Original Article

Data mining course choice sets and behaviours for target marketing of higher education

Received (in revised form): 26th June 2009

Alex D. Singleton
completed his PhD in November 2007 and is now a research fellow in the Department of Geography and Centre for Advanced Spatial Analysis at University College London.

ABSTRACT As the higher education (HE) sector has expanded, so has the variety of courses on offer, with applicants now choosing between greater numbers of potential options. Where applications to HE are administered through centralised admission services, applicants will often make multiple initial course choices, which offers an opportunity to examine systematic groupings of interest within course choice sets, and assess whether certain types of student are more likely to make concentrated or diffuse subject selections. Utilising a national database of an entire cohort's application behaviour, the empirical findings presented in this article indicate that there are clusters of subjects that are applied for in combination, and that certain ethnic minority, socio-economic groups and neighbourhood types are more likely to make more diffuse subject choices. This creates an information base of generalised course choice behaviours that HE institutions could utilise for targeted marketing, recruitment and selection activities, and additionally forms the basis of a decision support framework that could be implemented in a variety of online tools to help guide student courses.

Journal of Targeting, Measurement and Analysis for Marketing (2009) **17,** 157–170. doi:10.1057/jt.2009.13

Keywords: geodemographics, higher education, GIS, data mining, marketing

INTRODUCTION AND THEORETICAL FRAMEWORK

Whether higher education (HE) applications are made directly to an institution as in North America, or through a central clearing house such as the UK Universities and Colleges Admissions Service (UCAS), potential applicants face the complex challenge of narrowing their broad personal interests or aspirations into a single or set of HE course selections. Where applications to HE are made through a central clearing house, this offers an opportunity to examine national patterns of aggregate course choice behaviour. Using such a resource for the 2004 UK application cycle, there are two overarching aims of this article, which cumulatively builds to create an information base that could be used to reduce recruitment and marketing inefficiency[1] by enabling student consumers to make more informed decisions, and creating intelligence that enables institutional suppliers to better target their recruitment efforts. As such, the first aim of this research is to empirically investigate which course choices are typically made in combination across an applicant's choice set. The utility of such a

Correspondence: Alex D. Singleton
University College London, Department of Geography,
Pearson Building, Gower Street, WC1E 6BT, USA
E-mail: a.singleton@ucl.ac.uk

framework of associated preferences could be realised in an online service that guides applicants to discover courses or subject areas supplementary to their core interests, or inversely, be used to provide a national decision support tool for HE institutions engaging in strategic marketing and recruitment activities.

In addition to subject level insight, the second aim of this article is to examine differences in subject choice behaviour among different applicant groups. Specifically, these analyses aim to identify those groupings of students who make more diffuse course choices, that is, they have weak subject preferences across their six UCAS choices. From a marketing perspective, this further intelligence could be used to refine recruitment activities. For example, if a specific market segment is likely to make diffuse choices, then these potential students may be more responsive if offered a broader selection of courses as part of a targeting initiative.

The utility of marketing in HE, and specifically the role of segmentation, has an extended history of academic and practitioner interest[2,3] driven by a variety of policy changes that have created an increasingly competitive operational environment for HE institutions[4] and the perceived emergence of the student as consumer.[5,6] The focus of much early segmentation research was on revisiting previously established private sector techniques, such as geodemographics[7] or market surveys,[8] and discussing the challenges for their application in the public sector. However, more recent research in this area has moved away from operational feasibility into refining techniques, such as the creation of bespoke geodemographics[9] or deriving methods of better identifying the needs or desires of people within particular market segments.[10] Building further on this research base, this article explores a method of deriving new empirical evidence of course choice behaviour to better meet potential learner needs[11] through tailored marketing and centralised services that guide applicant choices. The importance of creating a better intelligence framework for guiding choices can be illustrated in a UK content. During 2007,[12] there were 454 148 home applicants; however, only 364 544 were accepted onto courses HE, leaving 89 604 people making applications that did not successfully gain or meet offers for places. Access to better advice and guidance during the recruitment phase could result in more offers being made to applicants applying to more appropriate courses, thus reducing wastage within the sector.

The emphasis of this article is not to hypothesise or evaluate those behavioural processes by which specific course choice sets are conceptualised by HE applicants, and indeed this area of research is thoroughly reviewed elsewhere utilising a variety of choice models,[13] but instead, examines inductively for the first time how a nationally extensive database can be analysed to derive an empirical evidence base of linked course preferences and behaviours. The implications for such an improved evidence base are far reaching, given that around 60 per cent of applicants to HE choose their institution on the basis that it offers their target course.[14] Additionally, the cumulative benefit of creating tools enabling more effective and efficient recruitment could have an impact on widening participation in HE. Knowledge of differences in HE participation rates between societal groups has a lengthy history,[15–18] and a comprehensive review of those changes in policy that have aimed to support a more egalitarian system can be seen elsewhere.[19,20] HE participation has been shown to stratify across a variety of segments, including ethnicity,[21,22] gender and social class,[23,24] neighbourhood type,[25,26] parental education[27] and schooling.[28] However, despite this extensive research and significant government funding interventions, a recent National Audit Office report highlights that some groups still remain significantly under-represented.[30] As such, it is argued that the new insights provided by this research could enable more effective targeting of courses to under-represented groups, thus with the potential of reducing access inequality.

Before introducing the operational context, data and methods, it is important to differentiate the contribution of this research from previous segmentation analyses that have focused purely on those subjects that different student consumer

groups will typically study. Subject-level analyses are limited in number; however, they have included Geography,[30] Physics and Chemistry,[31] Medicine[32,33] and Mathematics.[34] A variety of influences on choice are identified in the wider literature, ranging from parental guidance,[35,36] student peers,[37] league tables[38,39] through to university prospectuses and recruitment days.[40] Additionally, other externalities, such as the popular media and television programming appear to show influence over the appeal of courses. For example, there was a 57.5 per cent growth in Forensic and Archaeological Science acceptances between 2002 and 2006, which increased total student degree acceptances in this area by 674 places. Sir Howard Newby (Chief Executive of the Higher Education Funding Council for England) noted to a committee of MPs looking at declining applications to science degrees, 'There has been a big drive towards Forensic Chemistry, thanks to Amanda Burton,[41]'… 'I'm not quite sure who is going to employ all those Forensic Scientists'.[42] The research presented in this article differentiates itself from previous segmentation studies, firstly, by analysing choice set data with national coverage, thus encapsulating all students entering HE in the United Kingdom within a single year, and additionally detailing all courses rather than a specific subject or subset. This is the first time that these data have been made available for research purposes, and provide a unique and important empirical validation of cross-subject appeal, along with detailed analysis of choice behaviours among different student segments.

