Issue 21 Winter 2016

Research Matters

CAMBRIDGE ASSESSMENT

University of Cambridge
Local Examinations Syndicate

CAMBRIDGE ASSESSMENT

Citation

Articles in this publication should be cited as: Sutch, T., Zanini, N., & Vidal Rodeiro, C. (2016). The effect of specialism and attainment in secondary school on the choice of Higher Education institution and field of study. *Research Matters: A Cambridge Assessment Publication, 21,* 2–11.

All details are correct at the time of going to print in February 2016

If you would like to comment on any of the articles in this issue, please contact Sylvia Green – Director, Research Division. Email: researchprogrammes@cambridgeassessment.org.uk

The full issue of *Research Matters 21* and all previous issues are available from our website: www.cambridgeassessment.org.uk/research-matters

Research Matters : 21

A CAMBRIDGE ASSESSMENT PUBLICATION

Foreword

Those reading this issue of *Research Matters* should not tire of reading before they reach the final article by Bramley and Wilson. And of course, this is not a cue to skip every other article. Bramley and Wilson have continued assiduously to examine the assets and limitations of expert judgement – this article on judging item difficulty accompanies continuing work on the comparison between 'statistics' and 'judgement' in awarding. The properties of 'comparable outcomes' have been more widely understood by journalists, teachers and the wider population, not least due to the central role that comparable outcomes is likely to play in responding to the undoubted challenges posed by new GCSEs and A levels – these being phased in over a three-year period. But while the official requirements continue to emphasise 'judgement informed by statistics' at the heart of awarding, we should not depart from continued efforts to understand the precise characteristics of expert judgement, applied to patterns of outcomes, and to expectations (not least in the shape of exam questions). The acknowledgement from the Office of Qualifications and Examinations Regulation (Ofqual) of Cambridge Assessment's continuing work on the role which judgement can and should play may all too easily be forgotten as 'comparable outcomes' comes to dominate standards in the coming protracted period of intense change. If reminders that the comparable outcomes approach was designed as an approach to manage change are to be heeded (Bramley, Dawson & Newton, 2014), we must all remember that judgement must not permanently be laid to one side.

From the last article, to the first…

A series of articles from Cambridge has explored whether England's 'wide diet at 16, narrowing at 18' is both typical and educationally prudent. Our conclusion (through careful transnational comparison) is that far more countries have this pattern than is presumed. Part of our approach is to see this pattern as part of the total and delicate set of educational arrangements – culminating in high-quality, highly efficient, short-duration first degree programmes. The impact of specialism, choice and future route therefore is of profound interest; Sutch, Zanini and Rodeiro throw key patterns into sharp relief – a vital issue for both individuals, society and the economy.

Tim Oates, CBE *Group Director, Assessment Research and Development*

Editorial

The first four articles in this issue report on a range of studies in Higher Education (HE) contexts. Sutch, Zanini and Vidal Rodeiro, explore the way in which students' experiences in the secondary phase of education impact on the choices that they make when deciding where and what to study at HE level. These interactions have implications for future earnings and social mobility. In their research, Darlington and Bowyer focus on current undergraduate students of Architecture and the extent to which A level study in Mathematics and Further Mathematics prepared them for their studies. As well as providing insights into the perceived usefulness of A level study, this research highlights the need for detailed guidance so that students can be well informed when making A level choices according to their future ambitions. In his study of the Universities and Colleges Admissions Service (UCAS) tariff, Gill discusses the difficulties in comparing qualifications and the range of factors involved. He recognises that there are no easy solutions to the problems identified in his study given the extremely complex challenges involved. Carroll and Dodds discuss what we can learn about the concept and assessment of creativity through their research into Drama and Theatre in HE education. This is an interesting and perceptive study which provides useful insights into definitions and practices.

Johnson's article takes us in a different direction as he reports on a study into the way in which senior examiners communicate with their team members in a digital environment. His research highlights the importance of feedback, shared understanding and the relationships involved. In their article Wilson, Werno and Smith describe some encouraging findings in the context of the assessment of reading in reformed Modern Foreign Languages. The introduction of a wider range of text types, including more authentic texts, was welcomed by teachers involved in this study.

Research by Hughes and Shaw seeks to identify and explain the problem of under-used marks. This problem can have a detrimental effect on the reliability of measurement and discrimination. The authors make suggestions for strategies to address these problems and to support examiners in using the full mark range. In the final article, Bramley and Wilson describe two methods designed to support the maintenance of test standards. The evidence from their research suggests that their methods could be used to provide additional evidence about the difficulty of questions.

Sylvia Green *Director, Research Division*

The effect of specialism and attainment in secondary school on the choice of Higher Education institution and field of study

Tom Sutch, Nadir Zanini and **Carmen Vidal Rodeiro** Research Division

Introduction

Progression from secondary to Higher Education (HE) has direct implications on wage returns and social mobility. The recent expansion of HE in the majority of European countries has highlighted that returns to HE are not just associated with the decision to study at university rather than enter the labour market, but also with the choice of studying in a particular field (Chevalier, 2011; Greenwood, Harrison & Vignoles, 2011) at a specific HE institution (Chevalier & Conlon, 2003; Hussain, McNally & Telhaj, 2009; Walker & Zhu, 2013). Because the process of application and admission to universities in the United Kingdom (UK) places a strong weight on attainment, both overall and in specific subjects, the educational background of students is a key factor influencing progression from secondary education to specific fields of study and HE institutions (Vidal Rodeiro, Sutch & Zanini, 2015).

The aim of this article is to provide evidence about the relationship between educational background, measured by subject choice and attainment in the final years of secondary education, and HE participation in terms of institution attended and choice of the field of study, an area in which not much research has been carried out so far. This is a topic of particular interest especially in the United Kingdom (UK), where the HE sector is characterised by a vast subject offer and a substantial diversity among institutions. Most HE institutions are members of so-called 'mission groups', the most well-known and prestigious being the Russell Group of research-oriented universities, within which they share similar aims and practices.

General Certificate of Education (GCE) Advanced levels (A levels) are the most common qualification taken by 18 year olds in England, and are available in a wide variety of academic and applied subjects. Most students take A levels in three subjects (Gill, 2012) and study them over a period of two years. Over recent years the importance of subject choice at A level in preparing students for HE has been highlighted. Fazackerly and Chant (2008) drew attention to the disparity in the attitudes of HE admissions staff towards certain A level subjects, by showing the gulf in uptake of these subjects among students at prestigious HE institutions, compared to the national uptake. Individual institutions maintained their own lists of undesirable 'soft' subjects, some of which were more open than others, and in 2011 the Russell Group published *Informed Choices* (Russell Group, 2011) in an attempt to unify and simplify the message for the benefit of prospective applicants. In it, they advised students to study at least two from a list of 'facilitating subjects', which would leave their options open for a variety of courses. However, they acknowledged that this advice would not apply to all students, and those who were definitely intending to study certain specialist courses such as Music would be best served otherwise. The concept of 'facilitating subjects' has gained traction with the Government, which since 2012 has included statistics in its school performance tables on the percentage of students achieving good grades in facilitating subjects at A level[1]. However, there is evidence that the attitudes of admissions staff at Russell Group universities towards various subjects are more mixed than *Informed Choices* might suggest (Candy, 2013). Outside the Russell Group, there is less information available to prospective applicants to guide subject choices. Although requirements for each university and course are available to applicants via the Universities and Colleges Admissions Service (UCAS) website and others, and many of these websites have provided general advice to students wishing to study specific subjects[2], the underlying data on entry requirements is not freely available. As such, students are heavily dependent on guidance from schools and others in order to make sense of their options at age 16.

The recent expansion and diversification of the HE sector has led to many questioning whether the growth has been evenly felt and access has become more equitable, or whether some students are disadvantaged because of their background. Accordingly, studies have been carried out on the effect of socio-economic status (e.g., Gayle, Berridge & Davies, 2002; Chowdry, Crawford, Dearden, Goodman & Vignoles, 2013), vocational qualifications (e.g., Hoelscher, Hayward, Ertl & Dunbar-Goddet, 2008) or school background (e.g., Sutton Trust, 2009; Crawford, 2014) on progression to HE and subject of study. In these, attainment at A level is typically used as a controlling factor but it is not the primary variable of interest. Perhaps surprisingly, there has been little focus on attainment at A level itself, except in certain subject-specific studies.

Progression to university is a matter of interest not just from a research point of view but also for students, HE institutions, awarding bodies and policy makers. A better understanding of how students' choices are influenced by their educational background at A level is important, especially in a time of change in the English educational system. This work examined how students' choice of A level subjects and attainment influence their HE destinations by answering the following research questions:

1. is there a systematic relationship between A level background and progression to specific types of HE institutions?;
2. is there a systematic relationship between A level background and progression to specific fields of study?

1. Available at http://www.education.gov.uk/schools/performance/
2. For example, http://university.which.co.uk/advice/what-a-levels-do-you-need-to-study-history

Data and methods

Data and variables

The data for the analyses was provided by the Higher Education Statistics Agency (HESA)[3] and covers all full-time, first year undergraduates aged 17–19, domiciled in England, studying at UK universities in the academic year 2011/12. The dataset includes information on the HE course studied at university as well as qualifications obtained prior to starting the course. The analyses were restricted to 181,190 students with at least 3 A level passes (excluding General Studies[4]), around 72 per cent of the entire cohort of full-time, first year undergraduates. Although this group may not be representative of the whole student body, focusing on this particular qualification enables the use of a comparable measure of prior attainment.

In this research, the HE institutions were considered in mission groups: Russell Group, 1994 Group[5], University Alliance and Million+, which are mutually exclusive. Universities that did not belong to any of these groups were included in a separate group, labelled as Other. The Russell Group consists of research-led institutions, which tend to be the most prestigious and generally have the most competitive admissions criteria. Some other 'top' universities were included in the 1994 Group. University Alliance and Million+ bring together most of the newest universities and colleges, with the former including the more status-conscious 'post-1992' institutions (Scott, 2013).[6]

For each student, information on up to three subjects of study and the subject percentage (i.e., the relative contribution of that subject to the university degree) was available in the data. The subject of study was then aggregated into 20 broad subject areas[7] and analyses were carried out at this level. It should be noted that the subject area relates to the principal subject of study. For degrees with more than one subject (e.g., joint honours) the subject area corresponds to the subject with the largest percentage. If a student took a balanced combination, or a triple honours degree in three different subject areas, then the subject area was 'Combined'.

A level subjects were categorised as one of the following (Bramley, 2014): *Applied, Expressive, Humanities, Languages, STEM* (Science, Technology, Engineering and Mathematics). Students were then assigned to specialisms based on the subjects they passed (at grades A*–E). For example, if a student took two A levels in Humanities subjects and one Expressive subject, they would be classed as a Humanities specialist. Some students did not specialise in a particular area (e.g., taking one A level in each of three categories), or specialised in several areas (taking two subjects in each of two or more categories), and we assigned these to *None* and *Multiple* respectively. The distribution of specialism is shown in Table 1.

Table 1: Number and percentage of students with different A level specialisms

		Number	*%*
A level specialism	None	22,515	12.4
	Applied	7,520	4.2
	Expressive	3,620	2.0
	Humanities	87,040	48.0
	Languages	2,380	1.3
	STEM	54,950	30.3
	Multiple	3,165	1.7
Average A level attainment	Grade C or above	151,210	83.5
	Grade A or above	37,930	20.9
Total		**181,190**	

For A level attainment, the following measures were used:

a. The average grade of a candidate's A level passes, excluding General Studies (assigning the grades nominal values of A*=7, A=6, B=5, C=4, D=3, E=2);

b. Two binary variables indicating whether the candidate's average grade (across all A level passes, excluding General Studies) was:
   - C or above;
   - A or above.

c. Three binary variables for each of the five A level subject categories (Applied, Expressive, Humanities, Languages, STEM) indicating whether the candidate's average grade in this category was:
   - E or above;
   - C or above;
   - A or above.

The variables within each of b) and c) are not mutually exclusive, so a candidate with an average of an A grade in a particular subject category would satisfy all three of the conditions in c), for example[8]. For subject categories in which students had not taken any A levels, the variables within c) would all be zero. The average A level grade, described above, takes a limited number of values in practice (although we have treated it as continuous in our modelling), because most students included in the research take three or four A levels. The most frequent average grade was B, and the majority of students scored between a C and an A on average. This is shown in Table 1 which gives the breakdown of the discrete overall attainment variables described in b) above.

Detailed student-level information such as gender, the LSOA[9] of the student's home address and a previous institution identifier was also available. Information about the type of the previous institution was obtained from the National Pupil Database (NPD)[10] and matched to the HESA data using the previous institution identifier. The following previous institution types were considered: comprehensive schools, academies, independent schools, selective schools, sixth form colleges, further education (FE) colleges and other/unknown.

3. Source: HESA Student Record 2011/12. Copyright Higher Education Statistics Agency Limited 2012. HESA cannot accept responsibility for any inferences or conclusions derived from the data by third parties.
4. General Studies has been considered separately from other A levels in previous research and many universities do not include it in their offers.
5. The 1994 Group was disbanded in November 2013.
6. A full list of members and more detailed information about the scope and the practices of each group can be obtained from the groups' websites.
7. The university subject areas used are those defined by Version 2 of the Joint Academic Coding System (JACS). The subject group is denoted by the first letter in the JACS code.
8. Fails (grade U) could not be taken into account in calculating these attainment measures (to provide a correct denominator for the calculation of average grades) because they had previously been removed from the dataset. However, given that the analysis was restricted to those students with three A level passes (excluding General Studies) the effect of this should be minimal.
9. LSOA (Lower Layer Super Output Area) is a conglomeration of a number of census output areas (each output area has about 150 households). They usually have a minimum population size of 1,000 and an average of 1,500. There are over 34,000 LSOAs in England.
10. The National Pupil Database, compiled by the Department for Education (DfE), is a longitudinal database for all children in schools in England, linking student characteristics to school and college learning aims and attainment. The NPD holds pupil and school characteristics such as age, gender, ethnicity, level of deprivation, attendance and exclusions, matched to pupil level attainment data (Key Stage 2 to Key Stage 5 assessments and other external examinations).

A proxy for the students' socio-economic background was determined by the students' level of deprivation using the Income Deprivation Affecting Children Index (IDACI). This index, obtained from the Office for National Statistics, shows the percentage of children in the LSOA in which the student resides who live in families that are income deprived. We assigned students into one of three groups depending on the value of the IDACI index.

Methods

An assessment of the universities and subject areas in which different types of A level students are over- or under-represented can be made using the odds ratios derived from multilevel logistic regressions. The regression analyses allowed us to take into account students' background characteristics (e.g., gender, prior educational institution and socio-economic status) when looking at the probability of attending a specific university group or pursuing a specific course.

Multilevel models were proposed due to the hierarchical or clustered structure of the data (as students were grouped within schools). If this hierarchical structure were not recognised, then the standard errors of the regression coefficients would be underestimated, leading to an overstatement of the statistical significance. Detailed discussions of the implementation and outcomes of the multilevel logistic regression can be found in Goldstein (2011).

Two different specifications of multilevel logistic regression were considered. In the first (Model A) the dependent variable was the enrolment in a university group, whilst the dependent variable in the second (Model B) was pursuing a university course in a subject area. Separate models were fitted for each university group and course. Although the specifications of the regression models employed to study the two dependent variables were slightly different, in both models the explanatory variables included: gender, prior educational institution, socio-economic status, A level student category, and measures of attainment at A level.

The inclusion of these variables allowed us to control for several factors when investigating the type of university attended and the subject area studied by A level students. A detailed breakdown of the dependent and independent variables included in the regression models is presented in Table 2.

Table 2: Description of the variables included in the multilevel logistic regression models

Name	*Description*	*Range of values (Baseline value underlined)*	*Model*	
			A	*B*
DEPENDENT VARIABLES				
Candidate enrolled in a type of university	Indicator of university enrolment	Discrete variable: 0 was not enrolled at the university; 1 was enrolled at the university.	●	
Candidate pursuing a course in a subject area	Indicator of subject area uptake	Discrete variable: 0 did not take a course in the subject area; 1 took a course in the subject area.		●
INDEPENDENT VARIABLES				
Gender	Gender of the candidate	Discrete variable: male; female	●	●
Level of deprivation (IDACI)	Candidate level of deprivation based on the IDACI	Discrete variable: low, average, high	●	●
Centre type	Type of institution the candidate attended prior to university	Discrete variable: comprehensive; independent; academy; sixth form college; selective; FE college; other	●	●
Student category	A level subject specialism	Discrete variable: None; Applied; Expressive; Humanities; Languages; STEM; Multiple	●	●
University subject area	University subject area studied	Discrete variable: Architecture, Building and Planning; Biological Sciences; Business and Administrative Studies; Creative Arts and Design; Eastern, Asiatic, African, American and Australasian Languages, Literature and related subjects; Education; Engineering; European Languages, Literature and related subjects; Historical and Philosophical Studies; Law; Linguistics, Classics and related subjects; Mass Communications and Documentation; Mathematical and Computer Sciences; Medicine and Dentistry; Other/Combined; Physical Sciences; Social Studies; Subjects allied to Medicine; Technologies; Veterinary Sciences, Agriculture and related subjects	●	
A level score	Average grade across all A level subjects taken	(Quasi-) continuous variable: 2 represents an average of grade E; 7 represents an average of grade A*		●
Overall grade thresholds	Indicator of whether a candidate gained an average of C or above, an A or above, across all A levels	Two discrete variables: 1 if the student achieved an average A level grade greater than or equal to C/A; 0 otherwise	●	
A level specialism grade thresholds	Indicator of whether a candidate gained an average E or above (i.e., passed any A levels), a C or above, an A or above, in each of the five A level specialism categories	15 discrete variables: 1 if the student achieved an average A level grade greater than or equal to each threshold value in each A level specialism; 0 otherwise.		●

In modelling students' decisions to enrol at university we have assumed (following Maringe, 2006) that their choice of subject of study comes before their choice of institution; thus, we have allowed the university type to be influenced by the subject, but not vice versa.

The rationale for Model A is that university participation, in terms of the type of institution attended, might be expected to depend on a student's general academic ability and the subject area of study (as some subject areas are predominantly offered in certain university groups). A student's broad choice of A levels might also have an influence on the type of institution they attend, due to differing admission policies.

The rationale for Model B is that the subject studied at university might be expected to depend on a student's general academic ability, their broad choice of A levels, whether they have taken A levels in particular subject areas (represented by the 'E or above' dummy variables) and their A level grades in particular subject areas (represented by the 'C or above' and 'A or above' dummy variables).

Results of the estimated regression models are presented in the form of odds ratios for A level subject specialism and A level attainment[11].

Results

Descriptive analyses

Table 3 presents the destinations of A level students, in terms of HE mission groups, and shows a wide variation across specialisms at A level. Students who had specialised in Applied or Expressive subjects at A level were more likely to attend University Alliance or Million+ institutions. Linguists were particularly likely to go on to study at Russell Group universities (reflecting the concentration of Language degrees at these institutions), and those specialising in STEM or Multiple areas were also more likely than average to attend Russell Group universities. Humanities specialists were represented more evenly across all mission groups.

Students who had specialised in Expressive or STEM subjects at A level were most likely to have enrolled on a single honours degree, while linguists were much more likely than average to have gone on to study a joint honours course. Investigation of the joint honours courses (balanced and major/minor combinations) studied by Language specialists revealed that 94 per cent of these students took at least one Language subject in their degree, and 62 per cent of students took two Language subjects.

Table 4 shows the subjects that students with each A level specialism went on to study in HE. As might be expected, there is a strong relationship between A level subject choices and the subject area studied at university. For example, 63 per cent of students specialising in Expressive subjects at A level studied 'Creative Arts and Design' at university; likewise, 58 per cent of those specialising in Languages at

11. Odds ratios for the other independent variables are not reported as they are not the focus of the research, although they are available upon request. However, it is important to note that their inclusion in the regression models allows interpretation of the odds ratios for attainment and subject specialism accounting for their effect.

Table 3: University mission group by A level specialism (percentage of students in category)

A level specialism	*Russell Group*	*1994 Group*	*University Alliance*	*Million+*	*Other*
None	20.0	6.3	32.8	15.4	25.4
Applied	8.5	4.1	41.1	22.7	23.5
Expressive	8.5	4.4	36.3	18.4	32.4
Humanities	27.5	8.1	25.8	14.0	24.6
Languages	61.2	7.2	10.2	4.0	17.4
STEM	47.4	7.3	20.1	7.1	18.1
Multiple	54.2	7.9	14.4	6.5	17.0
All A level students	32.4	7.4	25.4	12.3	22.6

Table 4: University subject area by A level specialism (percentage of students in category)

University subject area	*A level specialism*							*All A level students*
	None	*Applied*	*Expressive*	*Humanities*	*Languages*	*STEM*	*Multiple*	
Architecture, Building and Planning	4.0	1.8	6.8	0.9	0.2	1.4	1.4	1.6
Biological Sciences	12.9	12.3	1.2	9.7	1.6	15.7	11.9	11.8
Business and Administrative Studies	17.5	40.5	4.8	8.3	4.4	3.9	6.0	9.3
Creative Arts and Design	13.8	2.4	63.3	8.7	2.7	1.5	5.8	7.8
Eastern, Asiatic, African, American and Australasian Languages, Literature and related subjects	0.3	0.0	0.0	0.6	2.2	0.1	0.5	0.4
Education	3.3	4.4	2.5	3.8	1.3	0.5	1.5	2.7
Engineering	2.5	1.5	3.3	0.4	0.2	14.3	3.0	5.0
European Languages, Literature and related subjects	3.0	0.1	0.2	1.1	57.5	0.3	7.4	1.9
Historical and Philosophical Studies	2.3	0.9	0.9	10.2	1.8	0.9	5.6	5.6
Law	4.1	9.3	0.2	6.6	2.8	1.2	6.2	4.6
Linguistics, Classics and related subjects	2.0	0.6	1.0	9.4	2.8	0.4	4.2	5.0
Mass Communications and Documentation	2.0	1.5	2.2	5.3	0.2	0.2	0.9	3.0
Mathematical and Computer Sciences	5.1	5.9	1.7	1.3	0.8	12.6	7.4	5.5
Medicine and Dentistry	0.0	0.0	0.0	0.0	0.0	8.3	2.6	2.6
Physical Sciences	3.6	1.1	0.4	2.8	0.8	14.8	6.6	6.5
Social Studies	7.3	6.4	1.1	14.2	3.6	3.5	11.9	9.3
Subjects allied to Medicine	5.3	3.5	1.0	2.8	0.9	12.0	4.4	5.9
Technologies	0.8	0.3	3.5	0.3	0.1	0.6	0.3	0.5
Veterinary Sciences, Agriculture and related subjects	1.3	0.7	0.6	0.5	0.1	2.0	0.4	1.1
Other/Combined	8.8	6.6	5.2	13.2	16.2	5.8	11.9	10.0

A level enrolled on degree courses in 'European Languages, Literature and related subjects'.

Students with multiple specialisms at A level were particularly well represented in 'Biological Sciences', 'European Languages, Literature and related subjects', 'Law', 'Mathematical and Computer Sciences', and 'Social Studies'. Students with no specialism, who had taken a mixture of subjects, were highly represented in 'Business and Administrative Studies' and 'Creative Arts and Design' courses at university.

A particularly strong association between A level subject choices and university subject area was found in 'Medicine and Dentistry'. While 8.3 per cent of STEM specialists, and 2.6 per cent of Multiple specialists went on to study a degree in this area, hardly any students from other specialisms did. This is because Medicine courses typically require Biology and Chemistry at A level, which would put students in the STEM category, or the Multiple category if they had taken more additional subjects to add breadth.

A surprising result at first glance is the high proportion of students in many categories going on to take courses in 'Biological Sciences'. This can be explained by the fact that, as well as Biology, this group includes courses in Psychology and Sports Science, and these subjects were classed as a Humanities and Applied subject respectively at A level.

Enrolment in type of university (Model A)

Table 5 presents the odds of attending each university group for the baseline case, alongside estimates of the odds ratios associated with the explanatory variables of interest: overall attainment and subject specialism at A level.

Looking first at the effect of overall attainment at A level, 9 out of 10 odds ratios were significant at the 5 per cent level, leading to the conclusion that, as expected, average attainment was a significant determinant of the institution group attended by students. The first column of odds ratios in Table 5 (alongside the baseline odds) relate to students graded C or above overall. With other conditions held fixed, students having an average attainment of grade C or above were much more likely to enrol in a Russell Group or 1994 Group university than students with an average attainment below grade C. The odds ratio for the Russell Group was particularly large, reflecting the low odds of students attending a university in this group with an average grade of below C at A level. Conversely, an average attainment above grade C reduced the likelihood of students attending HE institutions other than those in the Russell or 1994 mission groups.

The second column of odds ratios refers to students with an average grade at A level of A or above: these results should be interpreted as the odds ratios on top of those shown by the previous column, which referred to overall attainment of grade C or above. It is clear that students attaining an average grade of A or above were much more likely than students with grade C or above to attend a Russell Group university and less likely to enrol at universities in other groups; however, students with a grade A or above were still more likely than A level students in general to study at 1994 Group universities. For those students with an overall level of attainment above grade A, with respect to those below grade A, the likelihood of attending a University Alliance or Million+ institution was particularly low.

The remainder of Table 5 shows the odds ratios of the A level specialism category with respect to enrolment in each university group. In particular, specialising in STEM or Multiple areas greatly improved the likelihood of studying in a Russell Group university and reduced the chance of enrolling in other HE institutes. Students that had specialised in Humanities or Languages were also more likely to attend universities in the Russell Group, but the size of the association was smaller than in the former case. Conversely, students specialising in Applied and Expressive subjects at A level were less likely to attend Russell and 1994 Group universities. Finally, the likelihood of attending Million+ and University Alliance universities was higher for specialists in Applied and Expressive A level subjects than for those with no specialism.

Subject area at university (Model B)

Results from Model B are presented in Table 6 and Figure 1. Table 6 presents the odds associated with studying each subject area at university (for the baseline case) alongside odds ratios for each A level subject category in comparison to those with no specialism. Figure 1 shows the odds ratios for the achievement variables, both overall and in each A level subject category. In interpreting the results, two important dependencies should be borne in mind. Firstly, students specialising in a particular category must by definition have passed at least two A levels in this category, and so at least one of the odds ratios associated with A level attainment will also apply. For example, considering the odds of studying Historical and Philosophical Studies at university, the odds ratio associated with specialising in Humanities at A level is 1.95, but a further ratio from having an overall grade of at least an E in Humanities subjects, as shown in Figure 1, would also always be compounded. Secondly, there is a clear dependency of overall attainment on the individual subject

Table 5: Odds ratios for average A level attainment, and specialism category, in comparison to no specialism (Model A)

University Group	*Odds (intercept)*	*Odds ratios*							
		Average attainment at A level		*A level specialism category (with respect to no specialism)*					
		Grade C or above (with respect to all A level students)	*Grade A or above (with respect to 'C or above')*	*Applied*	*Expressive*	*Humanities*	*Languages*	*STEM*	*Multiple*
Russell Group	0.004	**37.81**	**7.92**	**0.67**	**0.61**	**1.12**	**1.47**	**2.19**	**2.75**
1994 Group	0.012	**7.98**	**0.77**	**0.77**	**0.82**	**1.08**	0.91	1.02	1.01
University Alliance	0.653	**0.87**	**0.10**	1.06	**1.14**	**0.89**	**0.63**	**0.69**	**0.54**
Million+	0.607	**0.24**	**0.11**	**1.11**	1.08	1.03	0.96	**0.64**	**0.64**
Other	0.366	0.97	**0.37**	**0.92**	0.98	1.00	1.00	**0.91**	**0.78**

Note: Significant estimates at the 5% level are presented in bold.

Table 6: Odds ratios for specialism category, in comparison to no specialism (Model B)

University subject area	*Odds (intercept)*	*Odds ratio for A level specialism category (with respect to no specialism)*					
		Applied	*Expressive*	*Humanities*	*Languages*	*STEM*	*Multiple*
Architecture, Building and Planning	0.015	**1.63**	**1.99**	1.11	**0.25**	0.92	1.24
Biological Sciences	0.093	1.04	**0.20**	**0.93**	**0.24**	**1.29**	1.05
Business and Administrative Studies	0.387	**3.04**	**0.74**	1.04	**0.76**	**0.56**	0.93
Creative Arts and Design	0.683	**0.29**	**2.24**	**0.58**	**0.28**	**0.19**	**0.71**
Eastern, Asiatic, African, American and Australasian Languages, Literature and related subjects	0.004	**0.13**	**0.00**	1.04	**1.72**	**0.14**	0.86
Education	0.284	**1.18**	**0.68**	0.93	0.73	**0.30**	0.75
Engineering	0.054	**0.66**	0.99	**0.38**	**0.17**	**4.58**	**1.67**
European Languages, Literature and related subjects	0.003	**0.17**	**0.17**	**0.44**	**4.39**	**0.15**	**1.48**
Historical and Philosophical Studies	0.002	**0.51**	**0.36**	**1.95**	**0.35**	**0.33**	1.01
Law	0.003	**2.17**	**0.29**	**2.16**	0.82	**0.69**	**2.15**
Linguistics, Classics and related subjects	0.000	**0.48**	**0.34**	**2.00**	**0.44**	**0.27**	0.96
Mass Communications and Documentation	0.374	**0.38**	**0.39**	**1.16**	**0.12**	**0.16**	**0.44**
Mathematical and Computer Sciences	0.023	**1.63**	0.88	**0.75**	**0.49**	**2.30**	**2.02**
Medicine and Dentistry	0.000	0.99	0.49	0.72	0.09	**49.54**	**15.59**
Physical Sciences	0.044	**0.39**	**0.17**	**0.81**	**0.29**	**1.95**	1.09
Social Studies	0.028	0.95	**0.29**	**1.48**	**0.46**	**0.49**	1.06
Subjects allied to Medicine	0.027	**0.77**	**0.40**	**0.63**	**0.35**	**2.12**	**0.88**
Technologies	0.091	**0.55**	**1.83**	**0.48**	0.38	0.99	0.67
Veterinary Sciences, Agriculture and related subjects	0.017	0.76	**0.55**	**0.53**	**0.28**	**1.38**	**0.41**
Other/Combined	0.069	**0.78**	**0.57**	**1.12**	**0.84**	**0.56**	0.93

Note: Significant estimates at the 5% level are presented in bold.

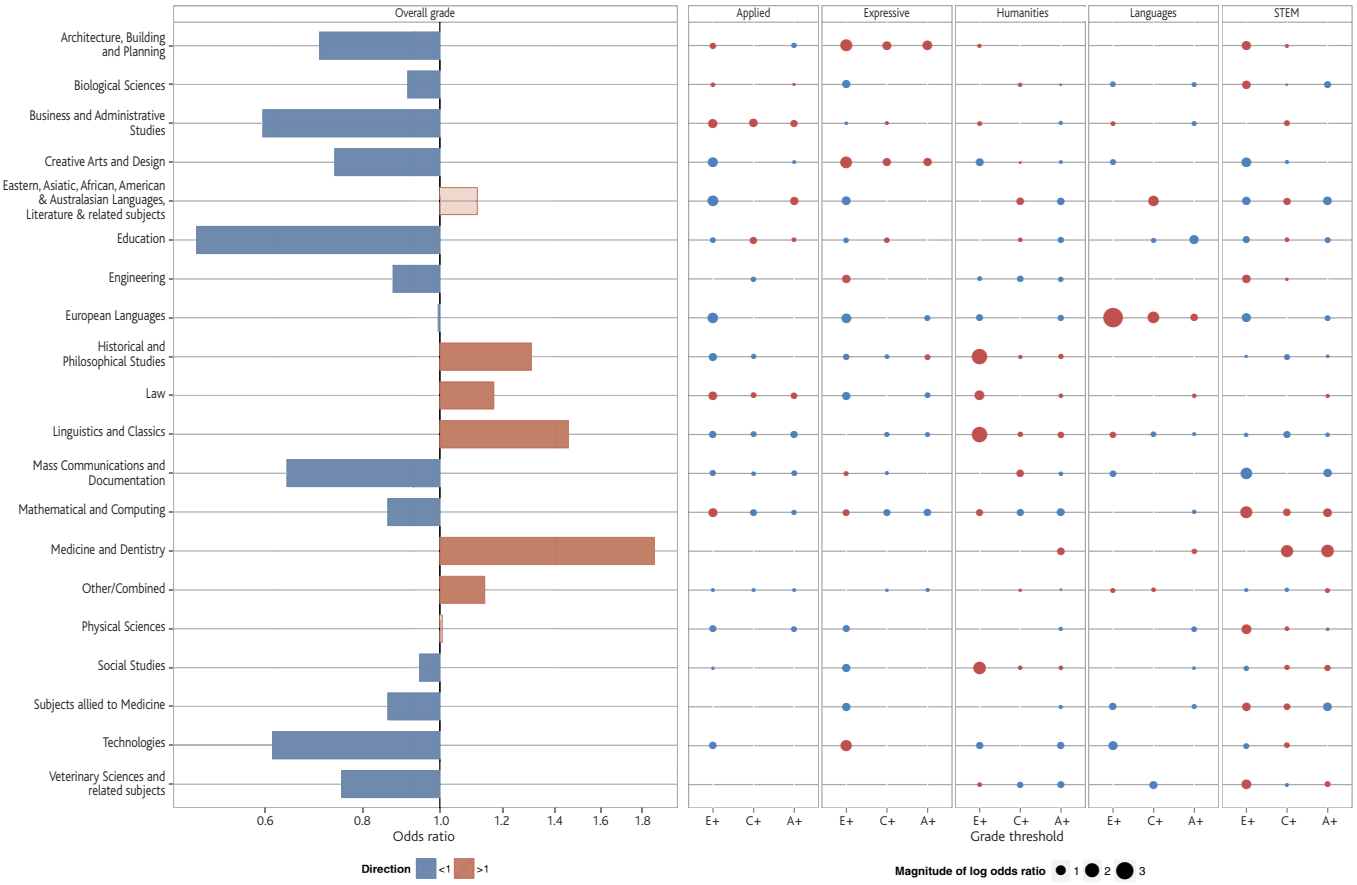


Figure 1: A level choice and attainment

category attainment variables: for a student gaining three A grades in Science subjects at A level, four odds ratios for attainment apply (overall grade and each of the binary variables relating to grades in STEM subjects), as well as STEM specialism. The breakdown of these odds ratios allows an assessment of the relative contribution from specialism and attainment.

The odds ratios in Table 6 show that subject choices at A level had a significant effect on the likelihood of studying in a particular subject area at university, even if students had not achieved very highly. For example, STEM specialists at A level were significantly more likely than average to study STEM subjects at university, and particularly 'Medicine and Dentistry'. Conversely, students who had specialised in Humanities subjects at A level were more likely to pursue courses in 'Historical and Philosophical Studies', 'Law', 'Linguistics, Classics and related subjects' and 'Social Studies'.

Students who specialised in Multiple areas at A level were significantly more likely to study 'Engineering', 'European Languages, Literature and related subjects', 'Law', 'Mathematical and Computer Sciences', and especially 'Medicine and Dentistry'. The Multiple specialists who went on to study 'Medicine and Dentistry' courses (of whom there were fewer than 100) were investigated: in most cases they studied two Humanities subjects as well as Biology and Chemistry. The most popular Humanities A levels taken by these students were History and English Literature.

