ETF. (2015). *Making maths and English work for all*. Retrieved from FE Week website: https://feweek.co.uk/wp-content/uploads/2015/03/Report.pdf

FFT Education Datalab (2018) *GCSE Results 2018: Four Things to Look out for on Thursday.* https://ffteducationdatalab.org.uk/2018/08/gcse-results-day-2018-four-things-to-look-out-for-on-thursday/

Hayward, G., & Homer, M. (2015). *Profile of the Mathematics Teaching Workforce in the College Sector in England*. Retrieved from Gatsby Charitable Foundation website: http://www.gatsby.org.uk/uploads/education/reports/pdf/profile-of-the-maths-teaching-workforce.pdf

Higton, J., Archer, R., Dalby, D., Robinson, S., Birkin, G., Stutz, A., . . . Duckworth, V. (2017). *Effective practice in the delivery and teaching of English and Mathematics to 16–18 Year Olds*. Retrieved from University College London, Institute of Education website: http://dera.ioe.ac.uk/30578/1/English_and_Mathematics_to_16-18_year_olds.pdf

Impetus–The Private Equity Foundation. (2016). *The road most travelled? The 16–19 journey through education and training*. Retrieved from https://impetus-pef.org.uk/assets/publications/Report/Impetus-PEF-Position-Paper-Final-draft-Web-individual-pages.pdf

Jerrim, J., Greany, T., & Perera, N. (2018). *Educational disadvantage: How does England compare?* Retrieved from Education Policy Institute website: https://epi.org.uk/wp-content/uploads/2018/04/Educational-Disadvantage-England-EPI-IOE.pdf

Maughan, S., Smith, J., Mitchell, T., Horrocks, N., & Taylor, A. (2016). *Improving Level 2 English and Maths Outcomes for 16 to 18 Year Olds: Literature Review*. Retrieved from Education Endownment Foundation website: https://educationendowmentfoundation.org.uk/public/files/Presentations/Publications/16-18_Literature_Review.pdf

MEI. (n.d.). *GCSE Mathematics retake for vocational students: GCSE teaching from 2015*. Retrieved from Mathematics Education Innovation website: http://mei.org.uk/files/pdf/GCSE-vocational-students.pdf

Moss, G., Duncan, S., Harmey, S., & Muñoz-Chereau, B. (2018). *Current Practice in Using a System of Phonics with Post-16 Learners*. Retrieved from Education & Training Foundation website: https://www.et-foundation.co.uk/wp-content/uploads/2018/06/ETF18002_PHONICS_REPORT_80PP_FINAL_AW.pdf

Nye, Philip (2018, August 20). GCSE results 2018: Four things to look out for on Thursday [Web log post]. Retrieved from https://ffteducationdatalab.org.uk/2018/08/gcse-results-day-2018-four-things-to-look-out-for-on-thursday/

Ofqual. (2015). *Improving Functional Skills Qualifications.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/398441/2015-01-27-improving-functional-skills-qualifications.pdf

Ofqual. (2017). *Grading new GCSEs*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/719124/Grading_new_GCSEs25.6.2018.pdf

Ofqual. (2018). *Guide to GCSE results for England, 2018*. Retrieved from https://www.gov.uk/government/news/guide-to-gcse-results-for-england-2018

Porter, N. (2015). *Crossing the Line: Improving success rates among students retaking English and maths GCSEs*. Retrieved from Policy Exchange website: https://policyexchange.org.uk/wp-content/uploads/2016/09/crossing-the-line.pdf

Robey, C., & Jones, E. (2015). *Engaging Learners in GCSE Maths and English.* Retrieved from Learning and Work Institute website: https://www.learningandwork.org.uk/wp-content/uploads/2017/01/Engaging-learners-in-GCSE-maths-and-English.pdf

Robey, C., Woodhouse, J., & Downes, S. (2016). *Functional Skills Reform Consultation: Findings of the learner focus groups*. Retrieved from Education & Training Foundation website: https://www.et-foundation.co.uk/wp-content/uploads/2016/12/Learner-focus-groups-report-FINAL.pdf

Searle, J., & Barmby, P. (2012). *Evaluation Report on the Realistic Mathematics Education Pilot Project at Manchester Metropolitan University*. Retrieved from Mathematics Education Innovation website: https://mei.org.uk/files/pdf/RME_Evaluation_final_report.pdf

Sezen, C. (August, 2018). 'GCSE resit data shows we need a policy rethink', *TES*. Retrieved from https://www.tes.com/news/gcse-resit-data-shows-we-need-policy-rethink

Skills Commission. (2016). *Spotlight On... young people with below average academic attainment and the skills sector*. Retrieved from Policy Connect website: https://www.policyconnect.org.uk/research/spotlight-young-people-below-average-academic-attainment-and-skills-sector

Vidal Rodeiro, C. (2018). Which students benefit from retaking Mathematics and English GCSEs post-16? *Research Matters: A Cambridge Assessment Publication, 25*, 20–28.

Williams, J., Hadjivassiliou, K., Marvell, R., Green, M., & Newton, B. (2017). *Effective curriculum practice at below Level 2 for 16/17 year olds*. Retrieved from Department for Education: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/628908/Effective_provision_for_low_attainers.pdf

Wolf, A. (2011). *Review of Vocational Education – the Wolf Report*. Retrieved from Department for Business, Innovation & Skills and Department for Education website: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/180504/DFE-00031-2011.pdf

Methods used by teachers to predict final A Level grades for their students

Tim Gill Research Division

Introduction

Prior to 2015, there was a requirement for teachers in centres in England to submit an estimated grade to the awarding organisation (AO) for all students undertaking Advanced (A) Level qualifications. This information was used as part of the evidence base for grading and for reviews of marking (Cambridge Assessment, 2013). Estimated grades are no longer collected by the AOs, but they still serve a number of purposes. Firstly, teachers are required to provide them as part of the university application process[1]. Secondly, estimated grades may be produced at several different points during an A Level course to monitor student progress, or serve as a motivational tool (Martinez, 2001). Finally, they may be used within the centre for teacher accountability purposes.

1. University admissions tutors use them to assess students' potential so that they can decide whether to make an "offer" of grades that the student needs to achieve to secure entry onto a course.

However, there is evidence that predicting A Level grades accurately is a task that teachers find difficult (see Gill & Rushton, 2011; Gill & Chang, 2013; Gill & Benton, 2015). This lack of accuracy may impact negatively on teachers' perceptions of the quality of marking for a qualification. The purpose of the research presented in this article was to understand more about how teachers go about the process of making grade predictions for their students, in order to help them make more accurate predictions. If teachers are able to make more accurate predictions, then this may increase their confidence in the reliability of marking.

