Predicted grades: accuracy and impact

A report for University and College Union

Dr Gill Wyness, UCL Institute of Education *December 2016*

Even among the elite group of students who do attend university, students from lower socio-economic backgrounds are less likely to attend a high-tariff university than their richer counterparts, even when they have similar A-level grades.

11. INTRODUCTION

Despite the possibility of high returns from a university education, many students do not go on to participate in higher education (HE). A great deal of policy interest has focused on the group of non-participants, and the reasons why they choose not to attend university. However, even among the elite group of students who do attend university, students from lower socio-economic backgrounds are less likely to attend a high-tariff university than their richer counterparts, even when they have similar A-level grades (Chowdry et al, 2013). One possible explanation, which has begun to receive increasing policy attention in the US concerns the decision-making process of those who do go on to attend. Research (Smith et al 2013) shows that some 40% of students in the US are 'undermatched' - ie their academic credentials would permit them to access a university that is more selective than the one they attend. This may be because students are badly informed about the benefits of certain universities or because of lack of knowledge of their own ability, or indeed because certain types of students may prefer to attend their local university. Whilst there is currently little research to help us understand whether this is a pressing problem in the UK, research (McGuigan, McNally & Wyness, 2016) has shown that students are fairly badly informed about how the costs and benefits of university differ by institution, and that this lack of knowledge is particularly pressing amongst disadvantaged students.

Another reason why students might make poor decisions regarding their choice of university stems from the UK's unique university applications process. Students in the UK still apply to university and receive initial offers based on their predicted A-level grades, rather than their actual results. Students must effectively commit to a particular university (or commit to a first and second choice) before they even sit their exams. This could result in students applying to universities that are a poor match (in terms of their academic attainment prior to university), and of course, universities missing out on talented students.

The system of predicted grades is inaccurate. Only 16% of applicants achieved the A-level grade points that they were predicted to achieve, based on their best three A-levels. However, the vast majority were over-predicted – ie their grades were predicted to be higher than they actually achieved.

The use of predicted grades has been widely criticised among policy makers and in the media (eg UCU, 2015; Wilson, 2015), but has not yet resulted in reform of the system. This area, meanwhile, has received little interest among the academic research community, most likely because of a lack of available data on applicants, their predicted and actual A-level grades and their subsequent university choices. applicants' actual and predicted grades and their university attended, as well as their background characteristics including level of disadvantage and school type. This data was obtained from the Universities and Colleges Admissions Service (UCAS), and is aggregate,[1] rather than individual-level, and hence restricted in a number of ways. Nevertheless, the data allow me to answer the following key questions:

1 How accurate are the predicted grades of university applicants?

2 How does grade accuracy vary according to student characteristics (gender, ethnicity, level of disadvantage) and the school type?

3 What is the impact of the predicted grade system on applicants' university choices?

I find evidence that the system of predicted grades is inaccurate. Only 16% of applicants achieved the A-level grade points that they were predicted to achieve, based on their best three A-levels. However, the vast majority (75% of applicants) were over-predicted – ie their grades were predicted to be higher than they actually achieved. Students from disadvantaged backgrounds and state schools are more likely to be over-predicted, whilst those at independent schools receive more accurate predictions. However, accuracy varies dramatically according to the A-level attainment of the student with lower attaining applicants far more likely to have their grades over-predicted. Therefore, after controlling for prior attainment and background characteristics, students from state schools are actually less likely to be over-predicted than those in independent and grammar schools.

Meanwhile, at the top of the attainment distribution, grades are slightly more likely to be under-predicted, and among these high-attaining students, applicants from low income backgrounds are significantly more likely to have their grades under-predicted than those from high-income backgrounds. This is important because under-predicted candidates are also more likely to apply to, and to be accepted to a university which they are overqualified for. This could in turn affect their future labour market outcomes.

The remainder of the report proceeds as follows. In section 2, I outline the data used in this analysis, and the methodologies employed. In Section 3, I present the main findings of the analysis. Section 4 concludes.

2. DATA

This project uses bespoke data provided by the University Colleges Admissions Service (UCAS). Therefore, the sample being used throughout this paper is of all applicants/acceptances to higher education courses (including those taught in colleges as well as universities) in the UK. We therefore cannot include students who did not apply for HE, such as those who chose to attend FE or to go directly into the labour market.

The data are aggregate (for reasons of privacy) to A-level points level, containing information on the number of university applicants, applications, and acceptances.

At the top of the attainment distribution, grades are slightly more likely to be under-predicted, and among these high-attaining students, applicants from low income backgrounds are significantly more likely to have their grades under-predicted than those from high-income backgrounds.

The variables included in the dataset are: A-level points (with points score defined by UCAS as the points score attached to the highest 3 A level grades achieved by the applicant, with the following points per grade used in the calculation: A*=6, A=5, B=4, C=3, D=2, E=1), the difference in achieved and predicted A-level score (with points score based on the same definition as previous), gender, ethnicity, most recent educational establishment type (which is generally school type but includes other routes such as FE college), Polar3 quintile (HEFCE's definition of disadvantage, which is defined according to the participation rate of the area the applicant is living in), cycle year (for 2013, 2014, 2015) and university tariff band (a measure of university quality based on the tariff score of accepted applicants, 'higher', 'medium' and 'lower'[2]). I also have information on the clearing routes of applicants (firm choice, insurance choice, main clearing, direct clearing or adjustment). In a final dataset I have information on the provider name, A-level points, predicted points again for applicants and acceptances. However because of the more granular level of university information, the only background characteristic included in this dataset is Polar3.

More information on the variables listed can be found in Appendix 1.