OPERATIONAL CONTEXT AND DATA SOURCES

The 2004 applications data used in this research are sourced from UCAS, the organisation that manage the application process for all full-time courses of HE in the United Kingdom. The UCAS nomenclature defines 'Applicants' as those individuals seeking entry to HE through UCAS. Applicants to HE can make an initial selection of up to six course choices[44] from around 12 000 different options across a plethora of subject areas. These choices (applications) additionally detail which institution and campus the course is taught from. UCAS collect these data alongside various other personal attributes of the applicant, such as age, gender and home address. The main application cycle begins in October of each year and runs through to a deadline at the end of June in the following year. During this period, applicants receive offers or rejections from their six or fewer applications. These decisions are communicated on behalf of the institutions through UCAS, as conditional offers, unconditional offers or rejections. Conditional offers usually specify a particular grade attainment or subject requirement. Unconditional offers are those with no conditions attached, and are usually made where an applicant already has prior qualifications, or where candidates have been judged to have exceptional promise or aptitude. Once institutions have responded to all applications from an applicant, the applicant can reply either accepting or rejecting the offers received, again communicating these decisions through UCAS. For undergraduate admissions, UCAS only collects data on full-time admissions, and only from those institutions within the UCAS scheme. Therefore, the database used in this analysis comprises of individual-level records for each successful applicant in the 2004 application cycle, detailing up to six initial course choices and a series of demographic characteristics.

In the UCAS database, courses are assigned a coded reference number. Thus, the course identifier code for Economics and Geography at University College London is LL17, and English at the University of Cambridge has the code Q300. Courses of HE can be a single subject or multiple combinations, such as Mathematics with Economic or Biology with Earth Sciences. Sometimes courses can also comprise more than two components, such as Geography, Environmental Science and French. In 2004 there were around 12 000 different course combinations available to applicants. For the majority of data analysis, individual course codes are far too diverse, and as such the joint academic coding system (JACS)[44] was developed to provide a

Table 1: JACS subject Line codes for courses within Group A – Medicine and Dentistry

Line code	*Group/line name*	*Line description*
A100	Pre-clinical Medicine	Vocational science of preventing, diagnosing, alleviating or curing disease in homo sapiens. Includes such areas as Anatomy, Physiology, Pharmacy and Nutrition that can be specialisms in their own right.
A200	Pre-clinical Dentistry	Vocational science concerned with the diagnosis and treatment of damage, disease and disorder to the teeth and gums of homo sapiens.
A300	Clinical Medicine	The observation, diagnosis and treatment of an illness or disease through direct interaction with human patients.
A400	Clinical Dentistry	The observation, diagnosis and treatment of disease or damage to teeth and gums through direct interaction with human patients.
A900	Others in Medicine and Dentistry	Miscellaneous grouping for related subjects that do not fit into the other Medicine and Dentistry categories.
A990	Medicine and Dentistry not elsewhere classified	Miscellaneous grouping for related subjects that do not fit into the other Others in Medicine and Dentistry categories.

framework that can be used to aggregate courses into sensible groupings based on their content. JACS is a hierarchical classification that groups courses of study into a fine level of 1281 'Lines', which aggregate up into 19 'Groups'. An example of those JACS Lines within the JACS Group code 'A – Medicine and Dentistry' are given in Table 1. JACS is a framework that allows future courses to be successfully incorporated within its schema, and as such, some Lines are not currently used.

UCAS lookup tables match each course code into the JACS framework. Single subject courses are usually represented by a single Line, for example, a course in English Literature fits into the Line Q320. However, where the course contains multiple subjects, such as Human Geography and Economics, these potentially fit within two Lines, that is, L700 (Human Geography) and L100 (Economics). UCAS have developed a system in which they report these as types of courses as 'combinations', for example, 'Combinations within Social Studies', where 'Social Studies' refers to the group in which both these Lines are found.

DATA MINING SUBJECT ASSOCIATIONS

Data mining is 'the science of extracting useful information from large data sets or databases',[45] and is used in the following analysis to examine those subject areas that nationally on aggregate are frequently applied for in combination across applicant choice sets. A structured query language (SQL) based algorithm was created that ran through the UCAS database and inspected the course code of each application (maximum of six per applicant) and matched these to an appropriate JACS Line. However, as previously illustrated, these courses could contain multiple JACS Lines, and therefore, the following weighting schemes were additionally applied to account for these balances. For courses that were a single subject (for example, Chemistry), the associated JACS Line was given a score of two, thus indicating a strong preference. For courses that contained two JACS Lines (for example, Chemistry and Business Studies), these were split into two, and each separate JACS Line given a score of one. For courses that contained more than two subjects (for example, French with Human Geography and Economics), the major component was converted into a single-subject JACS Line (for example, French) and was given a score of one, and the remainder was summarised as a 'combination' JACS Line, which in this example would be 'Combinations within Social Studies', and also given a score of one. Therefore, each application can have a maximum score of two, and because each applicant can make up to six applications, each applicant can contribute a maximum score of 12. These scores were computed and then stored in a database for each applicant. Using further

Table 2: An extract from the JACS Line/subject level analysis

Subject description/JACS Line code	*A1*	*A2*	*B0*	*B1*	*B2*	*B3*	*B4*	*B5*	*B6*	*B7*
A1 – Pre-clinical Medicine	**76.9**	0.3	0.0	2.2	2.6	0.0	0.1	0.4	0.1	0.4
A2 – Pre-clinical Dentistry	3.7	**72.0**	0.0	1.0	6.3	0.0	0.1	2.1	0.1	0.2
B0 – Subjects allied to Medicine: any area	0.0	0.0	**8.3**	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B1 – Anatomy, Physiology and Pathology	6.7	0.4	0.0	**49.3**	1.7	0.1	0.4	0.5	0.3	0.7
B2 – Pharmacology, Toxicology and Pharmacy	11.5	3.3	0.0	2.5	**49.4**	0.4	0.3	1.8	0.1	0.4
B3 – Complementary Medicine	0.8	0.2	0.0	3.8	3.5	**37.2**	0.7	0.2	0.4	0.7
B4 – Nutrition	1.2	0.3	0.0	2.6	1.4	0.2	**46.6**	0.4	0.6	1.8
B5 – Ophthalmics	7.7	4.9	0.0	2.3	7.4	0.0	0.4	**54.1**	0.7	0.5
B6 – Aural and Oral Sciences	1.7	0.2	0.0	2.1	0.3	0.1	0.5	0.7	**57.8**	1.7
B7 – Nursing	1.2	0.1	0.0	0.9	0.3	0.1	0.4	0.1	0.4	**78.2**

SQL, measures of aggregate associations between JACS Lines were calculated. The following calculation was run for each JACS Line in turn. All those applicants with at least one application score within a JACS Line were selected; the scores across the JACS Lines that made up the applicants remaining course choices (as extracted from applications) were each summed across the total population of the applicant database. The output from this analysis is a very large table, which contains 189 rows and 189 columns. These rows and columns relate to every JACS Line present in the 2004 UCAS database. A very small extract of this derived information is shown in Table 2 for the JACS Lines range A1–B7. The scores in the table are percentages of 'within' (bold) and 'outside' JACS Line applications, which when summed across all the columns in the full table equal 100 per cent. Thus, for applicants with at least one choice in 'A1 – Pre-clinical Medicine', 76.9 per cent of applicants have all their other choices within the same JACS Line. This is unsurprising given that this is a course that prepares students for a specific vocation, and a similar pattern can be observed for 'B7 – Nursing'. The other percentages across the columns represent the applications outside of the subject line, where at least one course choice was A1.