This research was a replication of a previous study undertaken by Cambridge Assessment (Child & Wilson, 2015) which used a survey and interviews to investigate how teachers of A Level qualifications made predicted grades for their students. A further aim of that study was to calculate the accuracy of those predictions, using data collected from the survey. Since the original work was undertaken, there have been some significant reforms to A Level qualifications (see Ofqual, 2016) which are likely to have had an impact on how grades are estimated and on the accuracy of these predictions.

Context: Reforms to A Levels

One of the most important changes brought about by the reforms relates to the connection between A Levels and another qualification: Advanced Subsidiary (AS) Levels. Prior to the reforms, an AS Level counted as half of an A Level (in the same subject), and was assessed at the end of the first year of a two-year A Level course. It was also available as a stand-alone qualification for students who did not want to go on to take a full A Level. The reforms led to AS Levels being "decoupled" from the A Level, meaning that currently they do not count towards an A Level. Centres, therefore, have the choice of no longer offering the AS Level for their A Level students, or getting students to take the AS Level in Year 12 and then the A Level in Year 13 (which means they will be reassessed on some of the same content as the AS Level). The structure of the new qualifications is such that co-teaching of the AS and A Level is (theoretically) possible, so that students in the same Year 12 class in a subject can be planning to take the AS Level only, the A Level only, or both qualifications. However, Vitello and Williamson (2018) found that only just over half of the heads of department they surveyed thought that AS Levels were *actually* co-teachable. There is also evidence of significant falls in both uptake and provision of reformed AS Level subjects after they had been decoupled (Vitello & Williamson, 2018).

The reforms are important in the context of making predictions of A Level grades because it means that some centres no longer teach AS Levels in the subject, and are therefore not able to use that information to help them make predictions. According to the previous study (Child & Wilson, 2015), the AS Level grade was an important source of information for estimating A Level grades. Some centres still offer AS Levels to their A Level students, which may be partly because AS Level exams are useful practice for taking A Level exams. Even so, the decoupling may have had an impact on the way in which results of AS Levels are used to help make predictions.

Previous research

There has been little previous research which looked at how teachers go about the process of making grade predictions in centres in England, apart from the original study on which the present research was based (Child & Wilson, 2015). The analysis of responses to that questionnaire found that teachers tended to combine data from several different sources to make their predictions. AS Level grades were used by 94 per cent of the respondents and were generally thought to be the best predictor of A Level grades. Other commonly used sources of information included observations of the quality of work or of student commitment and performance in coursework and mock exams.

There is some previous research investigating the *accuracy* of grade predictions in England. Several reports from Cambridge Assessment (Gill & Rushton, 2011; Gill & Chang, 2013; Gill & Benton, 2015) compared the A Level forecast grade which was submitted to the AO with the final grade achieved. The percentage of accurate predictions varied from 55% in the 2011 report to 43% in the 2015 report. Inaccurate predictions were much more likely to be optimistic (varying from 33% in 2011 to 43% in 2015) than pessimistic (12% in 2011 and 14% in 2015). Similar results were found in research undertaken by Universities and Colleges Admissions Service (UCAS, 2013) on behalf of the Department for Business, Innovation and Skills, which compared the predicted grades sent to UCAS as part of the university application process with the final grade. They found that grades were accurately predicted 42% of the time, with 48% of predictions being optimistic. In the study by Child and Wilson (2015) predictions were more likely to be optimistic than pessimistic, and some respondents revealed that this was deliberate, to provide motivation for students.

Hopkin (2011) found that just using the AS Level grade to predict the A Level grade produced more accurate results than the teacher predictions for the AO (see Gill & Chang, 2013; Gill & Benton, 2015). However, AS Level grades were still only accurate around 55 per cent of the time. There was no tendency (in contrast to the teacher predictions) for inaccurate predictions to be more optimistic than pessimistic. However, the reports are not entirely comparable with one another because of the use of different datasets (teacher predictions were for all students taking A Levels from a specific AO, whereas the AS Level predictions were based on data from all AOs, but restricted to students taking at least three A Levels).

Wyness (2016) investigated the accuracy of predicted grades for university applications. She found that only 16 per cent of applicants were predicted the same points score from their best three A Level grades as they actually achieved. Almost all of the remaining applicants (75 per cent) were over-predicted (i.e., achieved lower grades than predicted). She also found that lower ability applicants were more likely than higher ability applicants to be over-predicted. One possible explanation for this tendency to over-predict (particularly for lower ability students) is that teachers are using the predicted grade as an aspirational target for students to aim for, so that for students who they feel are on the borderline of two grades the teacher will tend to choose the higher grade.

The most up-to-date data on the accuracy of predictions comes from UCAS (2017). This included a comparison of the accuracy of predictions sent to UCAS for reformed subjects (first tranche only) and for non-reformed subjects. Throughout the period investigated (2012–2017) the accuracy was worse for the reformed subjects (including during the

pre-reform period). Post-reforms (i.e., 2017 only) the gap between reformed and non-reformed subjects was slightly wider, which might suggest that predicting A Level results is harder post-reform. However, the difference between the pre-reform and post-reform gap was only very small (exact figures were not available).

The main aim of the present study was to gather up-to-date information on grade predictions made in post-reform A Level subjects, in particular, the methods used by teachers to make grade predictions and the accuracy of the predictions. It was also hoped that more people would respond to the questionnaire than in the previous study, which would allow for more robust conclusions to be drawn. The subjects investigated in the current research were part of the first tranche of A Levels to be reformed, with first results in 2017. This meant that the centres contacted as part of the research were all teaching qualifications where the AS Levels had been decoupled from A Levels.

Methods

The methods for this research replicated those of the previous investigation by Child and Wilson (2015), by using a questionnaire sent to a large number of centres, followed by more in-depth interviews with teachers.

A Level subjects

We selected three A Level subjects offered by the Oxford, Cambridge and RSA (OCR) awarding organisation for this research. Two of these were the subjects that were used in the previous research (Chemistry and English Literature), so that direct comparisons could be made between these subjects pre- and post-reform. A third subject, Psychology, was also included because this is a very popular A Level which differs from the other subjects in that it does not include any non-exam assessment. This may have an impact on the way in which predictions were made.

Questionnaire

The first part of the data collection consisted of a questionnaire, to be filled in by centres offering OCR A Levels in either Chemistry, English Literature or Psychology.