3. RESULTS

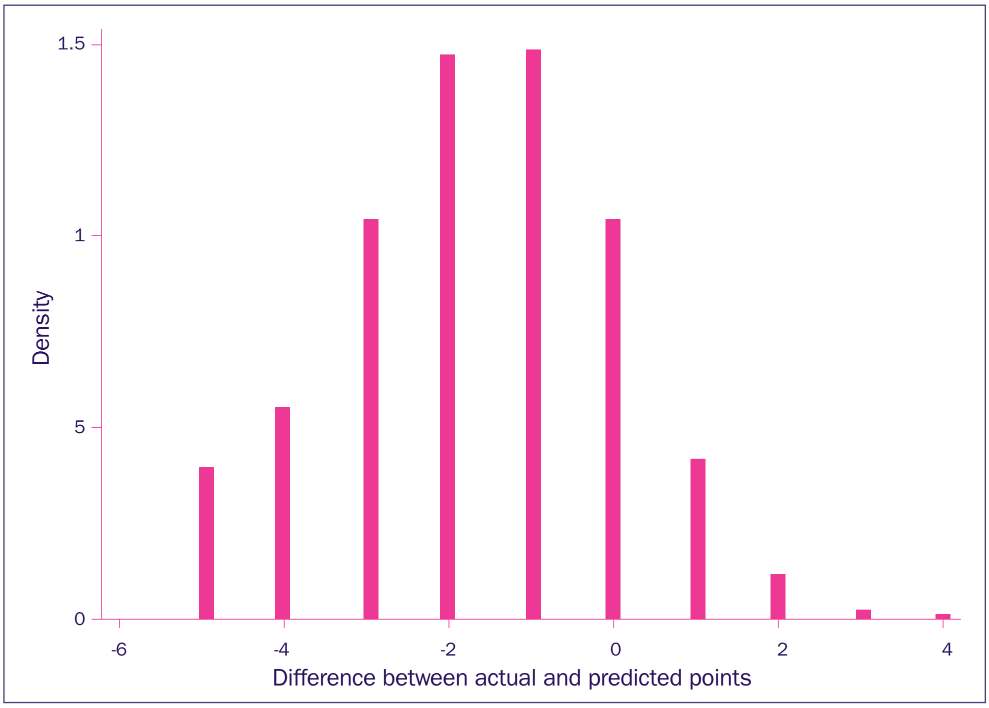
How accurate are predicted grades?

I begin by examining the accuracy of predicted grades among the population of university applicants. Figure 1 shows the distribution of the difference between actual and predicted grades for applicants between 2013-2015, where each point on the x-axis represents the achieved point score of the students minus the predicted point score of the students. In other words, a negative score means that the student achieved a lower grade than they were predicted to achieve (the student's grade was over-predicted), whilst a positive score means the student achieved a higher grade than they were predicted to achieve (the student's grade was under-predicted).

Figure 1 shows that, based on this particular measure,[3] the majority of predictions are over-predicted. In Table 1 and Figure 2, I explore this in more detail. Table 1 shows that accuracy is around 16% - ie 16% of applicants' grades were correctly predicted. A further 8.5% of applicants were under-predicted, whilst the vast majority (75%) of applicants' grades were over-predicted. This suggests a lower level of accuracy than previous research (Everett & Papageorgiou, 2011) which found that 41% of grades were over-predicted, 51.7% were accurate, and only 6.6% were under-predicted, however the Everett & Papagerogiou research uses a different measure of grade accuracy (based on the actual grades of each individual A-level, rather than based points assigned to the best 3 as in this paper), therefore is not directly comparable with the results presented in this paper. Nevertheless it is notable that in both their and my research, under-prediction is uncommon.

As table 1 also shows, the majority of predicted grades were within 1-2 points (where, as explained, 1 point equates to 1 grade, ie the difference between AAA=15 points and AAB=14 points, or DDD and DDE). Nevertheless, 6% of applicants' were over-predicted to the tune of 5 points – equivalent to an A-level at grade A.

Figure 1: Distribution of the difference between actual and predicted grades



Note 1: Each point on the x-axis represents the achieved point score of the applicant minus the predicted point score of the applicant

Note 2: Points score is defined by UCAS as the points score attached to the highest 3 A level grades achieved by the applicant, with the following points per grade used in the calculation: A*=6, A=5,B=4, C=3, D=2, E=1

Table 1: under/over prediction by A-level points

Difference between actual and predicted points	%
-5	6.03
-4	8.35
-3	15.88
-2	22.5
-1	22.65
0	16.05
1	6.45
2	1.82
3	0.26
4	0.01
Total	100

Does grade accuracy vary according to the 'ability' (as measured by prior academic attainment) of the student? Figure 2 provides the answer to this. The solid black bar shows the achieved A-level points score. The grey bar next to it shows the average point score prediction for applicants with that achieved points score. So for example, the first bar on the left shows that for applicants who actually achieved 3 points (eg EEE), the average prediction for such applicants was 8 A-level points (CCD) – a substantial over-prediction. It is notable from Figure 2 that over-prediction particularly occurs at the low end of the A-level attainment distribution, with those students achieving between 3-6 A-level points being over-predicted by upto A-level 4 points (a B-grade). It is also notable that as students' A-level attainment increases, so does the accuracy of their predicted grades. Indeed, only students with A-level points at 17 or above (A*A*A) appear to over-achieve on their predictions (though note, of course, that there are ceiling effects at play here; it is hard for students to over-achieve on very high predictions, and straight A* students cant, by definition, be over-predicted. Note also that this chart simply presents the average prediction for each A-level point score – there will still be a distribution of under/over predictions for each point score).

Whilst this research cannot uncover the causes of this inaccuracy, the fact that teachers tend to over-predict rather than under-predict grades (particularly at the lower end of the A-level attainment distribution) perhaps implies that they are using predicted grades as a target for students to aim for, rather than a prediction of how they will perform. Moreover, there would seem to be little incentive for teachers to under-predict a students' grade since this may encourage them to 'give up' or at least discourage them from aiming 'high', and will also likely increase their likelihood of an offer (UCAS, 2015). Another possibility is that AS-levels, upon which predictions are often based, are a poor predictor of subsequent A-level attainment; weaker students may perform well at (the relatively easier) AS level, but subsequently go on to perform badly at the more challenging A-level. It is also relevant that the AS is now being 'decoupled" from A-levels, meaning that it will no longer contribute to the final grade of a student. This effectively means that some students will no longer sit the AS-level, potentially making it even harder for teachers to make predictions. This may result in some teachers/schools using GCSE outcomes as predictors of A-level results, potentially introducing even more error into predictions.