The full table that was used to create the extract shown in Table 2 can be ranked into the top 20 subjects, with most homogeneous within JACS Line applications (See Table 3). For each of these JACS Lines, the percentages of other applications outside of the Line are presented in the third column. Thus, 67.3 per cent of applicants making applications with at least one choice to 'D1 – Pre-clinical Veterinary Medicine' applied only within this Line. Of those making other applications outside of this Line, 6.8 per cent were to 'D3 – Animal Science', 5.1 per cent were to 'C3 – Zoology', 2.9 per cent were to 'C1 – Biology', 2.8 per cent were to 'Y6 Combinations of medical/biological/agricultural sciences' and 1.7 per cent were to 'B9 – Others in Subjects allied to Medicine'.

These percentages are a measure of association between aggregate areas of subject interest, and can be better visualised by using network diagrams, such as those presented in Figure 1 for the JACS Line 'M1 – Law by Area' (67.6 per cent within Line), in which the strength of association is represented by the thickness of the connecting lines. The numbers on top of the lines are the percentage scores. The stronger the association between a JACS Line node (centre of the diagram – M1) and those adjoined nodes, the greater the probability that an applicant may be interested in other courses from within these alternate subject areas.

However, M1 is an example of a subject area in which applications come from applicants who typically make constrained choices, that is, they will typically have a strong interest in M1. In this instance, if a system were created to suggest other courses outside of M1 that might interest potential applicants, the probability of these suggestions being taken up would be low. However, this type of suggestion system may be more helpful for those other JACS Lines that typically attract more diffuse applicant behaviour. For example, Figure 2 shows a network

Table 3: Top 20 subjects where within JACS Line applications are most homogenous

Rank	*Subject description*	*Five highest outside Line applications by JACS Line codes*[a]	*Within Line %*
1	B7 – Nursing	B9 (2.2%), L5 (2%), C8 (1.3%), A1 (1.2%), B8 (1%)	78.2
2	A1 – Pre-clinical Medicine	B9 (5%), B2 (2.6%), C7 (2.3%), B1 (2.2%), C1 (1.8%)	76.9
3	A2 – Pre-clinical Dentistry	B2 (6.3%), A1 (3.7%), B9 (2.3%), B5 (2.1%), C7 (1.6%)	72
4	M1 – Law by Area	Y14 (2.5%), Y15 (2.5%), M2 (2.4%), M9 (2.4%), N1 (1.8%)	67.6
5	D1 – Pre-clinical Veterinary Medicine	D3 (6.8%), C3 (5.1%), C1 (2.9%), Y6 (2.8%), B9 (1.7%)	67.3
6	K1 – Architecture	K2 (6.3%), H2 (5%), Z (4.1%), W2 (3.8%), KK (1.3%)	64
7	V1 – History by Period	Y16 (9.2%), Y5 (5%), VV (3%), Y1 (1.9%), X1 (1.6%)	60.8
8	Q8 – Classical studies	V1 (10.6%), Y5 (6.6%), QQ (4.5%), VV (3.8%), Q3 (3.6%)	59.7
9	B8 – Medical Technology	B9 (5.7%), B1 (5.1%), B7 (2.7%), A1 (2.3%), B2 (2.1%)	59.5
10	B6 - Aural and Oral Sciences	B9 (4.7%), C8 (3.5%), B1 (2.1%), Q1 (2%), X1 (2%)	57.8
11	C8 –- Psychology	Y12 (8%), Y11 (3.5%), L3 (2.4%), M9 (1.9%), C6 (1.8%)	57.1
12	F3 – Physics	Y1 (5.6%), F5 (4.4%), G1 (3.2%), FF (2.4%), F1 (2.4%)	54.8
13	L5 – Social Work	X3 (4.9%), L3 (4.6%), C8 (4.5%), Y12 (2.9%), X1 (2.7%)	54.5
14	B5 – Ophthalmics	A1 (7.7%), B2 (7.4%), A2 (4.9%), B9 (3.6%), B1 (2.3%)	54.1
15	W4 – Drama	Y5 (7%), WW (5.5%), Z (4%), Y1 (3.8%), Q3 (2.8%)	54
16	H3 – Mechanical Engineering	H1 (7.6%), H4 (7%), H2 (3.3%), H6 (3%), H7 (2.9%)	53.9
17	G1 – Mathematics	Y9 (6.2%), GG (5.1%), Y1 (2.9%), Y8 (2.5%), N4 (2.2%)	53.1
18	H2 – Civil Engineering	K2 (5.7%), H3 (5.2%), K1 (4.7%), H1 (4.1%), Y13 (2.1%)	53
19	H4 – Aerospace Engineering	H3 (12.6%), H1 (4.6%), H6 (3.5%), HH (3%), H2 (2.7%)	52.4
20	W3 – Music	J9 (7.9%), W4 (6.7%), Z (4.7%), Y11 (3.9%), Y1 (2.8%)	51.6

[a]For a lookup between the JACS Line codes and their descriptions, see Appendix Table 1.

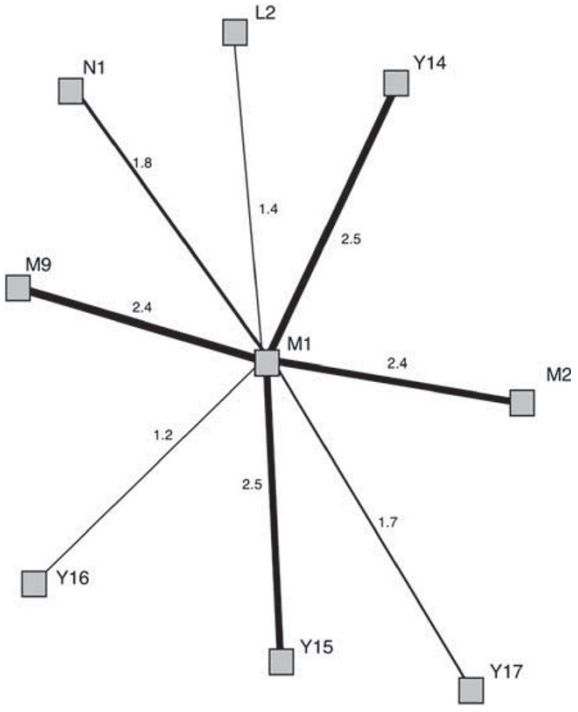

Figure 1: Network representation of linked JACS Line interests for 'M1 – Law by Area'.[46]