Most university subject areas were associated with higher odds in only one A level specialism (with the addition of students with Multiple specialisms in some cases), but 'Architecture, Building and Planning', 'Law' and 'Mathematical and Computer Sciences' were associated with two individual specialisms. Interestingly these subject areas all span the vocational/academic divide, and the mix may reflect the variety of courses available in these subject areas.

Figure 1 presents the odds ratios associated with attainment, under Model B. The bars on the left show the effect of increasing the overall average A level grade by one (e.g., moving from an average of D to C, or B to A) on the chances of studying a particular course. A subject area with an odds ratio greater than 1 denotes that students with a higher average A level grade were more likely than average to study in this area. Subject areas associated with significantly higher than average A level grades were 'Medicine and Dentistry', 'Linguistics, Classics and related subjects', 'Historical and Philosophical Studies', 'Law' and 'Eastern, Asiatic, African, American and Australasian Languages, Literature and related subjects' as well as 'Other/Combined'.[12] 'Education', 'Business and Administrative Studies' and 'Technologies' (all of which are vocational areas) were associated with lower than average A level grades. Despite the concentration of 'European Languages, Literature and related subjects' courses in Russell Group universities which might be expected to have higher admissions criteria, there was no significant effect for overall A level score for this subject area. This may be because grades in particular A levels, such as Languages, were of more importance.

The circles on the right of Figure 1 show the odds ratios associated with achieving threshold grades in each of the five A level subject categories, for each university subject area. Only statistically significant odds ratios are shown. The 'E or above' values (denoted as 'E+' for brevity) represent the change in the odds associated with having an A level pass in that subject area (at any grade). As with the overall score dummy variables in Model A, the 'C or above' and 'A or above' values show the *extra effect* of having these grades, compared to the previous attainment category. For example, the odds of a student enrolling on an 'Architecture, Building and Planning' course at university were significantly greater if they had an A level in an Expressive subject (at any grade). Scoring an average of grade C or above in their Expressive subjects at A level was associated with a significant increase in the odds. Finally, scoring an average of grade A or above was associated with a further significant increase in the odds.

There was generally a correspondence between the university subject area and A level subject category in terms of the sensitivity to grade. For example, students were more likely to follow a course in 'Creative Arts and Design' if they achieved good grades in expressive subjects at A level, but those with A levels in other subject categories, especially at higher grades, were less likely to pursue a course in this area. However, in some university subject areas, there was an association with A level grades in several categories. These tended to be more vocational areas (such as 'Engineering' and 'Technologies'), where A levels may not be the primary route to these courses, and also for 'Eastern, Asiatic, African, American and Australasian Languages, Literature and related subjects' where students might not have studied directly relevant subjects at A level (and which might encompass a variety of subject areas).

As expected, the sensitivity to A level grade (as measured by the 'C or above' and 'A or above' odds ratios) was particularly marked for courses in 'Medicine and Dentistry', which are highly competitive, concentrated in Russell Group universities and have stringent entry requirements. While taking at least two STEM subjects (Table 6) gave rise to a higher probability of studying 'Medicine and Dentistry' at university, having an average A level grade in these subjects of at least an A increased the odds by a further factor of 22. The highest increase in odds was found for 'European Languages, Literature and related subjects' courses. While the baseline odds of studying in this area were low, having an A level in one or more Languages increased the odds by 47 times, and scoring an average of grade A or above in Language A levels increased the odds by a further factor of 7 to 330 times. Specialising in Languages, that is, taking at least two Languages at A level, resulted in a further factor of 4.39.

Conclusions and discussion

In the current educational climate, it is important to better understand how A levels are used by students to progress to HE. This research aimed to provide quantitative evidence to show how different subjects and students' attainment channelled learners in particular directions.

We considered HESA data gathered at individual level which covered all full-time, first year undergraduates aged 17–19, domiciled in England, studying at UK universities in the 2011/12 academic year. We further filtered the data to consider students with at least three A levels (181,190 students, 72 per cent of the whole cohort of undergraduates). Considering data on undergraduates does not allow us to study the determinants of the probability of progressing to HE, but enables us to focus on university participation in terms of institution attended and subject chosen for the students who did progress. Along with socio-demographic characteristics, the data used throughout the article included information on each student's qualifications prior to starting the course, the subject studied at university and the HE mission group of the institution the student attended.

12. The odds ratio for 'Eastern, Asiatic, African, American and Australasian Languages, Literature and related subjects' is greater than 1 but not statistically significant, hence it is not fully shaded in Figure 1.

Together with simple descriptive statistics which show the popularity of A level subject areas in relation to university participation, multilevel logistic regressions were employed to study the likelihood of students with different combinations of A levels to study in specific HE institutions and subjects, once students' characteristics had been accounted for.

Importance of subject choice

It has frequently been argued (e.g., Purcell et al., 2008; Fazackerly & Chant, 2008; Russell Group, 2011) that careful choice of subjects post-16 is crucial to avoid students inadvertently closing their options down prematurely. Our research supports that view, as we have found that subject choice has a significant effect, not only on the subject area of study but on the institution studied at.

The modular A level system has allowed learners to drop a subject after one year (perhaps the one in which they performed least well), so affording a degree of flexibility and allowing deferral of final A level subject choice until shortly before applying for university. However, this may be more difficult as the A level is now reverting to a linear form, so it will be even more crucial that students choose their subjects wisely at age 16.

A level subject and the subject choice at university

It should not be surprising that A level subject choice is linked to the subject of study at university, as if a student is interested and able enough in a subject to pursue it at HE it is probable that their interest and ability was stimulated and developed by earlier study in that subject or a related area. We have found that the strongest link between A level subject choice and university subject area is in 'Medicine and Dentistry', where the specialist knowledge required means that students need to have specialised in Science (or Multiple areas including Science) at A level, taking at least two STEM subjects. Some university subject areas were linked to several fields at A level, for example 'Architecture, Building and Planning' which attracted students who had previously specialised in Applied and Expressive areas. These university subject areas span the vocational/academic divide, and the mixture of backgrounds may reflect the variety of courses available in these subject areas.

Those students who had not specialised in a particular field (e.g., choosing three A levels in different categories) were more likely to study in more vocational areas at university, such as 'Business and Administrative Studies' and 'Creative Arts and Design'. These students may have benefited from taking a traditional academic subject to support their interest in Applied or Expressive subjects. For example, taking Mathematics or Economics in addition to Business Studies A level.

Some students, who have been termed Multiple specialists, took a broad and yet still deep curriculum, studying two A levels in each of two subject areas. They were most likely to study Economics or Mathematics at university. The popularity of Mathematics among this group is probably due to the special case of Further Mathematics A level which is almost invariably taken alongside Mathematics: if a student took Mathematics, Further Mathematics and two non-Science subjects they would be classified as a Multiple specialist.

A level subject and the type of university

The statistical analyses carried out in this research revealed that there is a relationship between A level subject specialism and the type of university attended. In particular, it is interesting to note that specialising in STEM or Multiple subject areas greatly increased the likelihood of studying in a Russell Group university. This also holds for Languages and Humanities, although in this case the magnitude of the association is smaller.

Applied and Expressive A level specialisms, conversely, reduced the likelihood of attending a Russell Group university. Students taking A levels in these subjects were more likely to study at a University Alliance or Million+ institution. Humanities students were quite widespread across different types of institution, but together with the aforementioned positive association with the Russell Group, they seemed to be positively linked to the 1994 Group and negatively associated with the University Alliance, even if the magnitude of the association was barely significant.

These results contribute to the debate about the crucial role of subject choice, and not only the type of qualification taken, after age 16 in the future career opportunities of young students, because these associations hold also when controlling for other variables, such as level of attainment and prior institution attended. Although membership of the Russell Group is not necessarily important in determining the quality of undergraduates' university experiences, empirical evidence has shown that obtaining a degree from a Russell Group institution leads to a higher wage return in the labour market (Hussain et al, 2009; Chevalier & Conlon, 2003). In other words, even if our study is not exhaustive[13], there is some indication that A level choice might indirectly influence students' future career opportunities and their social and economic status after their university studies.

Attainment at A level and the subject choice at university

Attainment at A level, as measured by both average grade and grades in five subject areas, did have a significant effect on the subject area of study. The greatest effect of overall grade was observed for 'Medicine and Dentistry', which is a particularly competitive area as already discussed, and 'Linguistics, Classics and related subjects'. However, somewhat surprisingly we did not observe a similar effect for overall grade in 'European Languages, Literature and related subjects', courses which are offered predominantly by Russell Group universities; instead, attainment in Language A levels in particular was important. As with the overall choice of A levels, the areas in which students' performance had a significant effect on university subject area were closely linked to the university subject area. For example, students were more likely to study 'Historical and Philosophical Studies' if they had achieved higher grades in Humanities subjects at A level. Sensitivity to grades in specific subject areas was most marked in 'Medicine and Dentistry' (focusing particularly on STEM subjects).

Our research has confirmed that subject choice at university is linked to attainment at A level more generally as well as attainment in specific subject areas. Variation in admissions offers across subject areas is likely to account for much of this relationship, but it may also be the case that more academically able students favour certain subject areas.

Attainment at A level and the type of university

As discussed earlier, much of the literature on progression to HE does not focus on attainment at A level per se, but uses it as a controlling variable when investigating for other factors. There is thus an acknowledgement

13. The breakdown by HE mission group does not make allowances for the fact that reputation and prestige vary in different ways for many subjects. Notably, some courses at post-1992 universities (typically within the University Alliance or Million+ Group) within certain subject areas have excellent research ratings and reputations, and as such are in a position to make high offers to applicants. For these reasons, when the specific university is available it is possible to consider alternative measures. However, usually Russell Group universities are included among high status institutions (see, as an example, Chowdry et al. 2013).

that prior attainment does have a strong effect on HE destinations, and there is widespread knowledge that Oxford, Cambridge and other prestigious universities (most of which are in the Russell Group) have stringent entry criteria. Our research has confirmed that, after controlling for other factors, students with an average grade of C or above were much more likely to enrol in a Russell Group or 1994 Group university than students with a lower grade, and those with an average of grade A or above were even more likely to study at a Russell Group university. Conversely, high attaining A level students were least likely to attend Million+ institutions.

Without comprehensive information on the offers made to applicants, it is impossible to separate the variation in entry requirements specified by different institutions from the competition for places (whereby institutions can select the best of the applicants, who might achieve in excess of what is nominally required) or simply the types of institutions and courses favoured by students with a certain level of attainment. For example, courses in 'Medicine and Dentistry', predominantly offered at Russell Group universities, have high entry requirements and a high demand compared to the number of places available. As a result, admissions are competitive, leading to a relatively high proportion of non-placed applicants (Wilde & Hoelscher, 2007) and the average entry tariff associated with some institutions is far in excess of the typical offers quoted by these institutions.

Recent policy developments

This research is a snapshot of the distribution of prior qualifications in HE in one academic year, 2011/12. During recent years there have been many changes in education and assessment, particularly relating to Level 3 qualifications and university admissions policies, which could potentially affect some of our findings. Three in particular stand out.

Firstly, new qualifications have been introduced at Level 3 that aim to prepare learners for study at university (e.g., Cambridge Pre-U or the Extended Project), qualifications have been withdrawn (e.g., final awarding of diplomas took place in 2014) and other qualifications are being comprehensively reformed (e.g., Advanced Subsidiary level [AS level] and A levels). The uptake of these qualifications will probably fluctuate and therefore patterns of entry at university of undergraduates holding them could also vary in the next few years. In particular, the influence of subject attainment may differ for other qualifications such as the Cambridge Pre-U, and may change under the reformed A levels.

Secondly, the Government has recently relaxed controls on student numbers (which determine the number of funded university places). If institutions expand and adapt to attract more high-achieving students, this may affect the relationship between prior attainment and HE institution group.

Thirdly, from the academic year 2012/13, students attending universities in the UK have been charged new, higher university tuition fees of up to £9,000. While initial evidence (Higher Education Funding Council for England [HEFCE], 2013) suggests that this has not directly deterred applicants from disadvantaged areas, as some had feared, there are indications that students are more likely to choose courses in Science and Medicine, and less likely to choose Arts courses.

Within these considerations, however, the findings presented here can be beneficial for a better understanding of how the choice of studying in a particular field at a specific type of HE institution can be influenced by the subjects studied at A level and the level of attainment achieved. A thorough analysis of the pathways towards university participation is, in fact, a necessary step to ensure that policies and practices of educational organisations involved in the HE admission system are designed to guarantee equality of opportunities to students pursuing university studies.

References

Bramley, T. (2014). Multivariate representations of subject difficulty. *Research Matters: A Cambridge Assessment Publication*, *18*, 42–48.

Candy, S. (2013). *(Un)informed choices? University admissions practices and social mobility*. London: Pearson.

Chevalier, A. (2011). Subject choice and earnings of UK graduates. *Economics of Education Review*, *30*(6), 1187–1201.

Chevalier, A. & Conlon, G. (2003). *Does it pay to attend a prestigious university?* London: London School of Economics and Political Science.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. & Vignoles, A. (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(2), 431–457.

Crawford, C. (2014). *The link between secondary school characteristics and university participation and outcomes*. Report DFE-RR353. London: Department for Education.

Fazackerley, A. & Chant J. (2008). *The hard truth about 'soft' subjects*. London: Policy Exchange.

Gayle, V., Berridge, D. & Davies, R. (2002). Young people's entry into higher education: quantifying influential factors. *Oxford Review of Education*, *28*(1), 5–20.

Gill, T. (2012). *Uptake of GCE A level subjects 2011*. Statistics Report Series No. 42. Cambridge, UK: Cambridge Assessment.

Goldstein, H. (2011). *Multilevel Statistical Models (4th edition)*. Chichester: John Wiley & Sons.

Greenwood, C., Harrison, M. & Vignoles, A. (2011). *The labour market value of STEM qualifications and occupations*. London: Institute of Education.

HEFCE (2013). *Higher education in England: impact of the 2012 reforms*. 2013/03. Bristol: Higher Education Funding Council for England.

Hoelscher, M., Hayward, G., Ertl, H. & Dunbar-Goddet, H. (2008). The transition from vocational education and training to higher education: a successful pathway? *Research Papers in Education*, *23*(2): 139–151.

Hussain, I., McNally, S. & Telhaj, S. (2009). *University Quality and Graduate Wages in the UK*. IZA Discussion Paper n. 4043. Bonn: Institute for the Study of Labour.

Maringe, F. (2006). University and course choice. Implications for positioning, recruitment and marketing. *International Journal of Educational Management*, *20*(6): 466-479.

Purcell, K., Elias, P., Ellison, R., Atfield, G., Adam, D. & Livanos, I. (2008). *Applying for Higher Education: the diversity of career choices, plans and expectations*. Manchester: Higher Education Careers Services Unit.

Russell Group (2011). *Informed Choices: The Russell Group's guide to post-16 subject choices*. Retrieved from http://www.russellgroup.ac.uk/informed-choices.aspx.

Scott, P. (2013, March 4). University mission groups: what are they good for? *The Guardian*. Retrieved from http://www.theguardian.com/education/2013/mar/04/university-mission-groups-comment

Sutton Trust (2009). *Applications, Offers and Admissions to Research Led Universities*. BIS Research Report no. 5. London: Department for Business, Innovation and Skills.

Vidal Rodeiro, C., Sutch, T. & Zanini, N. (2015). Progression to Higher Education in the UK: The effect of prior learning on institution and field of study. *Research Matters: A Cambridge Assessment Publication*, 20, 13–21.

Walker, I. & Zhu, Y. (2013). *The impact of university degrees on the lifecycle of earnings: some further analyses*. London: Department for Business, Innovation and Skills.

Wilde, S. & Hoelscher, M. (2007). *Missed opportunities: non-placed applicants (NPAs) in the UCAS data*. Cheltenham: Universities and Colleges Admissions Service.

The Mathematics needs of prospective Architecture undergraduates

Ellie Darlington and **Jessica Bowyer** Research Division

Background to the study

The General Certificate of Education (GCE) Advanced level (A level) qualifications in Mathematics and Further Mathematics are being reformed for first teaching in England in 2017. All A levels are moving from a modular to a linear system, requiring students to take their examinations at the end of a two-year course, rather than throughout as is currently the case. Furthermore, the Office of Qualifications and Examinations Regulation (Ofqual), the regulator of qualifications in England, and the Department for Education (DfE) have introduced 100 per cent prescribed content for A level Mathematics, and 50 per cent prescribed content for A level Further Mathematics. Although this will help reduce the variability in students' mathematical backgrounds when entering university, the Applied Mathematics content (Statistics, Mechanics or Decision Mathematics) that students are able to study will inevitably be reduced.

These two qualifications prepare students for undergraduate study in a wide range of subjects, including Science and Social Science in addition to tertiary Mathematics. Consequently, the reforms will have implications for a large number of prospective students' readiness for undergraduate study. This article reports on current undergraduate architects' perceptions of the existing A levels as preparation for undergraduate Architecture, including the Applied Mathematics content they perceived to be most useful. Architecture was a chosen subject of focus as it is a field of study which requires some mathematical understanding, yet there is no existing research on Architecture undergraduates' mathematical abilities in the United Kingdom (UK) context.

Undergraduate Architecture

Undergraduate Architecture degrees are traditionally the first step in the process of becoming a professional architect. Some undergraduate courses form part of the formal training process as they are accredited by the Architects Registration Board (ARB), which ensures that they comply with particular skills requirements. Although not all undergraduate courses entitled 'Architecture' are officially accredited, there is nonetheless a direct link between the skills required in the Architecture profession and the content of Architecture degrees.

No university in the UK currently requires a post-compulsory Mathematics qualification for admission to study Architecture. The Universities of Bath and Cambridge both recommend A level Mathematics, whilst other universities set a minimum entry requirement of a grade C or above at GCSE[1].

Nonetheless, a relatively high proportion of students entering Architecture, Building and Planning[2] courses have taken at least A level Mathematics. In 2010, 42.8% of first year students in these disciplines had studied Mathematics at A level, and 3.9% had studied Further Mathematics (Vidal Rodeiro & Sutch, 2013). Consequently, whilst mathematical entry requirements may be limited, the high percentage of Architecture students with a post-compulsory Mathematics qualification suggests that students may perceive further study in Mathematics to be either relevant or helpful to their undergraduate aspirations.

Many universities require prospective Architecture students to have an Art qualification or to present a portfolio with their application. Perhaps reflecting this, the most popular A level subject amongst new Architecture, Building and Planning undergraduates was Art and Design (45.8%), followed by Mathematics and Physics (20.8%) (Vidal Rodeiro & Sutch, 2013).

Mathematics in Architecture

The disciplines of Architecture and Mathematics are considered to have a close relationship, predominantly because of the importance of geometry in architectural design. Traditionally there has been a focus on Euclidean geometry, although the rise of modernism in Architecture has led to a recent interest in newer topics such as fractal and topological geometry (Cikis, 2010; Megahed, 2013; Salingaros, 1999).

The Mathematics content in university Architecture courses can be broadly classified into three areas:

1. General Mathematics, based on calculus and algebra;
2. Applied Mathematics, predominantly related to building construction; and
3. Design-orientated Mathematics, including areas such as geometry and proportion.

The emphasis on Applied Mathematics, in particular the Mathematics needed in Building Design and Construction, is reflected in the incorporation of Architecture courses into Engineering faculties in countries such as Egypt. Cikis (2010) reviewed the mathematical content of Architecture courses in the United States of America (USA) and Europe and found that the most frequently occurring topics were calculus, descriptive geometry, geometry and analytical geometry, Applied Mathematics, and trigonometry (in decreasing order of frequency).

1. Correct at December 2015.
2. Architecture is grouped with Building and Planning by the Universities and Colleges Admissions Service (UCAS) when they supply applications data.

Additionally, the growing use of computer-aided design software (CAD) in architectural design has led some authors to argue that Architecture students should be aware of the mathematical principles behind the software they are using. In particular, an understanding of algorithms as well as parametric and linear modelling would be potentially useful skills for undergraduates (Cikis, 2010; Freiberger, 2007; Megahed, 2013).

There is, however, some indication that undergraduate architects may not necessarily be able to apply the mathematical content of their degrees directly to architectural design. Verner and Maor (2003) tested Architecture students in Israel and found that, whilst the students appreciated the relevance of Mathematics to design, they performed badly on a test which was designed to assess core mathematical concepts related to Architecture. Consequently, they designed a new Mathematics programme that focused on a problem-solving approach, which directly applied the mathematical concepts they were learning to architectural problems. Students who took this course performed better when re-tested than a control group who had not been taught the problem-solving approach. The importance and relative use of problem-solving in relating Mathematics to Architecture has been reiterated elsewhere, with Javier and Cepeda (2005) implementing a similar programme in an Architecture courses in Mexico. They found that students were more likely to engage with the Mathematics content of their courses when it was directly applied to architectural design.

Furthermore, Cikis (2010) raises a concern that, despite the historical relationship between Mathematics and Architecture, there is a disconnect between the Mathematics that students are required to study as an undergraduate and the Mathematics the undergraduates will eventually use in their professional careers. He argues:

> *The knowledge of Mathematics required by an ordinary architect to carry out his/her profession is at quite a simple level and, unless a very special situation arises, an architect can carry out all sorts of professional duties without resorting to any higher mathematical knowledge.* (p.106)

This is corroborated by the benchmark statement for Architecture (Quality Assurance Agency for Higher Education, 2010) and the ARB criteria for graduate architects, which make no reference to mathematical skills, apart from references to strategies for building construction and the ability to "critically examine the financial factors in varying building types, construction systems, and specification choices" (ARB, 2010, p.6).

The lack of an overt presence of Mathematics in professional Architecture may mean that potential students do not recognise the importance of Mathematics to either their undergraduate studies or their future career. However, there has not, as yet, been any research conducted assessing the mathematical preparedness of undergraduates for Architecture courses in the UK.

Content and structure of A level Mathematics and Further Mathematics

This article reports on findings from a large-scale project investigating current undergraduates' perceptions of existing A levels in Mathematics and Further Mathematics. This research was conducted in response to the forthcoming changes to these A levels from 2017 (DfE, 2013). As discussed previously, these reforms will entail significant changes to both the content and structure of these qualifications. This research thus investigated undergraduates' perspectives of the current A levels in order to inform development of the new specifications, as well as to consider the implications of the reforms for universities and prospective students. The current structures of A level Mathematics and Further Mathematics will now be discussed in more depth.

AS and A level Mathematics

At present, A level Mathematics comprises four compulsory Core Pure Mathematics units of equal weighting, with two Applied Mathematics units. These units may be chosen from one of three different strands:

1. Mechanics;
2. Statistics; and
3. Decision Mathematics.

Within each of these strands there are between two and five sequential units, depending on the strand and awarding body. The more advanced Applied Mathematics units (e.g., Mechanics 3 and above) can only be studied in AS or A level Further Mathematics.

Students are able to take either two units from the same strand (e.g., Mechanics 1 and Mechanics 2) or one from two different strands (e.g., Statistics 1 and Decision Mathematics 1). Hence, there are six[3] possible routes through A level Mathematics.

At AS level, students must take two compulsory Core Pure Mathematics units and one applied unit (Mechanics 1, Statistics 1 or Decision Mathematics 1).

It is not necessarily the case that students will be able to take the units that they want to. Restrictions on resources and timetabling within schools and colleges may mean that students are given a restricted choice, if at all.

AS and A level Further Mathematics

A level Further Mathematics comprises two compulsory Further Pure Mathematics units, plus four optional units. At AS level, students must take Further Pure Mathematics 1 and two optional units.

The optional units can be selected from any of the three standard strands offered within A level Mathematics[4] (Mechanics, Statistics and Decision Mathematics) or from an additional two Further Pure Mathematics units. There are therefore a large number of different routes through Further Mathematics.

Method and analysis

All universities which offer Architecture degrees were sent emails, using the contact details on departmental websites, requesting participation. Departments were asked to pass on the details of an online questionnaire aimed at students who fulfilled two criteria, namely:

- They must have been in their second year of study or above, in order that they could reflect on their experiences so far;
- They must have taken at least AS level Mathematics since 2006, when the qualification underwent its most recent restructuring.

Those who took International A levels were not permitted to take part.

3. These are: (1) M1 + M2; (2) S1 + S2; (3) D1 + D2; (4) M1 + S1; (5) M1 + D1; (6) S1 + D1.
4. Students are not allowed to take units as part of AS or A level Further Mathematics that they have already taken as part of AS or A level Mathematics.

The questionnaire surveyed students regarding:

- their **mathematical background** (e.g., highest Mathematics qualification, grades achieved, awarding body of Mathematics and/or Further Mathematics qualifications, units studied as part of AS or A level Mathematics and/or Further Mathematics);
- their **current studies** (e.g., university, degree type and title, year of study);
- their **perception of the A level(s)** as preparation for the mathematical component of their degree, both overall and by optional units;
- the **factors which motivated them to take Further Mathematics** (if applicable); and
- their **experience of Further Mathematics** (if applicable).

The questionnaire comprised a mixture of multiple choice questions, closed questions and open-ended questions. It was developed by the authors and an A level Mathematics expert, before being piloted by three recent graduates of mathematically-demanding degrees. Small changes were made in response to the piloting. The questionnaire was made available in an online format, and was open for responses between September and December 2014.

Analysis of quantitative data was conducted using SPSS (a software package used for statistical analysis), and qualitative responses to open-ended questions regarding the qualification(s) as preparation for undergraduate Architecture were coded using MAXQDA (a software package for qualitative and mixed methods data analysis). Thematic analysis was used in order to analyse and later describe participants' views of the qualification(s) and any suggestions they had for how they could have better suited their needs.

Data

Data were collected between October and December 2014, during the first term of the academic year.

Sample

After incomplete and inappropriate responses were removed as part of the data cleaning process, 37 students studying undergraduate degrees in Architecture had completed the questionnaire.

- **University:** Participants came from 7 different universities, all of which were ranked in the top 75 per cent of the 51 Architecture departments listed in the *Complete University Guide* (2015). All participants were studying for courses accredited by the ARB.
- **Year:** Half of the participants were in their second year of study, and the remainder in their third year.
- **Degree programme:** All students were studying for single honours degrees entitled 'Architecture'. Three were studying for (four-year) undergraduate Master's degrees, with the rest for (three-year) Bachelor's degrees.

A level results

All participants had studied more post-compulsory Mathematics than is required by any UK university for admission to Architecture. Most had studied a full A level in Mathematics, with over one-third (15 students) having also studied either AS or A level Further Mathematics (see Table 1).

Table 1: Participants' highest Mathematics qualification

Qualification	*No. participants*	*%*
AS level Mathematics	1	2.7
A level Mathematics	21	56.8
AS level Further Mathematics	6	16.2
A level Further Mathematics	9	24.3
Total	**37**	**100.0**

The majority of participants (58.3%) who took A level Mathematics achieved an A*. The rest achieved a grade A (19.4%), a grade C (16.7%) or a grade E (5.6%). In 2011, only 13.3% of Architecture undergraduates who had studied A level Mathematics achieved an A* (Vidal Rodeiro, 2012). The sample here is therefore skewed to the higher end of achievement compared to both all A level Mathematics candidates and undergraduate Architecture students.

All of the 15 participants who had taken at least AS level Further Mathematics achieved at least a grade A. This is obviously higher than the national figures, wherein 56.3% achieved a grade A or A* above in A level Further Mathematics (Joint Council for Qualifications [JCQ], 2015), further indicating that the students in the sample were particularly high-achieving students at A level.

A level units

Similar proportions of participants studied Mechanics to those who studied Statistics units as part of their Mathematics qualifications (see Figure 1). It was rare for participants to have taken more than two applied units in the same strand, although this is skewed by the fact that the majority of participants had not studied Further Mathematics. Decision Mathematics was the least commonly-taken optional applied unit amongst the participants.

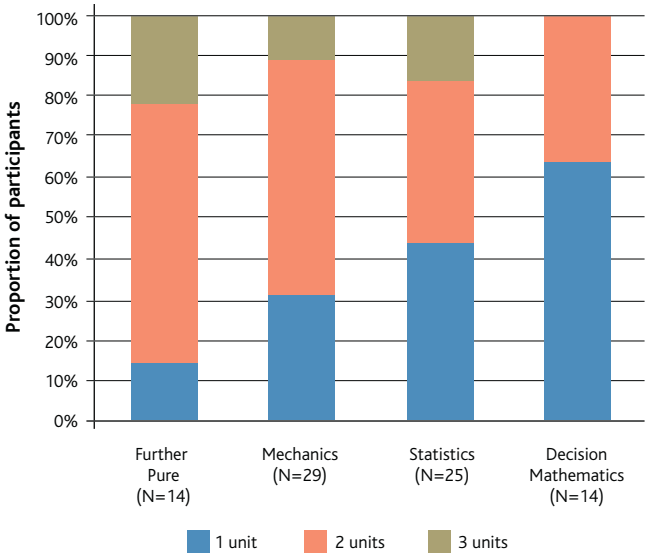


Figure 1: Number of optional units studied by Architecture students

It should be noted that the number of participants who studied Further Pure Mathematics units is low compared to the other strands because these units are only available as part of AS or A level Further Mathematics.

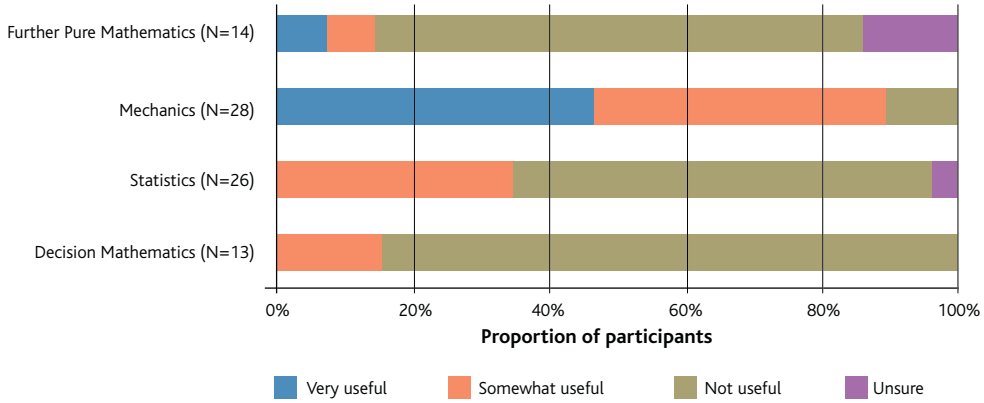


Figure 2: Architecture students' views of the utility of optional units

Experiences of non-compulsory A level units

The applied strand which participants considered to be the most useful preparation for Architecture degrees were Mechanics (see Figure 2). Nearly half (46.4%) of the participants described Mechanics as 'very useful', and only three participants declared it 'not useful'. No participants who had taken Decision Mathematics found it to be useful preparation for their degree, and Statistics was described as 'somewhat useful' (34.6% of participants) at best. For those students who had taken AS or A level Further Mathematics, Further Pure units were also less well-received, with only two participants describing it as either 'very' or 'somewhat useful'.

A level as preparation for the mathematical component of Architecture

A level Mathematics was described as good preparation for Architecture degrees by 75.7% of participants, and only one participant reported that it was bad preparation (see Figure 3). However, students were less positive about Further Mathematics. Of those who had studied Further Mathematics, less than half reported that it was good preparation for their degree.

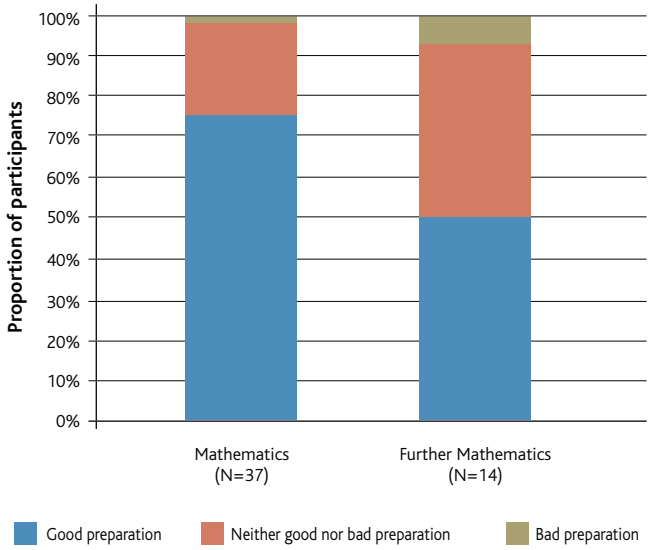


Figure 3: The A levels as preparation for undergraduate Architecture

This suggests that, though few students regarded Further Mathematics as bad preparation, Further Mathematics was perceived by the participants who took it to have had less additional benefit to A level Mathematics compared to studying other subjects. However, the low number of responses to this question, should be taken into account (*N*=14).

Improvements to A level Mathematics and Further Mathematics

Participants were also asked two open response questions:

1. Were there any topics that they would have found useful to have been included in A level Mathematics and Further Mathematics?
2. Could any improvements be made to these A levels in order to better prepare students for the mathematical components of Architecture degrees?

There were sixteen responses to the first question. The majority of students reported that there were no additional topics that would have been useful. They stated that A level study had prepared them well, particularly the Mechanics units. Indeed, a minority of participants indicated that they had not encountered any Mathematics in their degree which was more difficult than at A level. Suggestions for additional topics focused on increasing the proportion of Mechanics and structural Mathematics available to students and making this area compulsory. Nevertheless, two participants acknowledged that they had been unable to study Mechanics units because of restrictions at their school.

Twenty-four students commented on whether the A levels in Mathematics and Further Mathematics could provide better preparation for their undergraduate courses. The responses were very similar to those to the previous questions. Three students reported that their degree actually required very little Mathematics knowledge, and therefore A level Mathematics was sufficient preparation. Additionally, one participant who had studied A level Further Mathematics reported that it was enjoyable but largely irrelevant to their course.

Many reported that they would have preferred more Mechanics content at A level, as most students considered this strand to be the most relevant content for undergraduate Architecture and therefore the most useful preparation. The majority of students who offered suggestions for improvements stated that more practical applications and problem-solving at A level would be useful, as well as the use of real-world examples.

Discussion and conclusion

The data collected indicate that Mechanics units were considered by Architecture undergraduates to be the most useful optional units as preparation for the mathematical content of their university studies. The apparent utility of Mechanics correlates with the literature, due to the need for Architecture undergraduates to consider the Mathematics

involved in Building Construction. This type of Mathematics requires an awareness of forces and kinematics, as well as the mathematical modelling of real-life problems which has its foundations in Mechanics.

Conversely, participants were considerably more negative about the utility of Decision Mathematics and Statistics units. No student reported that they found either unit to be good preparation for their degree. The negative perception of the Statistics units may be unsurprising given that Statistics is not considered to have an essential role in architectural Mathematics. However, an awareness of probability and Statistics may be useful when considering factors such as construction programmes and building costs (Megahed, 2013).