Questionnaire design

The questions and structure of the survey were very similar to that used in the previous research project (Child & Wilson, 2015). There were two main sections:

- *Estimated grades for your students:* We asked participants to give their grade predictions for all their students who were completing their A Levels in 2018. They were also asked to give a ranking of where they believed students would reside within each grade (e.g., first, second).
- *How you decide estimated grades for your students:* We asked participants to say how important different sources of information were (from a list of options) and whether they asked anyone for advice in making their estimated grading decisions. We then asked them to describe in as much detail as possible their procedure for making estimated grading decisions. Finally, they were asked if there was any other information or support (not currently available) that would be useful.

Table 1: Number of centres invited, by subject

Subject	*Centres*
Chemistry	1,186
English Literature	508
Psychology	210
All	**1,904**

Participants

We recruited participants from several different lists of contacts provided by OCR. Where possible, we contacted the subject teacher directly. However, this was not always possible and for the majority of centres we used a general email address instead, with a request included to forward the email to a relevant teacher. The total number of centres contacted in each subject is shown in Table 1.

In May 2018, we sent an email to each centre to invite them to take part in the questionnaire. We advised participants that in order to complete the questionnaire they would need to know the estimated grades of their A Level students and that the questionnaire should take around 15 minutes to complete. We provided a link to the online questionnaire.

There were 54 respondents who completed the whole questionnaire, 38 for Chemistry, 8 for English Literature and 8 for Psychology. This was quite a low overall response rate (2.8 per cent). However, it is worth noting that a much larger number started it (went as far as to put in their name and their centre number), but did not finish. It seems likely that these people at least intended to complete the questionnaire. It was not clear if their decision not to continue was because they were concerned about revealing the candidate numbers of their students or because of the length of the task of putting in the predicted grades for all their students.

Interview

We asked questionnaire respondents if they were interested in taking part in a follow-up interview, after A Level results had been issued (in August 2018). The interview schedule had two sections:

- *How you decide and use predicted grades:* We asked interviewees about the relations between the different sources of information they said that they used in deciding predicted grades. This section included a specific question about the impact of the decoupling of AS Levels.
- *Questions on specific candidates:* For each interview, we identified three students who were of particular interest: one who achieved the same grade as their predicted grade; one who performed above predictions; and one who performed below predictions. We also asked if they had made any requests for a review of marking for any of their students.

Each interview took between 15 and 30 minutes to complete and was conducted by telephone. Each interviewee received a £20 book token as a thank you. In total, 45 of the 54 survey participants volunteered to take part in the follow-up interview, 32 for Chemistry, 7 for English and 6 for Psychology. Of these, we selected two for interview in each subject (six in total).

Results

Quantitative analysis of questionnaire responses

Comparison between predicted and actual grades

We asked respondents to list all students predicted to get each grade in turn (from A* to U), using their candidate number. This was so that after results had been issued we could compare this with the final grade for each student. We also asked respondents to rank students within each predicted grade; this enabled a complete ranking of all students in a centre to be generated based on their grade and their predicted position. We then compared this with the actual ranking (generated from total A Level marks).

There were a few issues with the recording of the predicted grades and ranking of students by respondents, perhaps because they did not understand the task they were being asked to do. Where the grades or rankings were duplicated or unclear, it was necessary to exclude the data. The remaining data was then merged with the final grade for each student. After doing this, there were 741 grade predictions, from 48 centres, with the breakdown shown in Table 2.

Table 2: Subject breakdown of matched grades

Subject	*Centres*	*Predicted grades matched to final grade*
Chemistry	33	524
English Literature	8	125
Psychology	7	92
Total	**48**	**741**

Table 3 presents a cross-tabulation of predicted and actual grades (all subjects together). This shows that for each predicted grade (apart from a grade U, which was only predicted for six students), the most common actual grade was the same grade (between 40 per cent and 50 per cent). For predicted grades A* to B the next most common actual grade was one grade below (i.e., over-prediction by one grade). However, for grades C to E the next most common actual grade was one grade above (i.e., under-prediction by one grade).

Table 3: Predicted vs actual grades (all subjects)

Predicted grade	*Students*	*Actual grade*						
		*A**	*A*	*B*	*C*	*D*	*E*	*U*
A*	105	**49.5**	41.0	9.5				
A	182	13.2	**42.9**	35.2	7.1	1.6		
B	205	1.0	20.5	**45.4**	26.3	5.9	1.0	
C	138	0.7	2.2	23.2	**40.6**	21.7	9.4	2.2
D	68			5.9	25.0	**47.1**	20.6	1.5
E	37		2.7	2.7	10.8	29.7	**45.9**	8.1
U	6					16.7	66.7	**16.7**

Figure 1 presents the accuracy of the predicted grades in each subject. It shows the percentage with exact agreement between predicted and actual grade, and the percentages which were optimistic or pessimistic and by how many grades. The numbers at the top of the bars show the absolute numbers of predictions in each category. This shows that for Chemistry and English Literature almost 50% of the predictions were accurate, but for Psychology only just over 30% were accurate. For Psychology around 45% were one grade optimistic, compared with about 25% for Chemistry and English Literature.

Tables 4 and 5 compare the accuracy of predictions from the current research with those from the previous analysis (Child & Wilson, 2015), for Chemistry and English Literature. This shows that in both subjects the levels of accuracy were much lower in the present study. In Chemistry, grade predictions in the present study were more likely to be either optimistic or pessimistic than in the previous study. In English Literature, predictions in the present study were more likely to be pessimistic, and were less likely to be optimistic.

Figure 2 presents a comparison of the predicted and actual rankings for each subject (all centres combined).

