Department for Education

Linked pair of GCSEs in mathematics (MLP) evaluation - Statistical annex

Research Report

February 2014

Dr Atai Winkler – Alpha*Plus* Consultancy Ltd.

Andrew Boyle – Alpha*Plus* Consultancy Ltd.

Contents

List of figures

List of tables

Executive summary

This is the statistical annex to the main report for the final wave of the independent evaluation of the Mathematics Linked Pair (MLP) pilot General Certificate of Education (GCSE) qualifications. It supplements the main evaluation report.

The statistical analysis used in this year's evaluation was based on several national data sets, which are listed in this annex.

A range of descriptive statistics were calculated. These showed the following:

- 839,407 candidates took single GCSE Mathematics in 2013. Gender split was very close to 50:50.
- Market share was different between single GCSE and MLP. Edexcel provided the biggest number of qualifications of the former type, whereas AQA provided the biggest proportion of the latter.
- 17,447 candidates completed both MLP qualifications in either 2012 or 2013 (i.e. they either did both in 2013, or one in 2012 and one in 2013).
- The percentage of males doing MLP was higher than the percentage of males doing single GCSE Mathematics (more than 52% for both Applications and Methods).
- The attainment of candidates taking MLP was higher than that of single GCSE candidates. This is true of Uniform Mark Scale (UMS), mean grade score (on the individual Applications and Methods qualifications) and mean 'best' grade score.
- The attainment of candidates taking the Edexcel MLP Applications qualification was high; almost a whole grade higher than scoring on the Edexcel single GCSE Mathematics qualification.
- Academy converter and community schools were the biggest groups taking single GCSE (more than 200,000 candidates). The numbers of candidates from other establishment types decreased quite rapidly.

Multiple regression models were developed to predict scoring on MLP Applications and Methods. These models were credible and had predictive power; explaining, respectively, 83 and 80 per cent of the variance in the data to which the models were applied. Further details are provided on p.16.

Introduction

Purpose of this document

The independent evaluation of the Mathematics Linked Pair (MLP) General Certificate of Secondary Education (GCSE) qualifications makes its judgements based on a wide range of evidence types (e.g. case-studies, interviews, focus groups, lesson observations, online surveys, etc.). The findings in the main evaluation report are based on that broad range of evidence. One important information source is quantitative data. This statistical annex to the main report seeks to convey insights gained from analysis of quantitative data.

Statistical analyses contribute to the evaluation's overall aims, which can be summarised in these research questions:

- How were the MLP qualifications being implemented?
- What impact have the MLP qualifications had on the teaching and learning of mathematics (including the impact on students' engagement, and on their skills, knowledge and understanding, in terms of the breadth and depth of their understanding of mathematics)?
- To what extent were the MLP qualifications appropriate for different student cohorts and different centres?
- What impact does the MLP have on students' participation, attainment and progression in mathematics?
- What has been the 'value' of the MLP qualifications over and above what is offered by the single GCSE?

Whilst the statistical analyses cannot answer all these research questions on their own, they are an important contributor to the evaluation overall.

Methodology

Use of national data sets

The statistical analyses in the evaluation sought insight by using several major national educational databases. In general terms, the data in these sources pertained to the following types of information:

- Examination candidates' attainment (their grades in mathematics GCSE, their (prior) attainment on Key Stage 2 (KS2) National Curriculum (NC) tests, etc.).

- Examination candidates' demographic status (their gender, ethnicity, age, Special Educational Needs (SEN) status, etc.).

- Schools' 'demographic and administrative' features (the type of school, its size, its location (rural/urban. etc.))

- Schools' effectiveness data (their Ofsted ratings, etc.)

Table 1 shows the databases that were used in the 2013 statistical analyses.

Table 1: The data sources used for the statistical analysis (England and Wales)

England	Wales
Data supplied by AQA, Edexcel, OCR and WJEC for all candidates on the MLP and all single GCSE	
Awarding organisation (AO) lists of MLP participant centres based on communication with centres supplied by AQA, Edexcel, OCR and WJEC	
National Pupil Database (NPD) Census (demographic) information about students in MLP and single GCSE	NPD Census (demographic) information about students in MLP and single GCSE
NPD KS2 Attainment information about students in MLP and MS	NPD KS2 attainment information about students in MLP and single GCSE (derived from teacher assessment)
School Section 5 inspection grades from Ofsted	Estyn did not supply school inspection data for Welsh schools.
School demographic and administrative information from Edubase	Some school demographic information from Edubase (which is incomplete for Welsh schools) plus information from DfES[1] regarding Welsh schools

[1] Department for Education and Skills, Welsh Government

Creating a master database and merging candidate information

Data files were provided by awarding organisations (AOs) (AQA, Edexcel, OCR and WJEC), and were read into the statistical analysis software product SPSS, now called IBM Statistics, and then merged to create a master database. All the analyses were carried out on this database or extracts from it using IBM Statistics.

The files from the AOs could not be simply merged to form the master file by matching them using UCI (Unique Candidate Identifier) as the key variable because the UCIs supplied by the AOs did not identify each candidate uniquely. This mis-match worked both ways: there were occurrences of different candidates, as defined by name and date of birth, having the same UCI; and occurrences of different UCIs referring to the same candidate as defined by name, date of birth and centre. Thus, another de-duping[2] method was developed. The de-duping results are shown below.

Table 2: Numbers of candidates in master data file using different 'de-duping' methods

AO	No. of candidates using UCI	No. of candidates using new de-duping method
AQA	178,324	178,254
Edexcel	631,948	632,425
OCR	67,117	67,224
WJEC	38,611	38,611

Each record in the AO files referred to one exam, so that the number of exams taken by a candidate is given by the number of records for him/her. This structure was not suitable for the analyses and so the file was restructured to a new form where *all* the exam data for a candidate are on one record. For example, if a candidate took four exams, his/her data were on four records in the original file but on one record in the new file. Several iterations of the de-duping algorithm were required until the results passed all the checks for consistency.

After the AO files were de-duped and merged to form the master file, all the information for each candidate was on one record, including candidates who sat exams from different AOs.