Figure 2: Accuracy of predicted grades, by applicants' prior A-level points score

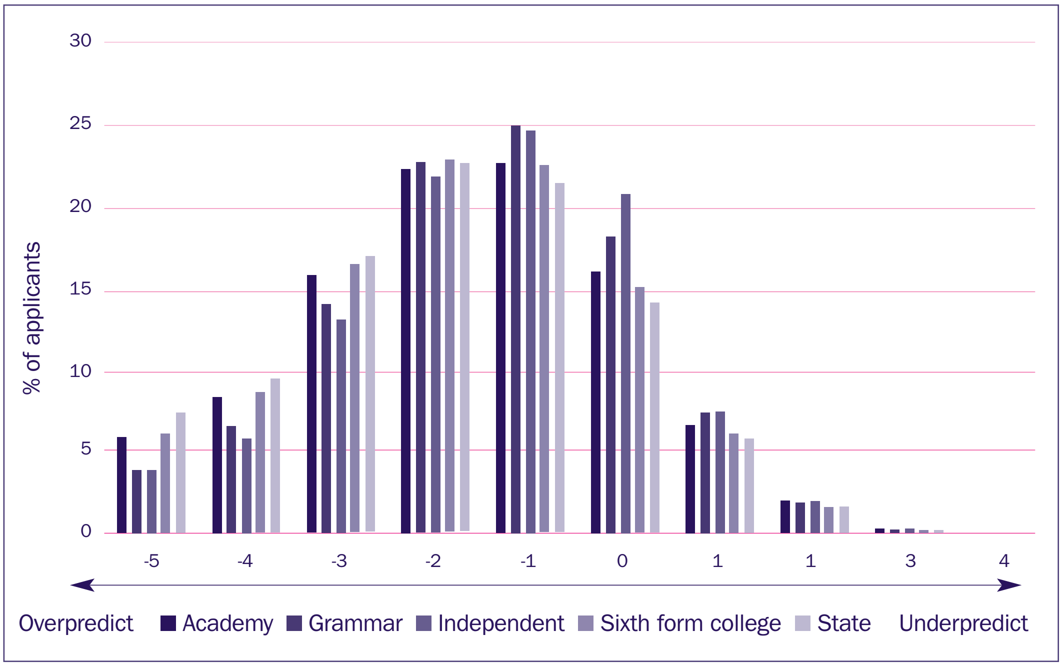
Note: Points score is defined by UCAS as the points score attached to the highest 3 A level grades achieved by the applicant, with the following points per grade used in the calculation: A*=6, A=5, B=4, C=3, D=2, E=1

How does accuracy vary by student characteristics and school type?

In this section I examine how accurate grades are according to the characteristics of students and schools, examining whether certain types of students, or school types are particularly likely to under or over-predict. This analysis is presented in figures 3-4.

As figure 3 shows, there is a good deal of variation in prediction accuracy according to school type. Independent schools appear to be the most accurate predictors – over 20% of applicants from independent schools' grades were accurately predicted. Meanwhile, academies, state schools and sixth form colleges are more prone to over-predicting their students' grades (as seen by the greater proportions from these school types with a difference between actual and predicted grades below zero). As is evident, there is a limited amount of under-prediction; however this will be explored in more detail in Section 3.

Figure 3: Difference between actual and predicted grades, by school type



Note 1: Each point on the x-axis represents the achieved point score of the applicant minus the predicted point score of the applicant

Note 2: Points score is defined by UCAS as the points score attached to the highest 3 A level grades achieved by the applicant, with the following points per grade used in the calculation: A*=6, A=5, B=4, C=3, D=2, E=1

Are applicants from disadvantaged backgrounds more likely to receive inaccurate predictions? The answer can be seen in Figure 4 which shows prediction accuracy according to the Polar 3 measure of disadvantage. As Figure 4 shows, that those applicants from the most disadvantaged backgrounds (ie Polar groups 1 and 2) are the least likely applicants to have their grades accurately predicted. They are also more likely to have their grades over-predicted than those from more advantaged groups. Indeed, the gradient in accuracy by advantage level is quite striking in this chart.

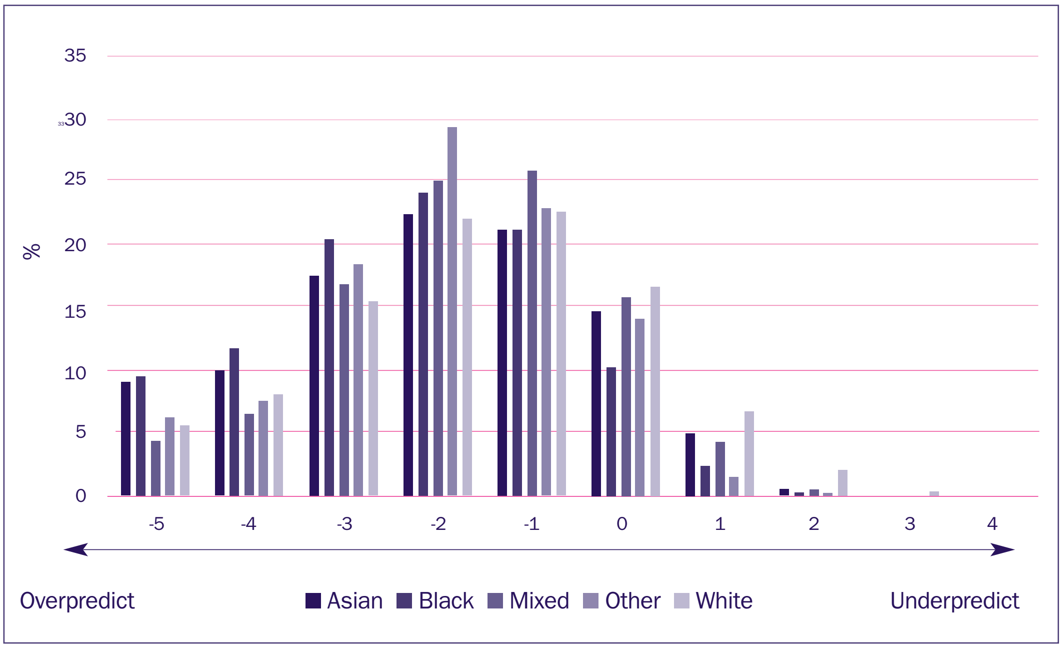
For example, applicants from the most disadvantaged backgrounds are more likely to experience moderate to severe over-prediction (ie from 2 to 5 points over-predicted) than those the most advantaged backgrounds. For example 7.5% of the most disadvantaged students were over-predicted by 5 points (equivalent to an A grade at A-level), compared with 5% of the most advantaged students. But the gradient reverses as accuracy improves, so that the most advantaged applicants are more likely than the most disadvantaged to experience relative accuracy (ie a difference in achieved and predicted between -1 to +1) in their predicted grades. Interestingly, however, for applicants whose grades are under-predicted by 2 points or more, there appear to be no differences by level of disadvantage.