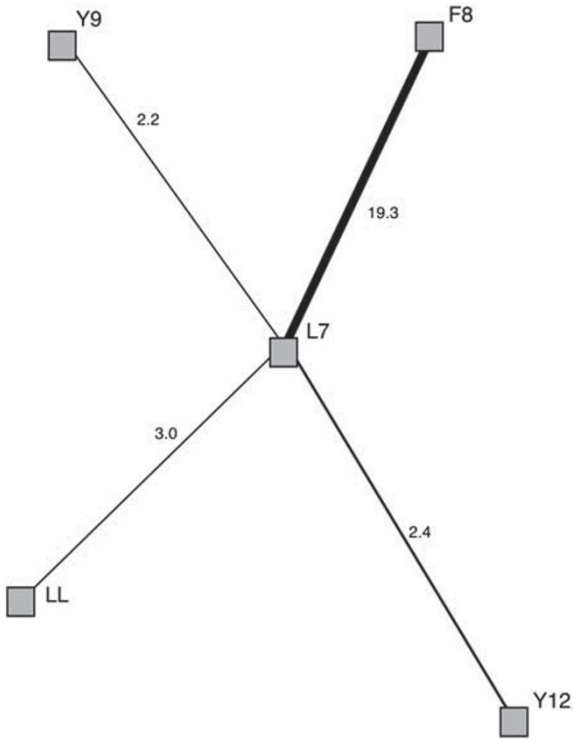

Figure 2: Network representation of linked JACS Line interests for 'L7 – Human and Social Geography'.

representation for the JACS Line 'L7 – Human and Social Geography' (50 per cent within Line). Half of the applicants making at least one selection inside L7 also make choices outside of the Line. Almost 20 per cent of these choices go to courses that appear in JACS Line 'F8 – Physical and Terrestrial Geography and Environmental Science'. As such an applicant interested in L7 might also be interested in courses in the Line F8.

DISAGGREGATING APPLICANT SUBJECT CHOICES

The analysis presented in the previous section showed how aggregate application behaviours are differentiated between subjects. However, not all applicants will exhibit the same course choice behaviour when applying to HE, and the degree to which choice sets are diffuse or constrained will likely be stratified by a variety of socio-economic factors. As such, the following analysis examines differences in aggregate application behaviour among ethnic, socio-economic and neighbourhood groups. In the previous section, applicant course choice sets were scored to indicate their JACS Line preferences, which when aggregated at subject level indicated those JACS Lines that are typically applied for in combination. In the following analysis, homogeneity of JACS Line preferences are analysed between societal groups. A measure of dispersal was calculated for each applicant by summing the total number of different JACS Line choices within an applicant's choice set. Hypothetically these scores could range from 1 to 12, as a maximum of two JACS Lines could be assigned to each potential course choice; however, in reality these scores ranged from 1 to 9, as the most diverse choices were not present in the data. It is important to differentiate between the frequency of weighted JACS Line scores and the frequency of applications made by an applicant (See Figure 3). The distribution of the JACS Line scores was aggregated by a geodemographic indicator, the National Statistics Socio-Economic Classification (NS-SEC) and by an ethnic group.

Geodemographic classifications link to applicants using their home address, and assign labels and descriptions that represent the aggregate characteristics of the neighbourhood in which they live. Although used prevalently in the commercial sector,[47,48] these types of classification have only recently experienced a renaissance for public sector service delivery,[49] with demonstrated applications in education,[9,25] health,[50] local government[51] and policing.[52] The geodemographic classification used for this analysis is the free National Statistics Output Area Classification (OAC).[53,54] This classification divides UK neighbourhoods into three hierarchically nested levels of 7 Super Groups, 21Groups and 52 Sub-Groups. For this analysis, the Group level was used, which is presented in Table 4.

Standardised index scores for the frequency of JACS Line choices by OAC Groups are presented in Table 5. In this table, a score of 100 represents the national average, a score of 200 is twice the national average and a score of 50 would be half the national average. Thus, those applicants living in the neighbourhood group 'Asian Communities (7a)' would be 1.3 times more likely than the national average to apply for six or more JACS Lines. Scores 115 or over are highlighted in bold to aid interpretation.

Several findings can be drawn from these data. First, there appears to be an economic dimension

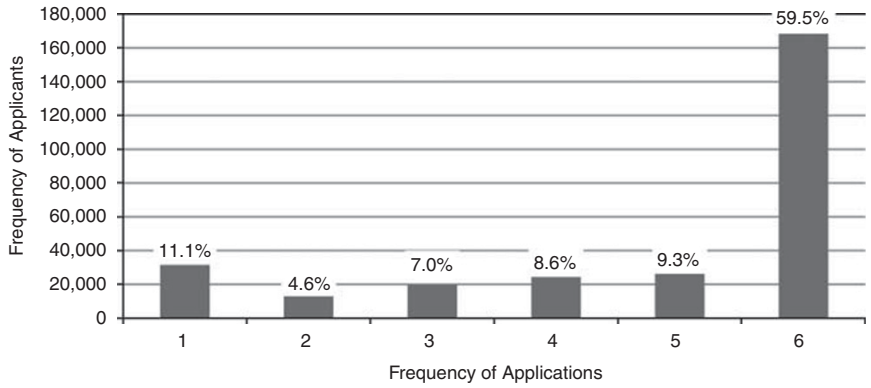

Figure 3: Frequency of applicants and applications.