The candidates' demographic data and the schools data were then added to the master file. This structure was necessary for the data to be analysed at candidate level and at

[2] Or 'removing duplicates'.

centre level. In contrast to previous years' analyses, the data collected and analysed for this report did not consider qualification unit level outcomes but focused on qualification level outcomes only.

Each candidate in the master data file satisfied one or more of the following criteria:

1. Took only one MLP qualification in 2012 (N = 6,443)

2. Took single GCSE mathematics in 2013 (N = 839,598)

3. Took at least one MLP qualification in 2013 (N = 24,809)

There were 858,850 people who matched these criteria[3].

Thus, we were able to include candidates who had taken one MLP qualification in 2012, and who were potentially completing the other in 2013. But our data collection did not include 2011 candidates, thus we were not able to calculate how many candidates had sat an MLP qualification through the entire duration of the pilot.

In our data structuring and analysis, and in the findings reported below, it was important to distinguish clearly the following concepts:

- A GCSE *candidate* is a person. He or she will have taken one or more mathematics qualifications.

- A *completion* is a completed 'sitting' of a GCSE mathematics qualification (either single GCSE or one of the two MLP qualifications) which results in the award of a grade from A* to U[4]. A *candidate* may have multiple *completions*.

- Generally speaking, when we write about an *MLP completion* we intend to refer to the completion of a particular qualification (either the Methods or the Applications – which we shorten to 'Apps' in tables and captions). If we intend to refer to incidences of persons who attain both Methods and Applications, we will be explicit in referring to 'the Pair' or 'the Mathematics Linked Pair'.

- It is also important, in interpreting results below, to attend to whether the results pertain to candidates taking examinations in 2013, or those taking them in either 2012 or 2013.

---

[3] The sum of categories 1, 2 and 3 does not equal 858,850. This is because the categories are not mutually exclusive (for instance, some people took one MLP qualification in 2012, AND at least one MLP qualification in 2013).

[4] 'X' grades do not count towards completion calculations.

Analytical procedures

This type of research exercise is perhaps atypical in that much of the resource expended in the project goes into obtaining, and then structuring, national databases[5]. Such databases, including ‘census’ type data sets, could, in general terms, be considered to contain all the relevant data from a population. The advantage of this was that it allowed the quantitative analyses to be relatively simple. For most of the findings that are set out below, it was sufficient to report unadorned descriptive statistics (counts of instances, means, percentages, etc.) because we had access to the population-level data. It was not necessary, for the most part, to report significance tests, and/or complex inferential indices. This is because we did not have to infer the extent to which a sample represented a population because our analyses were based on the population data.

Whilst this general rule remains true, there are some important points that must be discussed. Firstly, we did not consider the MLP candidates to be a sample drawn from the population of single GCSE mathematics candidates. A data sample can be defined as a set of data collected and/or selected from a statistical population by a defined procedure. In our case, MLP candidates are **not** a sub-set of single GCSE candidates (they are not a sub-set drawn from the bigger group, they are a **different** group of people). Thus, we can compare characteristics of MLP candidates and single GCSE candidates (we may be able to report that they are generally older, more likely to be male, attain higher grades on average, etc.). However, we do not consider it meaningful to discuss the extent to which MLP candidates **represent** the population of single GCSE candidates[6].

In addition to the descriptive statistics in the report, we carried out two analysis that investigated the relationship between input (‘independent’/’predictor’ variables) and a ‘dependent’ or ‘target’ variable. We did this in two ways; firstly, we calculated a simple correlation coefficient, showing the relationship between prior attainment (KS2 attainment) and MLP attainment[7].

Such a correlation coefficient has the advantage of being relatively straightforward and intuitive. However, it has the disadvantage of only including one variable as a predictor – when we have many potentially predicting variables in the data set. Further, such correlation coefficients may not account for sufficient variance in the data to explain output variable scoring.

[5] In many projects relatively more resource is used in either collecting new data and/or analysis.

[6] Equally, it would be hard to treat the dataset as a single population with other defined characteristics; for example, not all the people in the database are aged 15 or 16, or in Key Stage 4 (and so on).

[7] We used MLP UMS as the measure of MLP attainment. This variable is closer to being a continuous variable than grade score.

To mitigate such weaknesses, in one of our final analyses (see below, p. 16), we have developed a multiple regression model. In regression models of this type, several ‘predictor’ (or ‘independent’) variables are selected and modelled. The regression technique allows us to use modelled coefficients for the predictor variables to predict a value on a ‘target’ (or ‘dependent’) variable. In addition, the regression model allows us to estimate, with respect to predictor variables, their ‘sign’ (positive or negative) – so that positively-signed predictor variables tend to increase as the target variable increases and negatively-signed predictor variables tends to *decrease* as the target variable increases. We also estimate a ‘B co-efficient’, which measures how strongly each variable influences the target variable. Also, we estimate a Beta (standardised regression coefficient). The beta value is a measure of how strongly each predictor variable influences the target variable. The beta is measured in units of standard deviation. The higher the beta value the greater the impact of the predictor variable on the target variable.

In addition, we can report a *t*-statistic associated with each predictor variable and a significance level. In general, the higher the *t*-statistic, and the lower the significance level, the more likely it is that the predictor variable is truly associated with the target variable, and the less likely it is that the association is due to random chance. In common with many statistical applications, we interpret a significance level of less than 0.05 as being statistically significant. Further, we can calculate a squared correlation co-efficient (R-squared). This co-efficient tells us the amount of variance in data that the regression model accounts for.

In interpreting the plausibility of the regression model, we also need to consider residuals (‘error terms’) to ensure that no relevant variables have been left out of the model.

Findings

In the section that follows, we present a number of data tables that show various features of the populations sitting single GCSE Mathematics, and MLP qualifications.

Table 3: Count of numbers of candidates taking single GCSE in 2013

Gender	**AO name**					**Percentages**
	AQA	**Edexcel**	**OCR**	**WJEC**	**Total**	
Not stated	*0*	*2*	*0*	*0*	***2***	
male	69,283	301,590	28,615	17,358	**416,846**	**49.66%**
female	72,691	303,567	28,758	17,543	**422,559**	**50.34%**
Total	**141,974**	**605,159**	**57,373**	**34,901**	**839,407**	
Percentages	**16.91%**	**72.09%**	**6.83%**	**4.16%**		

839,407 candidates took single GCSE Mathematics in 2013. The gender split was very nearly 50:50. Edexcel had the biggest share of the market amongst AOs, providing qualifications to more than 70 per cent of all candidates.