Figure 4: Difference between actual and predicted grades, by level of disadvantage

Note 1: Each point on the x-axis represents the achieved point score of the applicant minus the predicted point score of the applicant

Note 2: Points score is defined by UCAS as the points score attached to the highest 3 A level grades achieved by the applicant, with the following points per grade used in the calculation: A*=6, A=5, B=4, C=3, D=2, E=1

Figure 5: Difference between actual and predicted grades, by ethnic group



Note 1: Each point on the x-axis represents the achieved point score of the applicant minus the predicted point score of the applicant

Note 2: Points score is defined by UCAS as the points score attached to the highest 3 A level grades achieved by the applicant, with the following points per grade used in the calculation: A*=6, A=5, B=4, C=3, D=2, E=1

Finally, in Figure 5, I examine predicted grade accuracy by ethnicity. Here we can see that, among all ethnic groups, Asian and Black applicants are more likely to be severely over-predicted. White applicants are most likely of all groups to be accurate (though at similar levels to black students) and are also most likely to be slightly over-predicted. Finally, I can also examine whether girls are more likely to have their grades over-predicted than boys. The results (not shown here) reveal no apparent difference in their rates of grade accuracy.

The results so far in this section have shown that applicants from state schools and disadvantaged backgrounds are more likely to be over-predicted. However, in Section 3 we learned that low ability students are more likely to be over-predicted than high ability students. But what about the interaction between the two? In particular, research from the US (Hoxby & Avery, 2012) has shown that high ability but disadvantaged young people very often do not apply to selective colleges or universities despite being qualified to do so, and despite the fact that these selective institutions would cost them less. Therefore, it is of interest to examine the accuracy of grade prediction among the group of disadvantaged but high ability students. The results are presented in Table 2.

Table 2: Accuracy of prediction, high ability applicants and level of disadvantage, proportion and frequency

	All	AAB, all applicants	AAB, most disadvantaged	AAB, least disadvantaged
Proportion of students				
Overpredicted (%)	75.41	46.46	44.03	47.4
Accurate (%)	16.05	32.17	32.31	32.29
Underpredicted (%)	8.54	21.37	23.66	20.31
Number of students (2013-2015)				
Overpredicted	647,680	107,440	5,145	47,845
Accurate	137,850	74,405	3,775	32,595
Underpredicted	73,305	49,425	2,765	20,500
Total	858,835	231,270	11,685	100,940

As can be seen, high ability applicants (defined here as those with 14 A-level points or more – equivalent to AAB or more at A-level) do tend to be more likely to be under-predicted than the average applicant, with 21% of AAB applicants having their points score under-predicted, versus 8.5% of all applicants. Among these AAB applicants, applicants from the most disadvantaged group are slightly more likely to have their grades under-predicted than those from the least disadvantaged groups (23.7% vs 20.3%) suggesting there is some evidence that high ability disadvantaged students are particularly likely to fall into the category of being under-predicted. Of course, as is widely known (eg Chowdry et al, 2013), only a small number of disadvantaged young people fall into the high-attainment category – as can be seen in the lower half of Table 2. Nevertheless, it can be seen that under-prediction affects some 2,700 high-ability disadvantaged students.

In table 3 I explore this in a regression framework, allowing me to examine whether certain groups are more or less likely to be under-predicted, whilst holding constant other factors such as ethnicity, gender or school attended, which also affect prediction accuracy.

Column 1 contains the raw results for the impact of disadvantage level on the probability of being under-predicted (which here, is defined as a binary variable=1 if under-predicted, and=0 if accurately or over-predicted), controlling only for academic year. As expected, those from the poorest quintile are less likely to be under-predicted (by 2.6 percentage points) than those from the richest quintile (the reference category). Moreover, it can also be seen that rates of under-prediction have been falling over time since 2013 (the reference year).

Table 3: Probability of grades being under-predicted

VARIABLES	(1) raw	(2) +chars	(3) CCC	(4) CCC-AAB	(5) AAB
Quintile of disadvantage (ref=richest)					
Poorest quintile	-0.026*** (0.001)	-0.024*** (0.001)	-0.003*** (0.001)	-0.001 (0.001)	0.011*** (0.004)
2nd poorest quintile	-0.019*** (0.001)	-0.016*** (0.001)	-0.003*** (0.000)	0.001 (0.001)	0.008** (0.003)
Middle quintile	-0.015*** (0.001)	-0.010*** (0.001)	-0.002*** (0.000)	0.001 (0.001)	0.013*** (0.003)
2nd richest quintile	-0.008*** (0.001)	-0.007*** (0.001)	-0.001*** (0.000)	0.000 (0.001)	0.002 (0.002)
School type (ref=state)					
Academy		0.013*** (0.001)	0.001** (0.000)	0.005*** (0.001)	-0.010*** (0.003)
FE college		-0.024*** (0.002)	-0.004*** (0.001)	-0.023*** (0.002)	-0.019*** (0.006)
Grammar		0.008*** (0.001)	-0.003*** (0.001)	-0.012*** (0.001)	-0.046*** (0.004)
Independent		0.012*** (0.001)	-0.004*** (0.001)	-0.015*** (0.001)	-0.065*** (0.003)
Other		-0.066*** (0.005)	-0.003 (0.002)	-0.051*** (0.005)	-0.089*** (0.023)
Sixth form College		0.005*** (0.001)	-0.000 (0.000)	0.003*** (0.001)	-0.017*** (0.003)
Ethnic group (ref=white)					
Asian		-0.037*** (0.001)	-0.001*** (0.000)	-0.013*** (0.001)	-0.100*** (0.003)
Black		-0.069*** (0.002)	-0.003*** (0.001)	-0.032*** (0.002)	-0.158*** (0.008)
Mixed		-0.050*** (0.002)	-0.005*** (0.001)	-0.033*** (0.002)	-0.107*** (0.005)

Other		-0.085*** (0.004)	-0.005* (0.003)	-0.046*** (0.004)	-0.195*** (0.013)
Male		-0.002*** (0.001)	0.003*** (0.000)	0.011*** (0.001)	-0.018*** (0.002)
Year=2014	-0.003*** (0.001)	-0.003*** (0.001)	0.000 (0.000)	-0.005*** (0.001)	0.009*** (0.002)
Year=2015	-0.007*** (0.001)	-0.006*** (0.001)	0.000 (0.000)	-0.007*** (0.001)	0.004** (0.002)
Constant	0.098*** (0.001)	0.099*** (0.001)	0.004*** (0.000)	0.053*** (0.001)	0.253*** (0.003)
Observations	858,720	858,720	146,890	480,610	231,220
R-squared	0.001	0.006	0.001	0.004	0.013

In the second column I add student background characteristics and school type. There is a small increase in the most advantaged coefficient, and an increase in the R-squared, indicating that background characteristics do explain some of the variation in prediction accuracy. In particular, we can see that, in comparison to those from state schools, those whose last educational establishment were independent schools, grammar schools or academy schools are more likely to be under-predicted, as seen above. We can also see that, all else being equal, white applicants are most likely to be under-predicted than all other ethnic groups, and boys are less likely to be under-predicted than girls.