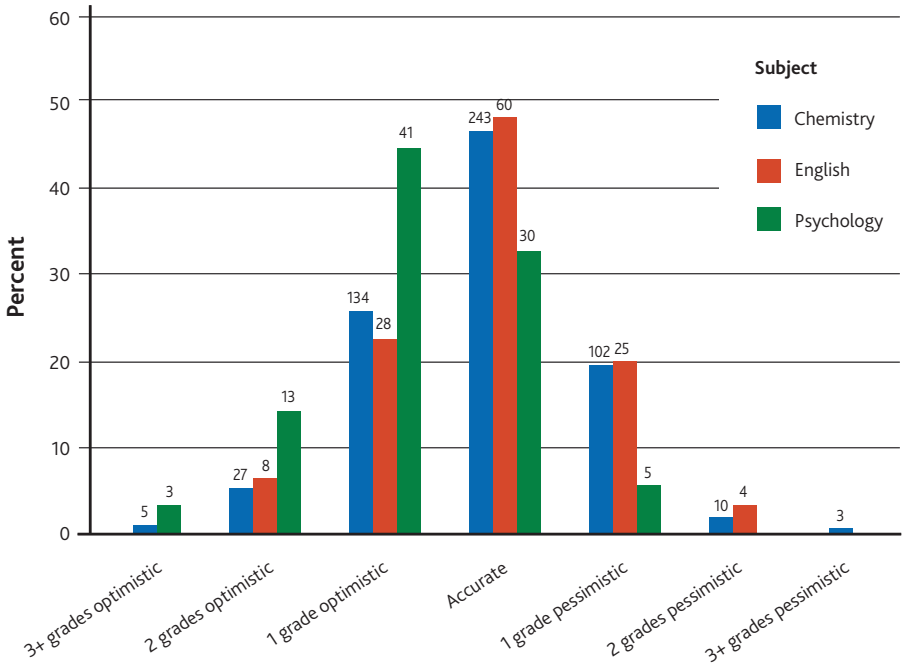

Figure 1: Breakdown of accuracy of predicted grade, by subject

Table 4: Comparison of predicted grade accuracy from current and previous study (Chemistry)

	N	*Accurate (%)*	*Optimistic (%)*	*Pessimistic (%)*
Previous study	106	54.7	26.4	18.9
Current study	524	46.4	31.8	21.4

Table 5: Comparison of predicted grade accuracy from current and previous study (English Literature)

	N	*Accurate (%)*	*Optimistic (%)*	*Pessimistic (%)*
Previous study	133	57.1	30.1	12.8
Current study	125	48.0	28.8	23.2

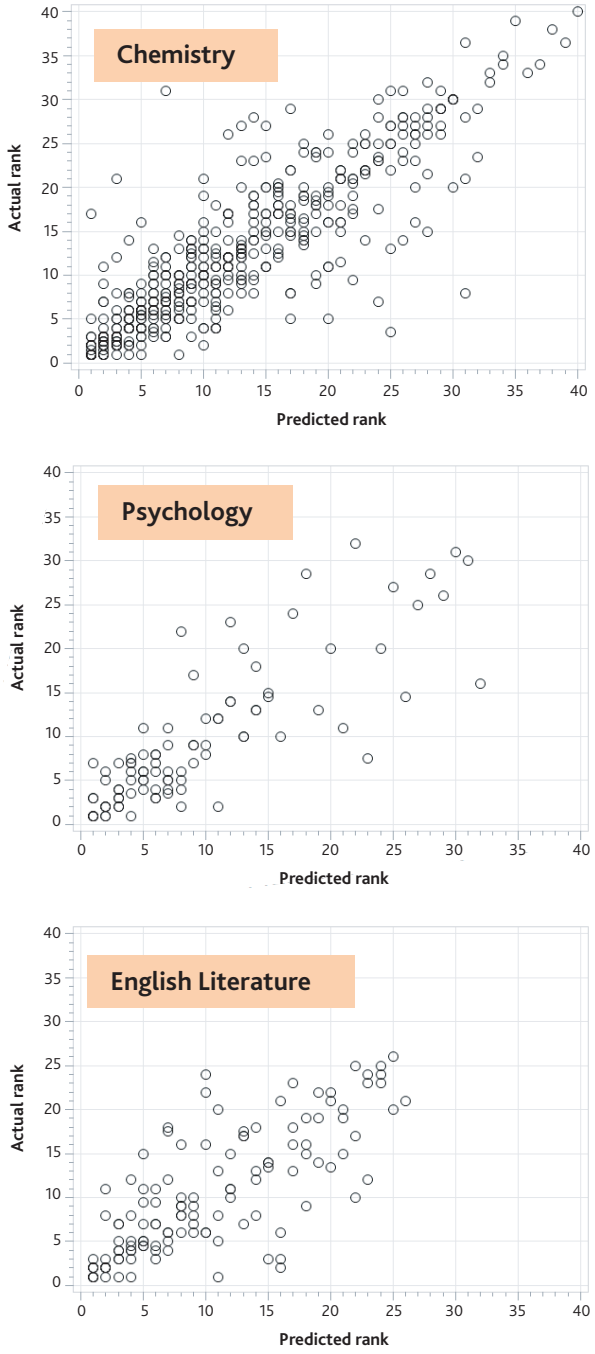

Figure 2: Predicted v actual rankings, by subject

The correlations were 0.87 for Chemistry, 0.76 for English Literature and 0.83 for Psychology. These results compare favourably with those reported in a meta-analysis into the accuracy of teachers' judgements (Machts, Kaiser, Schmidt, & Moller, 2016), which found average correlations of 0.61 between teachers' judgements of their students' cognitive abilities and their actual academic achievement. It is worth noting that the actual rankings were based on an overall mark for the qualification, and some students were very close in terms of marks. Therefore, it was not surprising that some teachers found it difficult to correctly rank these students.

How teachers decide the estimated grades for their students.

The first question in this section of the questionnaire asked how important different sources of information (as listed in the questionnaire) were in helping respondents make grade predictions. For each source of information they were given four possible options (*Very important; Somewhat important; Little importance;* and *Not used*).

There are a variety of different sources of information available to teachers to help them make estimates, which can be split into three categories, as outlined below:

1) *Statistical information:* A commonly used statistical method for tracking A Level students is the Advanced Level Information System (ALIS), which is provided by the Centre for Evaluation and Monitoring (CEM) at Durham University[2]. ALIS is an adaptive baseline test, which is usually taken at the start of Year 12. The results are used (alongside the results of General Certificate of Secondary Education (GCSE) exams, which are taken at the end of Year 11) to provide information on the level a student is working at, and a predicted AS or A Level grade (including an estimate of the probability of achieving each grade). Other similar packages are also available, including Active Learning Practice For Schools (ALPS[3]), the Cognitive Abilities Test (CAT)[4] and Fischer Family Trust[5].
2) *Assessment performance:* Performance in previous assessments is usually a strong predictor of performance in a later assessment. Therefore, teachers are likely to use the GCSE or AS Level grades (either in the same subject as the A Level, or as an overall average grade) to help them predict A Level performance for students. Other assessments within the course may also prove useful, such as coursework, practical endorsements and formative assessments.
3) *In-class judgements:* More qualitative factors may also be used by teachers, such as the perceived motivation of students, their interest in the subject and the day-to-day quality of their work.