Table 4: Count of number of candidates taking at least one MLP qualification in 2013[8]

	AO name			
	AQA	**Edexcel**	**OCR**	**WJEC**
Number of candidates	10,488	7,664	4,383	1,477
Percentages	**43.70%**	**31.90%**	**18.30%**	**6.10%**

Table 4 shows the numbers of candidates taking at least one MLP qualification in 2013. The market shares for MLP are very different to the shares for single GCSE. AQA has the biggest share of this market with over 40 per cent of candidates, whilst Edexcel has around 30 per cent of candidates.

[8] The gender split of MLP candidates can be found in Table 8 and Table 9.

There follow three tables (Table 5, Table 6 and Table 7) which throw more light on candidature patterns for the MLP qualifications. All three tables illustrate a count of candidates who took MLP qualifications in either 2012 or 2013. In each case, the counts of candidates by AO are cross tabbed against the tier that candidates were entered for. We have divided tiers into three categories: F stands for foundation, H for higher and M for mixed (i.e. candidates who took units from different tiers within their qualification).

Table 5: Count of candidates who only took MLP Apps in 2012 or 2013

	AO MLP Apps cross-tab					
Tier	**AQA**	**Edexcel**	**OCR**	**WJEC**	**Total**	**Percentages**
Not stated	*84*	*4*	*0*	*0*	***88***	***5.46%***
F	236	196	48	5	**485**	**30.07%**
H	341	490	151	6	**988**	**61.25%**
M	50	1	1	0	**52**	**3.22%**
Total	**711**	**691**	**200**	**11**	**1,613**	

Table 6: Count of candidates who only took MLP Methods in 2012 or 2013

	AO MLP Methods cross-tab					
Tier	**AQA**	**Edexcel**	**OCR**	**WJEC**	**Total**	**Percentages**
Not stated	*209*	*392*	*124*	*0*	***725***	***12.61%***
F	1,083	1,069	360	35	**2,547**	**44.30%**
H	763	1,253	336	4	**2,356**	**40.98%**
M	64	32	25	0	**121**	**2.10%**
Total	**2,119**	**2,746**	**845**	**39**	**5,749**	

Table 7: Count of candidates who took both MLP Apps and Methods in 2012 or 2013

	AO both MLP qualifications cross-tab					
Tier	**AQA**	**Edexcel**	**OCR**	**WJEC**	**Total**	**Percentages**
F	2,309	464	1,035	667	4,475	**25.65%**
H	3,513	3,345	1,696	730	9,284	**53.21%**
M	2,123	810	728	27	3,688	**21.14%**
Total	**7,945**	**4,619**	**3,459**	**1,424**	**17,447**	

The tiered entry patterns were quite different between the Applications and the Methods qualifications. For those who sat Applications only, roughly twice as many candidates entered for the higher as opposed to the foundation tier (61 per cent compared to 30 per cent, respectively). In contrast, entry for the Methods qualification (those who only took

Methods) was much more even between the foundation and higher tiers (44 per cent compared to 41 per cent, respectively).

The percentages of higher and foundation tier candidates in Table 7 reflect those seen in Table 5 and Table 6 (53 per cent were entered in the higher tier compared to 26 per cent in the foundation tier). The overall number of candidates taking both MLP qualifications in 2012 or 2013 (i.e. completing in 2013) was 17,447.

Table 8 and Table 9, respectively, cross tabulate the numbers of MLP Applications and Methods candidates against candidates' genders. This is done by AO and in total.

Table 8: Gender of candidates cross-tabbed against AO for MLP Apps

	AO MLP Apps cross-tab					
Gender	**AQA**	**Edexcel**	**OCR**	**WJEC**	**Total**	**Percentages**
Male	4,543	2,606	2,018	764	**9,931**	**52.10%**
Female	4,113	2,704	1,641	672	**9,130**	**47.90%**
Total	**8,656**	**5,310**	**3,659**	**1,436**	**19,061**	

Table 9: Gender of candidates cross-tabbed against AO for MLP Methods

	AO MLP Methods cross-tab					
Gender	**AQA**	**Edexcel**	**OCR**	**WJEC**	**Total**	**Percentages**
Male	5,298	3,723	2,376	778	**12,175**	**52.58%**
Female	4,766	3,644	1,929	685	**11,024**	**47.52%**
Total	**10,064**	**7,367**	**4,305**	**1,463**	**23,199**	

It is striking that both Applications and Methods have clear majorities of male candidates (over 52 per cent in both cases). This can be contrasted with the near fifty:fifty gender split that pertains in the case of single GCSE mathematics (see Table 3, above).

Other demographic characteristics of candidates were also examined. Table 10 counts the number of candidates on: single GCSE, MLP Applications alone, MLP Methods alone and both MLP Applications and Methods who were eligible for Free School Meals (FSM). Counts within these categories are expressed in absolute terms and as percentages.

Table 10: Free school meals eligibility of candidates doing various GCSE maths combinations

Qualification or combination	Type of quantity	FSM eligibility				
		Valid				
		Not eligible	Eligible	Total	*Not stated*	Total
Not stated	***Number***	*110*	*56*	***166***	*148*	***314***
	Percentage	*35.03*	*17.83*	***52.87***	*47.13*	***100.00***
Single GCSE	**Number**	486,684	94,485	**581,169**	*252,558*	**833,727**[9]
	Percentage	58.37	11.33	**69.71**	*30.29*	**100.00**
MLP Apps only	**Number**	820	67	**887**	*726*	**1,613**
	Percentage	50.84	4.15	**54.99**	*45.01*	**100.00**
MLP Methods only	**Number**	3,004	784	**3,788**	*1,961*	**5,749**
	Percentage	52.25	13.64	**65.89**	*34.11*	**100.00**
MLP Apps and MLP Methods	**Number**	7,668	737	**8,405**	*9,042*	**17,447**
	Percentage	43.95	4.22	**48.17**	*51.83*	**100.00**

The most striking thing about Table 10 is the large amount of missing data[10]. In some categories (MLP Applications and MLP Methods) over 50 per cent of the data are missing. Whilst it is likely that overwhelming majorities of such data will pertain to people who are not entitled to FSM, this cannot be assumed.