However, we have seen that the probability of being under or over predicted varies dramatically by student A-level achievement level. Therefore in the next 3 columns I attempt to account for this by sub-dividing the analysis by achieved A-level grades (CCC or less, CCC-AAB, and finally, AAB or more). Looking first at column 3, we can see that, after controlling for school type and student background characteristics, ‘low ability’, disadvantaged students are less likely to be under-predicted than ‘low ability’ but advantaged students. However, column 4 shows that, for ‘medium-ability’ (defined as achieving grades above CCC but lower than AAB) students there is little difference in probability of being under-predicted according to disadvantage. Finally, column 5 shows that ‘high ability’ (AAB or more) but disadvantaged students are significantly more likely to have their grades under-predicted than ‘high ability’ students from the most advantaged backgrounds, even after controlling for school type, gender, ethnicity and year. Although this is a small effect (of only 1 percentage point), it is nevertheless a worrying finding – it implies that some of our most high achieving students may be misinformed about their likely potential.

Table 3 also shows that, once we control for student prior attainment, those attending state schools are more likely to have their grades under-predicted than those from independent schools and grammar schools (columns 3-5). (Interestingly, the converse is also true; after controlling for ability and characteristics, students from state schools are less likely to be over-predicted than those from independent schools).[4] In terms of ethnicity, high achieving white students are more likely to be under-predicted than other ethnic groups. Boys are, however, still less likely to be under-predicted than girls.

If being under-predicted results in students applying to universities which they are over-qualified for, this could have a direct impact on social mobility, potentially skewing social representation across HE institutions. Moreover, it could result in students becoming de-motivated and dropping out of university altogether. On the other hand, students may not pay a great deal of attention to their grade point predictions, therefore the degree of accuracy may not actually affect students' application decisions at all. How grade accuracy impacts students' university choices is the question that I turn to in the next section.

What is the impact of the predicted grades system on students' university choices?

We can now turn to the important question of what is the impact of predicted grades – or rather inaccuracy of predicted grades – on students' university choices. In particular, the question I aim to ask here is whether students whose grades are under or over-predicted, apply to universities which they are under or over-qualified for, bearing in mind their academic credentials.

First, I examine the probability of applying to a high-tariff university based on accuracy of A-level point prediction. In other words, are applicants whose grades are under predicted less likely to apply to a high-tariff university, and are applicants whose grades are over predicted more likely to do so? The definition of a high-tariff university is again somewhat restrictive (see data section, and footnote 2); there are only three categories of university, therefore high-tariff universities represent the top third. The results can be found in Table 4 and again are split by A-level points achieved. It is clear that for all A-levels achieved, applicants whose grades are over-predicted are more likely to apply to a high-tariff university than those whose grades are accurate. And by contrast, applicants whose grades are under-predicted are less likely to apply to a high tariff university. The results also show that those from the poorest backgrounds, even within A-level ability groupings, are less likely to apply to a high tariff university, and applicants those from independent schools are most likely to apply to a high tariff.

What can we conclude from this analysis? The finding that high ability under-predicted applicants are less likely to apply to top universities (to the tune of 8.7 percentage points) is certainly of concern, and is evidence that grade prediction may be reducing the potential of these students. But what about students with lower grades? For example we find that students with CCC or below are more likely to apply to a high tariff university if their grades are over-predicted. Whilst some could argue this is a positive outcome, and that more holistic widening admissions policies have resulted in good progress in widening participation in recent years (Lasselle et al, 2014), a potential downside is that being over-predicted could result in student attending universities they are not qualified for, resulting in them struggling to complete their degrees or dropping out. However it is important to note that this report cannot examine the success rates of so-called 'mismatched' students; this is a particularly under-researched area in the UK.

Table 4: probability of applying to a high-tier university

VARIABLES	(1) CCC	(2) CCC-AAB	(3) AAB
Accuracy of prediction (ref=accurate)			
Over-predicted	0.110*** (0.006)	0.096*** (0.002)	0.036*** (0.002)
Under-predicted	-0.048*** (0.017)	-0.060*** (0.004)	-0.087*** (0.003)
Quintile of disadvantage (ref=poorest)			
Poorest quintile	-0.005 (0.003)	-0.045*** (0.003)	-0.061*** (0.005)
2nd poorest quintile	0.000 (0.003)	-0.043*** (0.002)	-0.061*** (0.004)
Middle quintile	-0.001 (0.003)	-0.035*** (0.002)	-0.055*** (0.003)
2nd richest quintile	-0.006** (0.003)	-0.026*** (0.002)	-0.038*** (0.003)
Schooltype (ref=state)			
Academy	0.002 (0.002)	0.026*** (0.002)	0.032*** (0.003)
FE college	0.018*** (0.005)	0.018*** (0.004)	0.024*** (0.007)
Grammar	-0.028*** (0.005)	-0.054*** (0.003)	-0.122*** (0.004)
Independent	0.045*** (0.005)	0.130*** (0.003)	0.156*** (0.003)
Other	-0.053*** (0.015)	0.004 (0.011)	0.052** (0.027)
Sixth form college	0.012*** (0.003)	0.020*** (0.002)	0.007** (0.004)
Ethnic group (ref=white)			
Asian	0.028*** (0.003)	0.030*** (0.002)	0.007** (0.003)
Black	-0.017*** (0.004)	0.019*** (0.003)	0.058*** (0.009)
Mixed	-0.047*** (0.007)	0.020*** (0.004)	0.086*** (0.006)
Other	-0.031* (0.017)	0.011 (0.009)	0.086*** (0.016)
Male	0.022*** (0.002)	0.039*** (0.001)	0.063*** (0.002)
Year=2014	0.013*** (0.002)	0.015*** (0.002)	0.008*** (0.002)
Year=2015	0.031*** (0.002)	0.025*** (0.002)	0.012*** (0.002)

Constant	0.014** (0.007)	0.200*** (0.003)	0.550*** (0.004)
Observations	146,890	480,610	231,220
R-squared	0.008	0.022	0.046

This raises the question, do students with inaccurate grade predictions apply to universities that are not appropriate for them, based on their realised A-level points? And importantly, do accepted applicants whose A-level scores were under-predicted, end up at universities which they are over-qualified for?