Table 4: Output area classification groups and descriptive labels

Group name	*Description*
Terraced blue collar (1a)	These neighbourhoods typically consist of high density publicly rented and/or terraced homes in areas with above average unemployment. Employed residents typically work in blue-collar manufacturing occupations. Those not in work often care for their young children, many of whom are below school age. There are high numbers of single parent households.
Younger blue collar (1b)	These neighbourhoods typically comprise households headed by young adults living in high-density publicly rented homes. Unemployment is above the national average and neighbourhoods might be considered to be 'deprived'. Residents in employment typically work in blue-collar manufacturing occupations. Those not in work spend much of their time looking after young children. Many households are headed by a single parent.
Older blue collar (1c)	The established households in these neighbourhoods typically live in publicly rented homes, in areas with average rates of unemployment. Many individuals work in blue-collar occupations, such as construction, or in agriculture.
Transient communities (2a)	These neighbourhoods are predominantly located in inner cities and are characterised by low-quality, high-density rented flats. They have multi-ethnic populations, many of whom are first generation immigrants. Other residents include students.
Settled in the City (2b)	These culturally diverse and high-density neighbourhoods are found in urban areas of England and Scotland. Many residents live in privately rented flats and are economically active, although there are also significant numbers of retirees, many of whom live alone. Students participating in higher education are also found living in these areas.
Village life (3a)	The residents of these affluent rural village areas often live in larger detached houses. Rates of unemployment are low, and they often work from home (sometimes in agriculture or related industries). There is low use of public transport in these areas, in part because of the high prevalence of households with more than one car.
Agricultural (3b)	The residents of these agricultural areas often live in larger detached farmhouses. Rates of unemployment are low, and many work in agriculture or related industries. There is low use of public transport in these areas, in part because of the high prevalence of households with more than one car.
Accessible countryside (3c)	These areas are mainly found in England and offer rural lifestyles in close proximity to towns and cities. Many properties are detached, and some are rented. Residents tend to be well educated.
Prospering younger families (4a)	These suburban neighbourhoods are often affluent, with families with young children living in large detached homes. Many residents have higher education qualifications and work in professional occupations.
Prospering older families (4b)	The residents of these areas are typically well educated and, having had successful careers, are now nearing retirement. They live in large detached houses, and children have often moved away from the family home.
Prospering semis (4c)	These well-educated and prosperous residents of suburban areas live in semi detached or detached houses, but have few dependant children. Many are approaching retirement age.
Thriving suburbs (4d)	These suburban neighbourhoods include some multi-ethnic areas. The majority of residents are highly educated and live in large detached homes. Many households have more than one car.
Senior communities (5a)	These neighbourhoods are typically found in Scotland and house a high proportion of elderly single people. The majority of residents live in flats rented from local authorities. Many of the residents that are not of retirement age are unemployed.
Older workers (5b)	Residents of these areas are typically single parents or pensioners living alone, with many living in flats or terraced houses that are rented from the local authority. There is a high incidence of unemployment in these neighbourhoods.
Public housing (5c)	Many residents of these neighbourhoods are single parents living in crowded local authority rented flats or terraced houses of mixed quality. Unemployment in these areas is high, and those that are employed work in routine or semi-routine occupations, such as the hospitality or catering industries.
Settled households (6a)	These neighbourhoods predominantly consist of aging terraced housing, as is typically found in and around Sheffield and Manchester. There is low unemployment in these areas.
Least Divergent (6b)	These areas are normally found in provincial towns. Many residents rent their accommodation from private landlords, and live in a range of different housing types. Some residents are elderly.
Young families in terraced homes (6c)	Young families predominate in high-density terraced housing. Many residents rent within the private sector, and there are large numbers of lone parents.
Aspiring households (6d)	The residents of these multi-ethnic neighbourhoods typically live in rented accommodation. They often have higher education qualifications. Many are quite affluent and rates of multiple car ownership are high.
Asian communities (7a)	The majority of residents are non-white, and many are first generation Asian immigrants that now have young families. Housing is high density, typically terraced and often split into flats. Many residents are hard-pressed, and unemployment is high. The cheap accommodation that is available in these areas also attracts students attending higher education.
Afro-Caribbean communities (7b)	The majority of residents of these areas are non-white, and many are first generation Afro-Caribbean immigrants that now have young families. Housing is high density, typically terraced, and often split into flats. Many residents are hard-pressed, and unemployment is high. The cheap accommodation that is available in these areas also attracts students attending higher education.

Table 5: Standardised index scores for course choice behaviour by OAC groups

OAC group	*Frequency of JACS Lines chosen/index scores*					
	1	*2*	*3*	*4*	*5*	*6 or more*
Terraced blue collar (1a)	101	91	96	105	114	**126**
Younger blue collar (1b)	105	93	99	99	103	99
Older blue collar (1c)	100	101	98	98	96	112
Transient communities (2a)	**124**	89	86	81	88	97
Settled in the city (2b)	**115**	100	91	85	82	84
Village life (3a)	101	102	99	98	98	95
Agricultural (3b)	93	95	100	107	**122**	**138**
Accessible countryside (3c)	103	108	99	92	87	84
Prospering younger families (4a)	93	100	104	106	108	107
Prospering older families (4b)	98	103	105	101	94	89
Prospering semis (4c)	95	104	104	101	103	90
Thriving suburbs (4d)	100	106	102	96	91	85
Senior Communities (5a)	111	96	90	95	87	106
Older workers (5b)	105	94	97	100	97	102
Public housing (5c)	103	92	94	102	**116**	**116**
Settled households (6a)	102	99	102	100	92	84
Least divergent (6b)	105	103	98	94	91	82
Young families in terraced homes (6c)	114	98	89	93	87	84
Aspiring households (6d)	103	103	100	98	89	84
Asian communities (7a)	89	95	102	112	**128**	**130**
Afro-Caribbean Communities (7b)	93	93	99	113	117	**144**

Table 6: standardised index scores for course choice behaviour by NS-SEC

NS-SEC	*Frequency of JACS lines chosen/index scores*					
	1	*2*	*3*	*4*	*5*	*6 or more*
Higher managerial and professional occupations	108	105	97	89	86	78
Intermediate occupations	99	100	101	102	103	98
Lower managerial and professional occupations	97	103	103	101	99	99
Lower supervisory and technical occupations	103	97	99	98	100	100
Semi-routine occupations	87	98	104	115	**122**	**125**
Small employers and own account workers	90	100	105	110	110	**118**

influencing course choice behaviour, with a number of the more deprived neighbourhood Groups (for example, 'Terraced Blue Collar (1a)', 'Agricultural (3b)', 'Public Housing (5c)') containing applicants who exhibit a higher propensity to make more diffuse applications. Second, two neighbourhood Groups ('Asian Communities (7a)', 'Afro-Caribbean Communities (7b)') who represent areas that contain higher numbers of ethnic minorities also make more diffuse choices. Finally, two neighbourhoods that are typically found within urban areas ('Transient Communities (2a)', 'Settled in the City (2b)') show a higher propensity for applicants making single JACS Line choices. A characteristic common across both these neighbourhood Groups is a high propensity for student accommodation, and as such, these scores may be attributable to reapplications through the UCAS scheme, in which previously admitted students switch courses within the same institution for re-entry in a new academic year. Indeed, 14.8 per cent of applicants from within 'Transient Communities (2a)' and 'Settled in the City (2b)' neighbourhoods only make a single application compared to 11.1 per cent in the total population of applicants.

Socio-economic status is assigned to HE applicants using the National Statistics Socio-Economic Classification.[55] This classification aggregates the applicant or parental occupation (for those under the age of 21) into one of six groups that characterise their socio-economic status. Table 6 shows the standardised index scores for the frequency of JACS Line choices by NS-SEC groups.