Within the valid data, we can observe some differences. For example, FSM entitlement is about seven percentage points lower amongst candidates taking both MLP Applications and Methods, as opposed to single GCSE[11].

[9] This total is lower than stated in earlier tables because some students took both single GCSE mathematics and one or more MLP qualification.

[10] This is due to data missing in the sources that were provided to us. It is beyond the scope of this project to establish where or why these data were not collected and/or recorded. The situation is similar for a number of indicators: SEN status, FSM eligibility, ethnicity, etc.

[11] As noted in the method section, no statement is made about the significance of this difference.

In Table 11, we count the ethnicities of groups in the various qualification combinations. In this table, counts of ethnicities are shown, with percentages in brackets.

Table 11: Count of ethnicities of candidates taking different GCSE maths qualifications

Ethnic group	*Not stated*	GCSE	MLP Apps only	MLP Methods only	MLP Apps and MLP Methods
Not stated	*148 (47%)*	*252,558 (30%)*	*726 (45%)*	*1,961 (34%)*	*9,042 (51%)*
Any other ethnic group	*3 (0%)*	8,366 (1%)	4 (0%)	50 (0%)	61 (0%)
Asian	*11 (3%)*	53,179 (6%)	29 (1%)	529 (9%)	514 (2%)
Black	*6 (1%)*	32,554 (3%)	37 (2%)	247 (4%)	241 (1%)
Chinese		2,059 (0%)	5 (0%)	10 (0%)	62 (0%)
Mixed	*14 (4%)*	22,651 (2%)	36 (2%)	167 (2%)	309 (1%)
Unclassified	*3 (0%)*	5,361 (0%)	8 (0%)	27 (0%)	84 (0%)
White	*129 (41%)*	456,999 (54%)	768 (47%)	2,758 (47%)	7,134 (40%)
Total	***314 (100%)***	**833,727[12] (100%)**	**1,613 (100%)**	**5,749 (100%)**	**17,447 (100%)**

As with Table 10, there is as much as 50 per cent missing data in some cells in Table 11. Beyond the large quantities of missing data, it is notable that the proportions of white candidates differ somewhat between the different qualification combinations. However, caution should be exercised; differences amongst such categories might be explained by differences in the amounts of missing data.

[12] This total is lower than stated in earlier tables because some students took both single GCSE mathematics and one or more MLP qualification.

Table 12 shows mean scores under a variety of conditions. Firstly, it shows UMS scores for the various GCSE mathematics qualifications. Then it tabulates mean grade scores for the various GCSE qualifications. Mean grade score is a numerical mapping against the letter grades, as follows:

A* = 8, A = 7, B = 6, C = 5, D = 4, E = 3, F = 2, G = 1.

Table 12: Mean UMS and grade scores for 2013 single GCSE, MLP Apps and Methods

Qualification and score type	N	Mean
UMS single GCSE 2013	839,407	132.46
UMS MLP Apps	19,060	186.29
UMS MLP Methods	23,196	185.10
Grade score single GCSE 2013	839,407	4.68
Grade score MLP Apps	19,060	5.22
Grade score MLP Methods	23,196	5.07

Thus, we can see that scoring in the two MLP qualifications is highly comparable, and both are well ahead of the single GCSE scoring. The two MLP means are a single uniform mark apart, and around 50 marks above mean single GCSE UMS score. The pattern on the grade score is similar; the two MLP qualifications' mean grade scores are in the C band (Applications being somewhat above Methods on average), whilst mean single GCSE grade score is within the D band.

As before, below there follow two tables (Table 13 and Table 14) which are best read as a pair. Table 13 shows the number of completions and the mean grade score for candidates who took single GCSE Mathematics, or MLP Applications or Methods in 2013. For comparison, Table 14 counts the number of candidates who completed **both** MLP Applications and Methods in 2012 or 2013.

Table 13: Completions, and mean grade scores for single GCSE, MLP Apps and Methods in 2013

	No. of completions	Mean grade score	MLP sub-totals	Each AO's percentage of whole (completions)	
				MLP	Single Maths
AQA single GCSE	170,697	4.85			**14.6%**
AQA MLP Apps	8,786	5.08			
AQA MLP Methods	9,575	5.08	**18,361**	**44.2%**	
Edexcel single GCSE	896,869	4.64			**76.7%**
Edexcel MLP Apps	5,316	5.56			
Edexcel MLP Methods	6,785	5.10	**12,101**	**29.1%**	
OCR single GCSE	60,548	4.70			**5.2%**
OCR MLP Apps	3,867	5.25			
OCR MLP Methods	4,200	5.13	**8,067**	**19.4%**	
WJEC single GCSE	41,571	4.58			**3.6%**
WJEC MLP Apps	1,502	4.79			
WJEC MLP Methods	1,555	4.59	**3,057**	**7.4%**	
Total no. of MLP completions in 2013	**41,586**				
Total no. of single GCSE completions in 2013	**1,169,685**				

Table 14: Mean best grade score for candidates taking both MLP qualifications

AO name	No of candidates	Mean best grade score
AQA	7944	5.44
Edexcel	4619	5.90
OCR	3459	5.50
WJEC	1424	4.96
Mixed[13]	*1*	*5.00*
Total	**17,447**	**5.53**

These tables show that scoring on MLP tends to be higher than scoring on single GCSE. Scoring on either MLP qualifications is higher than the same AO's score on single GCSE

[13] That is, one candidate took MLP Methods and Applications with different AOs.

in all cases (Table 13). The difference is lowest for WJEC Methods (one hundredth of a percentage point higher than that AO's single GCSE mean grade score), and highest for Edexcel MLP Applications; Edexcel's MLP Applications grade score is – on average – almost an entire grade higher than the grade for single GCSE mathematics.