To answer this question, I turn to a second dataset (again described in Section 2) which contains A-level points and predicted grades at every UCAS university. In order to perform this analysis I first calculate the average A-level point score at every university in my sample (by year, based on the A-level point scores of acceptances). I then define a student as over-qualified for the university they applied to if their own A-level achieved score exceeds the average for that university. Note, that this measure of over-qualification is, of course, subject to error. In particular, whilst some students may have A-level scores well above or below the mean for their university this could simply reflect the particular course that they are studying, since even within universities, courses have different entry requirements. As I do not have information on the course of study, I am unable to take this into account, and therefore these results can be seen as merely suggestive evidence of under/overqualification.

The results of this analysis can be found in Table 5. Again, I have grouped the data by student ability level. Note that for this analysis my set of variables is far more restrictive than above – in particular I am not able to control for any student characteristics, or school type. However, since the results in Section 3.2 were not affected by these controls by any great magnitude, we can assume with reasonable safety that this would not dramatically impact the findings.

In Table 5 I present two sets of results. In columns 1 and 2 I present analysis for unique applicants, in columns 3 and 4 I present results for acceptances at university. In column 1 we can see that under-predicted applicants are 10 percentage points more likely than applicants whose grades were accurate or over-predicted to have applied to a university that they are over-qualified for. In column 2 we can see that the most able applicants who under-predict are also more likely to apply to universities that they are over-qualified for. Column 3 now shows that among those accepted to university, having under-predicted grades increases the likelihood of being over-qualified for university (though perhaps unsurprisingly, the probability drops to just 1.5 percentage points). Finally in column 4 we can see that, among the most able students this is also true, and indeed being under-predicted increases the chances of attending a university which one is over-qualified for by 6.9 percentage points). Finally it is of interest to note that among accepted applicants from the most highly able groups (Column 4), the most disadvantaged students are most likely to be over-qualified for their university. This latter finding implies that the predicted grades system may be skewing the socio-economic mix of students at certain universities; if high ability in section 3), and these students in turn apply to universities that they are over-qualified for, then this could result in fewer disadvantaged students at top universities.

Table 5: probability of being over-qualified for university

VARIABLES	(1) raw	(2) ABB+	(3) raw	(4) ABB+
Under-predicted	0.102*** (0.001)	0.091*** (0.001)	0.015*** (0.002)	0.069*** (0.002)
Quintile of disadvantage (ref=poorest)				
Poorest quintile	-0.164*** (0.001)	0.003* (0.002)	-0.187*** (0.003)	0.014*** (0.004)
2nd poorest quintile	-0.116*** (0.001)	0.001 (0.001)	-0.123*** (0.002)	0.003 (0.002)
Middle quintile	-0.089*** (0.001)	-0.002 (0.001)	-0.087*** (0.002)	0.005** (0.002)
2nd richest quintile	-0.038*** (0.001)	-0.001 (0.001)	-0.036*** (0.002)	0.003* (0.002)
Year=2014	-0.012*** (0.001)	0.015*** (0.001)	-0.004** (0.002)	0.016*** (0.002)
Year=2015	-0.014*** (0.001)	0.010*** (0.001)	-0.009*** (0.002)	0.017*** (0.002)
Constant	0.600*** (0.001)	0.885*** (0.001)	0.663*** (0.002)	0.906*** (0.001)
Observations	2,842,190	701,070	509,820	130,610
R-squared	0.016	0.017	0.015	0.012

Finally, it is of interest to understand whether grade prediction affects students' acceptance route. In particular, it may be the case that students whose grades are under-predicted are more likely to have to go through clearing. I examine this in Table 6. As can be seen, the greatest proportion of acceptances (some 88%) were accepted at their firm or insurance choice institution, regardless of the accuracy of their grade point prediction (i.e the right-most column of Table 6). Interestingly, and perhaps not surprisingly a far greater proportion of over-predicted students went through clearing to obtain their place – though this may well reflect the fact that over-predicted students tend to have lower A-level scores (which I am unable to control for here). Whilst, those with accurate or under-predicted grades were extremely unlikely to go through clearing. We can also examine students' likeliness to go through 'adjustment' – ie where applicants who have met and exceeded the conditions of their firm choice offer choose to take up an alternative offer. As is evident from Table 6, only a tiny proportion of acceptances actually availed themselves of this opportunity, suggesting that knowledge of this particular aspect of the UCAS process is still rather limited – or perhaps that the most desirable places are already full by the time students reach this stage. Given that the adjustment procedure was created as a means to solve the problem of under-prediction, it is certainly a concern that so few individuals are actually using it.

Table 6: Grade accuracy and acceptance route

Acceptance Route	Over-predicted	Correct	Under-predicted	Total
Firm choice	70.14	93.2	96.4	76.78
Insurance choice	15.78	3.31	0.73	12.11
Main scheme Clearing	13.66	3.1	2.05	10.66
Extra	0.39	0.24	0.08	0.33
Adjustment	0.02	0.15	0.71	0.11
Total	100	100	100	100

Finally, it is of interest to examine whether certain groups of students are more or less likely to end up in clearing. As Table 7 shows, students from all SES groups appear to be almost equally likely to go through clearing – as is evident from the right-most column, 10% of the most disadvantaged students, and 10.6% of the least disadvantaged students take this route.

Table 7: Grade Accuracy and acceptance route, by level of disadvantage

Acceptance Route	Over-predicted	Correct	Under-predicted	Total
Most disadvantaged students				
Firm choice	72.44	94.48	99.21	77.96
Insurance choice	14.59	3.33	0.2	11.71
Main scheme Clearing	12.54	2.19	0.59	10
Extra	0.44	0	0	0.33
Adjustment	0	0	0	0
Total	100	100	100	100
Least disadvantaged students				
Firm choice	69.48	93.3	95.04	76.83
Insurance choice	16.25	3.36	1.02	12.13
Main scheme Clearing	13.93	2.86	2.75	10.58
Extra	0.31	0.28	0.07	0.28
Adjustment	0.03	0.21	1.12	0.18
Total	100	100	100	100

CONCLUSIONS

The UK's unique system of grade prediction has been widely criticised by policymakers and the media, yet the system has remained unchanged for many years. This report has used UCAS data to examine grade accuracy amongst university applicants and acceptances over 2013-2015.

I find a high level of inaccuracy of grade prediction. Among the best 3 A-levels students achieve, only 16% of higher education applicants' grades are accurately predicted. However, the vast majority of applicants actually receive predictions that are too optimistic for the grades they actually go on to achieve, with 75% of applicants achieving lower grades than predicted.