Students' negative perceptions of Decision Mathematics are especially relevant when considering the types of Mathematics that Architecture students will encounter during their degrees. The use of CAD software is becoming more widespread in undergraduate Architecture courses, and it may therefore be useful for Architecture students to understand the Mathematics behind this. Consequently, algorithmic thinking and linear programming have thus been identified as potentially useful mathematical skills for undergraduate architects (Freiberger, 2007; Megahed, 2013). Both of these areas currently are taught in the Decision Mathematics strand at A level. However, the perceived negative utility of Decision Mathematics by Architecture students may indicate that Architecture students struggle to see the relevance of these topics and apply these skills during their undergraduate studies. Additionally, Architecture students were least likely to take a Decision Mathematics unit than either Mechanics or Statistics, meaning that their exposure to these areas of Mathematics may be low. The extent to which A level reform will mitigate these concerns is currently unclear. There will be no Decision Mathematics in A level Mathematics, as a result of the recommendation of the A Level Content and Advisory Board (ALCAB) that these units should be removed from the reformed A levels. This recommendation was made based on universities' perceptions of existing Decision Mathematics units as irrelevant to undergraduate mathematical study and as 'soft' options (ALCAB, 2014, p.8). Nevertheless, the awarding bodies may opt to include some Decision Mathematics content in Further Mathematics. This means that in future, it will be very unlikely that potential architects will have any experience with areas such as linear programming before beginning university study.

Additionally, the data indicate that, whilst A level Mathematics is a useful subject for undergraduate Architects to have studied, there appears to be more limited use in having studied Further Mathematics. Fewer than half of participants who had studied either AS or A level Further Mathematics agreed that it had been good preparation for the mathematical content of their degree. The perceived lack of utility in Further Pure units may also be surprising, considering that Cikis (2010) found that the majority of the Mathematics components in Architecture courses in the USA and Europe focused on calculus. Calculus, principally more complex differentiation and integration, forms a significant part of Further Mathematics qualifications. It may therefore be expected that taking Further Mathematics would be useful preparation for prospective architects. However, geometry is also a key mathematical concept in Architecture courses and there is very little geometry in Further Mathematics. Therefore, it is likely that a large proportion of the content in Further Mathematics is irrelevant to the mathematical content in undergraduate Architecture. The subject criteria for the reformed A levels in both subjects indicate that this is unlikely to change in the future, as there is no geometry in the prescribed content for the reformed A level Further Mathematics (DfE, 2014).

Consequently, it is essential that prospective architects are given appropriate guidance about the post-compulsory Mathematics qualifications they should take. Whilst very few universities currently stipulate Mathematics A level as an entry requirement for Architecture, the data collected in this research indicate that it would be a useful qualification for students considering a degree in Architecture to have. Unfortunately, there is little literature regarding how and when students usually decide which subject to study at university in relation to when they choose their A level subjects. The decision could have been made prior to A level choices, and thus students may make an appropriate decision to study A level Mathematics and/or Further Mathematics. However, if students choose to study Architecture once they have already begun studying their A levels, then they may not be able to choose an appropriate post-compulsory Mathematics qualification. For the most part, students do not change their minds regarding what subject they wish to study at university after age 16 (Sutch, Zanini & Benton, 2015). Therefore, it is important that students are given appropriate advice regarding the most appropriate A levels to take as preparation for certain degree courses when making their subject decisions. It is not necessarily the case that the only useful A levels are those subjects that are explicitly required by university entry requirements. The introduction of the new Core Mathematics qualifications also warrants mention here, as this may be a useful qualification for prospective Architecture students who wish to continue studying Mathematics in Further Education but do not wish to study the A level.

For those students who do choose to study A level Mathematics or Further Mathematics, guidance on what optional units would be helpful mathematical preparation is very important. Prospective Architecture undergraduates would benefit from studying Mechanics content. Whilst Mechanics units are currently optional, the reformed A level Mathematics will have prescribed Mechanics content, meaning that all future Architecture students taking this A level will study Mechanics prior to beginning their university studies. Additionally, there currently appears to be limited use in taking Further Mathematics in its current form as preparation for the mathematical content in undergraduate Architecture. However, after the reforms, it will only be possible for A level students to specialise in Mechanics by taking Further Mathematics. Hence, admissions tutors, students and teachers should be made aware that Further Mathematics may serve a stronger purpose as preparation for the mathematical component of Architecture in the future.

References

ALCAB. (2014). *Report of the ALCAB panel on Mathematics and Further Mathematics*. Retrieved from https://alevelcontent.files.wordpress.com/2014/07/alcab-report-on-mathematics-and-further-mathematics-july-2014.pdf

Architects Registration Board. (2010). *Prescription of Qualifications: ARB Criteria at Parts 1, 2 and 3*. London: ARB.

Cikis, S. (2010). A Critical Evaluation of Mathematics Courses in Architectural Education and Practice. *International Journal of Technology and Design Education*, *20*(1), 95–107.

Complete University Guide. (2015). *University Subject Tables 2015: Architecture*. Retrieved from http://www.thecompleteuniversityguide.co.uk/league-tables/rankings?s=Architecture&y=2015

DfE. (2013). *Reformed GCSE Subject Content Consultation: Government Response*. London: DfE.

DfE. (2014). *Further Mathematics: GCE AS & A Level Subject Content*. London: DfE.

Freiberger, M. (2007). Perfect buildings: the maths of modern architecture. *Plus Magazine*. Retrieved from https://plus.maths.org/content/perfect-buildings-maths-modern-architecture

Javier, F., & Cepeda, D. (2005). Designing a Problem-Based Learning Course of Mathematics for Architects. *Nexus Network Journal*, *7*(1), 42–47.

Joint Council for Qualifications. (2015). A-Level Results. Retrieved from http://www.jcq.org.uk/examination-results/a-levels/2014/a-as-and-aea-results-summer-2014

Megahed, N. A. (2013). Towards Math-Based Architectural Education in Egyptian Engineering Faculties. *Nexus Network Journal*, *15*(3), 565–581.

Quality Assurance Agency for Higher Education. (2010). *Subject Benchmark Statement: Architecture*. Gloucester: QAA.

Salingaros, N. A. (1999). Architecture, Pattern, and Mathematics. *Nexus Network Journal*, *1*, 75–85.

Sutch, T., & Zanini, N. (2015). A Level reforms and subject choice. *Higher Education Bulletin*. Retrieved from http://www.ocr.org.uk/Images/233241-higher-education-bulletin-summer-2015.pdf

Verner, I. M., & Maor, S. (2003). The Effect of Integrating Design Problems on Learning Mathematics in an Architecture College. *Nexus Network Journal*, *5*(2), 103–115.

Vidal Rodeiro, C. L. (2012). *Progression from A level Mathematics to Higher Education*. Cambridge, UK: Cambridge Assessment.

Vidal Rodeiro, C. L., & Sutch, T. (2013). *Popularity of A level Subjects among UK university students: Statistical Report Series No.52*: Cambridge, UK: Cambridge Assessment.

Assessing the equivalencies of the UCAS tariff for different qualifications

Tim Gill Research Division

Introduction

In the United Kingdom (UK) the Universities and College Admissions Service (UCAS) provides the application process for most universities. The UCAS tariff points system is used by universities to help them select students for their courses. Each grade in eligible qualifications is allocated a points score, which can then be summed in order to provide an overall points score for each student. The allocation of points is such that, in theory, students with the same overall points score gained from different qualifications can be considered to be of equivalent ability or potential. The purpose of this article is to test whether this assumption works in practice, by calculating empirical equivalencies of the UCAS tariff for different qualifications.

In the past, UCAS has undertaken studies to try and determine what the tariff points scores should be for different grades achieved in any new qualifications to be considered under the tariff (e.g., UCAS, 2003; 2006). These included comparability studies carried out by an 'expert group', which compared the new qualification with a similar, benchmark qualification and provided recommendations for the number of tariff points allocated to each grade on the qualification. For example, the BTEC Nationals were first included in the tariff tables in 2003 following a comparability study with AQA's Advanced Certificates of Vocational Education (UCAS, 2003).

In these UCAS reports it is noted that a future review of the tariff points allocated to the qualification might be necessary once more evidence becomes available and once Higher Education (HE) admissions tutors have more experience in using the qualification to admit students. However, it is not clear how often this actually happens for individual qualifications. One study that did attempt to address this issue was undertaken by Green and Vignoles (2012). They used the future performance of students at university to make an empirical comparison between the tariff points allocated to A levels and the International Baccalaureate (IB) qualification. The present article seeks to update and extend their work by using more recent data and by also including BTEC qualifications in the comparisons.

One way of investigating the equivalence of tariff points for different qualifications is to compare the outcomes in terms of degree classification for students with the same UCAS tariff obtained from different qualifications. For example, Figure 1 shows the percentages of students achieving a First-class degree or at least an Upper Second-class degree, by their UCAS tariff score (tariff scores where fewer than 30 students achieved that score were excluded). Different lines are presented for students taking different qualifications (General Certificate of Education [GCE] Advanced levels [A levels] only, BTECs only, IB only or mixed).

This would seem to suggest that the current tariffs over-value BTECs and the IB compared with A levels, as the percentage of students achieving a First or at least an Upper Second is higher for A level students at any given UCAS tariff (except for IB students at the very top). However, this analysis does not take into account other factors that might have an influence on the probability of a good degree for a given UCAS tariff. These include the school and university attended and the degree subject, as well as student background characteristics such as gender and socio-economic background.

Data

The data for this research came from a linked dataset requested from the Department for Education. This request consisted of data from the National Pupil Database (NPD) and from the Higher Education Statistics Agency (HESA) student records' database, linked by a common student identifier. The data included information on:

- Degree studied by each student:
  - Institution identifier (anonymous)
  - Subject of study – subjects were classified into one of twenty different subject groups
  - Degree classification – First, Upper Second, Lower Second, Third (or below).

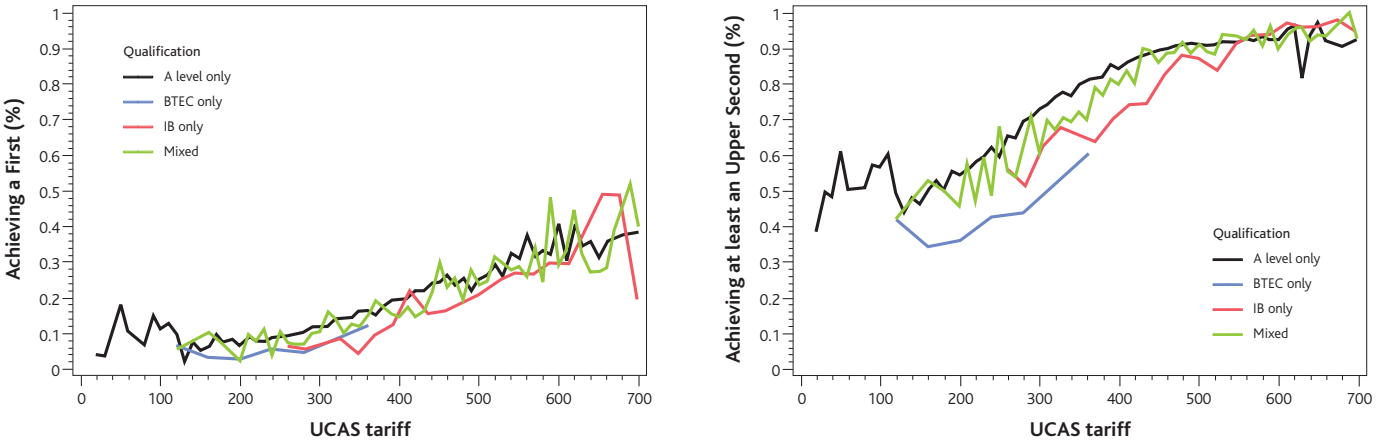


Figure 1: Percentage of students achieving a First- or Upper Second-class degree by UCAS tariff and prior qualification

- Students' prior attainment:
  - Qualifications taken at Level 3
  - Grades achieved in these qualifications
  - School type
- Students' background information:
  - Gender
  - Socio-economic classification, as determined by parents' occupation[1].

For the analysis undertaken in this article the data was limited to students on full-time, part-time or sandwich courses, achieving an honours degree. Two separate linkages of the HESA and NPD data were undertaken in order to capture two groups of students: those completing their degree three years after leaving school; and those completing their degree four years after leaving school. This second group of students included those taking four-year degrees and those taking a year out before university study. Thus the data consisted of students completing a first degree in 2012 or 2013, who were matched to prior attainment data from 2009.

The data from the two different groups was combined. Therefore, it was necessary to make the assumption that the standards in terms of degree class were the same in the two different years (2012 and 2013).

Method

This article investigates the accuracy of the equivalent UCAS tariff points allocated to different qualifications. The method used was to compare the performance at university of students with the same UCAS tariffs achieved through different qualifications. Initial investigation into the data revealed a very large number of different combinations of qualifications taken by students. In order to be able to make simple comparisons between different qualifications, it was decided to restrict the data to students taking one qualification only; that is, to exclude any students taking combinations of qualifications. The only exceptions to this rule were for students taking a combination of A levels and Advanced Subsidiary (AS) levels and students taking combinations of BTEC National qualifications (e.g., a BTEC National Certificate combined with a BTEC National Award). Following this restriction there remained only three qualifications with enough students (with a degree result) to allow for valid comparisons to be made. These were A levels (and AS levels), International Baccalaureate (IB) and BTECs.

The IB is an international qualification, which is studied in many different countries and recognised by universities world-wide. To achieve an IB Diploma, students generally study six different subjects, three at Higher level and three at Standard level. For each of these they receive a grade from 0 to 7. They are also required to write an extended essay and undertake a course in the theory of knowledge, for which they receive a combined grade of between 0 and 3. This gives a maximum Diploma score of 45 points, with 24 required to be awarded a Diploma. The UCAS Expert Group first met in 2004 to benchmark three IB subjects to the equivalent A levels (UCAS, 2006). From this comparison, each IB Diploma points score was allocated a UCAS tariff points score. A review was then undertaken in 2008, which led to a revised tariff from 2010 onwards (the tariff was reduced at each IB points score). With this new tariff an IB points score of 45 was allocated 720 UCAS points (equivalent to six A grades at A level). The comparisons made in this research were with the revised tariff.

The BTEC qualifications that were included in this analysis were BTEC Nationals at Award, Certificate and Diploma level. These are worth one, two and three A levels respectively. Students taking the Diploma tended not to take any other qualifications, whilst those taking an Award or Certificate tended to combine it with other qualifications (usually other BTECs or A levels). The highest grade for a Diploma (DDD) was allocated 360 UCAS points (equivalent to three A grades at A level).

Restricting to these qualifications meant that four different comparisons were made:

1. A levels only *v* IB
2. A levels and AS levels *v* IB
3. A levels only *v* BTEC Diploma only
4. A levels and AS levels *v* BTECs.

It was decided not to make comparisons between IB and BTECs because they tend to be taken by students at the opposite ends of the ability scale who generally study different subjects at university (Vidal Rodeiro, Sutch & Zanini, 2013). Therefore it seems unlikely that many admissions tutors will be comparing students with these qualifications.

1. Unless the student was aged 21 or over at the start of the course, in which case the classification refers to the student's occupation.

The method for making comparisons between the different qualifications was based on the method proposed by Green and Vignoles (2012). The basic idea was to use a statistical model to predict the probability of achieving a particular degree class (or higher), based on UCAS tariff. There should be a positive relationship, whereby a higher UCAS tariff is associated with a higher probability of a good degree. By including in the model an indicator of which qualification was taken, it was possible to determine whether students with a particular qualification had a higher probability of a good degree, for the same UCAS tariff. From the results of this model it was also possible to calculate equivalent tariffs for different qualifications; that is, for a student with a given UCAS tariff achieved through,(e.g., IB) what was the equivalent tariff achieved by a student through (e.g., A levels), such that the probability of achieving a good degree was the same for both students.

We used an ordered probit model, which allows for the rank of degree classification (First, Upper Second, Lower Second, and Third or Pass). The dependent variable was achieving each of these degree classes (or better) and the independent variables were the student's UCAS tariff[2], qualification taken, university attended, degree subject group, gender, socio-economic classification and school type. To account for possible differences in the relationship between UCAS tariff and degree classification between qualifications, an interaction term between qualification and UCAS tariff was also included. Finally, to account for the 'clustering' of students within schools, a multilevel model was used, with students nested within schools.

For each of the four comparisons being made, three separate models were run. The first of these included just the UCAS tariff, an indicator of the qualification taken and the interaction between them as the predictor variables. If the variable indicating the qualification taken had a parameter value that was statistically significantly different from zero then this means that, overall, the same tariff points had different predictions of degree performance depending on which qualification was taken. If such a difference was found between the qualifications then the inclusion of an interaction term allowed this to vary at different tariff levels. A second model included variables for the university attended and the degree subject group. Finally, student background variables (gender, socio-economic classification and school type) were added.

Initial data exploration revealed that in the data set there were some universities with only a handful of students. This meant that including the university as a variable in the model could have led to problems with the models converging. To overcome this issue, only universities where at least 10 IB students and 10 A level students attended were included.

Results

A levels only *v* IB only

Table 1 presents data on the distribution of UCAS tariff amongst students taking these qualifications, after restricting to institutions with at least 10 students from each group. This shows that IB students had a much higher mean UCAS tariff than those taking A levels only. Table 2 shows the distribution of degree class for students taking A levels only or IB. The IB students were more likely to achieve a First or an Upper Second than the A level only students.

2. This was calculated using the tariff tables available on the UCAS website: https://www.ucas.com/ucas/undergraduate/getting-started/entry-requirements/tariff/tariff-tables

Table 1: Distribution of UCAS tariff, by prior qualification

Qualification group	*N*	*Mean*	*SD*	*Min*	*Max*
A levels only	14,355	357.2	118.6	40	720
IB	1,135	504.4	109.8	260	720
All	**15,490**	**351.4**	**112.9**	**40**	**720**

Table 2: Distribution of degree class, by prior qualification

Qualification group	*N*	*First (%)*	*Upper Second (%)*	*Lower Second (%)*	*Third/ Unclassified (%)*
A levels only	14,355	17.3	57.8	21.8	3.1
IB	1,135	22.6	63.4	12.5	1.5
All	**15,490**	**17.7**	**58.2**	**21.2**	**3.0**

The results of the three different models are shown in Table 3. Statistically significant effects are shown in bold (standard errors in brackets). The parameters for the institution and subject group variables (included in Models 2 and 3) and student background variables (Model 3) are not included in this table because they are not directly relevant to the question being investigated. However, they are available on request. It is worth noting that all of these variables were statistically significant in the models.

Table 3: Model parameter estimates (A levels only *v* IB only)

Variable		*Model 1*	*Model 2*	*Model 3*
Intercept	First	**-3.828** (0.064)	**-4.627** (0.286)	**-4.948** (0.906)
	At least Upper Second	**-0.851** (0.055)	**-1.588** (0.283)	**-1.877** (0.905)
	At least Lower Second	**1.606** (0.068)	**0.908** (0.286)	0.641 (0.905)
UCAS tariff		**0.006** (0.000)	**0.007** (0.000)	**0.007** (0.000)
Qualification	[A levels only]			
	IB	**-0.224** (0.081)	**-0.313** (0.082)	**-0.281** (0.082)

In each model the variable for the qualification taken was statistically significant, suggesting that there was evidence of a difference in the predictions based on the UCAS points gained from the different qualifications. The qualification parameter estimate was negative for IB students compared to A level students, meaning that having the same tariff gained from IB was associated with a lower probability of achieving each degree class (or better). In other words, for a particular UCAS tariff, A level students were more likely to do better at university. The value of this parameter was higher for Models 2 and 3 than for Model 1. An interaction term between tariff and qualification was also included in each model to see if the effect was different at different tariff levels. However, this term was found to be non-significant.

We can use the parameter values (from the final model) to calculate the probabilities of achieving a First or at least an Upper Second, for students taking A levels or IB. The probabilities presented here were for students in the reference categories for institution (No.137), subject group (Education), gender (females), socio-economic classification (unemployed) and school type (other school). The results are shown in Figure 2.

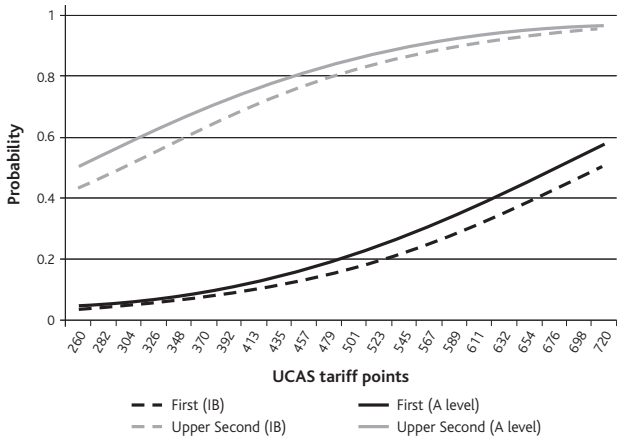


Figure 2: Modelled probabilities of achieving a First or at least an Upper Second-class degree, by UCAS tariff and prior qualification (A level only *v* IB only)

For an IB student with a UCAS tariff of 501 the probability of a First was 0.172, compared with 0.216 for A level students. For an IB student with a tariff of 720 the probability of a First was 0.507, compared with 0.576 for A level students.

The results of the models can also be used to calculate equivalent tariffs between the two qualifications; that is, for each IB tariff score (X), what is the A level tariff score (Y) which gives the same probability of achieving a First? Algebraically speaking we have the following equivalence:

P (First | AL tariff =Y) = P (First | IB tariff =X)

If we can find Y in terms of X then this will give us the A level tariff that is equivalent to each IB tariff. So, using the parameters in Model 3 we have:

-4.948 + 0.007*Y = -4.948 + 0.007*X – 0.281

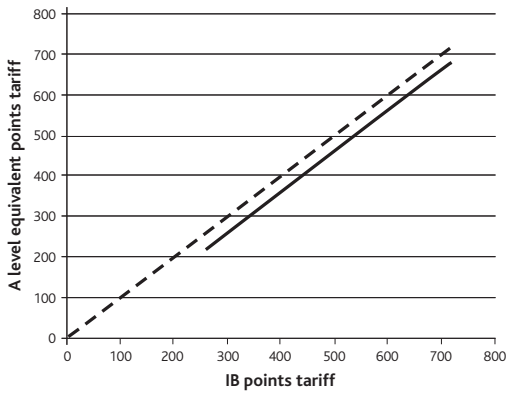
Re-arranging gives:

Y = (0.007*X – 0.281) / 0.007

Or

Equivalent A level tariff = (0.007*IB tariff-0.281)/0.007 =
IB tariff – 38.5

This equation produces the equivalent tariffs shown in Figure 3. The dotted line is the line of perfect equivalence between the two qualifications (y=x), the continuous line is the equivalence according to the model.



IB tariff	Equivalent A level tariff
260	221
282	243
304	265
326	287
348	309
370	331
392	353
413	374
435	396
457	418
479	440
501	462
523	484
545	506
567	528
589	550
611	572
632	593
654	615
676	637
698	659
720	681

Figure 3: Equivalent tariff scores (A level only *v* IB only)

This shows that, for example, the A level equivalent for IB students with 501 UCAS points was 462.

Thus, whichever model we use we find a similar effect, with the UCAS tariff from IB being over-valued slightly at all tariffs. The final model, controlling for institution, subject group and student background variables suggested that this was over-valued by 39 points.

A/AS levels *v* IB

The second set of models compared students taking combinations of A levels and AS levels (including A levels only) with IB students. Tables 4 and 5 present the UCAS tariffs and degree class distributions for students in the two groups. Students taking IB had a much higher UCAS tariff on average and were more likely to achieve a First-class degree.

Table 4: Distribution of UCAS tariff, by prior qualification

Qualification group	*N*	*Mean*	*SD*	*Min*	*Max*
A/AS levels	72,020	387.3	99.6	40	720
IB	1,135	504.4	109.8	260	720
All	**73,155**	**389.1**	**100.8**	**40**	**720**

Table 5: Distribution of degree class, by prior qualification

Qualification group	*N*	*First (%)*	*Upper Second (%)*	*Lower Second (%)*	*Third/ Unclassified (%)*
A/AS levels	72,020	18.5	61.0	18.2	2.3
IB	1,135	22.6	63.4	12.5	1.5
All	**73,155**	**18.6**	**61.0**	**18.1**	**2.3**

The results from the models are shown in Table 6. Again, the interaction between qualification and tariff was not significant.

There was a significant negative effect of having taken IB compared to A levels with AS levels. This effect increased slightly in Models 2 and 3, compared with Model 1. This was a very similar effect to that seen in the models using A levels only, although it was slightly larger here.

Table 6: Model parameter estimates (A/AS levels *v* IB)

Variable		*Model 1*	*Model 2*	*Model 3*
Intercept	First	**-4.211** (0.037)	**-4.675** (0.121)	**-4.928** (0.522)
	At least Upper Second	**-1.103** (0.033)	**-1.518** (0.119)	**-1.751** (0.522)
	At least Lower Second	**1.401** (0.039)	**1.012** (0.121)	**0.792** (0.522)
UCAS tariff		**0.007** (0.000)	**0.008** (0.000)	**0.008** (0.000)
Qualification	[A/AS levels]			
	IB	**-0.334** (0.076)	**-0.410** (0.076)	**-0.387** (0.076)

Figure 4 presents the probabilities for students in the two groups, using the final model parameter values (for a student in the reference categories for institution (No.204), subject group (Agriculture and related subjects), gender (males), socio-economic classification (unemployed) and school type (other school)).

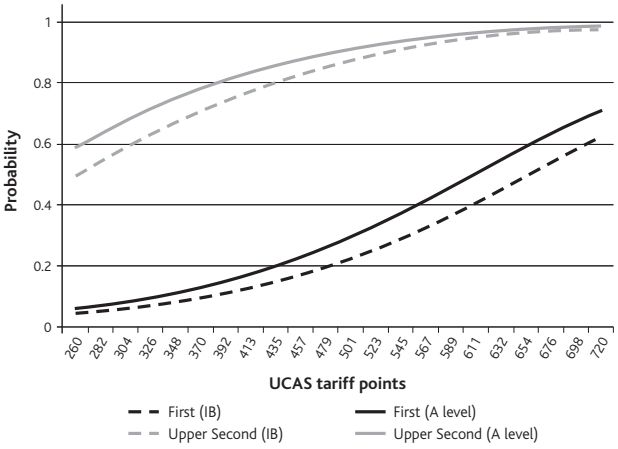


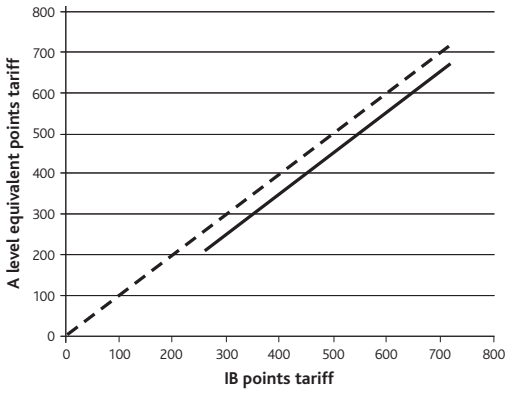
Figure 4: Modelled probabilities of achieving a First or at least an Upper Second-class degree, by UCAS tariff and prior qualification (A/AS level *v* IB)

For example, for an IB student with a tariff of 501 the probability of a First was 0.217, compared with 0.290 for an AS/A level student. The respective probabilities for students with a tariff of 720 were 0.618 and 0.704.

Equivalent tariffs were calculated using the following equivalence:

Equivalent A level tariff = (0.008*IB tariff-0.387)/0.008 =
IB tariff – 48.1

The equivalent tariffs are shown in Figure 5. This shows that, for example, a UCAS tariff of 720 achieved from IB was equivalent to a tariff of 672 from A levels (over-valued by 48 points).



IB tariff	Equivalent A level tariff
260	212
282	234
304	256
326	278
348	300
370	322
392	344
413	365
435	387
457	409
479	431
501	453
523	475
545	497
567	519
589	541
611	563
632	584
654	606
676	628
698	650
720	672

Figure 5: Equivalent tariff scores (A/AS level *v* IB)

A levels only *v* BTEC Diploma only

The next set of models compared students taking A levels only with those taking a BTEC Diploma only. As before, only universities where at least 10 BTEC students and 10 A level students attended were included. A further restriction was added because there were only very few BTEC students with a UCAS tariff above 360 (maximum tariff for a BTEC Diploma). Thus, to make the comparison as valid as possible, A level students with a tariff above 360 were also excluded. Table 7 presents some details of the distribution of UCAS tariff for students with each of the qualifications.

Thus, BTEC students had higher UCAS tariffs on average. The A level students included here had a much lower mean tariff than the group of students who were compared to IB students. This is because this group were restricted to those with a tariff of 360 or below. Table 8 presents the degree outcomes of the two groups. This shows that A level students had a higher probability of a First or an Upper Second-class degree.

Table 7: Distribution of UCAS tariff, by prior qualification

Qualification group	*N*	*Mean*	*SD*	*Min*	*Max*
A levels only	23,270	247.5	63.1	40	360
BTEC	9,770	296.0	69.8	120	360
All	**33,040**	**261.8**	**68.8**	**40**	**360**

Table 8: Distribution of degree class, by prior qualification

Qualification group	*N*	*First (%)*	*Upper Second (%)*	*Lower Second (%)*	*Third/ Unclassified (%)*
A levels only	23,270	11.2	51.4	32.2	5.1
BTEC	9,770	8.2	40.3	41.2	10.2
All	**33,040**	**10.3**	**48.1**	**34.9**	**6.6**

The results of the models are presented in Table 9. In each model there was a significant negative effect of having taken BTECs, meaning that the probability of each degree class was lower for students with the same UCAS tariff from BTECs than from A levels. This effect was smaller in Model 2 than in Model 1, but larger in Model 3. This time the interaction between UCAS tariff and qualification was also significant and negative, meaning that the BTEC effect was larger at higher tariffs.

Table 9: Model parameter estimates (A levels only *v* BTEC Diploma only)

Variable		*Model 1*	*Model 2*	*Model 3*
Intercept	First	**-4.193** (0.059)	**-4.658** (0.261)	**-4.836** (0.290)
	At least Upper Second	**-1.512** (0.054)	**-1.194** (0.260)	**-2.081** (0.288)
	At least Lower Second	**0.944** (0.055)	**0.589** (0.260)	0.440 (0.289
UCAS tariff		**0.008** (0.000)	**0.009** (0.000)	**0.010** (0.000)
Qualification	[A levels only]			
	BTEC Diploma only	**-0.731** (0.102)	**-0.693** (0.104)	**-0.812** (0.109)
UCAS tariff* Qualification	[A levels only]			
	BTEC Diploma only	**-0.001** (0.000)	**-0.002** (0.000)	**-0.002** (0.000)

The probabilities of a good degree (using Model 3 parameters) are presented in Figure 6 (for a student in the reference categories for institution (No.109), subject group (Subjects allied to Medicine), gender (males), socio-economic classification (Lower supervisory and technical occupations) and school type (grammar school)).

This shows a big difference in probabilities between the two groups of students, which also increased (albeit only slightly) with UCAS tariff. Thus, a BTEC student with a UCAS tariff of 360 had a probability of 0.051 of a First and 0.458 of at least an Upper Second-class degree, compared with a probability of 0.199 of a First and of 0.796 of at least an Upper Second for an A level student.

With the significant interaction effect we have the following equivalence:

-4.836 + 0.010*Y = -4.836 + 0.010*X – 0.812 – 0.002*X

Or

Equivalent A level tariff = (0.008*BTEC tariff-0.812)/0.010

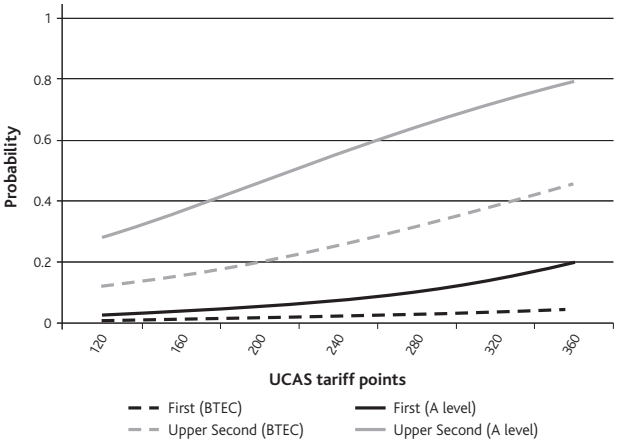
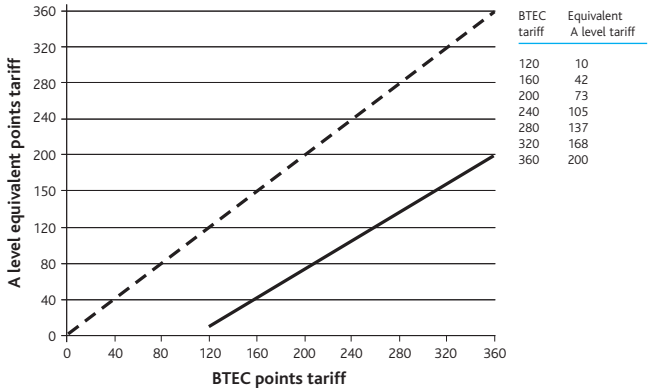


Figure 6: Modelled probabilities of achieving a First or at least an Upper Second-class degree, by UCAS tariff and prior qualification (A level only *v* BTEC Diploma only)

The equivalent tariffs are presented in Figure 7. This shows, for example, that a BTEC tariff of 120 was equivalent to an A level tariff of 10; a BTEC tariff of 360 was equivalent to an A level tariff of 200.



BTEC tariff	Equivalent A level tariff
120	10
160	42
200	73
240	105
280	137
320	168
360	200

Figure 7: Equivalent tariff scores (A levels only *v* BTEC Diploma only)

A/AS level *v* BTECs

The final comparison made was between students taking a mix of A level and AS level qualifications and students taking a mix of BTEC qualifications. Only universities where at least 10 BTEC students and 10 A/AS level students attended were included. However, for this there was no need to remove students above a particular UCAS tariff because there were enough BTEC students with tariffs up to 720.

Table 10 presents some details of the distribution of UCAS tariff for students with each of the qualifications. Thus, A level students tend to have higher UCAS tariffs on average than BTEC students. Table 11 presents the degree outcomes of the two groups. This shows that A level students had a higher probability of a First or an Upper Second-class degree.

These results again suggest that the UCAS tariff is not correctly aligned between BTECs and A levels, with a significant effect for the qualification variable. The effect was negative for BTEC students and there was also a negative interaction effect, meaning that the negative BTEC effect was larger at higher UCAS tariffs. The BTEC effect was lower for Models 2 and 3 than for Model 1, but the interaction effect was larger.

Figure 8 presents the probabilities of achieving a good degree for the two different groups of students (for a student in the reference categories for institution (No.109), subject group (Subjects allied to Medicine), gender (females), socio-economic classification (Higher managerial and professional occupations) and school type (grammar school)).