This pattern persists when one analyses those candidates who completed both MLP qualifications. WJEC has a mean best grade score slightly below the C boundary, whilst all three English boards have a mean best GCSE grade of above C for their candidates' better grade amongst their MLP grades. Edexcel is the highest with a mean of 5.90 – that is, approaching the B grade boundary.

To investigate this matter further we have created Figure 1 and **Error! Reference source not found.**. The former is a bar chart showing the numbers of candidates (across all AOs) achieving each grade. The latter is a table showing the percentage of candidates who achieved each grade; **Error! Reference source not found.** breaks down Figure 1 in that it shows the percentages by each separate AO.

Figure 1: Count of number of candidates scoring particular grades (combined/better MLP grade)

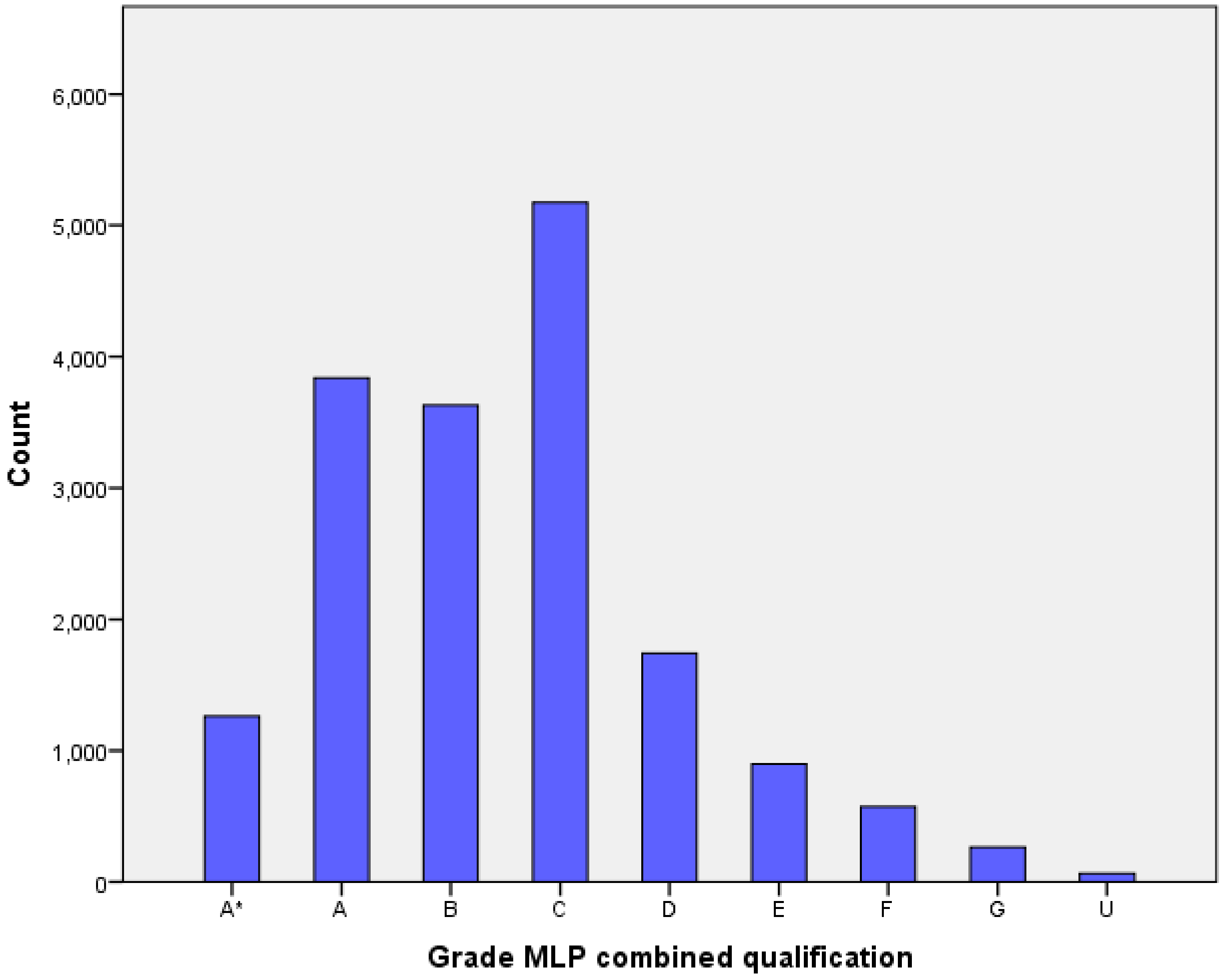


Table 15: Percentage of candidates for each AO scoring at each grade (combined/better MLP grade)

Grade	AQA	Edexcel	OCR	WJEC
A*	6.14	10.15	6.82	4.78
A	21.58	26.26	20.93	13.06
B	20.13	24.27	19.83	15.66
C	29.82	24.94	33.10	35.60
D	10.59	8.77	8.82	13.48
E	5.99	2.97	5.15	7.65
F	3.54	1.80	3.73	5.69
G	1.74	0.71	1.36	3.30
U	0.48	0.13	0.26	0.77
Total	**100.00**	**100.00**	**100.00**	**100.00**

Figure 1 shows that 'C' was the most frequently awarded grade across all the AOs. Next most frequent comes A – slightly ahead of B. **Error! Reference source not found.** casts further light on the higher scoring in the Edexcel MLP qualifications. Edexcel has higher numbers of candidates scoring A*, A and B grades; anywhere between four and ten percentage points, depending upon the grade and the other AO with which one is comparing Edexcel.

In addition to such descriptive statistics at the candidate level, we report some modelling of the relationship between variables that may impact on MLP scoring and MLP scoring itself. Firstly, we report the correlation between prior attainment – as exemplified by KS2 National Curriculum (NC) test scores – and MLP UMS. Correlation is strong for both MLP qualifications, being 0.695 in the case of MLP Applications, and 0.701 for MLP methods[14].

