Cambridge Assessment

OCR
Oxford Cambridge and RSA

An investigation of A level teachers' methods when estimating student grades

Research Report

Simon Child

Frances Wilson

October 2015

Author contact details:

Simon Child & Frances Wilson

Cambridge Assessment
The Triangle Building
Shaftesbury Road
Cambridge
CB2 8EA
UK

Child.s@cambridgeassessment.org.uk

Frances.wilson@ocr.org.uk (if applicable)

http://www.cambridgeassessment.org.uk

As a department of Cambridge University, Cambridge Assessment is respected and trusted worldwide, managing three world-class examination boards, and maintaining the highest standards in educational assessment and learning. We are a not-for-profit organisation.

Cambridge Assessment is committed to making our documents accessible in accordance with the WCAG 2.1 Standard. We're always looking to improve the accessibility of our documents. If you find any problems or you think we're not meeting accessibility requirements, contact our team: Research Division

If you need this document in a different format contact us telling us your name, email address and requirements and we will respond within 15 working days.

How to cite this publication:

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

Contents

List of Tables and Figures 3
Executive Summary 4
Acknowledgements 7
1. Introduction 8
1.1. Methods for estimating grades at A level 8
1.2. Accuracy of grade estimates 9
1.3. Aims of the present study 10
2. Methods 11
2.1. Questionnaire 11
2.1.1. Participants 11
2.1.2. Questionnaire design 11
2.1.3. Piloting 12
2.2. Interview 12
3. Results 13
3.1. Questionnaire analysis 13
3.1.1. Sample information 13
3.1.2. Grade prediction information used 13
3.1.3. Grade predictions 15
3.1.4. Grade outcomes 17
3.1.5. Student ranking analysis 19
3.2. Interview and qualitative data 22
3.2.1. Unease of using statistical modelling alone 22
3.2.2. The status of AS level 23
3.2.3. Support for future predictions 24
3.2.4. Student characteristics for non-predicted results 25
4. Discussion 26
4.1. Future work 27
5. References 29
Appendix A: Interview schedule for teachers 30
Appendix B: Mean absolute grade and ranking differences by subject and outcome grade 36
Appendix C: Percentage of candidates by outcome and estimated grades, split by subject 37
Appendix D: Ranking correlations by teacher 38

List of Tables and Figures

Table 1: Accuracy of estimated grades for final grade outcomes 9
Table 2: Information used in deciding predicted grades by subject 14
Figure 1: Predicted grades by subject 15
Figure 2: Grade prediction counts by teacher 16
Figure 3: Outcome grades by subject 17
Table 3: Percentage of candidates by outcome and estimated grades 19
Figure 4: comparison of predicted grades to outcome grades by subject 18
Figure 5: Scatter plot of estimated rankings vs actual rankings 20
Table 4: Correlation coefficients for estimated vs actual student rankings 21
Table 5: Absolute grade and ranking difference by prediction source type 22

Executive Summary

Introduction

There are several uses for teachers' estimated grading outcomes for students. Teachers may provide estimated grades to monitor student progress, to act as motivation, or to provide guidance for students making university applications.

There are several data sources that teachers can utilise in deciding estimated grades for students. These are broadly split into three areas: statistical methods; previous assessment performance; and in-class judgements.

Statistical information – methods include the Advanced Level Information System (ALIS), ALPS, the Cognitive Abilities Test (CAT) and Fischer Family Trust.

Other statistical information that might be utilised includes the AS level or GCSE performance of individual students. This could be related to the overall performance on the suite of qualifications taken by each student, or specifically related to performance in the equivalent subject at GCSE or AS level.

Assessment performance - teachers may use information gathered throughout each course to develop their grade estimates. This may include information gathered through formative assessments as part of the learning process, such as coursework or oral presentations.

In-class judgements – in making estimated grade decisions, teachers may also consider student motivation, student interest in the subject (e.g. as demonstrated by out of class reading) or day-to-day quality of work.