Figure 3 presents the results of the responses to this question, which suggest that the most important factors overall were the students' performance in mock (practice) A Level exams and observations of their quality of work and commitment. Each of these factors had more than 80% of respondents saying that they were very important or somewhat important. Despite the decoupling of AS Levels from A Levels for these subjects, over 60% of respondents said that AS Level performance in the same subject was very important or somewhat important. Two other factors ("AS Level mock examinations" and "other formative

2. https://www.cem.org/alis
3. https://alps.education/about-us/
4. https://www.gl-assessment.co.uk/products/cognitive-abilities-test-cat4/
5. https://fft.org.uk/fft/target-setting/

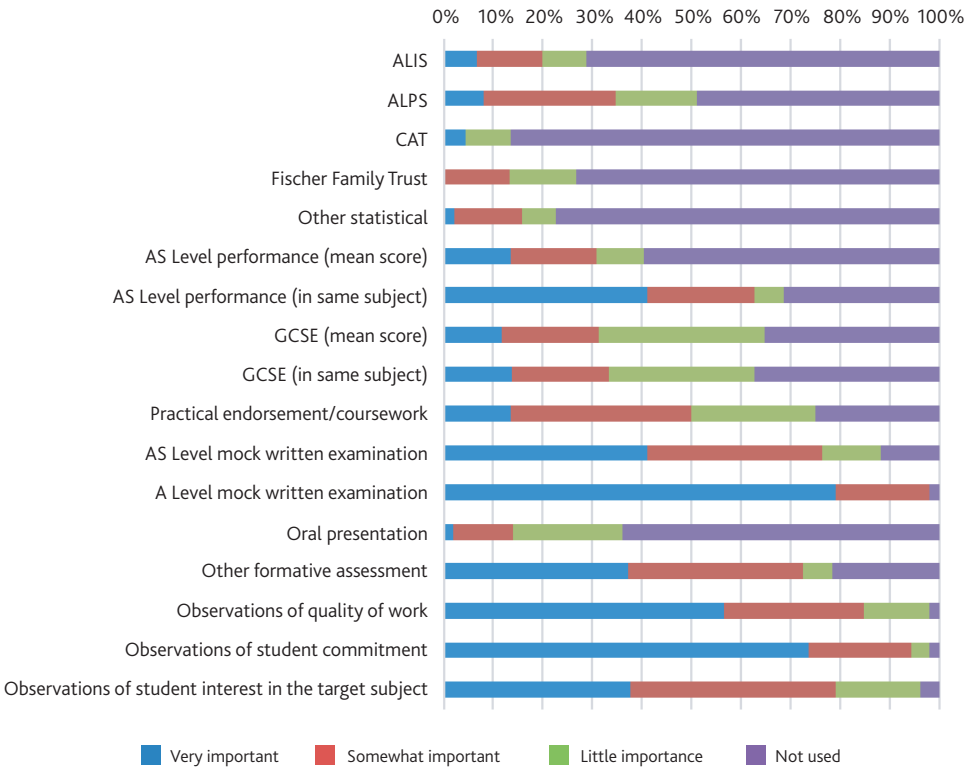

Figure 3: Importance of different sources of information in helping make grade predictions

assessments") were very or somewhat important for over 70% of respondents. Relatively few respondents said that they found statistical information (e.g., ALPS, ALIS) important. Another factor which was of little importance for most respondents was oral presentation. GCSE performance in the same subject or overall was only important to around a third of respondents.

Most respondents listed multiple sources of information in their answer to this question. The overall mean number of sources which were deemed to be very important was 4.3, and the mean number which was either very or somewhat important was 7.8. However, there was a significant variation in the number of different sources reported, varying from one to nine for very important and one to fifteen for very or somewhat important.

Respondents were asked to list any other sources of information that they used to help them estimate grades. There were 19 responses to this question. The most popular responses were around the use of end of topic/unit/chapter tests and the results of homework. Other responses included additional observations of students in the class, particularly in terms of their resilience/mental health and how well they responded to feedback from teachers about their assessments. Finally, one respondent mentioned that they used information on the progress of similar students from previous years.

The next question asked whether respondents asked for help in making predictions from other people or organisations. Again, they were given a list of possible sources of help, but were also able to add to this. Figure 4 presents the percentage of all respondents (n=54) saying they received advice from each of the different people. There were seven questionnaire respondents who did not give an answer to this question.

By far the most popular was other teachers in the same subject, which was selected by 74% of respondents. Head of department (30%) and teachers of a different subject (22%) were the next most popular selections. Very few respondents selected members of SLT, students, parents or the AO.

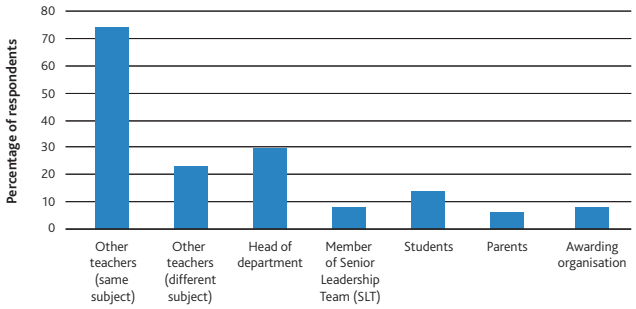

Figure 4: Sources of advice in making grade predictions

Qualitative analysis of questionnaire responses and interviews

There were a number of questions in the questionnaire and the interviews which required a more qualitative analysis. In the questionnaire, respondents were asked to explain in as much detail as possible how they go about the process of estimating the grade for a typical student. There were seven people who did not write anything for this question, leaving 47 responses in total. A number of related questions were then asked during the interviews to try and elicit further information about how predictions were made.

Sources of information for making predictions

Most respondents mentioned that they used several different sources of information, suggesting that it was a combination of factors which are taken into account. Two respondents explicitly noted that this was necessary because the individual sources of information were not reliable on their own. There were four main sources mentioned by respondents.

1) Results of internal tests/mocks

There was a fairly consistent message from respondents in terms of how they made their predictions. The most common sources of information mentioned were the results of end of topic or unit tests and the results of

mock exams. For some, this was then combined with information on the achievement of similar students in the previous year:

> *There are end of chapter assessments all the way through AS and A-Level. These, along with mock exams in January and past papers in April/May provide an overall picture of what level they are working at* (Chemistry, questionnaire).

> *In-class assessments, mock examinations and in-class performance are of the utmost importance. I use my years of experience with many examination classes to judge pupils' ability against past pupils and the grades that they managed to achieve* (Chemistry, questionnaire).

How well students responded to feedback following tests or mock exams was also important to some respondents:

> *...The main starting point is how students are doing on assessments, in particular, on weekly timed essays on each of the two main papers since the start of 2018; and then how much they are improving their scores on these week on week following teacher feedback* (English Literature, questionnaire).