Our final analysis at the candidate level is a multiple regression. This analysis (as described in the method section – above, at p. 7) inserts several 'predictor' variables into a model and attempts to predict a 'target' variable. We constructed one regression model for MLP Applications and one for MLP Methods. We used UMS as the target variable because it is a granular continuous variable and is therefore more suitable as a target variable than grades and grade scores

The model was developed from a carefully selected pool of candidate predictor variables by a combination of stepwise and entry methods. We did this in an iterative manner, repeatedly running the model until a credible solution ensued. Whilst such an approach runs the risk of being banal – plugging variables into the model opportunistically – we

[14] See also discussion of the benefits of multiple regression over 'simple' correlation analysis below.

believe that we have done this advisedly and that we have ended up with a model containing variables that have real-world consequence and intuitive meaning[15].

The variables in our multiple regression were of several types; most were scale variables (numerical) but some were nominal ('yes/no' or 'is/isn't'). The interpretation of nominal variables in the model requires explanation.

The predictor variables in our multiple regression model are named and described in Table 16:

Table 16: Definitions of predictor variables in multiple regression models

Variable name	Definition
Age MLP Apps (or Methods)	Age of candidate in months when they sat their MLP Applications (or Methods) qualification.
Gender	1= male; 2 = female
Ethnic group variables	*All ethnicity coefficients are relative to the coefficient for unclassified (= 0).*
AOEG	1 for 'any other ethnic group'; ELSE = 0
Asian	1 for Asian', ELSE = 0
Black	1 for 'black', ELSE = 0
Chinese	1 for Chinese; ELSE = 0
Mixed	1 for 'mixed' ethnicity; ELSE = 0
Main language	1 for main language is English; ELSE = 0
FSM eligible	1 if candidate eligible for FSM; ELSE = 0
SENs	1 if candidate noted as having SEN; ELSE = 0
Key Stage 2	Key stage 2 prior attainment measure
AO MLP Apps (or Methods) [AO name]	1 if candidate sat their MLP Applications (or Methods) qualification with AQA/Edexcel/OCR/WJEC; ELSE = 0
Tier MLP Apps (or Methods)[F: Foundation or H: Higher tier]	*Tier coefficients are measured relative to tier M (mixed).*

We developed two separate multiple regression models; one to predict the UMS scores for MLP Applications and one to predict the same score type for MLP Methods. Squared correlation co-efficients (R-squared) quantified the amount of variance explained by the respective models. That amount was 0.83 (MLP Applications) and 0.80 (for MLP Methods). These are high R-squared values for such a large and complex data set – suggesting that these multiple regression models are credible and have predictive power.

[15] To attempt to produce a coherent set of predictor variables, we considered only candidate-level variables and excluded school variables.

The model has several features over and above the correlation coefficients reported above:

- The amount of variance explained by the regression models' R-**squared** correlation coefficients is substantially higher than the variance explained by the values on the simple R correlation coefficients reported above.
- The multiple regression model allows us to observe how a basket of variables work singly and in combination to affect outcomes.
- The residual analysis that follows the reporting of regression coefficients gives us further insight into how the outcome variable (MLP attainment) is affected by predictor variables of diverse types (and at particular points in the ability continuum).

The respective regression coefficients for MLP Applications and Methods are shown in Table 17 and Table 18. These tables are identically structured. Each table contains: a co-efficient, Beta (standardised) coefficient, *t*-statistic and significance value (as described at p. 7, above). The rows contain the predictor variables as described in Table 16. In addition, the first row of values refers to a 'constant'. This is analogous to an intercept parameter; and would be equivalent to the point where a linear regression line crossed the *y*-axis in a simple regression model.

Table 17: Multiple regression predictor variable co-efficient values for MLP Apps

	Unstandardized Coefficients		Standardized Coefficients		
Model	**B**	**Std. Error**	**Beta**	***t***	**Sig.**
(Constant)	*-162.422*	*15.530*		*-10.458*	*.000*
Age MLP apps	.313	.066	.022	4.776	.000
Gender	2.142	.723	.013	2.964	.003
AOEG	19.333	6.016	.018	3.213	.001
Asian	13.972	4.054	.040	3.446	.001
Black	3.001	4.229	.006	.710	.478
Chinese	24.030	5.757	.024	4.174	.000
Mixed	5.106	4.050	.012	1.261	.207
White	2.036	3.622	.009	.562	.574
Main language	-2.164	1.789	-.007	-1.210	.226
FSM eligible	-10.289	1.298	-.036	-7.928	.000
SENs	-3.497	1.142	-.015	-3.063	.002
Key Stage 2	50.836	.677	.451	75.087	.000
AO MLP apps AQA	-14.708	7.155	-.059	-2.056	.040
AO MLP apps Edexcel	111.056	7.104	.686	15.633	.000
AO MLP apps OCR	46.499	7.106	.270	6.544	.000
Tier MLP apps F	-6.949	1.169	-.036	-5.947	.000
Tier MLP apps H	21.368	1.010	.132	21.156	.000

In Table 17, the following have **positive** signs (that is, as values on these variables increase, MLP Applications UMS scores increase). They are also all significant at the five per cent level:

- Age (tending to be older)
- Gender ('tending to be' female)
- The following ethnic groups: 'any other ethnic group', Asian, Chinese
- Having higher KS2 attainment
- Having either Edexcel or OCR as AO
- Being entered for the higher tier

However, the following variables have **negative** signs and are significant at the five per cent level (that is, as values on these variables increase, MLP Applications UMS scores *decrease*):

- Being eligible for FSM
- Having SEN
- Having AQA as AO
- Being entered for the foundation tier

The MLP Methods regression model is expressed in Table 18.