Disadvantaged students and students from state schools and academy schools are most likely to be over-predicted. However, I find that prediction accuracy and direction varies dramatically by student's prior A-level achievement. Low achieving applicants (who are disproportionately from poorer backgrounds and state schools) are far more likely to be over-predicted, whilst high achieving applicants are more likely to be under-predicted. Since low achieving applicants are more likely to be found at state schools, once I control for A-level scores and characteristics, I find that state school students are actually less likely to be over-predicted than independent and grammar school students.

When thinking about students university choices, however, we may be more concerned about under-prediction. And indeed I also find worrying evidence that, among high-achieving (ie AAB or more) applicants, disadvantaged students are more likely to be under-predicted than their more advantaged counterparts. Indeed almost 3,000 disadvantaged, high-achieving students (or 1,000 per year) have their grades under-predicted.

Whilst this report cannot study the consequences of under-prediction, I find evidence that applicants who are under-predicted are more likely to apply to, and attend, a university that they are over-qualified for (i.e. where their A-level points are higher than the average for that university). Assuming a link between university tariff and future earnings (Walker and Zhu, 2013), prediction accuracy could therefore have a negative impact on the labour market outcomes of high ability but disadvantaged students, as well as potentially skewing the socio-economic mix of high-tariff universities.

On the other hand there is also a correlation between being over-predicted and an applicant's likeliness to apply to a university that they are under-qualified for. Whilst some might argue that this is a positive aspect of the UCAS system, enabling lower achieving students to 'overshoot' and apply to a high-tariff university, it could potentially have negative consequences if such students go on to struggle during their degrees.

Moreover it seems highly inefficient to continue with a system in which life changing decisions are made, and scarce university places are allocated, on the basis of inaccurate information.

APPENDIX: TECHNICAL NOTES AND DEFINITIONS

Data Set 1

Statistics Requested: Number of Applications, Number of Unique Applicants, Number of Acceptances

Summary: Breakdown of Difference in Achieved and Predicted A level points score, Achieved A level point score, Provider Tariff Band, POLAR3 Quintile, Sex, Ethnic Group (Summary Level) , Educational Establishment Type for applicants domiciled in the UK (excluding Scotland) aged 18 years old.

Data Set 2

Statistics Requested: Number of Acceptances

Summary: Breakdown of Difference in Achieved and Predicted A level points score, Provider Tariff Band, POLAR3 Quintile, Sex, Ethnic Group (Summary Level), Educational Establishment Type, Acceptance Route for applicants domiciled in the UK (excluding Scotland) aged 18 years old.

Data Set 3

Statistics Requested: Number of Applications, Number of Unique Applicants, Number of Acceptances

Summary: Breakdown of Provider Name 2015, Achieved A level point score, Predicted A level points score, POLAR3 Quintile for applicants domiciled in the UK (excluding Scotland) aged 18 years old.

Years: 2013 - 2015. Year is defined as cycle year, when the application was processed, rather than entry year.

Cycle Reference Point: End of Cycle.

Applicant Domicile (High Level)

Applicant's area of permanent residence summarised at a high level. This variable is derived from domicile as declared by the applicant and does not guarantee fee status. Split by the following values: 'England', 'Northern Ireland', 'Scotland', 'Wales', 'EU (excluding UK)' and 'Not EU'. Please note: The Channel Islands and the Isle of Man have been assigned as 'Not EU'.

Age Band (5 Levels)

Derived from date of birth declared by the applicant, age is aligned with the cut off points for school/college cohorts within the different administrations of the UK. For England and Wales ages are defined on the 31 August, for Northern Ireland on the 1 July and for Scotland on the 28 February the following year. Defining ages in this way matches the assignment of children to school cohorts. For applicants outside of the UK the cohort cut off for England and Wales has been used. Age then is grouped in the following bands: '17 and under', '18', '19', '20', '21 and Over'. Please note: if the derived age is outside of the range 11-100 then it is set to 18 by default.

Sex

Sex as declared by the applicant.

Ethnic Group (Summary Level)

High level grouping of ethnic origin as declared by the applicant: ‘White’, ‘Black’, ‘Asian’, ‘Mixed’, ‘Other’, ‘Unknown’. Please note: Ethnic origin is captured for UK domiciled applicants only, therefore all non UK domiciled applicants are assigned as ‘Not Applicable’, and is not available within the current live cycle.

POLAR3 Quintile

Developed by HEFCE, POLAR3 classifies small areas across the UK into five groups according to their level of young participation in Higher Education. Each of these groups represents around 20 per cent of young people and is ranked from Quintile 1 (areas with the lowest young participation rates, considered as the most disadvantaged) to Quintile 5 (highest young participation rates, considered most advantaged). POLAR3 is based on the participation rates of young people between 2005 and 2009 who entered Higher Education between 2005-06 and 2010-11, therefore is most suitable for applicants aged 19 and under. These groups are assigned using the postcode declared by the applicant at the time of their application. If a UK postcode is invalid, considered unsafe for measurement or there is no link to Census geography possible then the applicant is classified as ‘Not Assigned’. Please note: POLAR3 is only relevant for applicants domiciled in the UK (as determined by EXACT definitions of domicile), therefore any applicants outside of this cohort are classified as ‘Not Applicable’. Following the inclusion of 2015 application cycle data a small change was made to allow for the assignment of POLAR3 to any new postcodes from 2011.

Educational Establishment Type

The type of the educational establishment most recently attended by the applicant. It is the type of the school or centre that the applicant is applying through, or, if an applicant is applying to UCAS directly, the type of the most recent school or centre as declared by the applicant. In the interest of keeping a consistent time-series, the most recent school or centre type for each school code held by UCAS is displayed across the time series, regardless of the school or centre type at the time of the application. For example, academies were introduced 2012. Therefore any schools that were formerly ‘Grammar’ (or any other type) and changed to an academy in 2012 will be displayed as ‘Academy’ across every application cycle going back to 2009. Please note: Each school or centre can take only value for their educational establishment type, so numbers are not reported twice if the centre is both an academy and sixth form college, for example. This value is usually assigned by the school or centre themselves. It is restricted to 19 year-olds and under, therefore any applicants outside of this cohort are classified as ‘Not Applicable’.

Provider Tariff Band

The grouping of providers based on the average levels of attainment of their accepted applicants (summarised through UCAS Tariff points) in a period of application cycles spanning from 2004 to 2011. Each group of providers accounted for around a third of all UK 18 year old acceptances in these cycles. Split by the following values: ‘Higher’, ‘Medium’, ‘Lower’.