Table 10: Distribution of UCAS tariff, by prior qualification

Qualification group	*N*	*Mean*	*SD*	*Min*	*Max*
A/AS levels	117,095	328.3	99.0	40	720
BTEC	12,395	289.7	88.9	40	720
All	**129,490**	**324.6**	**98.7**	**40**	**720**

Table 11: Distribution of degree class, by prior qualification

Qualification group	*N*	*First (%)*	*Upper Second (%)*	*Lower Second (%)*	*Third/ Unclassified (%)*
A/AS levels	117,095	16.5	56.6	23.7	3.3
BTEC	12,395	8.0	39.8	41.7	10.5
All	**129,490**	**15.6**	**55.0**	**25.4**	**4.0**

Table 12 presents the results of the models.

Table 12: Model parameter estimates (A/AS levels *v* BTECs)

Variable		*Model 1*	*Model 2*	*Model 3*
Intercept	First	**-4.128** (0.257)	**-5.102** (0.300)	**-5.169** (0.085)
	At least Upper Second	**-1.254** (0.225)	**-2.159** (0.300)	**-2.213** (0.084)
	At least Lower Second	**1.222** (0.253)	0.356 (0.300)	0.322 (0.085)
UCAS tariff		**0.007** (0.000)	**0.009** (0.000)	**0.009** (0.000)
Qualification	[A/AS levels]			
	BTEC	**-0.421** (0.064)	**-0.243** (0.066)	**-0.255** (0.068)
UCAS tariff* qualification	[A/AS levels]			
	BTEC	**-0.002** (0.000)	**-0.003** (0.000)	**-0.003** (0.000)

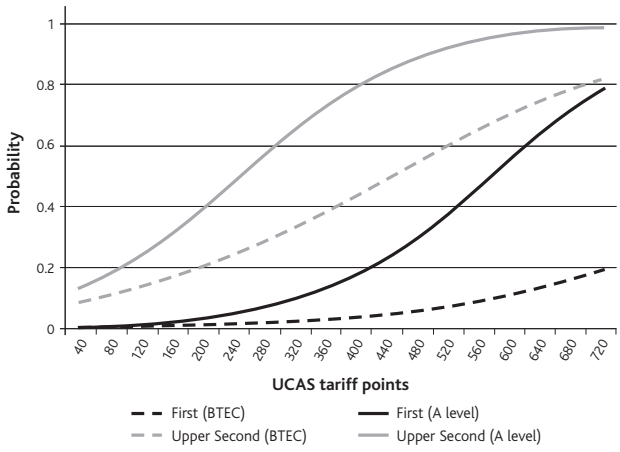
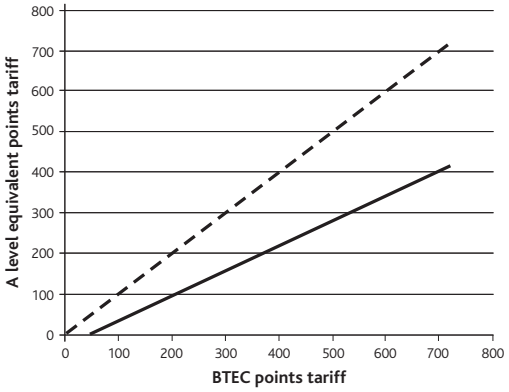


Figure 8: Modelled probabilities of achieving a First or at least an Upper Second-class degree, by UCAS tariff and prior qualification (A/AS levels *v* BTECs)

The differences in probabilities were large, particularly at the top end of the UCAS tariffs. A BTEC student with a tariff of 720 had a probability of a First of just 0.19, compared with a probability of 0.79 for an A level student with the same tariff. The equivalent tariffs were calculated, as follows:

Equivalent A level tariff = (0.006*BTEC tariff-0.255)/0.009

This produced the equivalent tariffs presented in Figure 9. Again, this demonstrates the large over-valuation of BTEC grades in the current UCAS tariff. For example, the A level equivalent for BTEC students with 200 UCAS points was 95, whilst for BTEC students with 720 UCAS points it was 416. However, there were very few BTEC students with UCAS tariffs above 500, so we need to exercise caution in making comparisons above this value.



BTEC tariff	Equivalent A level tariff
40	-4
80	21
120	46
160	71
200	95
240	120
280	145
320	169
360	194
400	219
440	244
480	268
520	293
560	318
600	342
640	366
680	391
720	416

Figure 9: Equivalent tariff scores (A/AS levels *v* BTECs)

Discussion

The claim made in UCAS Expert Group reports (which detail how new qualifications are allocated tariff points) is that one purpose of the UCAS tariff is to 'allow broad comparisons to be made between different types of achievement and different volumes of study' (UCAS, 2003). An implication is that, broadly speaking, students achieving the same UCAS tariff through different qualifications should have the same probability of achieving a good degree. This research has investigated whether this is true in practice, by making comparisons between A levels, IB and BTECs.

The results of the analyses presented here suggest that (after accounting for the institution attended, the degree subject studied and a number of student background characteristics) the current UCAS tariff slightly over-values the IB, in terms of students' probabilities of achieving a good degree, compared to the tariff for A levels. The models suggest that an IB points score is equivalent to an A level points score 39 points lower (for those taking A levels only) and 48 points lower (for those taking a mix of A levels and AS levels). An A level grade is worth 20 UCAS points, so this difference is equivalent to between two and two and a half A level grades.

This contrasts somewhat with the results of Green and Vignoles (2012). They found that the differences between IB and A levels varied across the ability range, with the current UCAS tariff over-valuing IB below an IB points score of 38 (tariff of 567), but under-valuing IB above this points score.

However, there is a potential issue with comparing directly the UCAS tariffs of A level and IB students, because almost all IB students took the same number of subjects (three at Standard level and three at Higher level), whereas the A level students with the highest tariffs were likely to have taken more A levels than those with lower tariffs. Because of this, there may be some attenuation of the worth of UCAS points at the top end for A level students. For example, a student with 5 A*s (700 points) may not be much better in terms of ability than one with 4 A*s (560 points). This suggests that the model used to predict the probability of a First/Upper Second-class degree for A level students may over-value the effect of more UCAS points at the top end. Therefore, it may be that a better predictor of university performance would be a mean A level measure. This suggestion is supported by research undertaken by the author which indicates that A level mean was a more influential measure of achievement than A level total tariff score, in terms of predicting university outcomes (Gill, 2015).

In the comparison of A levels and BTECs there was a much larger difference, with the current BTEC tariff being highly over-valued. There was also a significant interaction effect between the type of qualification and the UCAS tariff, which meant that the over-valuation increased at higher UCAS tariffs. According to the model, the maximum tariff for those taking a BTEC Diploma only (360) was equivalent to a tariff of just 200 from A levels. Similarly, for those taking combinations of BTECs, a tariff of 360 was equivalent to a tariff of just 194 from A levels. This suggests that a re-valuation of the tariff points allocated to BTEC Nationals may be necessary.

It should be noted that when UCAS determined the points scores for BTEC grades, no direct comparison was made with A levels (BTEC Nationals were compared to Advanced Vocational Certificates of Education [AVCEs] by the Expert Group). Indeed, given the very different nature of the two qualifications, it would probably be difficult to make such comparisons. Furthermore, it is likely that admissions tutors do not often have to make direct comparisons between BTEC students and A level students.

However, given that the UCAS tariff points are meant to be (broadly) equivalent for every qualification, some other method of calculating equivalent points scores might be advisable. An alternative might be to apply the equivalencies determined by a statistical model run on data from prior years (such as the one applied here) to the results achieved by students in the current year. For example, to make BTECs and A levels approximately equivalent we could give the maximum grade for a BTEC Diploma (DDD) 200 UCAS points, rather than the current 360. However, there are a number of drawbacks with this approach, mainly due to the significant time delay whilst the first cohort(s) of students taking a qualification complete their university studies.

Finally, it is worth considering the approach of admissions tutors in relation to the UCAS tariff. It may be that experienced admissions staff have perceptions of the relative worth of UCAS tariff scores achieved from different qualifications and adjust their offers accordingly. If these perceptions are accurate then this suggests that they can overcome any under- or over-valuation of particular qualifications. Green and Vignoles' (2012) analysis suggested that admissions tutors did have a different perception of the equivalence between IB and A level tariff points than the official tariff. However, they found that this perception did not agree with the equivalence derived from their modelling (admissions tutors rated points scores from the IB more highly at the bottom of the scale, but less highly at the top of the scale). Furthermore, for new qualifications or inexperienced admissions tutors, the official UCAS tariff may be the only source of information.

References

Gill, T. (2015). *Using generalised boosting models to evaluate the UCAS tariff.* Cambridge Assessment internal report. Cambridge, UK: Cambridge Assessment.

Green, F., & Vignoles, A. (2012). An empirical method for deriving grade equivalence for university entrance qualifications: An application to A Levels and the International Baccalaureate. *Oxford Review of Education*, *38*(4) 473–491. doi: 10.1080/03054985.2012.713858

UCAS (2003). *Expert Group Report for Awards Seeking Admission to the UCAS Tariff: BTEC National Awards, BTEC National Certificates, BTEC National Diplomas*. London: UCAS. Retrieved from https://www.ucas.com/sites/default/files/btec-nats-report.pdf

UCAS (2006). *Expert Group Report for Awards Seeking Admission to the UCAS Tariff: The International Baccalaureate*. London: UCAS. Retrieved from https://www.ucas.com/sites/default/files/ibreport.pdf

Vidal Rodeiro, C., Sutch, T., & Zanini, N. (2013). *Prior learning of undergraduates in UK higher education institutions*. Cambridge Assessment research report. Cambridge, UK: Cambridge Assessment

Taking risks and being creative: Assessment in Drama and Theatre

Prerna Carroll International House London and **Emma Dodds** Research Division (The study was completed when the first author was based in the Research Division)

Introduction

In this article we discuss the concept of creativity and its assessment. Creativity is critical to many subjects in secondary education, including Drama and Theatre, but is not easy to assess. Whilst there is a need for reliable assessments at General Certificate of Education (GCE) Advanced Level (A level), the validity and integrity of what is taught are also essential. We describe a small-scale study in which four course leaders at Higher Education Institutions (HEIs) were interviewed about Drama at undergraduate level. The aims of this study were to gain an insight into undergraduate assessment practices, and to identify any lessons to be learned and applied at A level.

Creativity as a concept

A varied range of creative subjects are on offer at GCSE and A level, and whilst the term 'creativity' is broadly understood and widely used in education, its precise definition has divided opinion for many years. In 1969, Barron defined creativity as "the ability to bring something new into existence" (as cited in Gallagher, 2007, p.1230). However, Bruner (1979) put forward the notion that creativity "confirms something that we already knew subconsciously" (as cited in Gallagher, 2007, p.1230). Gallagher (2007) discusses the concept of creativity in the context of Drama education, and after reviewing the literature in detail, still comes up short when trying to find an applicable definition, or in finding suitable studies that attempt to define creativity in Drama.

Politically, creativity gained importance during the rise of 'New Labour', when the economic benefits of creativity were highlighted (Buckingham &Jones, 2001) and the term 'democratic creativity' was introduced. This term was used by the National Advisory Committee on Creativity and Cultural Education (NACCCE) in its report *All Our Futures: Creativity, Culture and Education* which argued that creative and cultural education was the key to unlocking "Britain's economic prosperity and social cohesion." (NACCCE, 1999, p.5). The report defined creativity as "imaginative activity fashioned so as to produce outcomes that are both original and of value" (NACCCE, 1999, p.30). Within this definition, the NACCCE clarifies that creativity relates to four characteristics: *imagination*; *purpose*; *originality*; and *value*. *Imagination* refers to generating new ideas or to approaching ideas from a new perspective. *Purpose* refers to the creative process an individual undertakes when developing a product. The third characteristic, *originality*, is relative and refers to ideas or thoughts that are original compared to an individual's ideas, a group's ideas, or historic ideas. Lastly, *value* refers to the judgement of the work in relation to the purpose. It can refer to self or critical evaluation of the finished art form, and can be individual or shared. The NACCCE definition of creativity coheres more with Barron's definition than with Bruner's, focusing on the originality of ideas and their subsequent value.

Regardless of differences in definitions, the common themes revolve around new ideas or remodelled ideas that have value and purpose and are explored through a clear creative process (Beghetto, 2005). However due to its multi-faceted nature, creativity does not seem to be an easy skill to teach, let alone assess. Unlike the ability to add or subtract numbers, creativity cannot be taught explicitly, and is also difficult to measure systematically. The NACCCE draws on research by Woods (1995) to suggest that teachers can encourage creativity by ensuring autonomy and respect on both sides of the student-teacher relationship, authenticity in initiatives and responses, and fulfilment. Moreover, an element of trust is necessary, as the aims of so-called 'teaching for creativity' are to encourage self-confidence, independence of mind, and the capacity to think for oneself (Woods, 1995).

The teacher-student relationship is particularly important as it can foster creative thinking in students, enabling them to take risks and develop new and original ideas. The research literature indicates that positive perceptions of teacher support increase individuals' risk-taking in, and motivation towards, many subjects, including those that are not widely viewed as creative, such as Mathematics and the Sciences (Alonso-Tapia & Pardo, 2006; Kalchman & Koedinger, 2005; Nickerson, 1999). Those teachers who do not welcome students' ideas discourage students from taking risks and being creative in their classrooms (Kennedy, 2005). Furthermore, risk-taking and consequent creative

thinking also seem to be fostered by "positive competence-related feedback" (Beghetto, 2009, p.214). Taken with the definitions of creativity, this evidence suggests that formative assessment is integral to assessing creative subjects, as there is a need for continuous feedback during the process of creative production.

Assessing creativity

The concept of creativity focuses on a process leading to a unique outcome. Arguably, it therefore lends itself more to assessment for formative purposes than to assessment for summative purposes, and formative assessment is usually conducted internally rather than externally. Although the general criticisms of internal assessment such as its proneness to bias are well-rehearsed, there are still some important advocates in the research literature. For example, Beghetto (2005) explored the effects of assessment type on students' creativity in the context of American classroom assessment. He found that (internal) formative assessment was the best method of fostering both creativity per se, and the risk-taking that is associated with creativity.

In her review of assessment in Drama, Schonmann (2007) discusses the conceptualisation of Drama in the curriculum, and proposes two appropriate approaches to its assessment: a directive approach; and a dialectical approach. The directive approach evaluates individual achievements against predetermined criteria. The criteria are developed in relation to specific aims set for a successful performance. The dialectical approach aims to create a profile of a student's progress to becoming a professional artist and their knowledge and skills. This approach uses formative feedback and a set of introspective questions on the students' progress. Although Schonmann states that the two approaches are not mutually exclusive, she feels that they can appropriately reflect students' achievements based on their own journeys.

An important example of the use of large-scale external assessment in Drama and Theatre can be found by looking at the National Assessment of Educational Progress (NAEP) in the USA, which is a nationally representative and continual assessment of American students' knowledge and skills in various subject areas. In 1997 the NAEP conducted a large-scale assessment of students' achievements in the Arts, and in particular Theatre. Its assessment framework for Theatre identified two main themes: content; and process. Content referred to "knowledge and understanding of theatre and perceptual, technical, expressive, and intellectual/reflective skills" (Vanneman, 1998, p.2). Process skills, on the other hand, included "creating/performing and responding" (Vanneman, 1998, p.2). The NAEP used both paper-and-pencil and performance tasks, which were developed by the Educational Testing Service under the guidance of a committee of theatre education experts. The paper-and-pencil task assessed students' responses to Theatre and justifications for a variety of creative decisions. Students were exposed to Theatre through multiple media, such as video clips, photographs and paper excerpts, and responses were predominantly assessed through short and extended response questions. One question, for example, asked students to choose between an abstract or realistic set for a play they had been given, and to justify their choice.

The paper-and-pencil task appears to assess their knowledge and understanding of the technical aspects of Theatre and encourages reflection and evaluation. On the other hand, the performance task required students to work together in small groups to develop a short performance which was then videoed for assessment. Students were further encouraged to comment on their work, their achievements and their success. Due to the collaborative nature of the task, students' scores comprised of an individual score and a group score.

The NAEP has a longstanding reputation for using external assessment to determine national educational progress, and it is perhaps telling that, in order to effectively assess creative aptitude in Theatre, a paper-and-pencil task was not sufficient. Whilst it acknowledges that "creating and administering a national performance assessment was very challenging." (p.6), the assessment was developed by experienced professional and subject experts, and provides a useful, if somewhat rare example of how external assessment can be used to assess creativity in Theatre.

The assessment of Drama in England

England is currently in the midst of a series of major reforms to its general qualifications. Significant changes are being made to both GCSEs and A levels in terms of the content that students study, and how it is assessed. The Government has stated that the purpose of taking A levels is primarily for entry to university, and that changes are needed to ensure that students are better prepared for their undergraduate courses. Data from the Higher Education Statistics Agency (HESA) indicates that in 2011/12, 3,705 students were enrolled to study Drama at university – either as a single or joint degree. Among those, 2,625 students were enrolled to study Drama only[1]. 58.2 per cent of them had an A level in Drama and Theatre Studies and 2.2 per cent had an AS level only. Among the students with no A level in Drama and Theatre Studies, the most popular A level choices were: Performance Studies, English Literature and Media/Film/TV Studies[2]. Considering the high percentage of students pursuing Drama at university who have an A level in Drama, ensuring that the subject facilitates development at HE is clearly of great importance.

Oxford, Cambridge and RSA (OCR) is developing a new A level in Drama and Theatre to meet the new national criteria. It will be taught in schools and colleges from September 2016 onwards (OCR, 2015). The new A level will be fully linear; assessment of a student's knowledge and understanding of the whole course will take place at the end of two years of study. To support this development work, we conducted a small-scale study to ascertain the types of assessment that are used in Drama in HEIs and whether these practices are relevant to A level. The aim of the study was to address three main questions:

1. What assessment practices are used by HEIs in England?
2. Are written examinations used by HEIs, and what are their views on them?
3. What are HEIs' views on the skills that students with A levels possess?

Method

We contacted lecturers who teach Drama and Performance Arts at four highly reputable HEIs in England. Four course leaders for undergraduate Drama courses consented to be interviewed. Three of the four HEIs were

1. Numbers of students were rounded to the nearest multiple of 5, following HESA's rounding strategy.
2. Source: HESA Student Record 2011/12. Copyright Higher Education Statistics Agency Limited 2012. HESA cannot accept responsibility for any inferences or conclusions derived from the data by third parties.

self-described as schools that focussed on training students to become professional actors. The fourth HEI was more traditional and taught Drama in a more academic, rather than a vocational sense.

The interview schedule was semi-structured (see Appendix A). It comprised several questions addressing assessment practices within the HEI, including moderation, group performance and individual marks. We also asked for participants' views on external assessment; specifically, written examinations and their appropriateness in Drama. Lastly, we asked participants to comment on the skills acquired by students at A level and their relevance to undergraduate study. All questions were designed to elicit detailed responses and maximise discussion. We conducted the interviews face-to-face or by telephone. All four interviews lasted approximately an hour.

Analysis

The entire interview data generated was transcribed. We coded the transcripts into themes and analysed them qualitatively using MAXQDA (a software package for qualitative and mixed methods data analysis). The three main themes were aligned with the three research questions and the structure of the interview schedule:

1. assessment at undergraduate level;
2. written examinations; and
3. skills acquired during secondary education.

Subcodes within the first theme covered procedures for group performance assessment and moderation procedures, as well as participants' views on quantifying performance. Subcodes within the second theme covered skills valued by participants that could be assessed by a written examination, and participants' views on reflective writing, grades from written examinations, and learning specific texts. Subcodes within the third theme related to participants' opinions on examinations and the secondary education system, and skills that are detrimental to the further study of Drama.

Results

Assessment at undergraduate level

Group examinations were used regularly in the HEIs of the four lecturers interviewed, and usually individual students were marked on their own performance within the group. Generally, the 'marks' provided were descriptive and used as a basis for formative feedback. Unanimously, the lecturers felt that the size of the role did not make a difference to the mark, and stated that they chose particular performances so that no student would simply have one line in the whole production. The lecturers commented:

> *We mark every single assessment individually even though our students are often doing group performance work… We don't have any assessments where we say, "This is a group mark come what may so you will all get X." They're all individually marked against the criteria.*

> *What an actor is able to reveal will be as much revealed in a smaller role as it will in a bigger role.*

Assessment of performances at all HEIs in the study entailed second and usually third markers. There were multiple stages in the marking process, including markers' meetings and internal discussions. External examiners were also used to moderate performances.

> *I would then check [the marks] as a first point for quality as course group leader and then they all have to get sent to our external examiners.*

> *We have two people present at all of [the performance] assessments…and then we moderate and mark based on those things… If we can't come up with an agreed mark then it goes to the head of the department for mediation as it were and final agreement. If that doesn't work it would very unusually go to the external examiner for final arbitration.*

> *There will always be a minimum of two markers for any assessment… but often there will also be a third moderator there… We often have whole panels of markers so it's not uncommon to have five people marking an assessment all at the same time and then having a marks meeting afterwards.*

> *[The] External examiner is a professional director who has contacts with the school and who sees as many of our public performances as possible.*

The lecturers felt that it was not necessarily possible to put a quantifiable mark on a creative performance, and in some cases they thought grades were irrelevant. They also felt that marks would foster competition and remove focus from students' ability to develop creativity and originality. Success and failure were perceived in terms of employment and success in students' careers, as opposed to their mark for an assessment.

> *What is relevant [is] whether they manage to take on board enough of what we can teach to give themselves a chance outside. And I'm not going to put …a C- against something like that.*

> *If students start getting obsessed by what marks they are getting, then it creates competition… Rather than being open to take risks and develop, students start using the marks as an indicator of their progress.*

> *Essentially our students don't fail. That's not because we connive to pass them… failure is for them, if it's failure, several years down the line when they find nobody will employ them.*

Written examinations

The lecturers valued some skills that could be tested through a written assessment. The main skill they looked for in students was research and evaluative skills. They felt that these skills could be taught better at A level, and should require students to be more original and creative. However, the lecturers felt that a prescribed answer and a sense of correctness or incorrectness often removed students' abilities to take risks and generate individual thought.

> *When we need to assess their evaluative and analytical understanding of work … we really expect them to be drawing on and synthesizing research.*

> *[Written assessment] needs to test knowledge, understanding, their ability to critique their findings and those of others. It needs to demonstrate independent research and thinking… draw on analysis and evaluation… apply critical thinking to examples.*

They devise a research question, they do a load of research, they have to do a literature review, critique their research, do a bibliography… identify and select appropriate research sources… engage critically with sources… construct a persuasive argument … write in coherent prose with accurate referencing.

The lecturers were not enthusiastic about students producing reflective writing under examination conditions as they felt the time and pressure of such a setting is not conducive to reflection. They also felt that reflective writing is a self-development tool and should not be marked.

We don't reflect under pressure.

It's a reflective thing. So it's really for their own purposes [so that] as they are doing their work. They work very intensively. It's very demanding work, so we want them to have an opportunity to think as they go along about the stuff they are doing and how it's helping them or how it feels not to be helping them, potentially – the work that they really respond to, the work that they don't.

The lecturers felt that grades at A level were not a deciding factor in accepting students onto their courses. They were more interested in the abilities demonstrated by students in the audition process.

Any student coming here will not be at an advantage because they have done well in a written exam.

*So if we get a student coming in with three A*s, irrespective of what subjects they're in, it does say something about that student's academic potential for learning… that they can probably work independently, revise, work well under pressure, shape their thoughts, recollect information… However we would equally value someone with no A level grades who had a very different experience of education who could still demonstrate potential in the Arts.*

Similarly, the lecturers felt that students were not at an advantage for having learnt particular texts. Whilst they valued the extra information and knowledge that texts can bring, they were more interested in the students' ability in performance as witnessed in the audition process.

It would be really limiting if we said, "You can only come if you've read the six greatest Shakespeare plays and you know how to write about Hamlet," because immediately you will say, "All these people out here are off our radar. We're not interested in them. So all those people who haven't ever had the opportunities, let's keep if like that and make sure they don't get them and these people here who have had a very specific kind of relationship with the education system, we'll take them."

Skills acquired during secondary education

The lecturers were concerned about the reforms to A level because they felt they emphasised rote learning and discouraged creativity and risk-taking. Three out of four of the lecturers felt that they had to 'un-do' some of the learning taught in schools to encourage students to think innovatively and beyond 'right and wrong answers'.

The other thing that's very problematic with written assessments is they're predisposed to a right or a wrong answer, and that's reflected in the marking schemes that I see used for A level and GCSE… We spend a lot of time in the first year of university undoing an approach to learning that's been embedded in students through their GSCE and A level experience.

… because of an over-examination of children from a young age […] children and young people think that education is about the regurgitation of knowledge and they don't understand how to learn; they understand how to be taught… and what we need is to encourage our students to learn and to trust that they can learn and to not depend so much on us to be taught.

Furthermore, the lecturers felt that putting a student under examination conditions would result in reduced motivation and interest in the subject and the work studied. They felt that the subject of Drama allowed students to escape from the routine of traditional academic subjects and express themselves; where in other subjects they may not have had the opportunity to do so.

I think as soon as you say, "And now you have to write about love and revenge in Romeo and Juliet for 500 words at 9am on a Thursday morning", then you deaden that student's relationship with the work…It will stop them going anywhere near Drama because Drama is their opportunity to actually be expressive and for many students… it's the only thing that has got them through school… So I think if you then say, "Well, actually, now it's going to be the same as all the other subjects", then it's a serious misunderstanding of what Drama as a subject is…

Lastly, as a subject, lecturers felt that Drama was more vocational than academic, as many HEIs trained their students to be professional actors.

We're not in education. We're in training

What we are doing is introducing them to the industry… They have all had an opportunity to be seen by agents and casting directors in roles which sufficiently reveal their skills and aptitudes and castability.

Discussion

The assessment of creative subjects such as Drama is currently under scrutiny, with concerns being raised about internal assessment being prone to bias. Creativity as a concept in itself is defined in multiple ways, but common themes from definitions emphasise the artistic journey or process as well as the original product created at the end. The assessment of creativity should therefore assess both the process and the output, which in the case of Drama and Theatre, usually refers to a live performance.

The aims of this small-scale study were to gain an insight into the assessment practices used in undergraduate Drama courses at HEIs in England, and to identify any lessons to be learned and applied to the study of Drama and Theatre at A level. The four interviews provide a rare insight into the assessment of undergraduate Drama at HEIs in England. They also reveal that the views and experiences of the four lecturers overlapped extensively. Whilst it is difficult to generalise from such a small study, the findings hint at some likely generalities which could potentially be confirmed with further research.

The interview data suggests that stringent procedures are in place within HEIs to ensure that Drama students are graded fairly on their performance. Lecturers at all four institutions stated that they had two or three markers for all performances, plus external examiners to further moderate the marks given. Although group performances are common, marks are allocated individually and not for a group as a whole. Care is also taken to ensure students are not given minor roles. However, the lecturers at the three institutions that emphasised vocational training explained that marks were not treated with the same value as they are at

A level or in secondary education in general. Instead, they are provided as a means of feedback, and students are not told their marks unless they specifically request them. The lecturers felt that adding quantifiable values to students' performances would create unnecessary competition and not be reflective of their creative processes, which have peaks and troughs and cannot be measured at a single fixed moment in time.
This emphasis on the creative process and regular feedback reflects the NACCEE definition of creativity and how it should be taught.

When discussing written assessment, the four lecturers unanimously felt that a traditional examination setting was not 'fit-for-purpose' for the study of Drama. They felt that students cannot reflect under pressure and felt again that marks should be awarded not just on the final product but on the student's journey there. Even the course leader from the more academic HEI stated that there were plans to remove reflective writing from the Drama undergraduate course as the teaching staff found it not to be beneficial to the students and increasingly difficult to mark. This finding raises concerns over the current external written examinations of Drama offered by some awarding organisations; If true reflection cannot occur under timed conditions, then arguably this should not be attempted in A level Drama and Theatre. Perhaps an amendment to the examination paper that matches the style of the NAEP assessment would be more suitable, as it would require students to apply their technical knowledge of Drama and Theatre to a particular text or stimulus provided, rather than to reflect on their own or a professional performance.

Whilst reflective writing was not favoured in written examinations, the lecturers felt that students needed to develop their evaluative and research skills at school, as this is often something they were lacking when they started university. Furthermore, the HEIs in this study all used an audition process to shortlist students accepted onto a course. Therefore, they did not put as much value on the grades achieved through external examinations in Drama or other subjects, as other university departments might. The lecturers' main focus was on the applicants' demonstrated abilities on audition day, and whilst high grades in Science, Technology, and Engineering or Mathematics (STEM) subjects were indicative of their intellectual ability, they were not an indicator of their creative ability. As the institutions that the lecturers represented were highly inclusive, they did not feel students needed to learn particular texts prior to coming to university. They were more concerned with a student's ability to evaluate, critique and research a topic or text, and ensuring that they take risks to explore the subject. Furthermore, the lecturers felt that aligning Drama to other subjects in terms of assessment would reduce uptake and interest towards the subject for many students. The lecturers felt that often students of Drama found refuge in the subject due to its expressive and artistic nature, and that adding a traditional external examination to the subject would 'deaden' that relationship.

The lecturers also felt that current teaching styles at A level are too prescriptive and focus students on attaining the highest grades. As a result, students do not learn how to take risks, learn from mistakes and explore topics. Instead, they are taught to follow a set procedure, such as an essay format or argument, which hinders their creative development. This notion of risk-taking and creativity is echoed in the literature, and risk-taking has even been shown to decrease as students' progress further in education (Beghetto, 2009). This reduction in risk-taking could be due to the education system placing higher importance on attainment and creating a culture where there is a minimum tolerance for errors. As students strive for 'success' in these terms, taking risks and chances is less important as it may not secure them the best mark. This behaviour can often limit creativity in individuals and reduces the opportunity for independent thinking. The lecturers in this study felt that they had to undo this thinking in the first year of students' undergraduate courses, in order to enable students to perform to their best creative potential.

Appendix A: Assessing creativity: Higher Education interview schedule

Assessment at undergraduate level

- What types of assessments do you use at undergraduate level (performance, group, written, journal) and when (end-of-term, course, year)?
- Do you think that assessments are of equal difficulty/standards year-on-year?
- How do you ensure that assessments are of equal difficulty/ standards year-on-year?
- How do you assess group performances? How are these marked?
- How are group performances moderated (such as between examiners)?

Written examinations

- Do you use written assessment? What kind? Summative/Formative?
- If you use written assessment, how much emphasis or weighting is put on written assessment?
- If you use written assessment, how do you ensure the written examinations are of equal difficulty across years?
- What skills in Drama (if any) do you think are appropriate to test with a written examination?
- What skills aren't suitable for written assessment?

Skills acquired during secondary education

- Would you value A level grades based on performance in a written examination?
- What would a mark on a written exam at A level tell you about a prospective student?
- What are the basic requirements or skills new undergraduates need to meet/have?
- Is it important for students to have studied a specific text(s)/practitioner(s)/time period(s)?

References

Alonso-Tapia, J. & Pardo, A. (2006). Assessment of learning environment motivational quality from the point of view of secondary and high school learners. *Learning and Instruction 16* (2006), 295–309.

Barron, F. (1969). *Creative Person and Creative Process*. New York: Holt, Rinehart and Winston.

Beghetto, R. A. (2005). Does Assessment Kill Student Creativity? In *The Educational Forum* 69(3) 254–263.

Beghetto, R. A. (2009). Correlates of Intellectual Risk Taking in Elementary School Science. *Journal of Research in Science Teaching*, *46*(2), 210–223.

Bruner, J. (1979). *On knowing: Essays for the Left Hand*. Boston, MA: Harvard University Press.

Buckingham, D., & Jones, K. (2001). New Labour's Cultural Turn: Some Tensions in Contemporary Educational and Cultural Policy. *Journal of Education Policy*, *16*(1), 1–14. doi: 10.1080/02680930010009796

Department for Education (2015, February). *Reformed GCSE and A level Subject Content. Government Consultation Response*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/403347/Reformed_GCSE_and_A_level_subject_content_Government_response.pdf

Gallagher, K. (2007). Conceptions of Creativity in Drama Education. In *International Handbook of Research in Arts Education* (pp.1229-1245). Netherlands: Springer.

Kalchman, M., & Koedinger, K. R. (2005). Teaching and Learning Functions. In M.S. Donovan & J.D. Bransford (Eds.), *How Students Learn: History, Mathematics, and Science in the Classroom* (pp.351–393). Washington, DC: The National Academies Press.

Kennedy, M. (2005). *Inside teaching: How Classroom Life Undermines Reform*. Cambridge, MA: Harvard University Press.

National Advisory Committee on Creative and Cultural Education (NACCCE). (1999). *All Our Futures: Creativity, Culture and Education*. NACCCE Report

Nickerson, R. S. (1999). Enhancing Creativity. In R.J. Sternberg (Ed.), *Handbook of Human Creativity* (pp.392–430). New York: Cambridge University Press.

OCR (2015). *OCR Level 3 Advanced GCE in Drama and Theatre (H459) Specification*. Retrieved from http://www.ocr.org.uk/Images/242650-specification-draft-a-level-gce-drama-and-theatre-h459.pdf

Schonmann, S. (2007). Wrestling with Assessment in Drama Education. In *International Handbook of Research in Arts Education* (pp.409–422). Netherlands: Springer.

Vanneman, A. (1998). NAEP and Theatre: Framework, Field Test, and Assessment. *Focus on NAEP*, *3*(3).

Woods, P. (1995). *Creative Teachers in Primary Schools*. Buckingham: Oxford University Press.

All in good time: Influences on team leaders' communication choices when giving feedback to examiners

Martin Johnson Research Division

Introduction

In the Oxford, Cambridge and RSA (OCR) awarding body, senior examiners with responsibility for monitoring the marking performance of other examiners in a marking team are called team leaders. Prior to examiners being cleared to mark examination scripts, they undergo a standardisation process. This involves the most senior examiners aligning all other examiners to their decisions around how to apply a mark scheme. At the end of this standardisation process the team leaders verify that each examiner can apply the mark scheme appropriately.

Throughout standardisation and subsequent live marking some team leaders and examiners work remotely from each other in a digital marking environment. This environment supports a number of important marking quality assurance functions: Team leaders can see examiners' real time scripts and mark submissions; they can also easily compare examiners' marks with preordained definitive marks on special monitoring scripts to check marking accuracy. The digital marking system also allows team leaders to give examiners feedback on their marking.

My previous research has looked at some of the common and diverging characteristics of team leader feedback (Johnson & Black, 2012a; Johnson & Black, 2012b). In this article I take a closer look at some of the data from those studies to explore why team leaders choose different communication modes when giving feedback to examiners. I argue that these choices relate to the capacities of different modes to balance the needs of communication flow and to support the alignment of team leader intended meaning and examiner interpretation of feedback messages. As part of that discussion, I consider how these choices relate to communication theories, media richness, and the synchronous and asynchronous qualities of communication modes.

Why do team leaders give feedback to examiners?

The digital marking system that is used by OCR examiners supports the awarding body's marking quality assurance arrangements in a number of important ways. The ability to simultaneously distribute digitally scanned versions of common examination scripts across different examiners allows examiners' marks to be compared with each other in ways that were not practical prior to the introduction of the digital marking system. The use of common scripts supports the examiner standardisation training process by allowing common rationales to be shared with examiners on carefully chosen exemplar scripts. The system also allows team leaders (senior examiners who have the responsibility to monitor the marking performance of other examiners in their marking team) to oversee the quality of examiners' live marking in real time.