These data reiterate some of the findings from the geodemographic analysis. Those applicants in the highest NS-SEC group ('higher managerial and professional occupations') have an increased propensity to make more constrained course choices, and those in the lowest two groups ('semi-routine occupations', 'small employers and own account workers') have increased propensity to make more diffuse choices. There are a number of possible interpretations of these patterns. It could be that students of lower socio-economic status are less likely to have access to HE advice and guidance networks, and as such, may have less assistance when picking a single subject of interest. Alternatively, applicants from lower socio-economic status predominantly possess lower prior attainment scores, and as those institutions with lower entry requirements typically offer more diverse course portfolios, this could also increase the probability that applicants will make more diffuse JACS Line choices.

Table 7 shows standardised index scores for the frequency of JACS Line choices by ethnic groups. The geodemographic analysis indicated that applicants from predominantly Asian neighbourhoods had more diffuse course choice preferences, and although the analysis of ethnic groups broadly confirms these findings, there are differences between Indian, Pakistani and Bangladeshi applicants. Specifically, applicants with an Indian ethnicity tend to have less diffuse application behaviour. The ethnic group analysis also shows that the Black – African group make more diffuse choices than Black – Caribbean.

It is unlikely that through an online tool offering course or subject suggestions one would gather personal information about an applicant (for example, ethnicity or socio-economic status). However, geodemographics assigned by using a postcode appear to offer a good surrogate for this personal information and could prove useful in tailoring potential course suggestions in an online tool.

CONCLUSIONS AND WIDER IMPLICATIONS

This article has shown a method of linking HE subjects of study by their appeal across successful applicant course choice sets. This has created a national information base that can be used to inform HE stakeholders which courses appeal in combination, and, additionally, how course selection behaviours will likely differentiate between subjects, and also by demographic and socio-economic characteristics of the potential applicants.

It is proposed that the utility of this information could be demonstrated in a series of future web services oriented to the student, the centralised applications service/clearing house and the HE institution. For the student, an online tool that links subjects by their attractiveness

Table 7: Standardised index scores for course choice behaviour by ethnic groups

Ethnic group	*Frequency of JACS Lines chosen/index scores*					
	1	*2*	*3*	*4*	*5*	*6 or more*
Asian – Bangladeshi	73	105	109	**130**	**115**	**123**
Asian – Indian	90	101	104	109	113	101
Asian - Other Asian background (ex-Chinese)	100	102	102	97	104	83
Asian – Pakistani	77	100	111	**116**	**135**	**115**
Black – African	88	103	106	106	104	**121**
Black – Caribbean	96	97	104	113	91	102
Black – other black background	85	106	109	106	113	106
Chinese	93	103	106	107	91	110
Mixed – other mixed background	105	107	100	81	91	107
Mixed – White and Asian	109	104	98	91	85	77
Mixed – White and Black African	100	107	109	80	90	111
Mixed – White and Black Caribbean	98	96	115	96	97	86
Other ethnic background	102	97	99	91	112	110
White	108	106	102	91	81	72

could increase the visibility of supplementary subject areas. This segmentation based on subject interest alone could be enhanced through consideration of the geodemographic characteristics of the student, providing potential insight into the probable success rate of applicants taking up suggestions based on their selection of core subject area of interest. The empirical evidence created by this research could also enable HE institutions to make efficiency savings in a number of ways. For example, marketing or advertising efforts of specific course groupings could be combined for those areas that are known to have linked appeal. Additionally, based upon the patterns of applicant choice behaviours differentiated by geodemographic group, recruitment efforts could be targeted to those students who on aggregate are most likely to be responsive to alternate course suggestions. Such a resource may also be useful to an institution if a specific course became oversubscribed. Potential candidates could perhaps be recommended an intelligent alternative, decreasing the probability that they will make applications to other institutions. Finally, in a centralised admissions service, it is common practice to operate a clearing phase at the end of the application cycle in which attempts are made to match unfilled courses to applicants without confirmed offers. The information created by this research could help target these excess places more intelligently. For example, if a student had applied but failed to receive offers in one subject area by the end of the applications phase, they (additionally accounting for their geodemographic group) could be targeted for alternate courses sharing linked appeal.

The information base created and enabled by this research is the first step towards developing the services outlined above. However, although it has been argued elsewhere that course offerings govern institution selection for the majority of applicants,[14] the perceived accessibility of an institution, and therefore, those courses that are offered is likely to have some degree of interaction with applicant course choice behaviour. For example, applicants from very deprived backgrounds are unlikely to look at course offerings at those institutions far from their home location. Thus, the potential pool of perceived available courses may be modified as a function of distance from an applicant's home location. There is therefore a requirement for future research extending from this study to examine interactions between course and institution choice, and specifically investigate how a student's home location may limit or enhance the perceived availability of choice.

ACKNOWLEDGEMENTS

This research was conducted under an ESRC First Grant (RES-061-25-0303).

REFERENCES AND NOTES

1 Drummond, G. (2004) Consumer confusion: Reduction strategies in higher education. *International Journal of Educational Management* 18(5): 317–323.
2 Litten, L.H. (1979) Market structure and institutional position in geographic market segments. *Research in Higher Education* 11(1): 59–83.
3 Litten, L.H. (1980) Marketing higher education: Benefits and risks for the American academic system. *The Journal of Higher Education* 51(1): 40–59.
4 Bok, D. (2003) *Universities in the Marketplace*. Princeton: Princeton University Press.
5 Hill, F. (1995) Managing service quality in higher education: The role of the student as primary consumer. *Quality Assurance in Higher Education* 3(3): 10–21.
6 Maringe, F. and Gibbs, P. (2008) *Marketing Higher Education*. Berkshire: Open University Press.
7 Tonks, D. and Farr, M. (1995) Market segments for higher education: Using geodemographics. *Market Inteligence & Planning* 13(4): 24–33.
8 Borden, V. (1995) Segmenting student markets with a student satisfaction and priorities survey. *Research in Higher Education* 36(1): 73–88.
9 Singleton, A. and Longley, P. (2009) Creating open source geodemographics – Refining a national classification of census output areas for applications in higher education. *Papers in Regional Science*, doi:10.1111/j.1435-5957.2008.00197.x.
10 Rindfleish, J.M. (2003) Segment profiling: Reducing strategic risk in higher education management. *Journal of Higher Education Policy and Management* 25(2): 147–159.
11 Rogers, G., Finley, D. and Kline, T. (2001) Understanding individual differences in university undergraduates: A needs segmentation approach. *Innovative Higher Education* 25(3): 183–196.
12 Statistics on HE applicants and acceptances can be found using the UCAS Statistical Enquiry Tool: http://search1.ucas.co.uk/fandf00/index.html.
13 Foskett, N.H. and Hemsley-Brown, J.V. (2001) *Choosing Futures – Young People's Decision Making in Education*. London, RoutledgeFalmer: Training and Careers Markets.