Table 18: Multiple regression predictor variable co-efficient values for MLP Methods

Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	*t*	Sig.
(Constant)	*-305.830*	*13.580*		*-22.521*	*.000*
Age MLP methods	1.129	.053	.094	21.278	.000
Gender	3.604	.700	.022	5.146	.000
AOEG	18.235	5.337	.021	3.417	.001
Asian	9.440	3.980	.033	2.372	.018
Black	10.960	4.071	.026	2.692	.007
Chinese	27.365	5.905	.026	4.634	.000
Mixed	3.978	4.027	.010	.988	.323
White	2.646	3.656	.013	.724	.469
Main language	-5.037	1.625	-.020	-3.099	.002
FSM eligible	-9.870	1.104	-.040	-8.941	.000
SENs	-9.194	1.007	-.043	-9.133	.000
Key Stage 2	49.182	.636	.453	77.295	0.000
AO MLP methods AQA	-5.030	7.746	-.019	-.649	.516
AO MLP methods Edexcel	111.840	7.688	.687	14.547	.000
AO MLP methods OCR	53.647	7.696	.310	6.971	.000
Tier MLP methods F	-20.812	1.122	-.119	-18.543	.000
Tier MLP methods H	16.813	1.032	.105	16.284	.000

The B coefficients' signs and significances were slightly different to those in the MLP Applications model. In Table 18, the following have **positive** signs (that is, as values on these variables increase, MLP Methods UMS scores increase). They are also all significant at the five per cent level: (see list of variables above – same comments apply)

- Age (tending to be older)
- Gender ('tending to be' female)
- The following ethnic groups: 'any other ethnic group', Asian, Black, Chinese
- Having higher KS2 attainment
- Having either Edexcel or OCR as AO
- Being entered for the higher tier

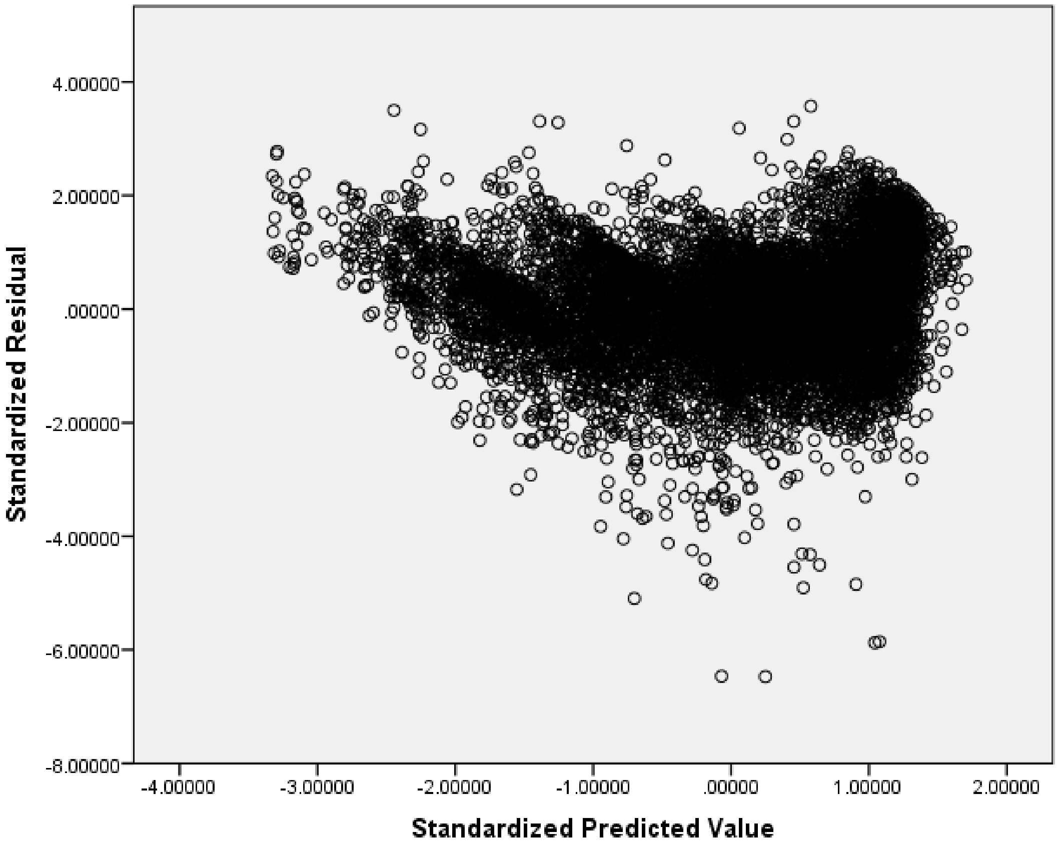
However, the following variables have **negative** signs and are significant at the five per cent level (that is, as values on these variables increase, MLP Methods UMS scores *decrease*):

- Having a language other than English as a first language

- Being eligible for FSM
- Having SEN
- Being entered for the foundation tier

To test the credibility and comprehensiveness of the multiple regression model, we studied the residuals. In Figure 2 we plot the standardised residuals ('error terms') against the standardised predicted. We show the MLP Applications plot – the Methods plot exhibits a similar pattern.

Figure 2: Residuals plotted against predicted values for MLP Apps regression model



It seemed fairly clear from this plot that residuals are not random. There is a clear group of candidates with residuals less than –2.5. On scrutinising this plot carefully, we find that the residuals in the lower right quadrant of the graph are associated with low grades.

Detailed analysis of the results showed that residuals smaller than about –3[16] tend to be associated with:

1. Students who are white.

2. Students whose main language is English.

3. Students who are not eligible for FSM.

The residuals for boys and girls have the same shape, showing that the model has considered gender correctly. However, the presence of non-random residuals suggests that the multiple regression model can be improved if another variable (which is not captured in national data sets) had been available for insertion into the model.

[16] Negative residuals show under performance, i.e. the observed values are smaller than the predicted values.

Next, we consider information about school types. Insight into establishment types' participation levels in the MLP pilot can be seen in Table 19. That table shows the numbers of candidates completing single GCSE, and it shows the numbers of candidates completing MLP Applications and MLP Methods. Further, the table is sorted in descending order on the MLP Applications column[17].