Another benefit of the digital marking system is that team leaders can engage more frequently with examiners in their marking team by giving them feedback on their recently completed marking. These benefits are reflected in an Office of Qualifications and Examinations Regulation (Ofqual) report on marking which states:

> *As well as its logistical benefits, on-screen marking should improve marking reliability by enabling more frequent and flexible monitoring of examiners by exam boards. Senior examiners review their team's marking almost in real time, ensuring that inconsistent or inaccurate marking is detected early*
>
> (Ofqual, 2013, p.12)

Previous research has started to elicit some information about team leader feedback practices (Johnson, 2015; Johnson, 2014; Johnson &

Black, 2012a; Johnson & Black, 2012b). These findings suggest that, as well as supporting crucial marking quality assurance functions, feedback also facilitates examiner learning and their membership of a professional community. This is because feedback gives team leaders and examiners direct and on-going access to each other's perspectives around shared areas of focus.

Regardless of the purpose of feedback, what is of interest for my analysis in this article is that all of these functions rely on an underlying mechanism that allows participants' intended and received meanings to be communicated effectively.

How does feedback communication work?

Distributed Cognition Theory (Hutchins, 1995) suggests that organisational intelligence grows through the push and pull of information across a professional community. These pushes and pulls occur when experts in a community decide that others need access to specific information, or where less expert members request something that they need to know in order for them to carry out an activity. This perspective on learning and development is supported by sociocultural learning theories which suggests that individual cognitive development is contingent on social interaction involving individuals who possess different levels of expertise (Roth & Lee, 2007; Vygotsky, 2014). Feedback communication is a form of social interaction that can allow recipients an insight into the perspective of experts, and therefore help to induct less experienced participants into a professional community (Wenger, 2000; Wenger, 1998).

For communication to be successful there must be alignment between the intended meaning and the received interpretation of any communicated message. Communication theory suggests that this alignment involves synchrony, which is described as 'the extent to which individuals have a shared focus' (Dennis & Valacich, 1999, p.5). In other words, participants in discourse need to have a shared common object if they are to be able to attend to each other's perspective. Shared focus may be fixed on either a tangible object (e.g., a mark scheme) or a semantic object (e.g., a mathematical concept). The important thing is that this shared focus gives participants common ground on which they can start to build a sense of meaning in relation to each other.

Synchrony may be influenced by communication mode. Research into digital communication categorises communication modes in terms of their media richness. This concept was developed by Trevino, Lengel, and Daft (1987), who describe rich media as those with a high bandwidth which are able to carry a relatively large amount of information. Rich media, such as face-to-face communication (Pfaffman, 2007; Hollan & Stornetta, 1992) facilitate instant feedback, allow both verbal and non-verbal cues, involve natural language, convey emotion, and are considered to be the best mechanisms for conveying ambiguous ideas or concepts (Cameron & Webster, 2005).

The richness of a communication mode also subsumes the concept of synchronicity (Whittaker, 2003). Asynchronous communication involves a disruption in communication, with perhaps the most common form of asynchronous disruption being temporal (e.g., where there is a gap between communication instigation and reply). Email is an asynchronous communication mode. On the other hand, synchronous communication occurs where there is a direct link between instigation and reply (e.g., a telephone conversation).

Research suggests that asynchrony and synchrony can impact the way that feedback communication influences learning. For example, a number of studies define effective feedback as having immediacy (Barton & Wolery, 2007; Ahmed et al., 2012; Chur-Hansen & McLean, 2006; Burke, Marks_Maran, Ooms, Webb, & Cooper, 2009; Hatzipanagos & Warburton, 2009; Mathieson, 2012; Li & De Luca, 2014). Archer (2010) and Cook, O'Shea, Young, and Stedmon (1999) report that synchronous feedback aids task completion, with information being incorporated within on-going learning processes (Panahi, Birjandi, & Azabdaftari, 2013). Similarly, learners perceive synchronous feedback to be more effective than asynchronous feedback (Owens, Hardcastle, & Richardson, 2009; Dennen, Aubteen Darabi, & Smith, 2007). On the other hand, Archer (2010) notes that asynchronous feedback best supports the transfer of knowledge, allowing time for reflection, and reinforces already learned skills (Barton & Wolery [2007].

Methods

The study of team leader feedback practice presents a number of challenges. These challenges relate to the scale and distribution of communication interactions in the digital marking system. In order to gain a picture of feedback practices and perspectives a four-phase matrix data collection plan was developed, spreading data collection across four subject areas and involving 22 team leaders and 6 examiners (Figure 1).

The first and fourth data collection phases used direct observation and follow up interviews to gather data about team leaders' feedback practices. These observation sessions involved two elements. The first element involved the team leader giving feedback to examiners in their team in the digital marking environment. During this time researchers remotely observed the team leader's on- and off-screen practice using Morae® usability software (TechSmith, 2011) (Figure 2). The second element of the observation session involved the team leader and the researchers jointly viewing the audio-visual recording of the feedback session and using it as a stimulated recall session. Stimulated recall is one of a set of introspective methods that elicit data about the thought processes that an individual uses when carrying out a task (Gass & Mackey, 2000).

The second data collection phase gathered information from examiners about their perspectives on team leader feedback through telephone interviews. Examiners were asked to talk through the feedback messages that they had received from their team leader. The third data collection phase surveyed 18 team leaders across the 4 subject areas. The survey focused on validating the themes identified during the observation sessions.

The next section outlines the themes that appear across the data analyses that relate to elements of media richness and synchrony.

Findings

The survey data show that all team leaders used a mixture of email and telephone modes to give feedback to examiners. The interview, survey and observation data provide a number of insights into the reasons why team leaders chose these differing communication modes, with communication management and synchrony appearing to be salient influences.

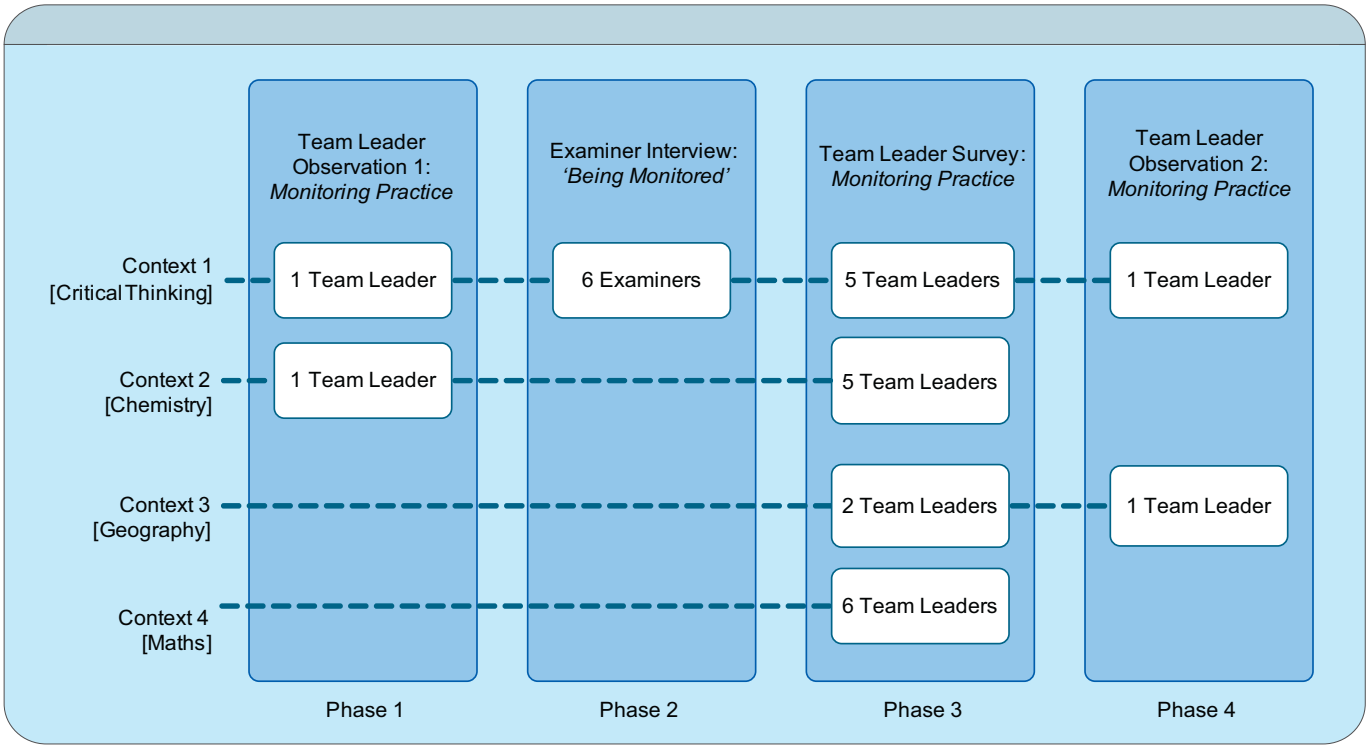


Figure 1: Project design phases

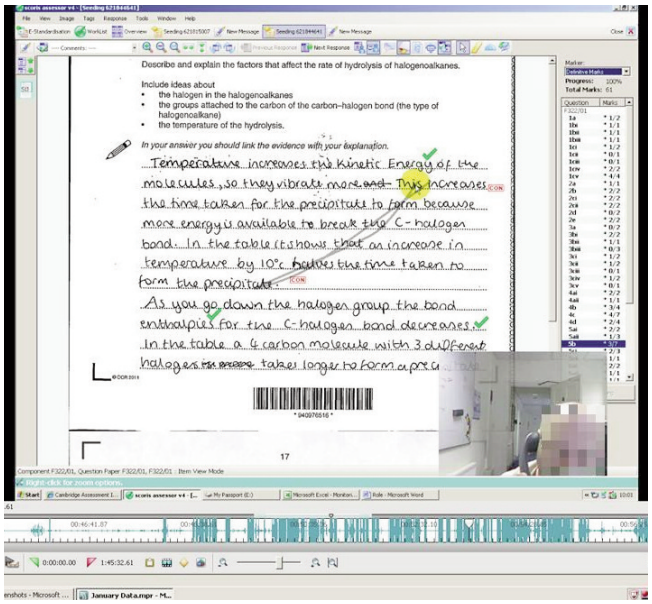


Figure 2: Feedback observation – the researchers' view

Analyses of the team leader survey responses (Table 1) suggest that email choice links with issues of communication speed and convenience. On the other hand, telephone communication links with issues of personalisation, information quality, and sense checking requirements.

Preferences for the telephone communication mode also appear to relate to the nature of the feedback being given. Team leaders prefer to use the telephone at times where the feedback being given is very detailed, or where it deals with comprehensive aspects of mark scheme application. There is also an indication that examiner effectiveness can be influenced by telephone communication:

> *Interestingly, (my most accurate examiner) has phoned me more (than other examiners) during the standardisation process.*
>
> (Team leader 1)

Table 1: Reasons for choosing communication mode

Reason for mode choice	*Mode*	
	Email	*Telephone*
Convenience	18	3
Speed	12	4
To personalise the message	4	13
Quantity of information	2	9
To check immediate understanding	3	11

The telephone mode also affords two-way discussion. Some examiners report that this helps them to better understand their team leader's intentions:

> *I think it speeds me up having [the feedback] on the phone because as we go through I have a chance to air further doubts ... which would otherwise require an email and then another one back.*
>
> (Examiner 5)

The digital marking system allows team leaders and examiners to remotely co-view scripts of interest on the computer whilst they discuss their perceptions of the quality of these scripts over the telephone. This type of activity is considered to be a positive strength of using the telephone communication mode. This is because it has a higher bandwidth than email and allows examiners to attend simultaneously to multiple sources of audio and visual information:

> *The fact that you can be talking about a script and a question and have it on screen and be talking on the phone at the same time. I know that is pathetic but I am not used to using technology in this way and I think my marking has improved since we have gone to the electronic version.* (Examiner 2)

Team leaders suggest that they prefer to use the telephone when they need to convey sensitive information. This is because there are concerns that a negatively received email may undermine examiner confidence, particularly when there are important and pressing issues with an examiner's marking to be dealt with. This is supported by data from the survey where three team leaders suggest that the telephone is useful for mediating a potentially negative examiner reaction to a serious message. In these instances team leaders are likely to adopt a more personal approach. The telephone is considered to be better for conveying an informal and friendly tone whilst communicating sensitive information to examiners.

Team leaders and examiners also recognise the benefits of using email to communicate feedback. Team leaders who prefer to use email highlight the way that it allows parcels of information to be communicated quickly to examiners. They also appreciate the way that it allows them to manage communication flow by being able to send and pick up messages when it is convenient for them to do so. They also recognise that this speed of information transfer can reduce cognitive load:

> *[Email] is quick and easy once it is working efficiently. Face-to-face standardisation can be very demoralising for examiners, especially if there is too much discussion and confusion can arise from this.* (Team leader 10)

> *[Email] gives faster responses … having to post 10 scripts and wait for the return is very time-consuming. Even continuing to mark while waiting for the return is difficult as changes will inevitably need to be made in the light of the team leader's comments. Being able to submit 3 scripts and get a response in 24 hours is more appropriate as you can remember the reasoning behind giving/not giving a particular mark.* (Team leader 12)

Email feedback also has the benefit of leaving a written record. Compared with the ephemerality of spoken communication, this feedback has a tangibility that can be used as a resource for examiner reflection at a later date:

> *Yeah, that was good because then I went back to it quite carefully and then I could make notes on my mark scheme as well then, that was good.* (Examiner 1)

Team leaders suggest that interpersonal issues can influence their feedback delivery, including communication mode choice. Team leaders and examiners acknowledge that prior co-working experience can influence how they interpret email messages, and support their development of a shared repertoire:

> *Again, if I did not actually know my team leader and know how he operates and how he traces things it might be a little more difficult to interpret the emails.* (Examiner 6)

In contrast, examiners who work with new team leaders report that a lack of any prior connection can affect the way that their work is interpreted:

> *I think he took it that I was rushing; actually I think I was over-thinking it.* (Examiner 1)

This issue also emerges in the team leader survey data. Out of the 18 team leaders, 14 report that their prior knowledge of an examiner influences the amount and/or the style of the feedback given.

Discussion

I argue in this article that the objective of team leaders when giving feedback is to construct messages that allow examiners insights into their thinking. This interaction is key to the development of an examiner's understanding of how to interpret and apply mark schemes in accordance with their team leader's views. For feedback to be useful, messages need to encourage synchrony. Drawing on the concept from Dennis and Valacich (1999), this synchrony involves the participants developing a shared focus so that both feedback message intention and reception are aligned.

My analyses suggest that team leaders consider a number of factors when choosing a feedback communication mode. These factors include the qualities of the feedback information being conveyed, consideration of the anticipated reaction of the examiner, and prior experience of co-working. When placed in the context of communication theory, these findings are not surprising. Media richness (Trevino et al., 1987) influences how communication works, and this richness includes synchronous and asynchronous dimensions (Whittaker, 2003). The data in this study suggests that this dichotomy only partially explains the complexity of the communication that takes place between team leaders and examiners.

I argue that synchrony is a nuanced concept which comprises two interacting types; logistic and semantic synchrony. Logistic synchrony describes the way that someone can arrange pieces of information so that they are coordinated, either physically (i.e., situated next to each other on a page) or temporally (i.e., situated next to each other in the course of a spoken conversation). Logistic synchrony can be influenced by the technology through which information is communicated, since different communication modes have different affordances (Sellen & Harper, 2002). For example, face-to-face communication affords participants the ability to respond to issues in conversation in an iterative, on-going fashion, whilst email conveys words that a recipient can reflect on at their convenience.

Semantic synchrony is a more abstract notion and describes the way that the meaning of a concept is coordinated between people (i.e., a commonly held shared meaning between individuals). The logistic arrangement of information can encourage semantic synchrony. This means that the two types of synchrony have a relationship, with the organisation of information influencing the development of meaning.

Drawing on theory that suggests that communication mode influences synchrony in general, I argue that each of these particular types of synchrony is afforded by different modes of communication. When team leaders choose a mode of communication, either synchronous or asynchronous, they are harnessing the logistic affordances of a communication mode to support the attainment of semantic synchrony. Figure 3 describes some of the ways through which the choice of communication mode influences logistic synchrony, and consequently supports semantic synchrony.

Telephone communication affords a different form of logistic synchrony from email in the way that it allows participants to manage the flow of communication in response to particular needs. For example, survey responses show that team leaders use the immediacy of the telephone communication mode to deal with any important and pressing issues with an examiner's marking. In terms of Distributed Cognition Theory, this synchronous communication mode allows information to be

	Synchrony type		
	Logistic	→	*Semantic*
Communication mode			
Synchronous (e.g., telephone)	• Push and pull based on demand	→	• Dual channel • Amount of information • Cross-reference and checking • Speed and quantity
Asynchronous (e.g., email)	• Push and pull based on convenience	→	• Words conveyed without loss • Workflow control • Review potential • Reduced 'MUM effects'[1]

Figure 3: Communication mode affordances and synchrony

1. The 'MUM effect' describes how, 'in general, individuals display greater reluctance to share bad news as compared to good news' (Dibble & Levine, 2010, p.3)

pushed and pulled between team leaders and examiners when they consider it to be necessary. As a result, the logistics related to the telephone mode allow information to be conveyed more quickly and in more detail. Studies report that verbal communication tends to include more words than text-based communication (King, McGugan, & Bunyan, 2008), due to the message composition time demands of typing. This affordance appears to be particularly valuable if there are serious marking issues to be dealt with. This finding tallies with other research which suggests that immediacy is a factor in feedback effectiveness (Barton & Wolery, 2007; Ahmed et al., 2012; Chur-Hansen & McLean, 2006; Burke et al., 2009; Hatzipanagos & Warburton, 2009; Mathieson, 2012; Li & De Luca, 2014).

This logistic synchrony, effectively co-locating the team leader and the examiner perspectives in the same time period, supports the strengthening of the semantic linkages between the participants. The timeliness afforded by the communication mode allows the participants to engage in discussion, where they can cross-check and clarify each other's intended and received meanings (Blair & McGinty, 2013). Team leaders and examiners also report that this dialogue is enhanced through the richness of the communication mode. Telephone use whilst viewing images of common scripts enables audio and visual information to be brought into connection with each other. This logistic positioning of information supports the building of meaning through the way that the messages from one information channel reinforce the message from another channel.

Examiners consider that timely delivery of information is a key characteristic of feedback that they feel has a positive effect on their marking. This message tallies with findings reported in other feedback studies (Owens et al., 2009; Dennen et al., 2007). One reason for this is that, where a feedback message is received by an examiner in the moment of marking, it is possible that they can factor the message into their practice without a great deal of extraneous mental load.

Asynchronous communication allows team leaders and examiners to manage their information flow differently from synchronous modes. The remote nature of examiner working means that it is sometimes difficult to coordinate work schedules so that conversations can take place. The use of an asynchronous communication mode such as email allows team leaders and examiners to send and pick up messages when it fits with their working arrangements. This element of control over information means that it is possible that the information recipient can also control the way that they fit the information into their task work. Through controlling information flow it is possible to ensure that distraction from extraneous information is minimised and does not interfere with the current focus of marking.

The use of email to asynchronously deliver feedback enables information to be delivered in a way that is not prone to listener bias since the writer controls the information included in the message. It also provides a written record of the interaction. This feature is a recognised strength of asynchronous communication, as it allows participants time for reflection when creating and interpreting information (Archer, 2010).

Finally, asynchronous feedback can overcome some interpersonal issues that can interfere with the accuracy of feedback information. Chur-Hansen and McLean (2006) observe that providing negative feedback is a demanding skill that requires participants to consider interpersonal issues when drafting feedback messages. This message is reinforced by Sussman and Sproull (1999) who draw attention to the finding in management research that information givers tend to distort feedback messages, particularly when giving bad news. Sussman and Sproull argue that this distortion is due to psychological anxiety surrounding the anticipated reactions of message recipients. This idea links the widely reported 'MUM effect' (Dibble & Levine, 2010).

Conclusions

The development of shared understanding across individuals is a key aspect of the development of expertise, helping to bring less experienced participants into the centre of a community of practice (e.g., Wenger, 2000; 1998). From a communication theory perspective, this shared understanding relies on the establishment of synchrony within interaction.

My interpretation of the study data suggests that when giving feedback, team leaders capitalise on the different technological affordances available to them as they attempt to build synchrony with examiners. In addition, the perceived nature of the information being conveyed and the interpersonal relations that a team leader has established with an examiner influence these choices.

An implication of the study findings is that communication systems need to allow some flexibility with regards to communication mode choice. The study data reinforce the point that understanding communication better means recognising that the 'text' of a communication is only part of the story, and that the mode of communication is also a contributory element to the meaning making process.

Synchronous and asynchronous communication modes afford logistic synchrony in different ways. Synchronous communication allows a greater quantity of information to pass between participants and is most responsive to push and pull demands. As a consequence, it is possible to cover more issues in less time. This suggests that a synchronous mode may be best suited to the initial marking stage when examiners have the most to learn. At the same time, there are potential weaknesses around some synchronous information modes. These weaknesses centre on the social anxiety of dealing with negative messages, and the ephemerality of verbal communication that leaves no record of interaction.

Asynchronous communication modes possess a logistic flexibility because they allow information flows to fit around busy work schedules. This is particularly useful in situations where it is not easy to coordinate work schedules across people. Feedback that is conveyed asynchronously

may be particularly suited to the purpose of reassuring examiners about their practice, conveying definitive interpretations of terminology, or passing on administrative arrangements. This is because the communication mode is most suitable for transmitting information and does not afford participants a great opportunity to immediately discuss or question the meanings implicit in the message, so the reception of message cannot be verified. Asynchronous communication supports semantic linkage through allowing participants time to reflect on the information.

References

Ahmed, M., Sevdalis, N., Paige, J., Paragi-Gururaja, R., Nestel, D., & Arora, S. (2012). Identifying best practice guidelines for debriefing in surgery: a tri-continental study. *The American Journal of Surgery*, *203*(4), 523–529. doi:10.1016/j.amjsurg.2011.09.024

Archer, J. C. (2010). State of the science in health professional education: effective feedback. *Medical Education*, *44*(1), 101–108. doi:10.1111/j.1365-2923.2009.03546.x

Barton, E. E., & Wolery, M. (2007). Evaluation of E-mail Feedback on the Verbal Behaviors of Pre-service Teachers. *Journal of Early Intervention*, *30*(1), 55–72. doi:10.1177/105381510703000105

Blair, A., & McGinty, S. (2013). Feedback-dialogues: exploring the student perspective. *Assessment & Evaluation in Higher Education*, *38*(4), 466–476. doi:10.1080/02602938.2011.649244

Burke, L., Marks_Maran, D. J., Ooms, A., Webb, M., & Cooper, D. (2009). Towards a pedagogy of work-based learning: perceptions of work-based learning in foundation degrees. *Journal of Vocational Education & Training*, *61*(1), 15–33. doi:10.1080/13636820902819917

Cameron, A. F., & Webster, J. (2005). Unintended consequences of emerging communication technologies: Instant Messaging in the workplace. *Computers in Human Behavior*, *21*(1), 85–103. doi:10.1016/j.chb.2003.12.001

Chur-Hansen, A., & McLean, S. (2006). On being a Supervisor: The Importance of Feedback and how to give it. *Australasian Psychiatry*, *14*(1), 67–71. doi: 10.1080/j.1440-1665.2006.02248.x

Cook, C. A., O'Shea, A. M., Young, A. L., & Stedmon, A. W. (1999). Helpful Feedback? Deriving Guidelines for Advanced Training Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(21), 1133–1137. doi:10.1177/154193129904302104

Dennen, V. P., Aubteen Darabi, A., & Smith, L. J. (2007). Instructor–Learner Interaction in Online Courses: The relative perceived importance of particular instructor actions on performance and satisfaction. *Distance Education*, *28*(1), 65–79. doi:10.1080/01587910701305319

Dennis, A. R., & Valacich, J. S. (1999). *Rethinking media richness: towards a theory of media synchronicity*. Presented at the 32nd Annual Hawaii International Conference on Systems Sciences, 5–8 January 1999, Washington, DC: IEEE Computer Society. doi:10.1109/HICSS.1999.772701

Dibble, J. L., & Levine, T. R. (2010). Breaking Good and Bad News: Direction of the MUM Effect and Senders' Cognitive Representations of News Valence. *Communication Research*, *37*(5), 703–722. doi: 10.1177/0093650209356440

Gass, S. M., & Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hatzipanagos, S., & Warburton, S. (2009). Feedback as dialogue: exploring the links between formative assessment and social software in distance learning. *Learning, Media and Technology*, *34*(1), 45–59. doi:10.1080/17439880902759919

Hollan, J., & Stornetta, S. (1992). Beyond being there. In P. Bauersfield, J. Bennett, & G. Lynch (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp.119–125). New York: ACM.

Johnson, M. (2014). A case study of inter-examiner feedback from a UK context: Mixing research methods to gain insights into situated learning interactions. *Formation et Pratiques D'enseignement En Questions*, *17*, 67–88.

Johnson, M. (2015). Articulation work: Insights into examiners' expertise from their remote feedback interactions. *Communication & Language at Work*, *1*(4), 28–52.

Johnson, M., & Black, B. (2012a). Feedback as scaffolding: Senior examiner monitoring processes and their effects on examiner marking. *Research in Post-Compulsory Education*, *17*(4), 391–407. doi:10.1080/13596748.2012.738965

Johnson, M., & Black, B. (2012b). What's Going On? Analysing Visual Data to Understand Context-Based Decision-Making Processes. *International Journal of Research & Method in Education*, *35*(3), 243–250. doi:10.1080/1743727X.2012.713222

King, D., McGugan, S., & Bunyan, N. (2008). Does it make a difference? Replacing text with audio feedback. *Practice and Evidence of Scholarship of Teaching and Learning in Higher Education*, *3*(2), 145–163.

Li, J., & De Luca, R. (2014). Review of assessment feedback. *Studies in Higher Education*, *39*(2), 378–393. doi:10.1080/03075079.2012.709494

Mathieson, K. (2012). Exploring Student Perceptions of Audiovisual Feedback via Screencasting in Online Courses. *American Journal of Distance Education*, *26*(3), 143–156.

Ofqual. (2013). *Review of Quality of Marking in Exams in A levels, GCSEs and Other Academic Qualifications: Interim Report* (No. Ofqual/13/5287). Coventry: Ofqual.

Owens, J., Hardcastle, L. A., & Richardson, B. (2009). Learning From a Distance: The Experience of Remote Students. *The Journal of Distance Education/Revue de l'Éducation À Distance*, *23*(3), 53–74.

Panahi, P., Birjandi, P., & Azabdaftari, B. (2013). Toward a sociocultural approach to feedback provision in L2 writing classrooms: the alignment of dynamic assessment and teacher error feedback. *Language Testing in Asia*, *3*(1), 1–10. doi:10.1186/2229-0443-3-13

Pfaffman, J. A. (2007). Computer-mediated communications technologies. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Roth, W.-M., & Lee, Y.-J. (2007). "Vygotsky's Neglected Legacy": Cultural-Historical Activity Theory. *Review of Educational Research*, *77*(2), 186–232. doi:10.3102/0034654306298273

Sellen, A. J., & Harper, R. H. R. (2002). *The Myth of the Paperless Office*. Cambridge, MA: MIT Press.

Sussman, S. W., & Sproull, L. (1999). Straight Talk: Delivering Bad News through Electronic Communication. *Information Systems Research*, *10*(2), 150–166. doi: 10.1287/isre.10.2.150

TechSmith. (2011). *Usability Testing for Software and Websites*. Okemos, MI: TechSmith Corp. Retrieved from http://www.techsmith.com/morae.asp

Trevino, L. K., Lengel, R. H., & Daft, R. L. (1987). Media Symbolism, Media Richness, and Media Choice in Organizations A Symbolic Interactionist Perspective. *Communication Research*, *14*(5), 553–574. doi:10.1177/009365087014005006

Vygotsky, L. S. (2014). Genesis of the Higher Mental Functions. In P. Light, S. Sheldon, & M. Woodhead (Eds.), *Learning to Think* (pp.32–41). Oxford: Routledge.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.

Wenger, E. (2000). Communities of Practice and Social Learning Systems. *Organization*, 7(2), 225–246. doi:10.1177/135050840072002

Whittaker, S. (2003). Theories and methods in mediated communication. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *The Handbook of Discourse Processes* (pp.243–286). Mahwah, NJ: Lawrence Erlbaum Associates.

Not dumbing down but stimulating up: Reading in the reformed GCSE Modern Foreign Languages classroom

Frances Wilson, Katherine Smith OCR and **Magda Werno** Cambridge International Examinations (The study was completed when the first and third authors were based in the Research Division)

Introduction

Modern Foreign Language (MFL)[1] skills in England have become a cause for concern. In a recent survey, only 36 per cent of employers were satisfied with school/college leavers' foreign language skills, while 70 per cent valued foreign language skills in their employees (CBI/Pearson, 2013). Within Europe, England has one of the highest percentages of citizens unable to hold a conversation in a language other than their native language (Coleman, 2009), indicating that many people in England are not benefiting from the economic and cultural benefits of being able to use a foreign language. Furthermore, uptake of MFL at General Certificate of Secondary Education (GCSE) and General Certificate of Education (GCE) Advanced level (A level) has been falling over the last 20 years (Malpass, 2014). This is partly as a result of societal and cultural factors, such as a shift in public opinion against greater European integration or the perception that English speakers do not need to learn a foreign language, and also due to the removal of MFL from the core curriculum at Key Stage 4 (KS4) in 2002 (Coleman, 2009; Macaro, 2008). Additionally, the transition between GCSE and A level is considered to be particularly difficult in MFL (Ipsos Mori, 2014), indicating a mismatch between the skills and knowledge taught at these levels. This may lead to fewer students deciding to continue the study of foreign languages post-16.

The current programme of reform for GCSEs and A levels in MFL aims to address these issues. New performance measures for schools, such as the English Baccalaureate (EBacc), a school performance indicator linked to GCSE, which requires students to study at least one language[2], have already led to an increase in the number of students studying languages at GCSE (Truss, 2013). Planned reforms to MFL GCSEs will increase the level of demand of the qualifications, and include changes to the content and assessment. In particular, the type and demand of texts which students are expected to read will change. It is hoped that these reforms will increase the foreign language competence of 16 year olds, and facilitate transition to A level. In this study we focus on this aspect of the reform to GCSEs: specifically, the increased focus on reading authentic texts at GCSE, and investigate how teachers may be supported to adapt to this change.

The planned reforms to GCSE aim to redress the balance between reading and the other skills of speaking, listening and writing. At GCSE, reading is often a neglected skill, which attracts less teaching time because it is considered to be easier than speaking, listening or writing (Brammell, 2011). Students often consider their reading skills to be stronger than other skills, even though their examination performance does not necessarily reflect that view (George, 2013). The Office for Standards in Education (Ofsted) has highlighted the teaching of reading to be a weakness in many schools, with schools often limiting reading materials to short texts found in textbooks or past examination papers (Ofsted, 2011). Currently the Office of Qualifications and Examinations Regulation (Ofqual) GCSE subject criteria specify equal weighting in the assessment to each of the four skills of reading, writing, speaking and listening (Ofqual, 2011). However, there is some flexibility. For example, the Oxford, Cambridge and RSA (OCR) MFL GCSE specification gives a weighting of 30 per cent each to speaking and writing, and only 20 per cent to reading and listening (OCR, 2012). Furthermore, currently, students are expected to read and comprehend a variety of forms of short text. These include signs, instructions, messages, emails, postcards, letters, internet sources, articles and brochures in the GCSE assessment. Although this may seem to be a wide range of text types, these texts are typically short, simple, and inauthentic (written for non-native speakers). The assessment model for the reformed GCSEs in MFL weights each of the four skills (reading, writing, speaking, listening) equally, and does not allow for any flexibility. The new subject criteria state that learners should:

> *Deduce meaning from a variety of short and longer written texts from a range of specified contexts, including authentic sources involving some complex language and unfamiliar material, as well as short narratives and authentic material addressing a wide range of relevant contemporary and cultural themes.*
>
> (Department for Education [DfE], 2014, p.6)

It seems, therefore, that students will be expected to read a much wider range of texts during their GCSE course, including some literary texts. Furthermore, these texts are much more likely to be authentic materials, originally written by native speakers for a native speaking audience, rather than texts targeted specifically at language learners at this level[3]. Although the weighting of reading (25 per cent) in the reformed GCSE will be relatively similar to the current GCSE, the change in the types of text which students will read is likely to require significant changes to the way in which reading is taught and learned, with implications for the whole MFL curriculum. This reform therefore provides an opportunity to re-examine the role of reading in the MFL curriculum, and to explore how different approaches to teaching reading might best support students' language learning. During times of change, curriculum support resources can provide opportunities for teacher learning which can help teachers to deepen their existing content and pedagogical knowledge, and enable them to adapt their existing knowledge and skills to navigate change (Loewenberg Ball & Cohen, 1996; Remillard, 2000).

In this article we first provide a brief overview of some psycholinguistic underpinnings of reading in a foreign language, and then relate this to the

1. We use the term 'Modern Foreign Language' to reflect current usage among awarding bodies, though acknowledge that the term 'modern language', 'living language', or merely 'language' is sometimes preferred.
2. This may be an ancient or a modern language.
3. See Gilmore (2007) for a discussion of the definition of an authentic text for use in an MFL classroom.

context of learning a language at GCSE and A level. We then present the findings from a focus group of MFL teachers which explored the use of reading resources at GCSE with respect to the reformed GCSE curriculum.

Reading and language learning

Reading in a foreign language

Reading is a complex skill, which requires the integration of lower and higher level cognitive processes to recognise words, and interpret the meaning of the text as a whole. Second language[4] (L2) learners are different from children learning to read, because L2 learners are typically already literate in their native language, but do not have fully developed knowledge of their L2. However, writing systems vary across languages. Where the native language (L1) and the L2 writing systems are similar, L2 learners might be able to transfer their L1 reading skills to the L2. For example, if both the first and second language use the same alphabet, and have similar sound-letter relationships ('bottom up' knowledge), then L2 learners may be able to use this L1 knowledge in their L2 (Frost, 2005; Goswami, 2008). Similarly, where the L1 and L2 are culturally similar, L2 learners may be able to apply background knowledge, and knowledge of text types ('top down' knowledge) to reading in their L2. L2 learners which have weak literacy skills in their L1 may need additional support to develop these skills in their L2. However, even where there are strong similarities between the L1 and the L2, L2 learners do still need to learn aspects of reading specific to their L2 (Nassaji, 2014).