14 Purcell, K., Elias, P., Ellison, R., Atfield, G., Adam, D. and Livanos, I. (2008) *Applying for Higher Education – the Diversity of Career Choices,Plans and Expectations: Findings from the First Futuretrack Survey of the 'Class of 2006' Applicants for Higher Education*. Manchester: Higher Education Careers Services Unit.
15 Dearing, R. (1997) *National Committee of Inquiry into Higher Education*. Norwich: HMSO.
16 Halsey, A.H., Heath, A.F. and Ridge, J.M. (1980) *Origins and Destinations: Family, Class and Education in Modern Britain*. Oxford: Oxford University Press.
17 Paterson, L. (1997) Trends in higher education participation in Scotland. *Higher Education Quarterly* 51: 29–48.
18 Rudd, E. (1987) The educational qualifications and social class of the parents of undergraduates entering British Universities in 1984. *Journal of the Royal Statistical Society. Series A (General)* 150(4): 346–372.
19 Archer, L., Hutchings, M. and Ross, A. (2003) *Higher Education And Social Class*. London: Routledge.
20 Gorard, S., Smith, E., May, H., Thomas, L., Adnett, N. and Slack, K. (2006) *Review of Widening Participation Research: Addressing the Barriers to Participation in Higher Education*. Bristol: HEFCE.
21 Archer, L. and Hutchings, M. (2000) 'Bettering Yourself'? Discourses of risk, cost and benefit in ethnically diverse, young working-class non-participants' constructions of higher education. *British Journal of Sociology of Education* 21(4): 555–574.
22 Cheng, Y. and Heath, A. (1993) Ethnic origins and class destinations. *Oxford Review of Education* 19(2): 151–165.
23 Ball, S.J., Davies, J., David, M. and Reay, D. (2002) 'Classification' and 'Judgement': Social class and the 'Cognitive Structures' of choice of higher education. *British Journal of Sociology of Education* 23(1): 55–72.
24 Reay, D., David, M. and Ball, S.J. (2005) *Degrees of Choice: Social Class, Race, Gender and Higher Education*. Stoke-on-Trent: Trentham Books.
25 Batey, P., Brown, P.J.B. and Corver, M. (1999) Participation in higher education: A geodemographic perspective on the potential for further expansion in student numbers. *Journal of Geographical Systems* 1(3): 277–303.
26 Tonks, D. and Farr, M. (2003) Widening access and participation in UK higher education. *The International Journal of Educational Management* 17: 26–36.
27 Burnhill, P., Garner, C. and McPherson, A. (1990) Parental education, social class and entry to higher education 1976–86. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 153(2): 233–248.
28 Vernon, G., Damon, B. and Davies, R. (2002) Young people's entry into higher education: Quantifying influential factors. *Oxford Review of Education* 28(1): 5–20.
29 Comptroller and Auditor General. (2008) *Widening Participation in Higher Education*. London: National Audit Office.
30 Longley, P., Goodchild, M., Maguire, D. and Rhind, D. (2005) *Geographic Information Systems and Science*. Chichester: John Wiley & Sons.
31 Elias, P., Jones, P. and McWhinnie, S. (2006) *Representation of Ethnic Groups in Chemistry and Physics: a Report Prepared for the Royal Society of Chemistry and the Institute of Physics*. London: Royal Society of Chemistry and the Institute of Physics.
32 Garlick, P.B. and Brown, G. (2008) Widening participation in medicine. *British Medical Journal* 336: 1111–1113.
33 Seyan, K., Greenhalgh, T. and Dorling, D. (2004) The standardised admission ratio for measuring widening participation in medical schools: Analysis of UK medical school admissions by ethnicity, socioeconomic status, and sex. *British Medical Journal* 328: 1545–1546.
34 Kitchen, A. (1999) The changing profile of entrants to mathematics at a level and to mathematical subjects in higher education. *British Educational Research Journal* 25(1): 57–74.
35 Brooks, R. (2003) Young people's higher education choices: The role of family and friends. *British Journal of Sociology of Education* 24(3): 283–297.
36 Brooks, R. (2004) 'My mum would be as pleased as punch if I actually went, but my dad seems a bit more particular about it': Paternal involvement in young people's higher education choices. *British Educational Research Journal* 30(4): 495–514.
37 Thomas, W. and Webber, D.J. (2001) Because my friends are: The impact of peer groups on the intention to stay on at sixteen. *Research in Post-Compulsory Education* 6(3): 339–354.
38 Bowden, R. (2000) Fantasy higher education: University and college league tables. *Quality in Higher Education* 6: 41–60.
39 Eccles, C. (2002) The use of university rankings in the United Kingdom. *Higher Education in Europe* 27(4): 423–432.
40 Maringe, F. (2006) University and course choice: Implications for positioning, recruitment and marketing. *International Journal of Educational Management* 20(6): 466–479.
41 Who is the lead character in the popular BBC forensic drama "Silent Witness".
42 Newby, H. (2005) More warnings over decline in science students. *The Guardian*, London.
43 An applicant can make further course choices at a later stage in the application cycle through the clearing process or UCAS extra scheme. These data have been excluded from the analysis as they operate at the end of the application cycle where candidates have either failed to secure an offer, or who applied late. In both cases the processes that lead to course choice in UCAS clearing or extra are likely to be different than those occurring at the start of the application cycle.
44 For further information on the use of the JACS course coding system, see the UCAS website: http://www.ucas.ac.uk/he_staff/datamanagement/jacs/course_coding_system/.
45 Hand, D., Mannila, H. and Smyth, P. (2001) *Principles of Data Mining*. Cambridge, MA: MIT Press.
46 The network diagram is drawn using the software NetDraw, which is available from: http://www.analytictech.com/.
47 Birkin, M., Clarke, G. and Clarke, M. (2002) *Retail Geography and Intelligent Network Planning*. Chichester: Wiley.
48 Harris, R., Sleight, P. and Webber, R. (2005) *Geodemographics, GIS and Neighbourhood Targeting*. Chichester: John Wiley & Sons.
49 Longley, P. (2005) Geographical information systems: a renaissance of geodemographics for public service delivery. *Progress in Human Geography* 29(1): 57–63.
50 Farr, M. and Evans, A. (2005) Identifying Unknown Diabetics using Geodemographics and Social Marketing. *Interactive Marketing* 7: 47–58.
51 Batey, P. and Brown, P.J.B. (2007) The spatial targeting of urban policy initiatives: A geodemographic assessment tool. *Environment and Planning A* 39(11): 2774–2793.
52 Ashby, D.I. and Longley, P.A. (2005) Geocomputation, geodemographics and resource allocation for local policing. *Transactions in GIS* 9(1): 53.

53 Vickers, D. and Rees, P. (2007) Creating the UK national statistics 2001 output area classification. *Journal of the Royal Statistical Society. Series A. Statistics in society* 170(2): 379.