Mock exam results were rated more important than in-class tests or homework by some respondents, because they are the closest the students get to a real exam.

> *We try and complete full exam papers, mark them using the mark scheme and then use the grade boundaries from the exam board to set our predicted grades… because the closer we are to how they are going to be examined at the end, the better the quality of our predictions.. if they've been very good at their homework, but…they've sat looking at their books and they've got help from their friends and they've gone on the internet… that's not reflective of how they're going to be tested when we get to the end of A Level* (Chemistry, interview).

In the interview, participants were asked about when during the course they undertook mock exams and end of topic tests, and also what the sources of these were (e.g., past papers, textbooks, online resources). All interviewees said that mocks were undertaken in Year 13, with dates varying from January to April, and these were usually the final opportunity to change the predicted grade. Mocks were also undertaken in Year 12, generally at the end of the year. However, two interviewees noted that these did not affect the final predicted grade. Most said that they used past papers for mocks, usually from the previous year because they are still secure. However, one issue with the Year 13 mocks was that having longer exams meant it was difficult for the teachers to find time for a full mock of all the exam papers. For one interviewee (Psychology), this meant "cobbling together" a mock from various previous papers, focusing on the long answer questions.

The sources of the end of topic tests included past papers, specimen papers, textbook exemplars and then various online resources such as OCR Exam Builder[6], *Kerboodle*[7] and *Doddle*[8]. Four of the interviewees said that they converted the results of these tests into grades for their students. There were two methods for doing this: either using the grade boundaries from previous years (converted to a percentage of total mark) or using a straight percentage conversion (90 per cent = A*, 80 per cent = A etc.).

6. https://exambuilder.ocr.org.uk/marketing/about-us/
7. https://global.oup.com/education/secondary/kerboodle/?region=uk
8. https://www.doddlelearn.co.uk/

2) Student characteristics

Most respondents to the questionnaire made it clear how important it was to consider the characteristics of the students themselves when making predictions. This can either be what was described as "commitment", "attitude" or "work ethic", or it could be in terms of students' ability to cope with the stress of exams. As these factors are not something that can easily be tested, this was often combined with the teacher's own professional judgement:

> *The estimated grade is based mainly on the AS grade achieved, with adjustments made on the professional judgement of the teachers of that student as to their commitment, ability and willingness to improve* (Chemistry, questionnaire).

> *…My grade predictions are informed by my assessment of students' resilience and capacity to cope with exam pressure. Often, with this exercise, it's a candidate's mental health which has dictated both the grade prediction and the position in the rank order* (English Literature, questionnaire).

3) Verbal discussions

Some questionnaire respondents also used verbal discussions to get an impression of students' true understanding of topics:

> *Frequent marking of homework allows progression to be tracked and verbal discussions with the pupil allow me to gauge true understanding* (Chemistry, questionnaire).

> *More and more in Year 13 the quality of verbal responses influences predictions as I can judge their ability to apply existing knowledge to new situations – which is after all what it all boils down to* (English Literature, questionnaire).

4) Statistical information

There was little mention in questionnaire responses of the use of statistical information, such as ALIS, in helping to make predictions. Where it was mentioned, it tended to be used as a starting point only, which could then be adjusted as students progress.

> *Obviously prior attainment is an indicator but, unless a student's work matches their ALPS predicted grade, I would never predict, say, an A simply because that's what ALPS says they should get or because they did well in English at GCSE.*

> *ALIS is a starting point for where a student should be. The main evidence has been in class summative tests using exam papers* (Chemistry, questionnaire).

In the interviews, both English Literature teachers mentioned that the statistical information could be misleading:

> *…That's the starting point, we look at the ALPS but then we look at the exceptions, because it may be that they're very good at English and not so good at other subjects and this might have a negative impact on what the ALPS target grade is* (English Literature, interview).

In contrast, another interviewee (for Psychology) revealed that in the future they were planning to rely only on the ALIS predicted grade, because it seemed to be more accurate than their predictions were.

Role of AS grades and the impact of decoupling.

As seen in Figure 3, over 60 per cent of respondents said that AS Level grades in the same subject were "somewhat" or "very" important in

helping to make grade predictions. However, in their responses to the open-ended questions, there were only a few references made to AS grades amongst Chemistry respondents, even by those who said it was an important factor. However, almost all of the Psychology respondents who said AS grade was important mentioned them when explaining how they make predicted grades. This suggests that, amongst those who answered this question, AS Level grade was more important for Psychology than for Chemistry. One Psychology respondent said that the removal of AS Level grades will make the predictions much harder.

> *This will become a much more difficult process when students are not sitting real AS examinations (from this Summer in my school) and in truth I have no idea at this point how I will choose predictions of my current Year 12s come September/October*
>
> (Psychology, questionnaire).

Interviewees were asked specifically about the effect of decoupling on making predictions. Only one of the centres still offered the AS Level in the subject, and this was only done on very rare occasions. Three interviewees said that the centre offered the AS Level for the first year or two following reform, but were no longer offering it, or were stopping this year.

Several interviewees agreed that they were less confident about their predictions following the decoupling, because in the pre-reform situation they would have had the results for 70 per cent of the course (AS exam plus coursework) when making their predictions. Additionally, one mentioned that the AS grade was used to inform predictions at the start of Year 13 because it was the "most real life exam they are going to sit." Any internal exam, however stringent, will not have the same high-stakes nature. However, one interviewee (English Literature) thought that decoupling was not going to make predictions harder, because some students coast during the first year, so their AS grade can be an underestimate anyway.

The final question in the questionnaire asked whether there was any further information or support that the respondents would find useful in making predictions. There were 12 responses to this question, which covered a range of different issues:

- Two respondents asked for more past papers but recognised that this was difficult with a new specification.
- Several mentions were made of A Level grade boundaries, and what each boundary should mean in terms of skills developed.
- Two teachers made reference to mark schemes (or guidelines on marking) and how they are applied.
- One respondent requested data on the performance on each question for students achieving different grades, to see "where each grade is typically gaining or losing marks."
- There was one request to provide more accurate sample scripts, with one per grade, per paper for each exam series. This respondent did not believe that the A*/A exemplars were written by students or written under exam conditions.

Reflections on results

Following the publication of results, the interviewees were asked to reflect on the performance of their students compared to the predictions. In terms of their overall perceptions, all but one of the interviewees thought that students had mostly performed to expectations. The one exception to this (Psychology) said that the students predicted to receive low grades tended to do better than expected and the students predicted an A* mainly achieved an A.