Benefits for language learning

Reading in a L2 may support language development, by providing input or exposure to the L2. This may be particularly important where spoken language input is limited. Additionally, reading may provide exposure to grammatical forms which are infrequent in spoken language. Written language typically uses a wider range of vocabulary than spoken language. Reading may facilitate the acquisition of vocabulary by providing exposure to vocabulary in context, allowing learners to develop breadth (quantity) and depth (knowledge about usage) of vocabulary. However, L2 learners are not always able to accurately infer the meaning of unfamiliar vocabulary items from context (Nassaji, 2003), and can only do this accurately when texts are targeted at learners' proficiency levels (Waring, 2006). Where texts are more demanding, there are benefits from instruction which provide an explicit focus on vocabulary learning, because incidental vocabulary learning is unlikely to occur. Additionally, L2 learners need extensive practice to learn to recognise words rapidly and automatically, using texts which are relatively undemanding, and so can be read fluently and easily. Nation and colleagues estimate that L2 learners should know approximately 98 per cent of vocabulary in a text for this type of reading to be possible (Hirsch & Nation, 1992; Hu & Nation, 2000). It is therefore important for teachers to provide their students with texts which are relatively undemanding, to allow the development of reading fluency, and provide explicit instruction to support vocabulary development for more demanding texts. Since it is important for L2 learners to have extensive practice of reading, choosing texts which motivate students to read is particularly important for reading and language development. If students are not provided with interesting texts, they are unlikely to read them.

GCSE and A level context

Ofsted (2011) has identified the teaching of reading as a key area for improvement in MFL at secondary level, because it believes that reading is frequently not taught well. Often, schools do not exploit the range of authentic reading materials which are available, and typically rely on short texts in textbooks or past examination papers. Furthermore, Ofsted found that opportunities for students to read for pleasure or develop intercultural understanding were rare. However, the blame for these perceived weaknesses should not necessarily be given solely to teachers. Teachers' practice is shaped by the GCSE course which they are teaching. If current GCSE specifications (and assessments) do not support good practice in teaching reading, then it will be difficult for teachers to build effective reading activities into their teaching.

The planned reforms to MFL GCSEs will require students to study a wider range of text types than at present, including some literary texts. These texts are likely to be longer, and more demanding, due to an increased emphasis on the use of authentic texts than those which are currently used at this level. Students will be expected to respond to these texts in a variety of ways. For example, they will be expected to understand general and specific details in texts, identify the overall message and themes in a text, be able to scan for particular information and draw inferences in context (DfE, 2014). In this section we evaluate current GCSE specifications with respect to their suitability for supporting the development of reading skills, and suggest ways in which the specifications could be reformed to support the new subject criteria for reading, with a particular focus on vocabulary.

Hirsch and Nation (1992) show that it is necessary to know approximately 98 per cent of vocabulary in a text for adequate comprehension, and that for English, a vocabulary of approximately 2,000 words is needed if the most frequently used words are taught and learned. However, more recent analysis indicates that this may be a conservative estimate. Instead, it has been suggested that 2,000 words would provide only 80 per cent coverage, and 6,000–8,000 words are needed to reach the coverage required for comprehension (Milton, 2007). The Common European Framework of References for Language Learning and Teaching (CEFR) places GCSE Foundation Tier at level A2 (Basic User, Waystage/Elementary), and Higher Tier at B2 (Independent User, Vantage/Upper Intermediate). For B1, a vocabulary of about 2,000 words is specified, and for A2, 1,000 (Council of Europe, 2001; Milton, 2007). Since the move from A2 to B1 marks a shift towards independent language use, this suggests that a vocabulary of 2,000 words can be considered a minimum for relatively independent text comprehension.

At GCSE, vocabulary lists which specify the vocabulary which students will be expected to know in the examination[5] are provided by awarding bodies (e.g., OCR, 2009). For Foundation Tier, students are expected to know 1,400 words, with an additional 520 for the Higher Tier, a total of 1,920 words. Although it is difficult to compare counts of vocabulary, because there may be differences in what are considered to be different vocabulary items for the purposes[6], for the Higher Tier, at least, this approaches the level of vocabulary specified by Hirsch and Nation (1992)

4. We use second language (L2) to refer to any additional language first learned after the native language (L1), and typically, but not necessarily after the onset of puberty.
5. At Higher Tier some unfamiliar vocabulary which students are expected to understand from the context is included in the assessment.
6. For example, masculine and feminine forms of professions may or may not be treated as one item.

and the CEFR. However, even if these levels of vocabulary are specified for the GCSE assessment, it is not necessarily the case that students will be taught this vocabulary. Tschichold (2012) analysed the vocabulary featured in a series of French textbooks, *Encore Tricolore*, which is widely used at Key Stage 3 (KS3) and GCSE. Overall, Tschichold found that learners were exposed to more than 2,500 word families (which include different morphological forms of the same root word), again indicating that learners potentially have the opportunity to acquire sufficient vocabulary for reading authentic texts. Since we might expect that students would be exposed to additional vocabulary in other contexts, such as other reading and listening materials, then this can be considered to be a conservative estimate of the vocabulary which students may encounter. However, the fact that students may have been exposed to a vocabulary item, does not entail that they will have learned it, particularly if they only see or hear it a small number of times.

Milton (2006) estimated the vocabulary levels of students in each year of French study in a school in England (Table 1). In this school, students start learning French in Year 7, and have the opportunity to continue to study French to A level. Vocabulary levels were estimated based on the probability of correct responses in relation to a list of the most frequent words in French. Since there are likely to be discrepancies between the input which students had received and the words which are most frequent in French, these may be conservative estimates of the vocabulary which students actually knew. However, since the estimate is based on the most frequent words in French, it provides an indication of whether students are developing a vocabulary suitable for reading authentic texts.

Milton (2006) found that student vocabulary levels do not increase at the same rate from year to year. He found that between Year 7 and Year 9, students on average acquire only about 150 new words, which is then followed by a larger increase over the GCSE course. However, on average, students taking the GCSE only know about 852 words, substantially less than the vocabulary list specified for the Foundation Tier, and less than the 2,000 words suggested by Hirsch and Nation (1992). Even the maximum vocabulary level reached by any student in his study, 1,800 vocabulary items, falls short of this level. Furthermore, vocabulary levels increase dramatically between Year 11 and Year 12, the start of A level and Advanced Subsidiary level (AS) courses. This indicates that one cause of transitional difficulties might be the differences in vocabulary required at each level.

If learners of French are exposed to about 2,500 word families by the end of their GCSE course, why, on average, do learners acquire less than half of these words? Repeated exposure to vocabulary is necessary for acquisition to take place (Schmitt, 2008). Tschichold (2012) found that for many vocabulary items there were insufficient opportunities for recycling of vocabulary to support acquisition. This is consistent with George (2013), who reported that pupils felt that they did not have sufficient opportunity to revisit vocabulary. Furthermore, Milton's (2006) estimate was based on word knowledge relative to a list of the most frequent words. This might suggest that the vocabulary which is commonly taught at GCSE does not correspond to the most frequently used words. Tschichold (2012) compared the vocabulary presented in *Encore Tricolore* with the *Français fondamental* (a list of words which are accepted as representing the most frequently used French words) with the limitation that the list is based on oral speech recorded several decades ago. Overall, while 65 per cent of the vocabulary in *Encore Tricolore* was included in the *Français fondamental*, 35 per cent was not. Additionally, 40 per cent of the vocabulary in *Français fondamental* was not covered by *Encore Tricolore*, indicating that *Encore Tricolore* does not provide good coverage of the most frequently used French words. Furthermore, this is not limited to French: Häcker (2008) conducted a similar analysis of German textbooks, and obtained very similar results.

It is perhaps not surprising that teachers do not use authentic reading materials more widely at GCSE, since it is unlikely that students' prior vocabulary learning would enable them to access many authentic texts easily.

Why don't GCSE courses teach students the most frequent vocabulary? Current GCSE courses are organised into topics, such as Health and Sport, or Travel and the Wider World (OCR, 2012), so students' vocabulary learning is focused on a limited set of contexts. Textbooks frequently present vocabulary in a series of mini-dialogues, such as in (1a) and (1b), where students are presented with what is effectively a list of vocabulary from the same semantic (sports) and syntactic (nouns) category (Häcker, 2008). It is highly unlikely that all of the sports that are presented are among the most frequently used vocabulary in a language.

(1a) *Machst du Sport? "Do you do sport?"*
(1b) *Ich spiele ...Fußball "I play football"*
...Basketball "basketball"
...Tischtennis "table tennis"
...Tennis "tennis"
...Volleyball "volleyball" and so on.

This approach does have some advantages. It is important for motivational reasons for students to be able to talk about themselves, and their own context, and allows students the opportunity for genuine communication in the target language (Harris, Burch, Jones, & Darcy, 2001). Such a list does provide a reasonable chance that most students will learn the vocabulary for the sport that they are interested in, although this is still somewhat limiting. Häcker (2008) suggests that textbooks are unlikely to present, for example, the word "*Spielfeld*" (playing field), "*Tor*" (goal), or "*Spiel*" (match or game), which restricts further communication on the topic. However, these vocabulary items (or similar items) are on the vocabulary list specified by OCR for GCSE German (OCR, 2009). Further work would be needed to determine the extent of any mismatch between textbook coverage and the GCSE specification. However, since textbooks are designed, at least to some extent, with the goal of preparing students for examinations (Ofqual, 2012), then it is plausible that a lack of coverage in the textbook may be related to what is perceived to be commonly assessed.

Teaching vocabulary in the form of a list of semantically related words, such as a list of sports or a list of pets, is intuitively appealing because students learn words which are related. However, presenting lists of semantically related words which share the same syntactic category (e.g., a list of nouns) can lead to lexical interference, making the vocabulary more difficult to learn. Instead, it is easier to learn words which are semantically related, but from a mixture of syntactic categories (Tinkham, 1997). As a result, the form of presentation of new vocabulary may be unintentionally impeding students' acquisition of vocabulary. However, there are different approaches to the presentation of new words. In many German Bundesländer, (federal state) foreign language teaching is based around the reading of longer texts than those used in foreign language teaching in England. For example, Gruber and Tonkyn (2013) found that the average length of text in a French textbook in

Germany was 1,394 words, compared to 727 words for a French textbook in England. Furthermore, these texts typically cover more cognitively challenging topics, and a wider range of vocabulary and syntax (Gruber & Tonkyn, 2013). Such an approach allows students to acquire vocabulary which extends beyond their own context (e.g., related to the other culture), but, depending on the text, may not help students to talk about their own interests or context (e.g., a text about visiting Bavarian castles may be culturally relevant, but students may not learn vocabulary relevant to their own leisure interests). Vocabulary is frequently presented in conjunction with a text (see, e.g., the Green Line textbook used in Bavaria [Beile, Beile-Bowes, & Dick, 2001]), such that vocabulary is semantically related, but from a mixture of syntactic categories (Gruber & Tonkyn, 2013). Presenting vocabulary in such texts may facilitate the acquisition of deeper lexical knowledge, such as collocations, and level of formality. In a comparison of the writing skills of 14 to16 year old English L1 and German L1 learners of L2 French, Gruber and Tonkyn (2013) found that the German L1 learners had a larger vocabulary, and showed greater lexical diversity, even once total learning hours had been taken into account. However, the syntactic complexity of writing was not significantly different across the two groups.

It seems therefore, that reforms to GCSEs will need to change to teach more, and more appropriate vocabulary to support reading development. Furthermore, if students are to read authentic texts *during* their GCSE course, then they will need to have acquired sufficient and appropriate vocabulary before the end of the course, so that reading activities do not become primarily focused on vocabulary. Additionally, if students are to develop reading fluency, then they need to be able to access at least some of the texts which they read relatively easily. However, not all authentic texts are equally demanding. It would be possible, for example, for authentic texts to be graded by level, to allow a progression of texts throughout the course. It is likely that teachers would need considerable support to compile a list of such texts, and maintain a list of texts which are up to date. Milton (2006) noted that students moving from GCSE to A level study showed a large increase in vocabulary level; if reforms to GCSEs include changes to the quantity and nature of vocabulary, then this may facilitate transition to A level.

Thus far we have focused on the role of vocabulary at GCSE, because previous work in this area has focused on vocabulary. However, it seems plausible that a similar analysis could be undertaken for morpho-syntax (grammar). This may be more dependent on text type. For example, if students are expected to read narratives, then it is likely that they will encounter different forms of the past tense more frequently than other tenses. In French, for example, the past historic, or passé simple tense is predominantly used in written narratives, and so may be particularly useful if students read this form of text frequently. In German, for example, a form of the subjunctive, *Konjunktive I*, is used for reported speech, and may support the comprehension of newspaper articles. However, these structures are not included in the subject criteria (DfE, 2014; Ofqual, 2011), so teachers may need to provide strategy instruction to help students to access texts which use these structures.

In summary, the current GCSE courses do not fully support the development of reading skills, by not providing students opportunities to acquire sufficient, and appropriate vocabulary to access authentic texts. If students are to read a wide range of authentic texts, then the reformed GCSE should take a different approach to vocabulary, focusing on the most frequently used vocabulary. However, reading more lexically diverse texts can, if appropriate support is given, support the acquisition of vocabulary. Similarly, the relationship between the morpho-syntax which is taught, and that which is likely to be encountered by students when reading should be considered.

Method

Participants

Nine teachers participated in all aspects of the research. The majority of teachers reported teaching more than one language. Across the group, seven taught French, three taught German, and six taught Spanish.

Materials

Resource sets

Wilson, Carroll, and Werno (2014) developed a typology of the various dimensions of reading activities, based on psycholinguistic and pedagogical aspects of L2 reading. This typology was used to develop five sets of resources which exemplified different aspects of the typology (Table 1). Each resource set contained an example of an approach to teaching reading in each of French, German and Spanish, to ensure that the texts and activities used in the resources were accessible to all participants.

Resource set 1: This resource is an adaptation of a traditional extensive reading activity. Students are directed to a newspaper/magazine website in the target language, and asked to select articles to read which interest them.

Resource set 2: This resource is taken from GCSE Foundation Tier papers. Students are given a short informational text, which is undemanding in terms of vocabulary and grammar, and answer comprehension questions on the text. This resource was used because it was assumed that it would represent a type of reading activity which all teachers would be familiar with.

Resource set 3: Students are provided with a set of five short newspaper articles. Each newspaper article has a headline and picture. After brief discussion about which articles look interesting, students choose two or more articles to read, and write short quiz questions about information in the texts. These questions are then used in a class quiz.

Resource set 4: Students are given some language focused activities, to pre-teach vocabulary, and so on, before reading a poem.

Resource set 5: Students are given reading strategy-focused activities before reading a short story.

Feedback forms

Participants were provided with a feedback form for each resource set. The feedback forms asked participants to provide their views on the resources, and comment on the type of text used, the level of demand, the usefulness of the resource, and whether it was similar or different to resources which they currently use.

Procedure

Feedback forms

The resources and feedback forms were sent to participants before the focus group. Their individual views on these resources were gathered through feedback forms on each resource set.

Table 1: Dimensions of reading resources

	Resource 1 *Extensive reading*	*Resource 2* *GCSE question (from Foundation Tier paper)*	*Resource 3* *Quiz questions from authentic texts*	*Resource 4* *Language focused poetry activity*	*Resource 5* *Short story, strategy-focused*
Type of text: • informational text • short story • poem	Any – students choose	Informational	Informational – authentic newspaper articles	Poem	Short story
Types of supporting activities: • language focused (e.g., to teach necessary vocabulary/structures) • strategy-focused (e.g., to develop reading skills)	None	None	Some strategy, to facilitate reading of texts	Teaching of vocabulary (e.g., from Harris et al., 2001, Chapter 2)	Strategy-focused – use of paralinguistic features, such as pictures. Teachers provided with information about strategies, and how to make them explicit
Demand – may vary in terms of vocab and (morpho-) syntax. Levels to include texts which are: • accessible after KS3 • targeting KS4 • beyond KS4	If authentic texts, likely to be high, but choice offers varying levels	Low	High	Medium to High	Medium to High
Surrounding activities: • pre-reading activities (e.g., to teach necessary vocabulary) • during and post-reading activities – to structure reading activity • post-reading activities – which give the reading purpose	No – or just a minor post-reading check that reading has occurred	Post-reading comprehension questions	During and post-reading activities – identifying information suitable for generating quiz questions, then using quiz questions in a class quiz	Pre-reading	Pre-reading
Length: • shorter • longer	Any	Short	Short	Any	Longer
Choice: • unlimited choice of text • limited choice • no choice	Unlimited choice	No choice	Limited choice	No choice	No choice
Paralinguistic features: • some texts (with e.g., pictures) • some without	Possibly	No, or very restricted	A picture for each text	No	Yes, pictures related to text
'Top down'/'bottom up'	'Bottom up'	Low demand, so may support development of fluency, but somewhat short for this purpose	Not specified, but may develop background knowledge for 'top down' processing	'Bottom up'	'Top down'

Focus group

The participants' views on the resources, and the teaching of reading were further explored in a focus group. The main facilitator was responsible for the discussion and the timing of each focus group, in addition to ensuring that all participants had an opportunity to respond to each question. The focus groups were audio recorded and participants provided their consent for this.

Analysis

Feedback forms

For each topic on the feedback form, participants' responses were coded as to whether they expressed a positive or negative view of the resource, or a view which was felt to be either positive or negative. One researcher coded the data initially. Subsequently, the second researcher reviewed the

coding. Any instances where there was disagreement between the two researchers were discussed and the coding amended.

Focus group

The audio recordings of the focus group were transcribed. Two researchers coded the transcriptions. Initially, the coding scheme was based on the structure and content of the questions specified in the schedule. This coding scheme was then modified to take additional themes into account. One researcher coded the data initially. Subsequently, the second researcher reviewed the coding. Any instances where there was disagreement between the two researchers were discussed and the coding amended.

Results and discussion

The planned changes to reading in MFL GCSE represent a major change to the curriculum. The teachers who participated in the focus group felt that these changes, and the resulting impact on the wider MFL curriculum, would have a positive impact on their subject. This is consistent with the views of the Association for Language Learning (ALL), which stated that the introduction of a wider range of authentic materials at GCSE would enrich students' linguistic and cultural knowledge, and had the potential to increase student motivation for language learning (ALL, 2014). Teachers said that they felt that the topics (e.g., environment) currently taught at GCSE were boring for both students and teachers. They felt unable to teach more interesting material beyond the syllabus because they are constrained by the pressures of achieving good grades for their students. They further thought that the emphasis on controlled assessment for speaking and writing meant that reading was often not given priority, consistent with the findings from the Ofsted (2011) review of MFL teaching.

In the focus group, teachers examined resources which used a range of different types of authentic text, including literary and journalistic texts. They were very positive about increasing the use of poetry in MFL teaching at GCSE, because it supported the development of language skills. Poems were also valued because they are short, and so can be read relatively quickly. The teachers indicated that using poetry would be a significant departure from their current practice at GCSE, and so would need to be introduced carefully, but that the benefits of reading poetry would make this worthwhile.

> *Because in a poem, essentially, the words are chosen so carefully, because there are so few of them, the language skill leads into the reading skill anyway.*

Participants were enthusiastic about the use of literature in general, because they can extend students' cultural experience. Additionally, literature can be used as a starting point for a wide range of activities in the MFL classroom.

> *It makes people realise that they can read literature, they can read books, they can read short stories and get something out of it.*

> *I think the charm of this resource is also that language is actually secondary in here and it's everything else that comes first, that is, enjoying literature, enjoying maybe different ways of exploiting a longer piece of work, well, not so long a piece of literature and do something with it, that is, you know, design a story board or read it aloud, act it out, work with the drama department and do something together or design, this is just an idea off the top of my head, but design a poster advertising the story as a play for instance, so many different ways of enjoying this piece.*

Overall, they felt that such texts would be demanding, particularly for less able students. However, somewhat encouragingly, they felt that with appropriate support, GCSE students would be able to read such texts successfully. Teachers noted that by reading authentic texts, students would be exposed to unfamiliar vocabulary which is not commonly taught at GCSE, consistent with Häcker (2008) and Tschichold (2012). It seems therefore that a focus on the most frequent vocabulary used in the target language would facilitate the inclusion of authentic texts in the curriculum, and may be particularly useful if students use or continue to study the target language after the GCSE.

Teachers also noted that some authentic texts, such as online news articles may have paralinguistic features such as pictures and diagrams which can help students to access texts. Such features can help students to understand the gist of a text, even where the level of language is relatively high.

> *I think the big possibility about it, is that it mixes and I found quite a lot of, what you would call prose narrative material, but it was backed with data, it might be graphs or pictures, which to me makes it accessible to more people, more instantly. On a very simple level, you read a sports report, the students may at least understand the score or the result and then that gives them immediate access to the writing, which may be above them in itself.*

Authentic texts may therefore require students to make greater use of 'top down' processing strategies. If students become accustomed to reading authentic texts, then they are more likely to develop reading strategies, and become more confident at applying such strategies when they don't understand everything in a text. However, teachers may need support to teach such strategies, because there may be less need to ensure students are able to use such strategies in the current GCSE courses. Furthermore, students who have not developed effective 'top down' strategies in their L1 may find this particularly difficult. The texts used currently in GCSE reading assessments, which are typically not authentic, do have some advantages for such students, because they are accessible, and may help to develop confidence.

> *The ones doing Foundation, reading and listening, they need very structured and limited text, they need to understand to get confidence. For some less able, it's good, they can do days of the week.*

However, this type of text was felt to be very limiting, both for students and teachers, because it is not very interesting, and does not allow students to develop as learners.

> *I think if teachers cannot come up with something more interesting than that, after teaching for ten years, then it's a shame on us really and it makes our life less interesting as well as the students'.*

Teachers were aware of the benefits of reading for language development more generally, and thought that reading authentic texts would help students to develop vocabulary skills and grammatical knowledge, which benefits their productive language abilities.

> *And of course it feeds directly into their spoken and written ability then, doesn't it?*

> *Yes, and they pick up the high frequency vocabulary that comes up and time and time again and the core vocabulary that transfers, they pick*

up that and they realise. I think it gives a bit more gravitas than just the teacher saying, 'You need to learn these ten words.' When they see it coming up again and again, they believe that rather than believing you as a teacher.

I think it just helps, it reinforces what they've learnt in the lessons, but it helps them, you know, really see the language patterns, because we only teach it in isolation, sort of, single sentences or maybe very, very short paragraphs, but the more that they read and the longer of the text that they read, it's constantly back in their faces, being reinforced over and over again.

The proposed reforms to reading in GCSE MFLs will require teachers to change the way that they approach reading. Students would need time to become familiar with the new approach, and would benefit from using shorter, relatively accessible texts at first. A role for awarding bodies to provide guidance in this area was also identified.

They do need training, because we are training them in a different way at the minute, because we have to meet everything that's on the exam and make sure that they can pass the exam and cover the spec as best that we can.

Increasing the use of authentic texts in the MFL classroom would provide greater opportunities for cross-curricular work. In particular, the potential for forming greater links with English departments to support the development of literacy skills was seen as advantageous. However, teachers indicated that at present there are relatively weak links between MFL and English departments, and little sharing of expertise as a result.

Although this is quite sad, I'm looking at the German resource, a sad poem, it's open for, you know, it provides a lot of opportunity to go deeper into all sorts of topics and also cross curriculum with History, English, PSHE, Citizenship, you name it and it depends what you do with it.

How closely does your MFL department collaborate with the English department, because in schools I've worked in, they are two entirely separate entities which never work together, which rarely sit together, which rarely pull resources or even compare resources … We are helping them with their English skills, we are teaching them the rudiments of the language, which they may not be doing in English anyway. Looking at how English teachers teach English and reading English would help us enormously and vice versa, I think, because I don't think they are particularly well married.

Teachers identified some challenges associated with the teaching of reading in general, including the fact that some students do not read in any language, and the challenge of finding texts which will motivate students to read.

And the mental barrier from the students, they are not used to reading in any language, so why in Spanish?

Teachers found that textbooks were useful for covering the material which students would need for their GCSE examinations, however, textbooks were considered to be expensive and boring. Identifying suitable alternative materials was considered to be beneficial, but time consuming.

A lot of the reading to do is based on textbook, because, one, it's there, they have spent thousands of pounds on them, what's the point to never use them and, two, they are tailored to the exam and that ultimately does end up being the be-all and end-all and not getting people into A level because it's boring.

The thing I find is, if you're trying to be interesting and if you are trying to move away from the textbooks, you are working more hours than if you just stick to the textbook, do you know what I mean? It's so much harder work, if you are trying to be creative and trying to use authentic materials, that's why a central pool from OCR would be a really good idea, rather than each individual reinventing the wheel all the time.

However, there was consensus that moving towards an approach to reading at GCSE that encourages students to read in the target language could have long term benefits.

We do realise that we have to encourage reading in the target language as much as possible. That's going to help them with their language learning and hopefully the love of learning the language throughout the time at school.

Limitations

This study aimed to explore the use of reading resources at GCSE with respect to the reformed GCSE curriculum. However, it should be noted that only nine teachers participated in the focus group, so the extent to which it is possible to generalise to the wider population is limited.

Conclusions

The introduction of a greater focus on the use of authentic reading materials at GCSE presents an opportunity and a challenge, for awarding bodies and teachers alike. Although there is clear consensus that the current reading curriculum and assessments neither support good language learning nor motivate students to study languages, teachers indicated that there are some challenges, such as finding appropriate materials, associated with the reform. However, teachers welcomed the opportunity to teach a wider range of texts at GCSE level, which they felt would be motivating for students, support language learning, and literacy development more generally. It is therefore incumbent on awarding bodies and the developers of teaching resources to design assessments and resources which will facilitate the implementation of this reform.

References

Association for Language Learning. (2014). Announcement on GCSEs and A levels. *New A Level and GCSE content*. Retrieved from http://www.all-languages.org.uk/news/news_list/announcement_on_reformed_gcses_and_a_levels

Beile, W., Beile-Bowes, A., & Dick, G. (2001). *Learning English, Green Line, Ausgabe für Bayern, Tl.2, Schülerbuch, Klasse 6*. Stuttgart: Ernst Klett Verlag

Brammell, G. A. (2011). *Pupils' reading strategies in the higher tier teading tests of the GCSE German examination*. (PhD), University of Manchester, Manchester.

CBI/Pearson. (2013). *Changing the pace: CBI/Pearson education and skills survey 2013*. Retrieved from http://www.cbi.org.uk/media/2119176/education_and_skills_survey_2013.pdf

Coleman, J. A. (2009). Why the British do not learn languages: myths and motivation in the United Kingdom. *The Language Learning Journal*, *37*(1), 111–127. doi: 10.1080/09571730902749003

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Department for Education. (2014). *Modern languages. GCSE subject content*. London: Department for Education.

Frost, R. (2005). Orthographic systems and skilled word recognition processes in reading. In M. S. Snowling & C. Hulme (Eds.), *The science of reading: A handbook*. (pp.272–295). Oxford: Basil Blackwell.

George, L. (2013). An analytical approach to improving GCSE reading skills. In R. Churches (Ed.), *The quiet revolution: transformational languages research by teaching school alliances* (pp.17–21). Reading: CfBT Education Trust.

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, *40*(02), 97–118. doi:10.1017/S0261444807004144

Goswami, U. (2008). The Development of Reading across Languages. *Annals of the New York Academy of Sciences*, *1145*(1), 1–12. doi: 10.1196/annals.1416.018

Gruber, A., & Tonkyn, A. (2013). Writing in French in secondary schools in England and Germany: are the British really 'bad language learners'? *The Language Learning Journal*, 1–20. doi: 10.1080/09571736.2013.856456

Häcker, M. (2008). Eleven pets and 20 ways to express one's opinion: the vocabulary learners of German acquire at English secondary schools. *The Language Learning Journal*, *36*(2), 215–226. doi: 10.1080/09571730802393183

Harris, V., Burch, J., Jones, B., & Darcy, J. (2001). *Something to say? Promoting spontaneous classroom talk*. London: The National Centre for Languages (CILT).

Hirsch, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, *8*(3), 689–696.

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430.

Ipsos Mori. (2014). Why is the take-up of Modern Foreign Language A levels in decline? *A report for the Joint Council for Qualifications*: JCQ.

Loewenberg Ball, D., & Cohen, D. K. (1996). Reform by the Book: What Is – or Might Be – the Role of Curriculum Materials in Teacher Learning and Instructional Reform? *Educational Researcher*, *25*(9), 6–8, 14.

Macaro, E. (2008). The decline in language learning in England: getting the facts right and getting real. *The Language Learning Journal*, *36*(1), 101–108. doi: 10.1080/09571730801988595

Malpass, D. (2014). *Chapter 1 The Decline in Uptake of A-level Modern Foreign Languages: Literature Review*: JCQ.

Milton, J. (2006). Language Lite? Learning French Vocabulary in School. *Journal of French Language Studies*, *16*(02), 187–205. doi:10.1017/S0959269506002420

Milton, J. (2007). *French as a foreign language and the Common European Framework of Reference for Languages*. Paper presented at the Crossing Frontiers: Languages and the International Dimension, Cardiff.

Nassaji, H. (2003). L2 Vocabulary Learning From Context: Strategies, Knowledge Sources, and Their Relationship With Success in L2 Lexical Inferencing. *TESOL Quarterly*, *37*(4), 645–670. doi: 10.2307/3588216

Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching*, *47*(01), 1–37. doi:10.1017/S0261444813000396

OCR. (2009). *Vocabulary List GCSE German*. Retrieved from http://www.ocr.org.uk/images/68532-vocabulary-list-by-topic.pdf

OCR. (2012). *GCSE Languages Specification*. Retrieved from http://www.ocr.org.uk/Images/82560-specification.pdf

Ofqual. (2011). *GCSE Subject Criteria for Modern Foreign Languages*. Coventry: Ofqual.

Ofqual. (2012). *Textbooks: Risks and Opportunities. Research on the Risks and Opportunities Arising from Current Arrangements for Provision of Textbooks and Other Learning Resources for GSCEs and A levels*. London: Ofqual.

Ofsted. (2011). Modern languages Achievement and challenge 2007–2010. Manchester: Ofsted.

Remillard, J. T. (2000). Can Curriculum Materials Support Teachers' Learning? Two Fourth-Grade Teachers' Use of a New Mathematics Text. *The Elementary School Journal*, *100*(4), 331–350. doi: 10.2307/1002146

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. doi: 10.1177/1362168808089921

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*(2), 138–163. doi: 10.1191/026765897672376469

Truss, E. (2013). *Thousands more pupils studying rigorous subjects*. Retrieved from https://www.gov.uk/government/news/thousands-more-pupils-studying-rigorous-subjects

Tschichold, C. (2012). French vocabulary in Encore Tricolore: do pupils have a chance? *The Language Learning Journal*, *40*(1), 7–19. doi: 10.1080/09571736.2012.658219

Waring, R. (2006). Why extensive reading should be an indispensable part of all language programs. *The Language Teacher*, *30*(7), 44–48.

Wilson, F., Carroll, P., & Werno, M. (2014). *Not dumbing down but stimulating up: Reading in the reformed GCSE modern foreign languages classroom*. Retrieved from http://www.cambridgeassessment.org.uk/Images/206450-not-dumbing-down-but-stimulating-up-reading-in-the-reformed-gcse-modern-foreign-languages-classroom..pdf

Why do so few candidates score 4 out of 8 on this question? The issue of under-used marks in levels-based mark schemes

Sarah Hughes and **Stuart Shaw** Cambridge International Examinations

Introduction

Marks on a question which are rarely achieved by students are 'dead marks' or 'under-used' marks. Under-used marks may have a detrimental effect on reliability and can reduce the discriminative powers of a test (Bramley, 2001). It is necessary to ensure, therefore, that the full range of marks is used.

This study aimed to identify any under-used marks that occur in a History examination for 16 year olds. It explains this occurrence and presents recommendations to ensure that under-used marks are minimalised.

Context

The focus of the study was the Cambridge IGCSE®[1] (International General Certificate of Secondary Education) History Paper 1 (June 2013). The Cambridge IGCSE History syllabus looks at some of the major international issues of the nineteenth and twentieth centuries, as well as covering the history of particular regions in more depth. The emphasis is on both historical knowledge and on the skills required for historical research. Paper 1 contains 25 optional questions. Students are expected to answer three (two questions from Section A – 'Core Content' and one question from Section B – 'Depth Studies'). Each question comprises three parts: a, b and c, with maximum marks of 5, 7 and 8 respectively. The questions in Paper 1 are differentiated by outcome. Student responses are marked using a levels-based mark scheme. The levels of performance in the mark scheme relate to a progression of skills which are summarised in Table 1.

Three issues within the literature are relevant to the under-use of marks in a levels-based mark scheme:

1. The number and width of levels

The number of levels and the width of each level in a levels-based mark scheme can enable or hinder accurate ratings (Shaw & Weir, 2007). Shaw and Weir suggest that each mark point must be defined to clearly and unequivocally embody differing and distinct levels. If this is possible, then the more levels there are, the more precise the rating scale will be. However, markers must be able to clearly and consistently distinguish all of the different levels defined. Pollitt (1991) has argued that it is optimistic to even claim five reliable bands of performance (although this will depend on the target ability of the candidature and the construct being assessed).

Table 1: Skills assessed using the levels-based mark scheme

Question part	*Level and marks available*	*Skills rewarded at each level*
Part a	Level 0 (0 marks) Level 1 (1 mark) Level 2 (2–5 marks)	Answer lacking specific contextual knowledge Description
Part b	Level 0 (0 marks) Level 1 (1 mark) Level 2 (2–3 marks) Level 3 (4–7 marks)	Answer lacking specific contextual knowledge Description/identification Explanation
Part c	Level 0 (0 marks) Level 1 (1 mark) Level 2 (2 mark) Level 3 (3–5 marks) Level 4 (5–7 marks) Level 5 (8 marks)	 Answer lacking specific contextual knowledge Description/identification Explanation of one side of the argument Explanation of both sides of the argument Evaluation

Ahmed and Pollitt (2011) argue that it is more problematic to distinguish between marks within a level than between levels. Shaw and Weir (2007) report that markers seem to be able to effectively distinguish between three levels of performance within a band. Fowles (2009) found that, where there were many marks in a GCSE English mark scheme, markers under-used the extreme marks in a band, and differences between markers were exaggerated. Fowles concluded that fewer marks in a greater number of levels may result in greater marking consistency.

The levels-based mark schemes in the Cambridge IGCSE History examination paper have between three and six levels, with each level containing up to four marks.

2. Range of performance within a level

Ahmed and Pollitt (2011) argue that decisions about whether a response is very good or very poor (i.e., which level to apply) are easy judgements to make; it is decisions about which mark within a level to apply that are more difficult. Consequently, they propose that a mark scheme should help markers to score consistently those responses that are close to the extremes of a level. This suggests that descriptions at the extremes of the bands would be most useful.

Some levels in the Cambridge IGCSE History mark schemes function as points-based mark schemes. For example, in Level 2 in Part a and Part b questions which award description of events, one mark is awarded for

1. Cambridge International Examinations offers the International General Certificate of Secondary Education (IGCSE), which is a two-year qualification aimed at 14 to 16-year-olds. The Cambridge IGCSE encourages learner-centred and inquiry-based approaches to learning. It has been designed to develop learners' skills in creative thinking, inquiry and problem-solving, giving learners a sound preparatory basis for the next stage in their education. More than 70 subjects are available for study, and schools may offer any combination of these subjects. In some Cambridge IGCSE subjects, there are two course levels, known as the 'Core Curriculum' and the 'Extended Curriculum'. The 'Extended Curriculum' includes the material from the 'Core Curriculum', as well as additional, more advanced material.

each point given. In these cases the differentiation between each mark within the level is precisely described. In some levels in Part b and Part c questions there is less prescription and markers are required to make a judgement between marks within the level by following the marking guidance: "Where a band of marks is indicated for a level these marks should be used with reference to the development of the answer within that level."

