*	*10%*	*3%*	*11989*	*7770*
Medium Mathematical Demand										
Biological sciences	31274	22%	78%	2%	15%	27%	23%	8%	24493	15568
Business and administrative studies	30100	21%	79%	1%	9%	26%	27%	12%	23737	16056
Social studies	24422	25%	75%	2%	12%	24%	23%	10%	18375	11274
Subjects allied to medicine	17560	26%	74%	2%	13%	22%	24%	11%	12944	8173
Computer science	10418	25%	75%	0%	6%	22%	30%	15%	7765	5385
Education	10149	8%	92%	1%	9%	29%	36%	15%	9364	6653
Architecture, building and planning	4832	32%	68%	1%	9%	23%	23%	7%	3270	2249
Medicine & dentistry	4552	72%	28%	10%	11%	2%	1%	1%	1293	159
Agriculture & related subjects	2558	14%	86%	1%	13%	27%	29%	13%	2201	1435
Veterinary science	516	59%	41%	9%	17%	5%	3%	1%	209	42
Total	*136380*	*24%*	*76%*	*2%*	*11%*	*24%*	*25%*	*11%*	*103651*	*66994*
Low Mathematical Demand										
Creative arts and design	32134	8%	92%	2%	11%	27%	32%	18%	29600	18864
Languages	17812	15%	85%	7%	21%	27%	17%	5%	15063	7762
Historical and philosophical studies	12901	19%	81%	5%	19%	26%	17%	4%	10491	5609
Law	10258	16%	84%	4%	16%	28%	23%	10%	8577	5237
Mass Communications and Documentation	8672	4%	96%	1%	9%	28%	37%	19%	8322	5663
Combined/general subject unspecified	1885	17%	83%	2%	13%	24%	26%	14%	1571	946
Total	*83661*	*12%*	*88%*	*3%*	*15%*	*27%*	*26%*	*12%*	*73624*	*44080*
Total Overall	**253557**	**25%**	**75%**	**2%**	**12%**	**24%**	**23%**	**10%**	**189264**	**118844**

Table 1: Mathematics attainment for all honours students from the English GCSE cohort of 2007/2008, who completed A-level between 2008/2009 and 2010/2011, and attended a UK university between 2010/11 and 2012/13.Joint honours students have been weighted by proportion of time spent studying the subject. Rounding errors apply. Some GCSE grades are missing, so grades to not sum to total GCSE, but the total GCSE column includes these students where no post-16 mathematics qualification has been recorded.

Subject Grouping	Total Number of Students	Post-16 Mathematics (AS, A2, IB)	Maximum Pre-University Mathematics Qualification: All GCSE	A*	A	B	C	D-G	Total GCSE	Total with GCSE B or C
Biology	4765	39%	61%	4%	20%	21%	12%	3%	2903	1564
Biology & Biological Sciences	195	85%	15%	1%	4%	3%	6%	1%	28	17
Genetics	278	54%	46%	5%	21%	14%	4%	2%	129	50
Microbiology	367	45%	55%	4%	18%	20%	10%	5%	201	108
Molecular biology, biophysics & biochemistry	2151	53%	47%	4%	17%	15%	8%	2%	1001	480
Others in Biological Sciences	1069	38%	62%	2%	16%	20%	17%	7%	666	397
Psychology	11413	15%	85%	3%	20%	32%	23%	6%	9667	6275
Sport & exercise science	10022	8%	92%	1%	12%	29%	34%	15%	9185	6304
Zoology	998	29%	71%	4%	22%	24%	13%	5%	704	367

Table 2: Mathematics attainment of all honours students in biological sciences. See note to Table 1 for details.

Subject Grouping	Total Number of Students	Post-16 Mathematics (AS, A2, IB)	Maximum Pre-University Mathematics Qualification: All GCSE	A*	A	B	C	D-G	Total GCSE	Total with GCSE B or C
Accounting	3991	54%	46%	1%	10%	19%	13%	4%	1836	1270
Business studies	10970	15%	85%	1%	12%	28%	30%	12%	9283	6379
Finance	1708	57%	43%	1%	12%	19%	12%	3%	727	528
Hospitality, leisure, sport, tourism & transport	4627	5%	95%	0%	7%	25%	39%	21%	4388	2967
Human resource management	449	6%	94%	0%	11%	26%	38%	21%	420	286
Management studies	4960	20%	80%	2%	16%	26%	23%	11%	3951	2412
Marketing	3329	8%	92%	1%	12%	32%	34%	13%	3078	2180

Table 3: Mathematics attainment of all honours students in business and administrative studies. See note to Table 1 for details.

Appendix 1: Methods and data-cleaning

Our analysis makes use of two linked sources of administrative data - the National Pupil Database (NPD) and the Higher Education Statistical Agency database (HESA). We focus on the age cohort of English-domiciled students who completed GCSE in 2007/8 at age 16, who took a Level 3 qualification between 2009/10 and 2011/12, and who entered university between 2010/11 and 2012/13.

The HESA database holds information on a substantial number of factors such as student demographics, higher education institution attended, four subject variables to capture both single and joint-honours students including proportion of time spent studying each subject, mode and level of study, and degree classification.

We made use of the "Unique Pupil Reference Number" which is assigned to children when they enter school, to track individuals (even though completely anonymised) from their Key Stage 4 institution and qualification outcomes through to our KS5 extract (A-level age 18/19). This was then further linked to our HESA extract containing data on undergraduate students and their studies. Our data focuses on linking those who completed GCSE in 2007/8, their A-levels between 2009/10 and 2011/12 and attended university between 2010/11 to 2012/13.

Data cleaning

We followed a data cleaning and processing strategy previously adopted in Adkins and Noyes (2016). We conducted all data cleaning and analyses in *R*. The original text data files provided by the DfE, HESA and UCAS were at various stages of matching and particularly with the DfE and HESA, these data files contain a substantial number of duplicate cases which needed to be addressed to provide as realistically accurate descriptive analyses as possible.

Our process was as follows. Our KS4 and KS5 data files were already matched by the DfE, and so we stripped out the school contextual variables from KS4 and KS5, keeping the student ID, qualification route, and all examination results for A-level, AS-level and GCSE. Next we merged all KS5 years from 2009, 2010, and 2011, and then recoded all grade data into point scores - 0-8 for GCSE, 0-5 for AS-level and 0-6 for A-level. This then provided the basis for addressing the duplicate data issue. We subsetted the dataset 5 times, each time only keeping those cases which were duplicates (meaning that for some students there were up to five records each). We then recoded all the variables to allow the duplicate datasets to be matched to the original and merged back in. To then update our original records, we conditionally recoded them keeping the highest grade point achieved and deleted the duplicate variables from the dataset.

HESA records were provided for all students for all years of study. The HESA dataset was straightforward to prepare. We merged the data frames and then recoded the course aim (the qualification to be obtained), Joint Academic Coding System (JACS subject) codes to create subject area and detailed subject categories, and finally the HE_UKPRN (institutional identification). After we had merged the HESA records with the prior qualification data we then duplicated records where students were enrolled on joint honours programmes and multiplied any numerical statistics generated by the proportion of time spent studying the subject.