Interviewees were also asked about the performance of specific students and, if possible, to give reasons for why they performed as they did. For each interview, three students were identified: one who achieved the same grade as their predicted grade, one who performed above predictions and one who performed below predictions. For those whose final grade matched their predicted grade, there were only two reasons which were mentioned by the interviewees. Firstly, the students in question had performed consistently at that grade throughout the course, or throughout Year 13. Secondly, the statistical prediction (ALIS or ALPS) for those specific students had been correct.

For students whose predicted grade was higher than their actual grade, there were several reasons given by teachers for the inaccurate prediction. Some of these related to events occurring after predictions had been made. For example, personal circumstances, or the student prioritising Science subjects (over English Literature) because they wanted to do Physics at university. Other reasons related more to unexpected performance on the day, such as poor performance on one paper, and being only a few marks below the grade boundary. One interviewee talked more generally about under-performing students rather than describing a specific example. For him, there was a certain amount of looking back with hindsight:

> *I think our grade boundaries have been a bit too generous, so we've sort of got students predicted into a 'C' where actually they've ended up with a 'D' and if we looked back at what they did when we predicted them a 'C', it probably was a 'D'* (Chemistry, interview).

This interviewee also suggested another reason, which came from his perspective as the head of department:

> *Based solely on January mock exams we were fairly close to actual outcomes, but when you looked at the final predicted grade which had a bit of teacher input, so if you like, he's been doing well for the last three months, so he got a 'C' in the exam, but actually I think he might be a bit better than that, they tended to be a bit inflated*
>
> (Chemistry, interview).

It is worth noting that this interviewee was under the mistaken belief that they were still required to send in predicted grades to the AO and that these might be used in determining a final grade for a student who missed an exam. This led him to say that they were "more positive, rather than negative" in their predictions. Finally, there was one interviewee who could not explain why the prediction was wrong.

For students who achieved a higher grade than predicted, there were again a number of different reasons suggested. One interviewee for English Literature had not predicted a higher grade (A*) for the student because they felt their "simplistic" writing style would hold them back. In contrast, the other English teacher said that their student wrote very fluently, but was worried about a slight "superficiality" in her writing, which was "disguised by the quality of her prose." Another reason given was that the student in question worked incredibly hard towards the end of the course, motivated by a university offer. One of the interviewees for Psychology did not predict A* grades as a rule and seemed to think that the new A Levels meant that it was harder to get an A*. Finally, there were two interviewees who could not explain why their predictions were wrong.

The final question asked the interviewees whether the results for 2018 would affect their approach to making predictions in the future. Almost

all of them said that they would alter their approach to some degree. Two interviewees mentioned changing their approach for specific grades following under- or over-prediction at that level:

> *Where we've really had a problem this year…is the C/D and the D/E borderline, I think we've been too generous there…so we reviewed that for this year and we're gonna [sic] sort of concentrate on making sure [the grade boundaries] are higher* (Chemistry, interview).

> *We're going to…hold back on predicting A stars because we think that's where we got caught out the most this year, is the number of A stars we predicted and didn't get* (Psychology, interview).

One interviewee said that in the future they will rely on ALIS grades only to make predictions, as this year these were more accurate than the centre predictions. Other future changes in the approach to making predictions included more rigorously keeping data from exam questions, and reviewing the use of Kerboodle and using more past papers instead.

Discussion

Teachers' predictions of their students' A Level grades were, on the whole, fairly accurate. Across all three subjects, 44.9% of predictions were correct and 90.1% were within one grade. Where they were not correct, the predictions tended to be optimistic (35.0%) rather than pessimistic (20.1%). This pattern fits in with previous research (Child & Wilson, 2015; Gill & Benton, 2015). However, it is worth noting that the level of complete agreement was lower in this study than in the previous analysis undertaken by Child and Wilson (2015) three years ago, which found 54.7% of grades were correctly predicted in Chemistry and 57.1% in English Literature. The equivalent figures in the current study were 46.4% and 48% respectively. It is not known why the accuracy was lower, but it may be due to the decoupling of AS Levels, meaning that for many centres they did not have the AS Level result to help in the prediction. It may also be because teachers were still getting to know a relatively new specification.

Breaking the analysis down by subject, it was found that the accuracy of predictions was very similar in Chemistry and English Literature, but was considerably lower in Psychology, where only 32.6 per cent of predictions were correct. Over-predictions were also much more likely in Psychology than in the other two subjects. It is not clear why accuracy was so much lower in Psychology, but it may be related to the lack of non-exam assessment, which potentially made predictions more difficult.

In terms of the accuracy of the rankings data, there was a reasonably good association between the predicted and actual ranking of students, with correlations of between 0.76 and 0.87. This was despite the fact that, according to respondents to the previous survey (Child & Wilson, 2015), they did not usually rank students in making their predictions, so this process may have been a new experience for them. However, it may be that the respondents were, in fact, undertaking a ranking of sorts, even if they were not aware of it. According to Laming, "there is no absolute judgement. All judgements are comparisons of one thing with another" (Laming, 2004, p.9). In the context of making grade predictions, this might mean that teachers were comparing students in their current cohort with an internalised standard of what, for instance, an "A grade" student looks like, or with "A grade" students from previous years.

The findings from the questionnaire and interviews were quite consistent in terms of which factors were most important in helping teachers make their predictions. These were mock exams, in-class tests, and student characteristics such as attitude and how they cope with the stress of exams. Mock exams were seen as particularly important because with the removal of AS Levels in many centres, these were the closest experience that students had to a real exam. Examples were given of students who performed at a particular grade on in-class tests or homework but were given a lower predicted grade because of poorer performance in mocks. However, there was still concern from some respondents about the lack of high stakes for mock exams, which might mean that students did not always treat them as seriously as the real exams. Statistical predictions such as ALIS or ALPS were used as a starting point for predictions by some teachers, but always with the belief that the prediction could be over-written if the evidence of in-class work or mocks pointed to a different grade.

The interviewees revealed that mock exams and in-class tests were usually based on questions taken from past papers, or specimen papers. Requests were made for more of these to be made available and, by one interviewee, for an OCR specific source of questions. Therefore, it is important for OCR to continue to ensure the quality of these resources, in terms of alignment with content coverage and topics in the actual exam paper, and providing guidance for teachers to help them make reliable marking judgements. For example, the extra resources requested by respondents were mainly related to understanding how to assess students work better, either by improved understanding of the mark scheme or better understanding of what performance at each grade should look like.