3. A priori versus empirically-derived levels

Levels-based mark schemes for many general qualifications have been developed using an a priori approach based on the judgement and experience of expert syllabus developers and question writers (Lumley, 2002). Alternatively, some mark scale developers propose an empirical approach to developing level descriptors informed by the analysis of actual student performance (e.g., Milanovic & Saville, 1996; Weir, 2003). Upshur and Turner (1995) argue that an empirical method almost certainly guarantees that the whole range of the rating scale is employed thereby eliminating any under-used marks.

Research questions

The two research questions addressed by this study were:

1. Are any marks within the Cambridge IGCSE History paper under-used?
2. What factors impact on the occurrence of under-used marks?

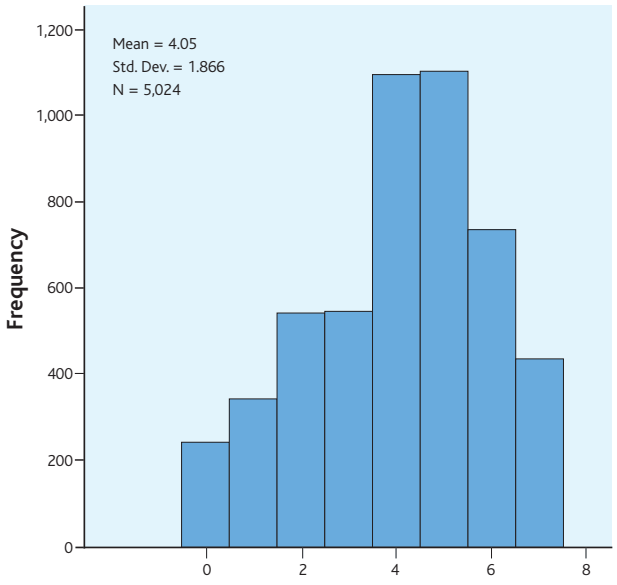


Figure 1: Mark distribution in which no marks are under-used

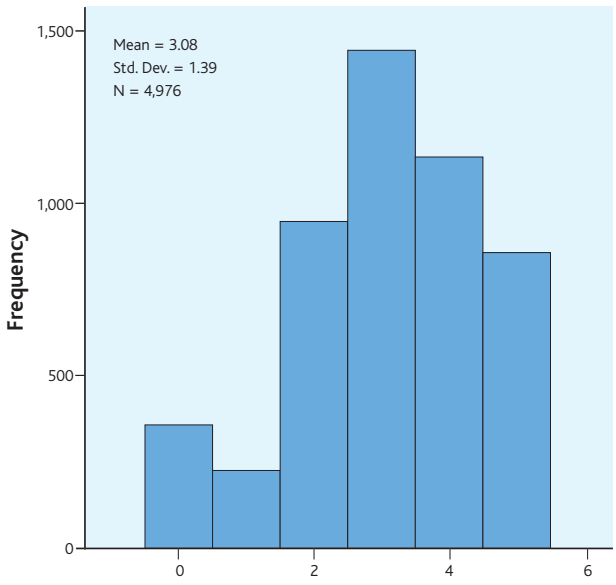


Figure 2: Mark distribution for Part a questions in which a mark of '1' is under-used

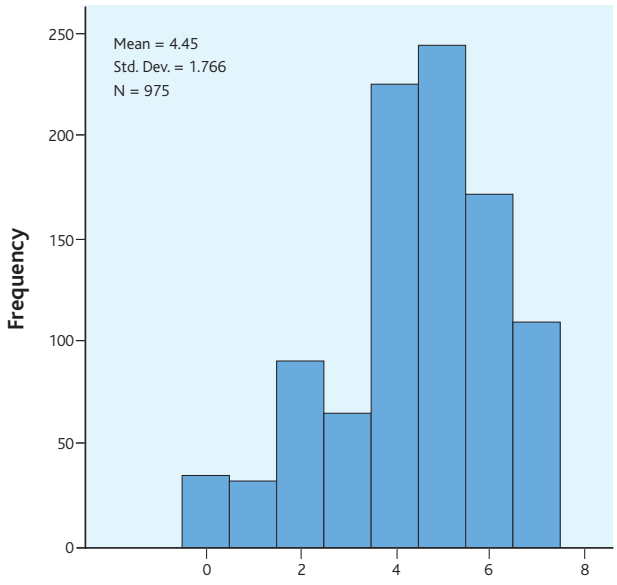


Figure 3: Mark distribution for Part b questions in which marks of '1' and '3' are under-used

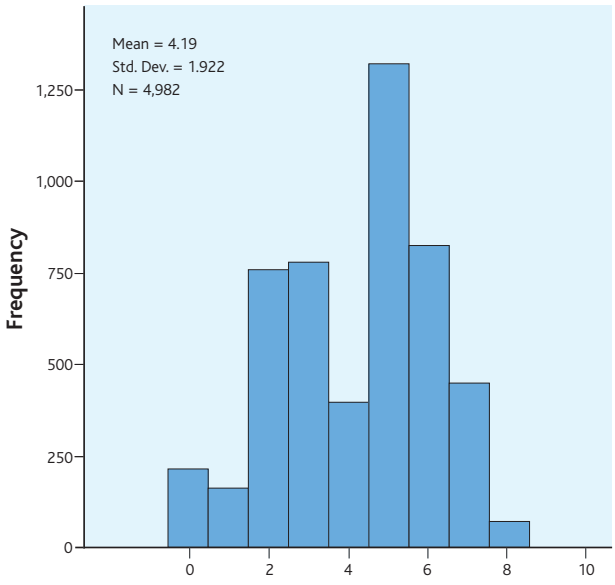


Figure 4: Mark distribution for Part c questions in which a marks of '1' and '4' are under-used

Methodology

Research question 1: Are any marks within the Cambridge IGCSE History paper under-used?

Traditional item analyses

Traditional item analyses were carried out. Analysis included estimates of Backhouse P[2] (a measure of internal consistency using average correlation of items) and histograms showing mark frequency distributions for each of the questions. The data set included 8,144 candidates who took the Cambridge IGCSE History Paper 1 in June 2013.

Rasch analyses

The data was analysed using the Rasch partial credit model (Masters, 1982) with FACETS software (Linacre, 2005). Three separate models were fitted: one examining Part a questions, one looking at Part b and one looking at Part c. Within each of these models, data from any of the 25 optional questions that were answered by at least 100 candidates were included. It is not necessary for every person to have attempted every question for the software to be able to estimate the person and item parameters, but it does require sufficient overlap of persons and questions such that there are no subsets of questions that have only been attempted by a subset of the persons. Separate overall difficulty parameters were estimated for each question. However, across the different questions, the sizes of the differences in difficulties between each successive mark (i.e., the category thresholds) were assumed to be constant. As illustrated later, these category threshold estimates were used to identify potentially under-used marks.

Research question 2: What factors impact on the occurrence of under-used marks?

Repertory Grid analyses

Structured interviews with four interviewees were carried out using the Repertory Grid Technique (Fransella, Bell & Bannister, 2004). This technique identifies the ways that a person construes (interprets or gives meaning to) his or her experience. The Repertory Grid Technique is underpinned by Personal Construct Theory, developed by George Kelly (1955/1991).

Four markers were interviewed either face-to-face or by telephone. Markers were given copies of six examination questions containing under-used marks and asked to consider two exam questions at a time. In order to elicit marker's constructs relating to a number of examination questions with under-used marks, markers were provided with the following prompts (Landfield, 1971):

- *Think of these two exam questions and why the under-used marks were rarely awarded.*
- *Are the two questions alike in terms of why the under-used mark is rarely awarded? If so, how are they alike?*
- *Are the two questions different in terms of why the under-used mark was rarely awarded? If so, how are they different?*

Inductive coding (using codes generated by the researcher) was adopted. Jankowicz (2004) suggests that inductive coding requires that the researcher:

- Identifies themes in the data
- Allocates each segment of data to a theme (or to more than one theme)
- Defines the themes
- Finds examples of each theme
- Finds the frequency of each theme.

Analysis of qualitative data was facilitated using MAXQDA, software for qualitative and mixed methods data analysis.

Findings

Research question 1: Are any marks within the Cambridge IGCSE History Paper under-used?

Traditional item analyses

The measure of internal consistency (Backhouse P) of 0.92 suggests that the questions on the paper are measuring the same construct. Figure 1 shows, for illustration, a mark distribution where no marks are under-used. Figures 2, 3 and 4 show examples of Part a, b and c questions (respectively) which exhibit under-used marks.

These findings were triangulated with those from the Rasch analyses.

Rasch analyses

Not unsurprisingly, score frequencies for the items included in the Rasch analyses (Table 2) show the same pattern or under-use as the mark distributions, that is: a mark of '1' in Part a questions, a mark of '3' in Part b questions, a mark of '4' in Part c questions and possibly a mark of '1' in all question parts.

Table 2: Score frequencies for Part a, b and c questions

Score	*Part a*		*Part b*		*Part c*	
	No.	*%*	*No.*	*%*	*No.*	*%*
0	1,409	6	786	3	816	3
1	977	4	861	4	1,048	4
2	3,188	15	2,149	9	3,576	15
3	5,051	23	2,529	11	3,759	15
4	5,138	23	5,055	21	2,287	9
5	6,195	28	5,384	23	5,599	23
6	-	-	4,526	19	3,782	16
7	-	-	2,529	11	2,983	12
8	-	-	-	-	479	2

'Category Probability Curves' showing the relation between the probability of a given category as a function of person location (in logits) were generated using the FACETS software and are shown in Figures 5, 6 and 7. Figure 5 indicates that for the least able students (with an ability measure on the x axis of -3 logits) the most likely outcome (with a probability of about 0.9) is a mark of '0'. As ability increases, the probability of getting no marks reduces. For students with ability of between about 0 and +0.8 logits, the most probable outcome is a mark of '3'.

Under-used marks are defined here as those which are not the most probable outcome at any point on the ability scale. Adams, Wu and Wilson (2014) propose that marks which are not most probable are not

2. Backhouse P is a measure of reliability (internal consistency) for tests with optional questions. Values range from 0 to 1, where higher values indicate more reliable tests.

necessarily evidence of a problem, but may be an indication of the relative number of respondents in each category. Nonetheless, Adams et al. (2014) recognise that these may be an indication that an item is not functioning as intended and may indicate issues with the discrimination of the question.

In Figure 5 a mark of '1' is not the most probable outcome for any ability. This is evidence that a mark of '1' is under-used. Figures 5, 6 and 7 show that:

- A mark of '1' is under-used in all question parts (a, b and c) indicating that it is rare for a student to be awarded the one available mark for an 'answer lacking specific contextual knowledge'.
- A mark of '3' is under-used in question Part b. This mark is awarded for description/identification.
- A mark of '4' is under-used in question Part c. This mark is rewarded for an explanation of one side of the argument.

Research question 2: What factors impact on the occurrence of under-used marks?

Four themes were identified by markers as prominent within the data: 1) the skills assessed; 2) marking issues; 3) questions features; and 4) topic content. Frequencies of each theme manifest in the data are shown in Table 3. It is interesting to note from Table 3 that, in terms of references to themes, 'Skills assessed' and 'Question features' were mentioned far more often than 'Marking issues' or 'Topic content'.

Table 3: Frequency of markers' references to themes

Theme	*No.*
1. Skills assessed	
Evaluation	5
Explanation	20
Description/identification	14
Balance of argument	14
2. Marking issues	
Overlapping marks	2
3. Question features	
Question language	12
Familiarity of question type	7
4. Topic content	
Familiarity of topic	2
Question parts	3

1. Skills assessed

A mark of '1' is rewarded for a response 'lacking in specific contextual knowledge'. The reasons for under-use of this mark appear to relate (in part) to marker expectations. Expectations are partly set by knowledge of which question part is being attempted. Part a questions, for example, are described by one marker as containing "less difficult content" and by another as "easier than b or c". Part c questions were described as neither easy nor simple, but as demanding. The following comments were illustrative of this point:

- this is a more difficult area to study,
- a 'sophisticated question' which 'ramps up' from identification skills to explanation skills.

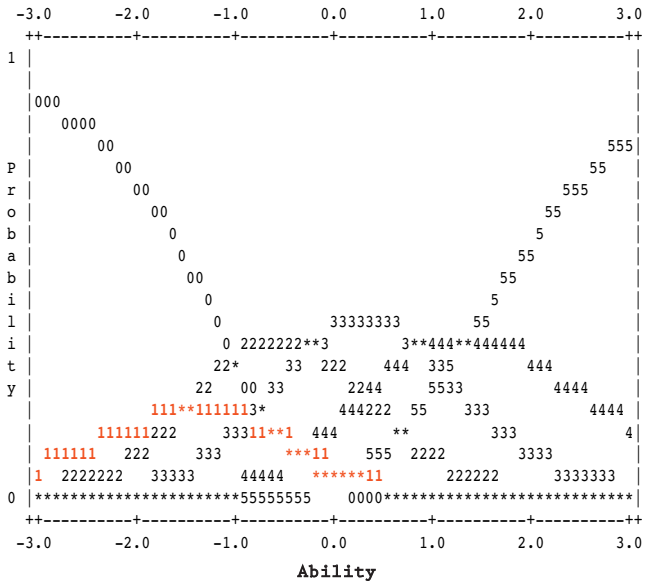


Figure 5: Category Probability Curves for Part a questions

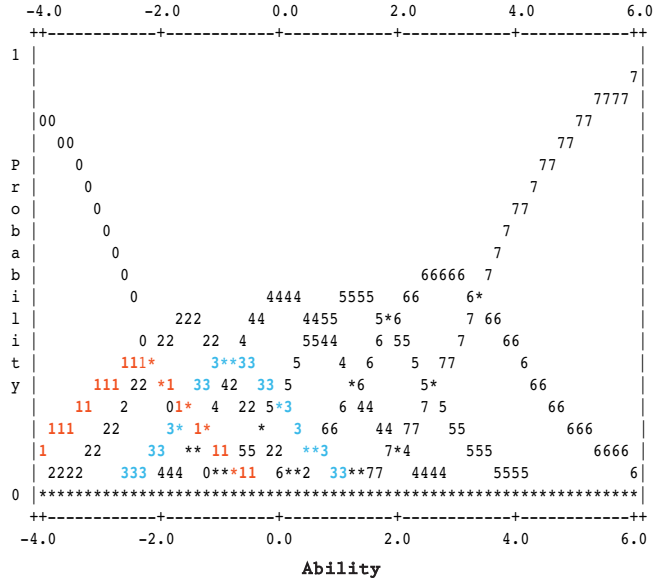


Figure 6: Category Probability Curves for Part b questions

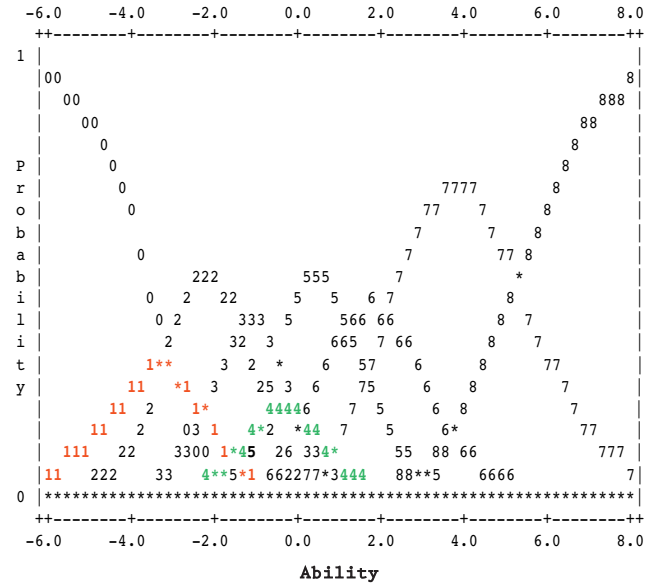


Figure 7: Category Probability Curves for Part c questions

Very few candidates provided an answer 'lacking specific contextual knowledge' and markers indicated that in Part a questions they expect candidates to achieve the highest available level (Level 2). Such an expectation may contribute to the under-use of a mark of '1' in Level 1.

The under-use of a mark of '3' in Part b questions is described as relating to progression from mark scheme Level 2 – which rewards students' ability to describe or identify, to Level 3 – which rewards students' ability to explain.

ABILITY TO DESCRIBE OR IDENTIFY

The stimulus material for one question is a picture of Hungarian refugees fleeing from Austria after the failure of an uprising. Markers reported that candidates find it easy to identify with people involved in historical events leading them to describe participants' experience of those events, rather than explain or evaluate the impact of those events. However, even students working at Level 2 (identifying and/or describing) are highly likely to achieve one mark in Level 3. This is because, when providing a narrative, students typically gain one explanation mark raising their performance to the bottom of Level 3. Markers will usually manage to identify some explanation in a response that is mainly descriptive. For example, one marker described how candidates might provide significant amounts of 'floundering description' and 'happen across' one explanation mark moving them from Level 1. This indicates that markers can recognise explanation in what is mainly a descriptive answer. Markers also described how this may account for students achieving a mark of '4' in Level 3 (in Part b questions) leaving a mark of '3' in Level 2 under-used.

ABILITY TO EXPLAIN

Markers described two features of questions which help candidates move from providing a descriptive response to providing an explanation:

1. Some topic areas foster explanatory responses. Markers described some topic areas which lend themselves to explanation, for example, The Cold War or the advantages of Stalin's economic policies. Such topics prompt explanatory rather than descriptive responses from candidates.
2. The teaching of higher order skills in some schools and colleges with particular focus on explanation and evaluation facilitates progression through the levels. One marker explained how some schools and colleges support candidates' progression through levels by teaching "the difference between description and explanation". Understanding the difference between description and explanation on the part of the candidate supports progression to Level 3 in Part b questions (for a mark of '4') and could, therefore, reduce the occurrence of a mark of '3' in Level 2.

ABILITY TO PROVIDE A BALANCED ARGUMENT

The Part c mark scheme contains five levels. In particular, Level 3 ('3'–'5' marks) rewards explanation of one side of the argument; Level 4 ('5'–'7' marks) rewards explanation of both sides of the argument. Whether a response includes a balanced argument or not appears to be the most significant factor in the under-use of a mark of '4' in Part c questions. Whilst not all students necessarily provide a balanced or two-sided argument, they are able to refer to both sides of the argument (even if one side is weak) and so achieve marks in Level 4, leaving at least one mark in Level 3 under-used. Markers explained why even an unbalanced argument is unlikely to be totally-one sided:

- *Teaching of both sides of an argument:* Students are likely to be trained to address both sides of any argument (but despite training, in an exam situation students may forget to focus on both sides). Each side of an argument may be given different treatment by teachers. For example, one marker said that "they teach Germany better than the Soviet Union". This may be because there are more knowledge or resources available to teach one context compared to another, or because contexts may differ in their complexity.
- *Distinct sides of an argument:* Where the information relating to the two different sides of the argument is distinct, markers reported that candidates are better able to provide a two-sided argument. One marker, for example, pointed to the clear benefits of the Nazi-Soviet pact to Germany on the one hand and the Soviet Union on the other.
- *Stimulus material:* Most stimulus materials are balanced in the view they present enabling candidates to see the two sides of the argument and achieve marks in Level 4 (giving a two-sided argument). Occasionally, stimulus material was described as having a bias towards one side of the argument (e.g., a quote from Hitler rallying his generals in May 1939 or Anthony Eden supporting the League of Nations).
- *Question wording:* Question wording supports candidates' engagement with both sides of the argument. In only one question did markers identify question wording which could be biased towards one side of an argument.

2. Marking issues

OVERLAPPING MARKS

A mark of '5' in Part c questions is an overlapping mark and is gained at the top of Level 3 or at the bottom of Level 4. Markers suspected that a mark of '5' in Level 3 is under-used in the same way as a mark of '4' in Level 3 is under-used . One marker estimated that about 80 per cent of students gaining a mark of '5' are doing so in Level 4. This suggests that, although a mark of '5' was not under-used overall, it may be under-used in Level 3. Of the three marks available ('3', '4' or '5') in Level 3, it is likely that both a mark of '4' and a mark of '5' are under-used. Since this study, the overlapping mark has been removed from the mark scheme.

The finding that both a mark of '4' and a mark of '5' are under-used[3] may shift the focus from the under-use of a single mark to the under use of Level 3 as a whole. Level 3 rewards candidates who present a developed one-sided argument and the findings suggest that students rarely give a developed one-sided argument.

3. Question features

- *Content compatible/obligatory language: Content-obligatory* language is content- or discipline-specific and academic in nature and it is necessary for learning key concepts in the subject (Fortune & Tedick, 2008). *Content-compatible* language goes beyond the student's subject learning. In the Cambridge IGCSE History paper, for example, the non-historical language of the question stem and instructions is *content-compatible*. *Content-compatible* language is uncomplicated and enables candidates to access the higher levels of the mark scheme and so leaves some marks under-used. Conversely (and much more rarely in the Cambridge IGCSE History paper)

3. Item Level Data does not differentiate between the two routes to a mark of '5'.

complex language provides a barrier to progression through the levels.

- *Familiarity of question types:* For a topic area that is regularly set, the question writer needs to devise new topic-related questions which are as yet unseen. Three of the four markers described how this may lead to "obscure", "sophisticated" or "unfamiliar" question types. This is problematic because the setter needs to devise novel questions which are neither obscure nor overly sophisticated. Questions which were described as set "in a new way" could hinder a candidate's ability to show his or her skills and thereby impede progress through the levels.

4. Topic content

- *Topic familiarity:* Markers made frequent references to familiar topic content (as opposed to unfamiliar topic content). For example, those which are "regularly" set topics which "occur every year", are "similar to past questions", are "well taught" and candidates "thoroughly know it". Another marker described how candidates "will have had experience of [the topic] because past papers include it and it is probably well prepared for". Findings suggest that familiar topics tend to result in better quality answers than new or rarely assessed content areas, and that familiar content allows candidates to move up the levels in the mark scheme taking them beyond performance that might occur with a less familiar topic.
- *Question parts:* Certain topics are associated with particular question parts. The question type may be unfamiliar to candidates because content usually assessed using a Part a or a Part b question, for example, is assessed using a Part c type question. This changes the skills being assessed in relation to content: Part c questions require candidates to provide a balanced argument and to evaluate the argument, neither of which are required in Part a or b questions.

Conclusion

In some cases the under-use of marks might be prompted by the accessibility of the questions (in the form of familiar topics and question types, predictable skills, simple non-historical language and clear and readable stimulus material). These features enable students to perform at their potential level of ability without being distracted by irrelevant (non-historical) demands in the questions, and so leave marks in lower levels under-used.

Accessible questions are not necessarily 'easy' questions; accessible questions can assess high-level skills and demanding content. What makes a question accessible is that it assesses the target skills and content without assessing factors irrelevant to the intended construct(s) (e.g., question wording or layout). Accessibility is desirable (in that it minimises 'construct irrelevant variance').[4] If a consequence of accessibility is that some marks are under-used, does it matter?

Adams et al. (2014) use the term 'middle score categories' for marks in a question that appear in the middle of the available mark range. Where few respondents achieve middle score categories, these marks are not useful and may indicate issues with the discrimination of the question (Adams et. al., 2014). As such, under-used marks may threaten validity.

This study shows that two of the three marks available in one of the mark scheme levels for the Cambridge IGCSE History Paper 1 are under-used. This raises questions about the purpose of this level and the validity of a mark scheme level which is rarely awarded. An empirical approach to developing level descriptors based on student performance, rather than a declared construct of performance (Upshur & Turner, 1995) could help reduce the number of under-used marks.

The study also suggests a need for clarification of the level descriptors which would support examiners making judgements between single marks (Ahmed & Pollitt, 2011; Shaw & Weir, 2007). Special attention also needs to be given to the setting of questions on any over-exposed topics which require novel questions to prevent repetition over time.

Research relating to levels-based mark scheme development and application would suggest a number of features that characterise effective practice which inform the construction and continuing improvement of general qualifications such as the Cambridge IGCSE History mark scheme. Almost all best practice described in the literature is already applied to the Cambridge IGCSE History mark schemes:

- using positively worded levels (Galaczi, ffrench, Hubbard & Green, 2011);
- providing indicative content (Tisi, Whitehouse, Maughan & Burdett, 2013);
- having a number of levels that markers can effectively distinguish (Ahmed & Pollitt, 2011);
- articulating clear and precise definitions of the distinction between different levels and between marks within a level (Shaw & Weir, 2007);
- reducing the use of relative adjectives (e.g., very frequent, fairly frequent, some) to differentiate descriptions of performance (Galaczi et. al., 2011);
- including examiner training and standardisation as part of the marking process (Baird, Beguin, Black, Pollitt & Stanley, 2011);
- ensuring expectations are made clear to students and teachers about the skills being assessed and the assessment model used to assess them (Sweiry, Crisp, Ahmed & Pollitt, 2002).

As a future line of research inquiry, one potential area of interest relates to the use of empirical evidence to establish the construct being assessed in each level (Upshur & Turner, 1995). This practice is not generally employed in the development of mark schemes for general qualifications as there is no pre-testing of the papers. However, it may be possible in an examination like Cambridge IGCSE History which uses similar mark scheme structures over time, to analyse student performance in one year and apply lessons to future papers and mark schemes.

The research reported here highlights concerns which were already articulated by senior examiners and which resulted in a re-designed mark scheme for the Cambridge IGCSE History paper ready for the June 2015 examination. The new mark scheme aims to support examiners judging the quailty of answers at the top of Levels 3 and 4 in Part c questions. The revised mark scheme has eliminated overlapping marks. Further research could usefully focus on monitoring student outcomes in Levels 3 and 4 in the future to evaluate this new mark scheme structure.

4. Variability in performance which is not attributable to the construct being assessed.

References

Adams, R. J., Wu, M. L., and Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement*, *72*(4), 547–573.

Ahmed, A. and Pollitt, A. (2011) Improving marking quality through a taxonomy of mark schemes. *Assessment in Education: Principles, Policy and Practice*, *18*(3), 259–278.

Baird, J. A., Beguin, A., Black, P., Pollitt, A. and Stanley, G. (2011). The Reliability Programme: Final Report of the Technical Advisory Group. In: *Ofqual's Reliability Compendium* (Chapter 20). Coventry: The Office of Qualifications and Examinations Regulation.

Bramley, T. (2001) The Question Tariff Problem in GCSE Mathematics. *Evaluation and Research in Education*, *15*(2), 95–107.

Fortune, T. W. and Tedick, D. J. (Eds.) (2008) *Pathways to multilingualism: Evolving perspectives on immersion education*. Clevedon, England: Multilingual Matters, Ltd.

Fowles, D. (2009) 'How reliable is marking in GCSE English?' *English in Education*, *43*(1), 49–67.

Fransella, F. Bell, R. and Bannister, D. (2004) *A Manual for Repertory Grid Technique (2nd Edition)* Chichester, UK: John Wiley and Sons.

Galaczi, E. D., ffrench, A., Hubbard, C. and Green, A. (2011) Developing assessment scales for large-scale speaking tests: a multiple-method approach, *Assessment in Education: Principles, Policy and Practice*, *18*(3), 217–237.

Jankowicz, D. (2004) *The Easy Guide to Repertory Grids*. Chichester, UK: John Wiley and Sons.

Kelly, G. A. (1955/1991) *The Psychology of Personal Constructs*. London, UK: Routledge.

Landfield, A. W. (1971) *Personal Construct Systems in Psychotherapy*. Chicago, USA: Rand McNally.

Linacre, J. M. (2005). *A User's Guide to FACETS Rasch-Model Computer Programs*. Available at: www.winsteps.com

Lumley, T. (2002) Assessment criteria in a large-scale writing test: What do they really mean to raters? *Language Testing*, *19*(3), 246–76.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika 20*(2), 149–174.

Milanovic, M. and Saville, N. (1996) Performance testing, cognition and assessment: selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem. *Studies in Language Testing 3*. Cambridge, UK: University of Cambridge Local Examinations Syndicate/Cambridge University Press.

Pollitt, A. (1991). Response to Alderson, Bands and scores. In J. C. Alderson & B. North (Eds.), *Language Testing in the 1990s*. London, UK: MacMillan.

Shaw, S. D. and Weir, C. J. (2007) Examining Second Language Writing: research and practice. *Studies in Language Testing 26*. Cambridge, UK: University of Cambridge Local Examinations Syndicate/Cambridge University Press.

Sweiry, Z., Crisp, V., Ahmed, A, and Pollitt, A. (2002) *Tales of the Expected: The Influence of Students' Expectations on Exam Validity*. British Educational Research Association Conference, Exeter, UK, September 2002.

Tisi J., Whitehouse, G., Maughan S. and Burdett, N. (2013) *A Review of Literature on Marking Reliability Research* (Report for Ofqual). Slough: National Foundation for Educational Research (NFER).

Upshur, J. and Turner, C. (1995) Constructing rating scales for second language tests. *ELT Journal*, *49*(1), 3–12.

Weir, C. J. (2003) A survey of the history of the Certificate of Proficiency in English (CPE) in the twentieth century. In C.J. Weir & M. Milanovic (Eds.). Continuity and Innovation: A History of the CPE Examination 1913–2002. *Studies in Language Testing 15*. Cambridge, UK: University of Cambridge Local Examinations Syndicate/Cambridge University Press.

Maintaining test standards by expert judgement of item difficulty

Tom Bramley Research Division and **Frances Wilson** OCR (The study was completed when the second author was based in the Research Division)

Introduction

This article describes two methods for using expert judgements about examination questions (items) to arrive at a cut-score (grade boundary) on a new examination paper where none of the items has been pre-tested. We wanted to see if we could exploit the wealth of data about item difficulty that has been available in the years since the majority of papers have been marked (scored) on-screen.

The General Certificate of Secondary Education (GCSE) and the General Certificate of Education GCE Advanced level (A level) are high-stakes curriculum-based examinations taken at age 16 and 18 respectively by pupils in England. They are offered by three Awarding Organisations (AOs), and schools can decide which AO's exams they enter their pupils for. Outcomes are reported on a grade scale (A* to G at GCSE; A* to E at A level, with U indicating 'ungraded' for both). From 2017, reformed GCSEs in England will be graded on a 1–9 scale. The full assessments normally consist of several components (e.g., written examination papers, practical or coursework assessment, portfolios, speaking tests, musical performances etc.). The assessments are usually graded at component level, and the overall grade is determined by aggregation rules which can vary considerably depending on the structure of the assessment (e.g., whether the assessment is 'linear', where all components are taken at the end of the course, or 'modular', where assessment units can be taken at various stages throughout the course). At component level, the grading process involves establishing the cut-scores (grade boundaries) on the raw mark scale that define the ranges of raw scores mapping to each grade.[1] A regulatory code of practice (Office of Qualifications and Examinations Regulation [Ofqual], 2011) sets out the mandatory aspects of this process, which requires the AOs to consider a variety of sources of evidence. Benton and Bramley (2015) show that these sources of evidence can be broadly classified as: i) evidence about the ability of the cohort of examinees; ii) evidence about the difficulty of the examination; and iii) evidence about the quality of work produced in the examination.

Setting the grade boundaries is essentially a standard-maintaining process (as opposed to a standard-setting process) where the aim is for

1. Only particular 'key boundaries' are established by the 'Awarding Committee' – the other boundaries are derived from these by interpolation rules. At A level, the key boundaries are at grades A and E.

the resulting grades to have the same meaning over time and across AOs. However, the decision on where to locate the grade boundaries has to combine the evidence from the three sources – which is not an easy task since they can relate to different conceptions of what it means to maintain a standard. The conceptual and practical problems created by this are well documented (e.g., Baird, Cresswell & Newton, 2000; Newton, Baird, Goldstein, Patrick & Tymms, 2007; Coe, 2010).

Traditionally the first and third of these sources have been the most dominant, and in recent years the first source (in the form of the 'comparable outcomes' method [Benton & Lin, 2011; Taylor, 2014]) has particularly constrained the possible locations of the boundaries. The second source of evidence (about the difficulty of the examination) has played a more minor role. This is partly because the high-stakes nature of the assessment makes pre-testing and re-use of items[2] impractical for security reasons (ruling out statistical evidence about item difficulty), and partly because there seems to be some scepticism about the ability of experts to provide accurate and reliable information about difficulty based on their informed judgements about examination questions.

This scepticism is based on the well-known method-dependence in the results of various item-based standard-setting methods (e.g., Glass, 1978); the variability in results within a given standard-setting method attributable to the judges (e.g., Clauser, Margolis & Clauser, 2014); the possibility that the expert judgements have a different implied scale unit from the student responses (Humphry, Heldsinger & Andrich, 2013); and the fact that there is often poor absolute agreement between judged item difficulty and empirical item difficulty (e.g., Bejar, 1983; Impara & Plake, 1998). However, there is also evidence in the research literature that in some circumstances there can be reasonable agreement between judged and empirical difficulty, particularly when judgements of experts are pooled; when those making the judgements are properly trained; when there is empirical data for judges to 'anchor' their judgements; and when judgements of difficulty are relative, rather than absolute (e.g., Brandon, 2004; Hambleton & Jirka, 2006; Attali, Saldivia, Jackson, Schuppan & Wannamaker, 2014). Most of the research on judgement of difficulty has been in the context of standard-setting methods for tests comprising objective (usually multiple-choice) dichotomous items. It is therefore an open question as to whether, and how best, expert judgement of difficulty can be used in the standard-maintaining context of GCSEs and A levels for those components where the majority of items are polytomous, short answer questions.

In this standard-maintaining context, a large and as yet largely untapped source of relevant information is available to guide the experts in their judgements of difficulty – statistical information about the performance of examinees on each item of previous versions of the component. The study reported here involved using this information in two different ways to derive estimates of the grade boundaries.

The first way was closely related to the Angoff standard-setting method and its extension to polytomous items – the 'mean estimation' method (e.g., Loomis & Bourque, 2001). If experts are able to estimate the mean score of examinees on the borderline of a particular grade for each item, then summing these estimates across the items would give an estimate of the grade boundary for that grade. It is possible to provide the experts with the actual mean scores of boundary examinees for each item on previous versions of the relevant component. Therefore their estimates of mean scores for boundary examinees on the new version of the component can be guided by their judgements of similarity of new items to previous items. The advantage of this method is that it can be applied well before the examination is taken – that is, it does not require any statistical information from the examination itself. A potential disadvantage is that it still requires numerical estimates from the experts which, as we have shown, there are reasons to doubt that they can make reliably enough.

The second way was technically more complex, but required less from the experts. It was based on an idea first suggested in Bramley (2010):

> *On [this] approach ... the awarding panel would identify questions on the current paper which are similar enough to questions on a previous paper for it to be reasonable to expect performance on them to be equivalent. Now the argument would be along the following lines: 'Last year the borderline grade C examinees (with a test score of 40) averaged 1.2 out of 2 on question 7a, which required them to label a diagram of a cell. This year's question 3b was practically identical, and examinees who averaged 1.2 out of 2 scored 42 on the test overall, suggesting a mark of 42 would be appropriate for this year's boundary.*
>
> (Bramley, 2010, p.35).