54 For further information about this classification see the website: www.areaclassification.org.uk.

55 Rose, D., Pevalin, D. and O'Reilly, K. (2001) *The National Statistics Socio-economic Classification*. London: Office of National Statistics.

APPENDIX

See Table A1.

Table A1: JACS subject Line lookup

JACS Line	*Subject description*
A1	Pre-clinical Medicine
A2	Pre-clinical Dentistry
B0	Subjects allied to Medicine: any area
B1	Anatomy,Physiology and Pathology
B2	Pharmacology,Toxicology and Pharmacy
B3	Complementary Medicine
B4	Nutrition
B5	Ophthalmics
B6	Aural and Oral Sciences
B7	Nursing
B8	Medical Technology
B9	Others in Subjects allied to Medicine
BB	Combinations within Subjects allied to Medicine
C1	Biology
C2	Botany
C3	Zoology
C4	Genetics
C5	Microbiology
C6	Sports Science
C7	Molecular Biology, Biophysics and Biochem
C8	Psychology
C9	Others in Biological Sciences
CC	Combinations within Biological Sciences
D0	Veterinary Science, Agriculture and related: any area of study
D1	Pre-clinical Veterinary Medicine
D2	Clinical Veterinary Medicine and Dentistry
D3	Animal Science
D4	Agriculture
D5	Forestry
D6	Food and Beverage studies
D7	Agricultural Sciences
D9	Others in Veterinary Science, Agriculture and related subjects
DD	Combinations within Veterinary Science, Agriculture and related subjects
F0	Physical Sciences: any area of study
F1	Chemistry
F2	Materials Science
F3	Physics
F4	Forensic and Archaeological Science
F5	Astronomy
F6	Geology
F7	Ocean Sciences
F8	Physical and Terrestrial Geography and Environmental Science
F9	Others in Physical Sciences
FF	Combinations within Physical Sciences
G0	Mathematical and Computer Science: any area
G1	Mathematics
G2	Operational Research
G3	Statistics
G4	Computer Science
G5	Information Systems
G6	Software Engineering
G7	Artificial Intelligence
G9	Others in Mathematical and Computer Science
GG	Combinations within Mathematical and Computer Science
H0	Engineering: any area of study
H1	General Engineering
H2	Civil Engineering
H3	Mechanical Engineering
H4	Aerospace Engineering
H5	Naval Architecture
H6	Electronic and Electrical Engineering
H7	Production and Manufacturing Engineering
H8	Chemical, Process and Energy Engineering
H9	Others in Engineering
HH	Combinations within Engineering
J1	Minerals Technology
J2	Metallurgy
J3	Ceramics and Glass
J4	Polymers and Textiles
J5	Materials Technology not otherwise spec
J6	Maritime Technology
J7	Industrial Biotechnology
J9	Others in Technology
JJ	Combinations within Technology
K0	Architecture, Building and& Plan: any area
K1	Architecture
K2	Building
K3	Landscape Design
K4	Planning (Urban, Rural and Regional)
K9	Others in Architecture, Building and Plan
KK	Combinations within Architecture, Building and Plan
L0	Social Studies: any area of study
L1	Economics
L2	Politics
L3	Sociology
L4	Social Policy
L5	Social Work
L6	Anthropology
L7	Human and Social Geography
L9	Others in Social Studies
LL	Combinations within Social Studies
M1	Law by Area
M2	Law by Topic
M9	Others in Law
MM	Combinations within Law
N0	Business and Admin studies: any area

Table A1: Continued

JACS Line	*Subject description*
N1	Business Studies
N2	Management Studies
N3	Finance
N4	Accounting
N5	Marketing
N6	Human Resource Management
N7	Office Skills
N8	Tourism, Transport and Travel
N9	Others in Business and Admin Studies
NN	Combinations within Business and Admin Studies
P0	Mass Communications and Documentation: any area
P1	Information Services
P2	Publicity studies
P3	Media studies
P4	Publishing
P5	Journalism
P9	Others in Mass Communications and Documentation
PP	Combinations within Mass Communications and Documentation
Q0	Linguistics,Classics and related: any area
Q1	Linguistics
Q2	Comparative Literary studies
Q3	English studies
Q4	Ancient Language studies
Q5	Celtic studies
Q6	Latin studies
Q7	Classical Greek studies
Q8	Classical studies
Q9	Others in Linguistics, Classics and related
QQ	Combinations within Linguistics, Classics and related
R0	European Languages, Literature and related: any area
R1	French studies
R2	German studies
R3	Italian studies
R4	Spanish studies
R5	Portuguese studies
R6	Scandinavian studies
R7	Russian and East European studies
R9	Others in European Languages, Literature and related
RR	Combinations within European Languages, Literature and related
T1	Chinese studies
T2	Japanese studies
T3	South Asian studies
T4	Other Asian studies
T5	African studies
T6	Modern Middle-Eastern studies
T7	American studies
T8	Australasian studies
T9	Others in non-European Languages and related
TT	Combinations within non-European Languages and related

Table A1: Continued

JACS Line	*Subject description*
V0	History and Philosophical studies: any area
V1	History by Period
V2	History by Area
V3	History by Topic
V4	Archaeology
V5	Philosophy
V6	Theology and Religious studies
V9	Others in History and Philosophical studies
VV	Combinations within History and Philosophical studies
W0	Creative Arts and Design: any area
W1	Fine Art
W2	Design studies
W3	Music
W4	Drama
W5	Dance
W6	Cinematics and Photography
W7	Crafts
W8	Imaginative Writing
W9	Others in Creative Arts and Design
WW	Combinations within Creative Arts and Design
X0	Education: any area of study
X1	Training Teachers
X2	Research and Study Skills in Education
X3	Academic studies in Education
X9	Others in Education
XX	Combinations within Education
Y1	Combs of Arts/Humanities
Y10	Combinations of Physics/Mathematical sciences
Y11	Combinations of Science/Engineering with Arts/Humanities/Languages
Y12	Combinations of Science/Engineering with Social Studies/Business/Law
Y13	Combinations of Sciences with Engineering/ Technology
Y14	Combinations of Social Studies/Law
Y15	Combinations of Social studies/Law with Business
Y16	Combinations of Social Studies/Business/Law with Arts/Humanities
Y17	Combinations of Social Studies/Business/Law with Languages
Y2	Combinations of Engineering/Technology/ Building Studies
Y3	Combinations of Engineering/Technology
Y4	Combinations of Languages
Y5	Combinations of Languages with Arts/ Humanities
Y6	Combinations of Medicine/Biology/Agricultural sciences
Y7	Combinations of Medicine/Biology/Agricultural sciences with Physics/Mathematical Science
Y8	Combinations of Physics/Mathematical Science with Arts/Humanities/Languages
Y9	Combinations of Physics/Mathematical Science with Social Studies/Business/Law
Z	Other courses not elsewhere classified