Table 19: Numbers of candidates from establishment types (sorted by MLP Apps)

Establishment type	No.s of candidates			Single GCSE rank order
	Single GCSE	MLP Apps	MLP Methods	
Not stated	*62,063*	*2,068*	*2,058*	
Academy converter	279,784	9,300	11,487	**1**
Community school	206,075	3,110	3,812	**2**
Foundation school	90,242	1,801	2,573	**3**
Voluntary aided school	63,904	1,151	1,321	**4**
Other independent school	26,640	866	785	**7**
Academy sponsor led	37,305	621	1,067	**6**
Offshore schools	2,501	96	29	**11**
Community special school	2,242	23	38	**12**
Pupil referral unit	4,031	12	11	**10**
Other independent special school	1,062	11	11	**14**
Miscellaneous	130	1		**23**
Free schools	321			**18**
City technology college	1,897			**13**
Service children's education	426			**15**
University technical college	197			**21**
Voluntary controlled school	9,591			**8**
Academy 16-19 converter	62			**25**
Sixth form centres	171			**22**
Further education	41,192		3	**5**
Welsh establishment	8,157			**9**
Special college	9			**26**
Non-maintained special school	318			**19**
Studio schools	65			**24**
Academy special converter	411			**16**
Academy alternative provision	321		1	**17**
Foundation special school	286			**20**
Academy special sponsor led	4			**27**
Total	**839,407**	**19,060**	**23,196**	

[17] This essentially arbitrary decision was taken because the MLP Applications and MLP Methods sort orders were almost identical.

By and large the rank order of number of candidates by establishment type does not change much between single GCSE and MLP Applications. The first four most populous establishment types remain the same in the table (academy converter, community school, foundation school and voluntary aided school). The most notable 'absentee' in the MLP Applications rank order is 'further education' which had the fifth most candidates in the single GCSE rank order.

Academy converter and community schools are the two biggest types of establishment participating in the single GCSE. They both had over 200,000 candidates in 2013. The number of candidates from other establishment types decreased quite rapidly.

Table 20 takes the analysis of establishment types further. It tabulates the mean grade scores on the three types of mathematics GCSEs by establishment type. It is sorted by single GCSE mean grade score (descending).

Table 20: Mean grade scores for single GCSE, MLP Apps and Methods, by establishment type

Establishment type	Mean grade score			No.s of candidates		
	Single GCSE	MLP Apps	MLP Methods	Single GCSE	MLP Apps	MLP Methods
Not stated	*4.57*	*4.82*	*4.73*	*62,063*	*2,068*	*2,058*
Other independent school	6.03	6.39	6.60	26,640	866	785
Free schools	5.58			321		
City technology college	5.29			1,897		
Service children's education	5.10			426		
Academy converter	4.99	5.51	5.34	279,784	9,300	11,487
University technical college	4.93			197		
Voluntary aided school	4.83	4.96	4.96	63,904	1,151	1,321
Offshore schools	4.80	5.50	5.76	2,501	96	29
Voluntary controlled school	4.73			9,591		
Academy 16-19 converter	4.71			62		
Community school	4.51	4.86	4.70	206,075	3,110	3,812
Foundation school	4.33	4.64	4.43	90,242	1,801	2,573
Sixth form centres	4.32			171		
Further education	4.28		2.67	41,192		3
Miscellaneous	4.21	3.00		130	1	
Academy sponsor led	4.18	4.81	4.73	37,305	621	1,067
Welsh establishment	3.90			8,157		
Special college	3.33			9		
Non-maintained special school	3.22			318		
Other independent special school	3.18	4.45	4.27	1,062	11	11
Pupil referral unit	2.65	4.08	4.27	4,031	12	11
Studio schools	2.58			65		
Academy special converter	2.52			411		
Academy alternative provision	2.33		0.00	321		1
Foundation special school	2.33			286		
Community special school	2.29	2.96	3.21	2,242	23	38
Academy special sponsor led	1.50			4		
Total	**4.68**	**5.22**	**5.07**	**839,407**	**19,060**	**23,196**

'Other independent school' is ranked highest on mean GCSE score, with an average grade being above the B boundary. Free Schools, City Technology Colleges and service children's education all score above C on average, whilst the rest are below that boundary.

Further analysis of scoring by establishment type, is presented in Table 21. This table is similar to Table 20, except that: the mean grade scores were sorted in descending order by MLP Applications score, the numbers of candidates were omitted (they were the same as in Table 20), and – by way of comparison – establishment types' rank order (based on single GCSE mean grade score) was listed in the right-hand column.

Table 21: Mean grade scores, sorted by MLP Apps (descending)

Establishment type	Mean grade score			Single GCSE rank order
	Single GCSE	MLP Apps	MLP Methods	
Not stated	*4.57*	*4.82*	*4.73*	
Other independent school	6.03	6.39	6.60	**1**
Academy converter	4.99	5.51	5.34	**5**
Offshore schools	4.80	5.50	5.76	**8**
Voluntary aided school	4.83	4.96	4.96	**7**
Community school	4.51	4.86	4.70	**11**
Academy sponsor led	4.18	4.81	4.73	**16**
Foundation school	4.33	4.64	4.43	**12**
Other independent special school	3.18	4.45	4.27	**20**
Pupil referral unit	2.65	4.08	4.27	**21**
Miscellaneous	4.21	3.00		**15**
Community special school	2.29	2.96	3.21	**26**
Free schools	5.58			**2**
City technology college	5.29			**3**
Service children's education	5.10			**4**
University technical college	4.93			**6**
Voluntary controlled school	4.73			**9**
Academy 16-19 converter	4.71			**10**
Sixth form centres	4.32			**13**
Further education	4.28		2.67	**14**
Welsh establishment	3.90			**17**
Special college	3.33			**18**
Non-maintained special school	3.22			**19**
Studio schools	2.58			**22**
Academy special converter	2.52			**23**
Academy alternative provision	2.33		0.00	**24**
Foundation special school	2.33			**25**
Academy special sponsor led	1.50			**27**
Total	**4.68**	**5.22**	**5.07**	

The rank order sorted by MLP Applications mean grade score is similar to that of single GCSE (whilst the scores were higher throughout in absolute terms). The exception to this similarity is that several of the higher scoring establishment types in the single GCSE are absent from the MLP Applications list: Free Schools, City Technology College, service children's education, University Technical College and voluntary controlled schools were not part of the MLP pilot. This difference is more likely affected by (non) participation in the pilot, rather than by attainment.

Department for Education

Reference: DFE- RR328B

ISBN: 978-1-78105-305-8