Thus here the task was merely to find item(s) on previous versions that were similar or identical to each item on the current version. Empirical item characteristic curves (EICCs) were created for each item on previous versions of the component, and for each item on the new component (as soon as the data was available). These EICCs were smoothed plots of item score against total score using the TRANSREG procedure in SAS® software with a smoothing parameter of 50.

Estimates for a particular grade boundary on the new version were derived from the following steps:

1. Find (from the relevant EICC plot or by tabular interpolation) the item score corresponding to the grade boundary on each previous item judged to be similar or identical to a new item;
2. Find (from the relevant EICC plot or by tabular interpolation) the total score on the new test corresponding to each item score identified in step 1;
3. Average the total scores obtained in step 2.

The process is illustrated graphically for a single item in Figure 1.

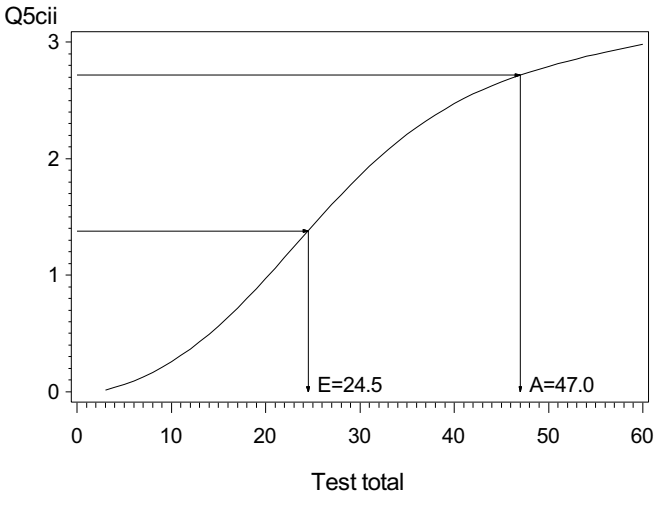


Figure 1: Smoothed plot of item score against test score for the new test

2. In this article, an 'item' is a subpart of a larger question. For example, Q1a (part of Q1) would be considered to be an item. 'Question' and 'item' are used interchangeably depending on the context.

In Figure 1 the previously obtained (smoothed) mean score on the item judged to be similar/identical to Q5cii for grade A boundary examinees was 2.72 and for grade E examinees it was 1.38. Reading across from these values on the y-axis and down to the x-axis gives the A and E boundaries implied by performance on this 'pseudo-anchor' item. In Figure 1 these are 47.0 and 24.5 respectively. Averaging these implied boundaries at each grade across all similar/identical items identified by the judges gives the A and E boundaries produced by this method. Since this method requires statistical data from the new test, it can only be carried out after the examination has been taken and marked (scored).

The present study applied these methods to an A level Chemistry component in order to estimate grade boundary locations for the June 2014 paper. Interest focused on the variability of results across experts and the agreement of the resulting grade boundaries with the actual grade boundaries that were eventually set. The procedures and results are described below. The discussion relates the above methods to existing methods in the standard-setting and maintaining literature and considers further their strengths and weaknesses.

Method

The A level Chemistry component was chosen for the research because it had a large, stable entry, with reliable statistical data at item level available for more than six previous versions.[3]

Participants

Because of the need to maintain security of the examination materials, only two experts – the Principal Examiner (Expert 1 [Ex1]) and the Chief Examiner (Expert 2 [Ex2]) – were used. They had already seen the June 2014 paper (because they were involved in setting the questions) so the exercise could be completed before the date of the live examination. The second author (Au) also completed the task to allow comparison between expert and non-expert judgements; her highest qualification in Chemistry was A level, though she had recently worked on a number of research projects relating to Science qualifications.

Materials

The experts were sent the following materials:

- Past question papers and mark schemes (scoring rubrics) from each previous version from January 2011 to June 2013 (six papers in total);
- The question paper and mark scheme to be taken in June 2014;
- A spreadsheet which listed the (smoothed) mean mark achieved by examinees on the grade A and grade E boundary on each subquestion on each previous paper. The specification reference[4] was provided for each subquestion;
- A spreadsheet listing each subquestion and the specification reference for each subquestion on the Summer 2014 paper, for participants to fill in their responses.

Task

The experts were asked to estimate the mean marks that would be obtained on each subquestion of the June 2014 paper by examinees on the grade E boundary, and examinees on the grade A boundary.

They were instructed to use the specification reference information provided for the Summer 2014 paper to identify past questions which assessed similar or identical material to use as the basis for their judgement. If they could find a question that was identical, or nearly identical, in the past papers to the Summer 2014 question, they were instructed to use its previous empirical values of facility for A and E boundary examinees as their estimate. If questions which were similar, but not identical, could be found, then they were instructed to use those past values as a basis for their estimate, but to modify their estimate according to their judgement of the effect of any differences on difficulty. Where no similar questions could be found, they were asked to use their own judgement. They were asked to explain the rationale for their decisions.

Results

Of the 33 subquestions on the June 2014 paper, there was only 1 where neither expert was able to identify anything similar or identical in any of the previous 6 papers. For 2 of the subquestions, 10 and 11 similar previous subquestions were identified. Most commonly between two and six similar subquestions were identified. The judges differed considerably in how many similar questions they identified in total – Ex1 found 38, Ex2 found 90, and Au found 30.

Table 1: Agreement between the experts in number of instances of similar questions identified

Judge	*No. of questions*
Ex1 only	4
Ex2 only	44
Au only	3
Ex1 and Ex2*	29
Ex1 and Au*	16
Ex2 and Au*	30
Ex1, Ex2 and Au	14

*Regardless of whether the third judge identified that question.

Table 1 shows that, given that Ex2 identified many more similar questions than the other two judges, it was rare for the other two to find a question that he had not identified. There were 14 instances where all 3 judges agreed, representing 12 questions on the June 2014 paper. (For two questions, all three judges agreed there were two similar questions that had been asked before).

Mean estimation method

Figure 2 shows the agreement between the two experts' judgements at grade A and E. The estimates of mean marks have been scaled by the number of marks available for the subquestion (in other words the graphs show the estimated facility values) in order to highlight more clearly where there were differences between the experts. It is clear from Figure 2 that Ex1 generally estimated higher values than Ex2 at both A and E, although the judgements at both grades were reasonably well correlated (see Table 2).

3. This component was available to examinees in January or in June. A completely new version was created each time.
4. A code indicating the area of the specification (syllabus) relevant to the question. For example, 1.2.1j referred to the subsection of the specification about electron structure (1.2.1) and the 'assessable learning outcome' j which was 'classify the elements into s, p and d blocks'.

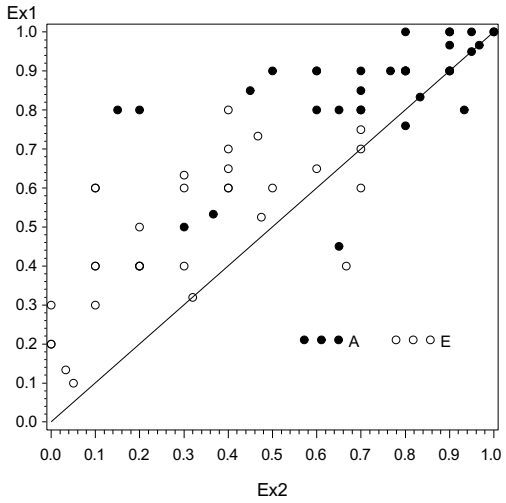


Figure 2: Estimated facility value for Expert 1 *v* Expert 2 at grade A (dots) and grade E (circles)

Table 2: Inter-correlations of estimated and actual facility values for boundary examinees. Grade A above and right of the main diagonal, grade E below and left. (N=33)

	Ex1	*Ex2*	*Au*	*Actual*
Ex1		0.60	0.20	0.41
Ex2	0.77		0.13	0.66
Au	0.51	0.56		0.14
Actual	0.72	0.76	0.52	

Table 2 shows that there was better agreement among the judges at grade E than grade A. It is particularly noticeable that there was a higher correlation between the two experts at both grades than between either of them and Au, although this represents expert agreement about relative difficulty rather than absolute difficulty, as seen in Figure 2. The correlation of the expert judgements with the actual values later obtained when the paper 'went live' was quite high for Ex2 at both grades, but only at grade E for Ex2, whose correlation with the actual values was only 0.41 at grade A.

Table 3: Grade boundaries implied by the judgements

Judge	*Grade A sum*	*Grade A*	*Grade E sum*	*Grade E*
Ex1	50.6	51	30.2	30
Ex2	44.2	44	21.8	22
Au	49.7	50	27.2	27
Mean (all)	48.2	48	26.4	26
Mean (experts only)	47.4	47	26.0	26
Actual boundary		46		26

At grade A the boundaries implied by the judgements of the two experts were 7 marks apart and at grade E they were 8 marks apart, a rather discouraging finding given an average grade bandwidth (difference between grade boundaries on the raw mark scale) of around 5 marks on previous versions of this component. However, the mean of their judgements did equal the eventual actual boundary at grade E and was only 1 mark too high at grade A. The boundaries implied by the researcher's (non-expert) judgements were between those of the experts, and did not significantly affect the mean at grade E, but raised it at grade A to a value 2 above the actual boundary.

Similar items method

Applying the second method for deriving grade boundaries involved deciding which items on the June 2014 test should be deemed similar enough to previous items to justify using the previous statistics. An initial list contained 15 items from the June 2014 paper where both experts and the researcher had identified the same similar previous question. The first criterion we used for selecting similar questions from this initial list was to choose questions with the same maximum mark as the previous question and where at least one of the expert judges had used the same value as the previous value as their estimate (i.e., at least one expert thought the difficulty would be the same). This criterion produced slightly different lists of questions for grade A and E (because the experts could have used the same value as the previous question for one grade but modified the value for the other). At grade A, 9 items worth 17 marks met the criterion, and at grade E 8 items worth 15 marks did.

Next, we tried a stricter criterion for anchor item selection, choosing only those items where all three judges had agreed and where all had used the same values as previously for both the A and E boundaries. This only identified two items worth 4 marks in total.

Finally we tried just using the judgements of Ex2 (who had identified the most similar items, and whose correlations with the actual values were highest at both grades), taking only those items where he had used the same value as the previous statistics (i.e., that he judged to be of identical difficulty to a previous question). This gave 11 items worth 20 marks in total. The text of these items, but not their layout, is shown in Table 5 in the Discussion section.

Table 4: Grade A and E boundaries implied using three different criteria for identifying similar items

	No. of similar items	*No. of marks*	*A*	*A (rounded)*	*E*	*E (rounded)*
Criterion 1	8 (A)/9 (E)	15/17	46.3	46	24.2	24
Criterion 2	2	4	48.7	49	27.0	27
Criterion 3	11	20	46.8	47	25.4	25
Actual				46		26

Table 4 shows that all three criteria for identifying similar previous items led to similar estimates for the grade boundaries, and that in all cases they were close to the actual boundary. This suggests that the second method of deriving boundaries may be a better way to use the data available.

Discussion

The aim of this study was to investigate two methods for deriving grade boundaries on an exam paper, using expert judgements about the questions. The first method could be characterised as statistically-informed expert judgement about question difficulty. It transplants the Angoff standard-setting method into a standard-maintaining context where the experts can use statistical information about performance of grade boundary examinees on items in past versions (forms) of the test to inform their judgements about the likely performance of grade boundary examinees on the new test. The main advantage of the method is that it provides a source of evidence about the difficulty of the new test which is independent of any data about the performance of

examinees on the new test, and hence can be applied before the test is taken. Furthermore, the previous statistical information can be used intelligently by the experts to guide their judgements based on how similar they think each new item is to one that has been asked in the past. The results of this study suggested that the first method would need to involve more expert judges to reduce the impact of variability among the judges on the final outcome. Although the outcomes from this study were close to the actual boundaries, this may have been due to luck considering how far apart the two experts' judgements were (in absolute terms).

The second method could be characterised as non-parametric Item Response Theory (IRT) common item equating using expert judgement of item similarity to define pseudo-anchor items. As far as we are aware, this is a new method, although the idea of using smoothed EICCs has appeared in a recent article by Zu and Puhan (2014), who described a method for test equating without IRT. (In the Zu & Puhan research the context was more directly analogous to IRT equating – no judgements of difficulty were made, and all the items on both the 'old' and the 'new' test had been used before).

The main extra assumption needed here (apart from the obvious one that examinees on the grade boundary would have the same expected score as previous boundary examinees on the items judged/deemed to be identical) is that the grade boundaries were set in the 'correct' place on all previous versions of the component. The standard-maintaining procedures for A level examinations focus (rightly) on outcomes at the aggregate level. This means that anomalies and discrepancies can arise at unit/component level, particularly in assessments with a modular structure. That is, that the grade boundaries on units taken in January are not aligned with those taken in June (e.g., Black, 2008; Bramley, Dawson & Newton, 2014). However, this potential drawback would not apply to assessments with a more simple structure, and will be less relevant in future to A levels and GCSEs in England, where the reformed versions of both qualifications will be linear.

Table 5: Content of the questions for which an effectively identical counterpart was identified by Expert 2

June 2014 Question	*Max mark*	*Content of June 2014 question*	*Content of a previous question*
Q1bi	1	Antimony exists as a mixture of isotopes. What is meant by the term *isotopes*?	Tungsten has many isotopes. Explain what is meant by isotopes.
Q1biii	1	Complete the table below to show the atomic structure of ^{121}Sb. (Table with heading 'protons' 'neutrons' and 'electrons').	The mass number of one isotope of tungsten is 184. Complete the table below to show the atomic structure of this tungsten isotope. (Table with heading 'protons' 'neutrons' and 'electrons').
Q1ci	3	The relative atomic mass of antimony is 121.8. Define the term *relative atomic mass*.	Define the term *relative atomic mass*.
Q1dii	2	SbCl3 molecules are polar. Explain why.	Molecules of BF3 contain polar bonds, but the molecules are non-polar. Suggest an explanation for this difference.
Q2a	2	A compound used as a fertiliser has the following composition by mass: C, 20.00%; H, 6.67%; N, 46.67%; O, 26.66%. Calculate the empirical formula of this compound.	A compound containing magnesium, silicon and oxygen is also present in rock types in Italy. A sample of this compound weighing 5.27g was found to have the following composition by mass: Mg, 1.82g; Si, 1.05g; O, 2.40g. Calculate the empirical formula of the compound. Show your working.
Q4ai	2	H2O has hydrogen bonding. Complete the diagram below to show hydrogen bonding between the H2O molecule shown and one other H2O molecule. Include relevant dipoles and lone pairs. Label the hydrogen bond.	The solid lattice structure of ammonia, NH_3, contains hydrogen bonds. Draw a diagram to show hydrogen bonding between **two** molecules of NH_3 in a solid lattice. Include relevant dipoles and lone pairs.
Q4b	1	Draw a *'dot-and-cross'* diagram to show the bonding in CO2. Show outer electrons only.	Draw a *'dot-and-cross'* diagram to show the bonding in a molecule of CH_3Cl. Show **outer** electrons only.
Q5a	3	The Periodic Table is arranged in periods and groups. Elements in the Periodic Table show a periodic trend in atomic radius. State and explain the trend in atomic radius from Li to F. *In your answer you should use appropriate technical terms, spelled correctly*. (Answer space with one line for 'trend' and six lines for 'explanation').	Periodicity is a repeating pattern across different periods. First ionisation energy shows a trend across Period 2. The first ionisation energies of lithium, carbon and fluorine are shown in **Table 5.1** below. (Table giving the 3 values). Explain the trend across Period 2 shown in **Table 5.1**. *In your answer you should use appropriate technical terms, spelled correctly*.
Q5biii	1	A student adds a small volume of aqueous silver nitrate to an aqueous solution of bromide ions in a test-tube. The student then adds a similar volume of dilute aqueous ammonia to the same test-tube. Write an ionic equation for any precipitation reaction which occurs in the student's tests. Include state symbols.	A student was provided with an aqueous solution of calcium iodide. The student carried out a chemical test to show that the solution contained iodide ions. In this test, a precipitation reaction took place. Write an ionic equation, including state symbols, for the reaction that took place.
Q5cii	3	Under different conditions, chlorine reacts differently with aqueous sodium hydroxide. A disproportionation reaction takes place as shown below. (Chemical equation given.) State what is meant by disproportionation and show that disproportionation has taken place in this reaction.	The hydrides of Group 5 elements all exist as gases at room temperature. Phosphine gas, PH3, can be prepared by adding phosphorus, P4, to warm concentrated aqueous sodium hydroxide as shown in the equation below. (Chemical equation given.) Using oxidation numbers, explain why this is a disproportionation reaction.
Q6ai	1	Group 2 carbonates undergo thermal decomposition. Write the equation for the thermal decomposition of calcium carbonate. Include state symbols.	Magnesium carbonate, $MgCO_3$, is present in dolomite [...] A student collected two equal-sized samples of dolomite. These samples were put into two labelled test-tubes, **A** and **B**. Tube **A** was heated until there was no further change in mass and was then allowed to cool. Tube **B** was left unheated. Write the equation for the action of heat on the magnesium carbonate present in tube **A**.

Unlike the first method, this method requires item level score data from the new test, so cannot be used before the test is taken (although the judgements of item similarity can of course be made in advance). However, the results of this study suggest it may be a more stable method in the sense of being less susceptible to differences among the judges, because the outcomes did not vary much when the criteria for identifying similar items were varied and different subsets of items were used to derive the boundary on the new test. Boundaries derived by this method were within 3 marks of the actual boundary in all cases, even with a very strict criterion for identifying pseudo-anchor items which identified only two items (worth 4 marks). Relaxing the criterion to include 20 marks worth (a third of the paper) of pseudo-anchor items produced boundaries only 1 mark away from the actual boundaries. Future work could explore whether it is better to use a few highly similar or identical questions as pseudo-anchors, or a larger number of less similar questions.

A further notable advantage of the second method is that it does not require the experts to make any judgements about mean scores of examinees, but just requires them to identify similar or identical questions. This should help to strengthen stakeholder confidence in the results by removing doubts about the ability of judges to make absolute (or indeed relative) judgements of difficulty. It also arguably allows the experts to give a more objective rationale for why they have deemed questions to be similar or identical, one which is more open to public scrutiny. For example, in a context such as in England where exam papers are published after they have 'gone live' the AO could publish the list of questions and their previous similar/identical counterparts that were used to derive the boundaries. Such a list is provided for this study in Table 5.

A limitation of this study is that it only involved two expert judges. This was necessary to meet the strict security conditions surrounding research in a 'live' setting. However, the expertise of the judges was as high as it would be possible to achieve, involving as it did the most experienced and senior examiners involved in setting the examination. It is an open question whether widening the pool of judges would improve the estimates (by reducing random error) or make them worse (by introducing bias and/or random error from relative lack of expertise).

Further work is needed to determine how well the findings from these two methods will generalise to other assessments than the one studied here. It seems reasonable to expect that judgements about question difficulty or similarity are better suited to exams consisting of relatively objective shorter answer questions where 'question difficulty' is a more tangible concept, and where it may be easier to define the knowledge and skills required to answer a question. The component used in the study reported here had a very large entry with hundreds of examinees on each mark point in the score distribution. More technical work could focus on the numbers of examinees needed to allow satisfactory estimates of the EICCs and experiment with varying the smoothing parameter to see what effect it has on the results.

In conclusion, both methods show promise for use in operational standard-maintaining procedures in contexts where tests are constructed to the same general specifications, but there is no possibility for pre-testing or re-use of items. In the context of GCSEs and A levels in England, these methods could provide a good source of relatively independent evidence about the difficulty of the questions, which could complement the existing evidence about the ability of the examinees and the quality of their work in the examination.

Acknowledgements

We would like to thank Gwen Low, formerly of OCR, and the two examiners who acted as expert judges for their support with this research.

References

Attali, Y., Saldivia, L., Jackson, C., Schuppan, F., & Wanamaker, W. (2014). Estimating item difficulty with comparative judgments. *ETS Research Report Series, 2014*(2), 1–8. Available online at: doi: 10.1002/ets2.12042

Baird, J.-A., Cresswell, M., & Newton, P. (2000). Would the real gold standard please step forward? *Research Papers in Education*, *15*(2), 213–229. Available online at: doi: 10.1080/026715200402506

Bejar, I. I. (1983). Subject Matter Experts' Assessment of Item Statistics. *Applied Psychological Measurement*, *7*(3), 303–310. Available online at: doi: 10.1177/014662168300700306

Benton, T. & Bramley, T. (2015). *The use of evidence in setting and maintaining standards in GCSEs and A levels*. Cambridge: Cambridge Assessment. Retrieved from http://www.cambridgeassessment.org.uk/Images/204310-maintaining-standards-discussion-paper-tom-benton-and-tom-bramley.pdf

Benton, T., & Lin, Y. (2011). *Investigating the relationship between A level results and prior attainment at GCSE*. Coventry: Ofqual.

Black, B. (2008). *Using an adapted rank-ordering method to investigate January versus June awarding standards*. Paper presented at the Fourth Biennial EARLI/Northumbria Assessment Conference, Berlin, Germany. Available online at: http://www.cambridgeassessment.org.uk/images/109767-using-an-adapted-rank-ordering-method-to-investigate-january-versus-june-awarding-standards.pdf

Bramley, T. (2010). 'Key discriminators' and the use of item level data in awarding. *Research Matters: A Cambridge Assessment Publication*, *9*, 32–38. Available online at: http://www.cambridgeassessment.org.uk/Images/109986-research-matters-09-january-2010-.pdf

Bramley, T., Dawson, A., & Newton, P. E. (2014). *On the limits of linking: experiences from England*. Paper presented at the 76th annual meeting of the National Council on Measurement in Education (NCME), Philadelphia, PA.

Brandon, P. R. (2004). Conclusions about Frequently Studied Modified Angoff Standard-Setting Topics. *Applied Measurement in Education*, *17*(1), 59–88. Available online at: doi: 10.1207/s15324818ame1701_4

Clauser, J. C., Margolis, M. J., & Clauser, B. E. (2014). An Examination of the Replicability of Angoff Standard Setting Results within a Generalizability Theory Framework. *Journal of Educational Measurement*, *51*(2), 127–140. Available online at: doi: 10.1111/jedm.12038

Coe, R. (2010). Understanding comparability of examination standards. *Research Papers in Education*, *25*(3), 271–284. Available online at: doi: 10.1080/02671522.2010.498143

Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement*, *15*(4), 237–261. Available online at: doi: 10.1111/j.1745-3984.1978.tb00072.x

Hambleton, R. K., & Jirka, S. J. (2006). Anchor-based methods for judgmentally estimating item statistics. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp.399–420). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Humphry, S., Heldsinger, S., & Andrich, D. (2013). Requiring a consistent unit of scale between the responses of students and judges in standard setting. *Applied Measurement in Education*, *27*(1), 1–18. Available online at: doi: 10.1080/08957347.2014.859492

Impara, J. C., & Plake, B. S. (1998). Teachers' Ability to Estimate Item Difficulty: A Test of the Assumptions in the Angoff Standard Setting Method. *Journal of Educational Measurement*, *35*(1), 69–81. Available online at: doi: 10.1111/j.1745-3984.1998.tb00528.x

Loomis, S. C., & Bourque, M. L. (2001). From Tradition to Innovation: Standard setting on the National Assessment of Educational Progress. In G. J. Cizek (Ed.),

Setting Performance Standards: Concepts, Methods and Perspectives (pp.175–217). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Newton, P. E., Baird, J.-A., Goldstein, H., Patrick, H., & Tymms, P. (Eds.). (2007). *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.

Ofqual. (2011). *GCSE, GCE, Principal Learning and Project Code of Practice*. Coventry: Ofqual. Available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/371268/2011-05-27-code-of-practice.pdf

Taylor, M. (2013). *GCSE predictions using mean Key Stage 2 level as the measure of prior attainment*. Report to Joint Council on Qualifications (JCQ) Standards and Technical Advisory Group (STAG). Revised 26/06/13.

Zu, J., & Puhan, G. (2014). Preequating with empirical item characteristic curves: an observed-score preequating method. *Journal of Educational Measurement*, *51*(3), 281–300. Available online at: doi: 10.1111/jedm.12047

Research News

Karen Barden Research Division

Conferences and seminars

European Conference on Educational Research (ECER)

The ECER conference took place in Budapest, Hungary in September, under the theme of *Education and Transition – Contributions from Educational Research*. Nadir Zanini, Research Division, presented a paper on *The importance of teaching styles and curriculum in Mathematics: Evidence from TIMSS 2011*. The paper was co-authored with Tom Benton, Research Division.

Simon Child, OCR, presented a paper co-authored with Research Division colleagues Prerna Carroll and Ellie Darlington on *The role of assessment in facilitating student transition to 'active' citizenship*.

Royal Statistical Society (RSS)

The RSS 2015 Annual Conference took place in Exeter in September. Now in its 23rd year, the RSS conference has gained prestige for its focus on current statistical issues, how it fosters the exchange of ideas and information, and the quality of its speakers. Tom Benton, Research Division, presented a paper on *How statistics determine examination results in England*.

British Educational Research Association (BERA)

Held in September at Queen's University, Belfast, Northern Ireland, the BERA Annual Conference was an opportunity to develop new research ideas, and to build new research relationships within the research education community. Based on work undertaken by the Research Division, Cambridge Assessment colleagues presented the following papers:

Carmen Vidal Rodeiro, Research Division: *An investigation into the numbers and characteristics of candidates with incomplete entries at AS/A level*.

Simon Child, OCR, Ellie Darlington and Tim Gill, Research Division: *An investigation of the motivations underpinning student and teacher topic choice in History qualifications*.

Jessica Bowyer (née Munro), Research Division: *The assessment of creativity and innovation in Design and Technology*.

Martin Johnson, Research Division: *Reading between the lines: exploring the characteristics of feedback that support examiners' professional knowledge building*.

Tim Gill, Carmen Vidal Rodeiro and Nadir Zanini, Research Division: *Students choices in Higher Education*.

Jackie Greatorex, Lucy Chambers, Filio Constantinou and Jo Ireland, Research Division: *Piloting a method for comparing examination question paper demands*.

Jackie Greatorex, Tom Sutch, Jessica Bowyer, Karen Dunn, Research Division, and Magda Werno, Cambridge International Examinations: *Investigating a new method for standardising essay marking using levels-based mark schemes*.

Victoria Crisp, Research Division: *Validity and comparability of assessment: how do these concepts relate?*

Magda Werno, Cambridge International Examinations, Frances Wilson, OCR, and Prerna Carroll, Research Division: *Translation in the reformed ancient languages GCSEs*.

Gender differences – the impact of secondary schooling – boys or girls, who's winning?

A Cambridge Assessment conference on 'Gender differences' took place in London in October. The conference brought together more than 600 experts from within the education and assessment community both at the conference and online, with over 30 countries represented. The audience heard from speakers from around the world who unpacked the complex range of issues that surround gender differences in secondary education and how they might be tackled to attempt to remove, or at least start to reduce, the gap between girls and boys. Presentations included the following papers:

Tim Oates, Assessment, Research & Development: *An analysis of the gender divide – from primary school to workforce*.

Tom Benton, Research Division: *Attitudes to learning – questioning the PISA data*.

Tom Bramley, Carmen Vidal Rodeiro and Sylvia Vitello, Research Division: *Gender differences at GCSE*.

Agnieszka Walczak and Ardeshir Geranpayeh, Cambridge English Language Assessment: *The Gender Gap in English Language Proficiency? Insights from a Test of Academic English*.

Further details of the conference, videos of the proceedings and additional resources can be found on our website at: http://www.cambridgeassessment.org.uk/events/gender-differences-conference-2015/

International Association for Educational Assessment (IAEA)

The 41st IAEA Conference was held in Kansas, USA in October and gathered educational leaders from around the world to share innovative ideas of educational assessment techniques aimed at improving quality in education. The conference theme was *The Three Most Important Considerations in Testing: Validity, Validity, Validity*. Stuart Shaw, Cambridge International Examinations, presented papers on *What Makes for a Sound Validity Argument? Exploring Criteria for Evaluation – the Strength of Validation Evidence; Critiquing Kane's argument-based approach to validation;* and *Language Rich: Insights from Multilingual Schools* based on work co-authored with his colleagues Helen Imam and Sarah Hughes.

Stuart also presented a plenary session on *Testing as a Positive Force: Changing the Reality and the Perception*. Neil Wade, OCR, presented a paper on *Validity Issues in the Reform of Practical Science Assessment: An English Case Study*.

Association for Educational Assessment – Europe (AEA-Europe)

The 16th AEA-Europe Annual Conference took place in Glasgow, Scotland in November under the theme of Assessment and Social Justice. Several colleagues from Cambridge Assessment attended the conference and the following papers were presented:

Carmen Vidal Rodeiro and Nadir Zanini, Research Division: *The A* grade at A level in England: an evaluation of its use as a selection tool for HE courses*.

Ellie Darlington, Research Division: *The value of Mathematics: optimising opportunity in higher education and beyond*.

Jackie Greatorex, Lucy Chambers, Filio Constantinou and Jo Ireland, Research Division: *Piloting a method for comparing examination question paper demands*.

Tim Oates, Assessment, Research & Development: *To catch moving standards; how small do the holes in the regulatory net need to be?*

Tom Bramley, Research Division, and Chris Wheadon, No More Marking Limited: *The reliability of adaptive comparative judgment*.

Tom Bramley, Research Division, and Frances Wilson, OCR: *Maintaining standards by expert judgment of question difficulty*.

Sarah McElwee, Cambridge English Language Assessment: *Widening participation and positive impact in the design and administration of university admissions tests: A case study of the BioMedical Admissions Test*.

Sarah Hughes and Stuart Shaw, Cambridge International Examinations: *Why do so few candidates score 4 out of 8 on this question? The issue of fairness and under-used marks in levels-based mark schemes*.

Stuart Shaw, Cambridge International Examinations, and Martin Johnson, Research Division: *Insights into teacher-developed Pre-U assessment: opportunities and challenges*.

Stuart Shaw, Cambridge International Examinations: What makes for a sound validity argument? *Exploring criteria for evaluating the strength of validation evidence*.

The following posters were also presented:

Sarah Hughes and Stuart Shaw, Cambridge International Examinations: *The issue of fairness and under-used marks in History examination questions*.

Stuart Shaw, Helen Imam and Sarah Hughes, Cambridge International Examinations: Language Rich: *Insights from Multilingual Schools*.

Stuart and Sarah also led a pre-conference workshop on *Issues around how best to provide evidence for assessment validity: the challenge of validation,* and a discussion group on *Fairness in Educational Assessment*.

Further information on all conference papers can be found on our website: http://www.cambridgeassessment.org.uk/our-research/all-published-resources/conference-papers/

Publications

The following articles have been published since *Research Matters*, Issue 20:

Darlington, E. (2015). The Mathematics Needs of Higher Education. *Sigma Network Newsletter*, 9, (December). Available online at: http://www.sigma-network.ac.uk/wp-content/uploads/2015/12/sigmanewsletter9_Dec15.html#MathsHE

Greatorex, J., Rushton, N., Mehta, S. and Grayson, R. (2015). Do experts' views of specification demands correspond with established educational taxonomies? *Online Educational Research Journal*. Advance online publication available at: http://www.oerj.org/View?action=viewPaper&paper=210

Vidal Rodeiro, C.L. and Zanini, N. (2015). The role of the A* grade at A level as a predictor of university performance in the United Kingdom. *Oxford Review of Education*, (*41*)5, 647–670. Available online at: http://www.tandfonline.com/doi/full/10.1080/03054985.2015.1090967#abstract

Shaw, S., Imam, H. and Hughes, S. (2015). *Language Rich: Insights from Multilingual Schools*. Cambridge: Cambridge University Press.

Vergolini, L. and Zanini, N. (2015). Away, but not too far from home. The effects of financial aid on university enrolment decisions. *Economics of Education Review*, *49*, 91–109. Available online at: doi:10.1016/j.econedurev.2015.08.003

Wilson, F., Child, S., Mehta, S., Evans, S. (2015). Purpose and structure of science qualifications for 16-year olds in England: teachers' and employers' perspectives. *School Science Review*, (*97*)358, 99–106. Available online at: http://www.ase.org.uk/journals/school-science-review/2015/09/358/

Further information on all journal papers and book chapters can be found on our website: http://www.cambridgeassessment.org.uk/our-research/all-published-resources/journal-papers-and-book-chapters/

Reports of research carried out by the Research Division for Cambridge Assessment and our exam boards, or externally funded research carried out for third parties, including the regulators in the United Kingdom and many ministries overseas, are also available from our website: http://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-reports/

Data Bytes

A new Data & Analytics team has been created within the Research Division, consisting of three Data Scientists. The team is responsible for providing data services and consultancy to the Cambridge Assessment Group. They also produce a regular series of graphics highlighting the latest research findings and trends in education and assessment, known as 'Data Bytes'. These can be found on our website at: http://www.cambridgeassessment.org.uk/our-research/data-bytes/

Statistical Reports

The Research Division

Examinations generate large volumes of statistical data (approximately 800,000 candidates sit general qualifications each year in the United Kingdom). The on-going Statistics Reports Series provides statistical summaries of various aspects of the English examination system. The objective of the series is to provide statistical information, such as trends in pupil uptake and attainment, qualifications choice, subject combinations and subject provision at school. The reports, mainly produced using national-level examination data, are available in both PDF and Excel format on our website: http://www.cambridgeassessment.org.uk/our-research/all-published-resources/statistical-reports/

The most recent additions to this series are:

- *Statistics Report Series No.89: The accuracy of forecast grades for OCR GCSEs in June 2013*
- *Statistics Report Series No.90: The accuracy of forecast grades for OCR A levels in June 2014*
- *Statistics Report Series No.91: The accuracy of forecast grades for OCR GCSEs in June 2014*
- *Statistics Report Series No.92: Candidates awarded the A* grade at A level in 2014*
- *Statistics Report Series No.93: Re-sitting patterns at GCSE across subjects in 2013 and 2014*
- *Statistics Report Series No.94: Age distribution of GCSE candidates in England in 2014*
- *Statistics Report Series No.95: Uptake of international GCSE subjects 2014*
- *Statistics Report Series No.96: Provision of international GCSE subjects 2014*
- *Statistics Report Series No.97: A level Uptake and Results, by Gender 2005–2014*
- *Statistics Report Series No.98: GCSE Uptake and Results, by Gender 2005–2014*
- *Statistics Report Series No.99: A level Uptake and Results, by School Type 2005–2014*
- *Statistics Report Series No.100: GCSE Uptake and Results, by School Type 2005–2014.*

Build your expertise in assessment

Cambridge Assessment provides training events, courses and bespoke programmes for assessment professionals and organisations in the UK and internationally.

Our free seminars are a catalyst for topical professional debate, bringing authoritative voices and the wider education community together.

Join us **www.canetwork.org.uk**

the NETWORK
BUILDING EXPERTISE IN ASSESSMENT

CONTENTS : Issue 21 Winter 2016

Cambridge Assessment
1 Hills Road
Cambridge
CB1 2EU
UK

Tel: +44(0)1223 552666
Fax: +44(0)1223 552700
Email: researchprogrammes@cambridgeassessment.org.uk
www.cambridgeassessment.org.uk

7284150116

ISSN: 1755–6031