It was not surprising that most of the interviewees said that the decoupling of AS Levels made them less confident about the accuracy of their predictions, particularly as all but one of them had stopped offering the AS Level in their subject. Prior to reforms, the AS Level was worth half of the A Level and was, as one interviewee put it "the most real life exam they are going to sit" prior to taking the A Level exam(s). It is worth noting that the predictions in both Chemistry and English Literature were less accurate than when the same analysis was undertaken prior to the decoupling (Child & Wilson, 2015). We cannot know for certain that the decoupling caused this fall in accuracy, but it was likely to have had a negative impact. It was unfortunate that none of the interviewees were in centres which continued to offer the AS Level, so that the effect of the decoupling in these centres could have been investigated, in terms of the way this information was used in making predictions.

There were many different reasons given for why students did not perform as expected. Some of these related to factors which influenced performance after predictions were made, and therefore there was no way that the teachers could have predicted these. This worked in both ways, with better than expected performance from a student who worked very hard at the end of the course, and worse than expected performance by students who did little work before the exam. Other reasons acknowledged the fact that students sometimes just perform unexpectedly (well or badly) on the day, and there were some instances where the teacher was at a loss to explain their students' performances. This highlights that grade predictions are inherently unreliable, because of the multitude of factors which are beyond the teacher's control. In considering why predictions were inaccurate, it is worth noting that there was no mention made of poor marking by OCR. Indeed,

one interviewee commented that changes to marks after a review of marking were rare because the marking was generally accurate in the first place. Even the respondent who felt that, overall, their students had not performed to expectations thought that this was an error on their part and talked about the need for them to improve predictions.

Most of the interviewees said that they would change their approach to making predictions in the future, following the results for their students in 2018. Some were making quite radical changes, such as relying on ALIS predictions only. Others were focused on changes to predictions at particular grades. We were encouraged to see that teachers were continually adjusting their approach in an effort to improve their predictions. However, we hope that these decisions were not knee-jerk reactions to inaccurate predictions. There are many reasons why student performance can be unexpected in any one year (e.g., marking reliability, exam coverage of curriculum, personal circumstances) and therefore, it would be wise for centres to look at the accuracy of their predictions over several years, rather than just the previous year. Furthermore, it may be that predictions were influenced by the mistaken assumption that results within a centre should be reasonably stable year-on-year. Previous research (Crawford & Benton, 2017) has shown that some centres can experience high levels of volatility in (GCSE) results between consecutive years. It may be that results in a centre which differ from predictions may be partly due to this natural variation in year-on-year results. A further area of research would be to look at whether predictions are less accurate when performance in a centre in a particular year is unexpected compared to previous years.

Finally, it is worth noting one limitation with the results presented in this article. Responses to the questionnaire were more likely to come from smaller centres. For each subject, the mean value of centre "size" (number of students finishing the A Level in 2018) was higher in the whole cohort (19.7 for Chemistry, 18.6 for English Literature and 25.8 for Psychology) than in the sample of respondents (16.0, 15.6 and 13.1 respectively). Furthermore, the maximum centre size amongst the sample data was only 40 for Chemistry (compared with 423 amongst all centres), 26 for English Literature (compared with 180) and 32 for Psychology (compared with 378).

The most likely reason for this pattern was because of the size of the task for a larger centre (listing predicted grades for all students). This issue was highlighted by an email from one teacher stating that they would not consider completing the questionnaire because they had 260 students sitting the exam[9]. This may have been an issue if, as one teacher commented, it was easier for smaller centres to make predictions because they know individual students better.

In hindsight, it probably would have been better to reorder the questionnaire so that the questions about how the respondents made their predictions came before the questions about the predicted grades for each student. As previously noted, there was a large number of people who started the filling in the questionnaire, but did not get past this section. Changing the order may have led to more responses to the questions about methods, particularly from larger centres. Another way to encourage responses from larger centres would have been to say that it was only necessary to enter the predicted grades for one or two classes, rather than for all A Level students.

In conclusion, the research reported in this article found that teachers were reasonably accurate in their predictions of final A Level grades for their students, and were more likely to be optimistic than pessimistic. Prediction is clearly a complex process, involving the weighing up of different factors, which has seemingly become more difficult since A Levels were reformed. Therefore, it is of added importance for AOs to provide the best available support to teachers.

9. The teacher was encouraged to complete the questionnaire anyway, referring to one or two classes only, not the whole cohort.

References

Cambridge Assessment. (2013). *A simple guide: Assuring OCR's marking accuracy*. Retrieved 24th August 2018 from http://www.cambridgeassessment.org.uk/Images/143058-a-simple-guide-to-marking-and-grading.pdf

Child, S., & Wilson, F. (2015). *An investigation of A level teachers' methods when estimating student grades*. Cambridge Assessment internal report. Cambridge, UK: Cambridge Assessment.

Crawford, C., & Benton, T. (2017). *Volatility happens: Understanding variation in schools' GCSE results*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Gill, T., & Benton, T. (2015). *The accuracy of forecast grades for OCR A levels in June 2014. Statistics Report Series No. 90*. Cambridge, UK: Cambridge Assessment.

Gill, T., & Chang, Y. (2013). *The accuracy of forecast grades for OCR A levels in June 2012. Statistics Report Series No. 64*. Cambridge, UK: Cambridge Assessment.

Gill, T., & Rushton, N. (2011). *The accuracy of forecast grades for OCR A levels. Statistics Report Series No. 26*. Cambridge, UK: Cambridge Assessment.

Hopkin, R. (2011). *Predicting A level grades using AS level grades. Statistics Report Series No. 29*. Cambridge, UK: Cambridge Assessment.

Laming, D. (2004). *Human Judgment: The Eye of the Beholder*. London, UK: Thomson.

Machts, N., Kaiser, J., Schmidt, F. T. C., & Moller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review, 19*, 85–103.

Martinez, P. (2001). *Great expectations: setting targets for students*. London, UK: Learning and Skills Development Agency.

Ofqual. (2016). *Your qualification, Our regulation: GCSE, AS and A level reforms in England*. Coventry, UK: Office of Qualifications and Examinations Regulation.

UCAS. (2013). *Investigating the accuracy of predicted A level grades as part of the 2010 UCAS admission process*. London, UK: Department for Business, Innovation and Skills.

UCAS. (2017). *End of cycle report 2017: Qualifications and competition*. Cheltenham, UK: UCAS.

Vitello, S., & Williamson, J. (2018). *AS level trends 2017*. Cambridge Assessment Research Report, Cambridge, UK: Cambridge Assessment.

Wyness, G. (2016). *Predicted grades: accuracy and impact: A report for University and College Union*. London, UK: University and College Union.