Provider Name 2015

A mapped view of the Higher Education Provider (HEP) as-at 2015. In the interest of keeping a consistent time series the HEP displayed is mapped from a previous HEP if a merger has occurred or the HEP has been renamed. E.g. In 2013 HEPs K05 and H50 merged, so all instances of H50 prior to 2013 will be reported as K50.

Predicted A level points score

A score attached to the predicted A level grade profile declared by the applicant on the application. The highest 3 grades are considered only, adding up the following points per grade: A* = 6, A = 5, B = 4, C = 3, D = 2, E = 1. AS Levels are not included. This variable is currently only available for 18 year-old applicants domiciled in England, Northern Ireland and Wales, any applicants outside of this cohort will be assigned the value ‘Not applicable’. These scores are calculated purely for analytical purposes and do not relate to the UCAS tariff.

Achieved A level point score

A points score attached to the highest 3 A level grades achieved by the applicant as identified either on application or through awarding body linkage process. The following points per grade are used in the calculation: A* = 6, A = 5, B = 4, C = 3, D = 2, E = 1. A level Double Award grades are included, AS Level grades are not included. The score represents the qualifications recorded as held by the applicant at the end of the cycle to which the statistic relates. This variable is currently only available for 18 year-old applicants domiciled in England, Northern Ireland and Wales, any applicants outside of this cohort will be assigned the value ‘Not applicable’. These scores are calculated purely for analytical purposes and do not relate to the UCAS tariff.

Difference in Achieved and Predicted A level points score

The difference between the points score attached to the A level grades achieved by the applicant, as identified either on application or through awarding body linkage process, and the score attached to the predicted A level points score declared by the applicant on the application. This is only applicable to those with at least 3 predicted and achieved A level grades. The highest 3 grades are considered only, adding up the following points per grade: A* = 6, A = 5, B = 4, C = 3, D = 2, E = 1. AS Levels are not included. Due to relatively small frequencies of more extreme values, differences of 5 points or more, positive or negative, have been given the value of plus or minus '5 or more points'. These scores are calculated purely for analytical purposes and do not relate to the UCAS tariff. Zero points refers to those who were predicted the same number of points as they achieved. This variable is currently only available for 18 year-old applicants domiciled in England, Northern Ireland and Wales, any applicants outside of this cohort will be assigned the value ‘Not applicable’.

Number of Applications

Application is defined as a choice to a course in higher education through the UCAS main scheme. Each applicant can make up to five choices, which was reduced from six in 2008. The number of applications does not include choices made through the following acceptance routes: Clearing, Extra, Adjustment and RPAs.

Number of Unique Applicants

The number of applicants making at least one choice through the UCAS main scheme (which excludes Clearing, Extra, Adjustment and RPAs). Each cell gives the number of unique applicants. E.G. If split by subject group an applicant is counted once within each subject group of the course applied to.

Number of Acceptances

Acceptance is defined as an applicant who has been placed for entry into higher education. RPAs are included in the total. An RPA (record of prior acceptance is an application submitted to UCAS by an institution when an unconditional firm has already been offered and accepted by the applicant.

Acceptance Route

The acceptance route with the following values attached to placed applications. Firm Choice: where the applicant has been accepted to their first choice. Insurance choice: where the applicant has been accepted to their second choice. Main Scheme Clearing: where an applicant was unsuccessful in the main scheme (i.e. applied before 30 June) and subsequently found a place using Clearing. Direct Clearing: where the applicant has applied via Clearing without an initial application through the main scheme. Adjustment: where applicants who have met and exceeded the conditions of their firm choice choose to take up an alternative offer - introduced in 2009. Extra: where applicants who held no offers after using all of their main scheme choices, make additional choices. RPA: where an application is submitted to UCAS by an institution when an unconditional firm has already been offered and accepted by the applicant. Please note: ‘Insurance choice’ and ‘Firm choice’ values are based on the applicant’s response to an offer as-at June deadline. There are 10,000 to 20,000 acceptances to a main scheme choice each year where the applicant has not responded, or is awaiting an offer, by June deadline. These acceptances are classified as ‘Other Main Scheme Choice’.

Disclosure Controls

Disclosure controls have been applied to the data to reduce the risk of disclosing personal data about identifiable individuals. These controls include reporting each cell to the nearest five. In particular cell counts of 1 and 2 are reported as 0. Rows that only report 0 are omitted from the output. These controls are applied to each cell independently so this may result in instances where totals do not equal the sum of the components

UCAS 2016

NOTES

[1]The main dataset used in this analysis is aggregated to predicted* actual* gender* ethnicity* polar3* university type; see Section 2 for more details.

[2] Given that this measure of quality is based on universities' published tariff points, this measure of quality may be prone to error; it is well known that institutions accept students with grades below the advertised requirement (Times Higher Education, 2016).

[3]It is important to note that this measure of under-over prediction is rather blunt. The measure is based on the total points achieved from the applicants' best 3 A-levels. This means that each teacher in each subject would have to correctly predict the applicants' grade for the total points score to be accurately predicted. There is also room for error in the measure: a student could be predicted to achieve BBB=12 points, and actually achieve AAD=12 points, and be accurately predicted.

[4]Not shown, but results available on request

REFERENCES

Chowdry, H, Crawford, C, Dearden, L, Goodman, A and 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, 431–457.

Everett & Papageorgiou (2011), 'Investigating the Accuracy of Predicted A Level Grades as part of 2009 UCAS Admission Process'

Havergal, C (2015), 'Elite universities lower entry standards to expand, report reveals', Times Higher Education, online, (accessed October 2016)

Hoxby & Avery (2012), 'The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students', NBER working paper No. 18586, National Bureau of Economic Research

Lasselle, L, J McDougall-Bagnall, & Smith I, (2014), 'School grades, school context and university degree performance: evidence from an elite Scottish institution', Oxford Review of Education, Volume 40, 2014 – Issue 3

McGuigan, M, McNally, S, & Wyness, G (2016), 'Student Awareness of Costs and Benefits of Educational Decisions: Effects of an Information Campaign', Journal of Human Capital (forthcoming)

Smith, J, Pender, M, & Howell, J, (2013). 'The Full Extent Of Student-College Academic Undermatch'. Economics of Education Review 32: 247-261.

UCU (2015), 'Undergraduate application and admission survey', University and College Union

Wilson, J, (2015), 'Students should apply to university once they've received their grades, not before', The Guardian Higher Education Network, blog post, (accessed September 2016)

Produced by University and College Union
Carlow Street, London NW1 7LH
T: 020 7756 2500 E: mwaddup@ucu.org.uk W: www.ucu.org.uk December 2016