To our knowledge, no previous research has attempted to investigate how teachers arrive at their estimated grading decisions. With the wealth of statistical and observational information available to them, teachers are likely to display considerable variability in how they synthesise this information. The present study aimed to:

(1) Explore the strategies teachers utilise to decide overall grade estimates at A level, and map this information onto their actual grade estimates for a cohort of students.
(2) Investigate how closely predicted grades matched with outcome data for candidates, and how teachers reconciled these differences.

Methods

To meet the research aims, a questionnaire was developed that gathered information on the real time predictions made for a future live session (June 2015) for two A level subjects (English literature and Chemistry). This was followed up by the collection of interview data on how centres reconcile the differences (if any) between the predicted grade and the outcome grade for students.

Overall, 16 participants completed the questionnaire, making a total of 239 overall grade predictions. Five of the participants that completed the questionnaire volunteered to be interviewed.

The interview schedule comprised the following structure:

- *How you decide and use predicted grades* – this asked participants to detail how the relations between the different sources of information they said that they used in deciding predicted grades
- *Questions on specific candidates* – for each interview, the team identified three candidates that were of particular interest. This may have been if the student performed above predictions, below predictions, or if there was a particular movement in their ranking. They were also asked if they intended to make any EAR requests
- *Predicted grades in the future* – this set of questions asked participants if changes to the structure of examinations, or the change in requirements from examination boards, has changed how they will use predicted grades in future sessions.

Results

The most commonly used statistical information in predicting the outcome grade was the AS level grade in the relevant subject. GCSE grade in the same subject was used by less than half of the teachers surveyed. ALIS was used by over half of the teachers in both subjects.

Popular assessment measures included coursework and mock examinations, both used by over 80% of teachers overall. Interestingly there were subject-related differences, with coursework marks being used by approximately half of Chemistry teachers but all of the English Literature teachers. This may reflect differences in the assessment models for the two subjects.

Overall the teachers tended to overestimate their grade predictions for their students. There was a comparable degree of inaccuracy between the two subjects. The overall mean absolute grade difference was slightly below .50 for both subjects. When this was split by outcome grade, grade B was found to have the largest average absolute grade difference for both English literature (mean = .56) and chemistry (mean = .69).

A correlational analysis showed that teachers were reasonably effective at ranking their students, although they appeared to be slightly more accurate at the top end of the ranking scale.

From the interview and questionnaire qualitative data, four themes were identified by the research team: unease of using statistical modelling; the status of AS level; support for future predictions; and student characteristics for non-predicted results.

Discussion

The participants stated that their predicted grades were often deliberately optimistic. This was to provide motivation for students and to reflect the reality that students are competing for university places.

The results from the correlational analysis suggest that teachers are able to make sound relative judgements about candidates' abilities even if their representation of how these judgements convert into a grade is influenced by factors related to student motivation and competition for university places. Interestingly, the teachers stated that they did not usually rank students in making their predictions.

The teachers regarded the AS level to be a crucial staging post in making sound predictions for their students. In the reformed suite of A level qualifications, it is possible for AS levels to be co-taught. Some of the teachers anticipated they would offer the reformed AS level as a means to track the progress of students. Otherwise they would be reliant on effective mock examinations and assessment materials to be provided by awarding bodies.

Future work could investigate whether teachers that co-teach the AS level with the A level in the reformed qualifications provide significantly more accurate predictions of overall A level performance, compared to those that do not. Future work could also explore whether students that take AS levels perform over and above their predictions, compared to those that did not take it. This research could potentially inform the strategy for engaging teachers with the reformed AS level qualifications.

Acknowledgements

We are grateful to OCR subject specialists Stephen Diston, Rachael Tomkins, Natasha Chowdhury, Daniele Gibney, Sophie Maloney, Tony Fahy, and Mary Seddon for their helpful assistance in recruiting teachers and providing feedback on early drafts of data collection materials. We are also grateful to Sylvia Green and Jo Ireland from ARD Research Division for their advice on research design and questionnaire development.

1. Introduction

Estimated grades (up until 2015) were collected approximately a month before the June series (and during the same month as other series) at the unit level and overall grade level for GCSEs and A levels (AQA, 2015). There are several uses for teachers' estimated grading outcomes for students. First, estimated grades were submitted to exam boards to form part of the evidence base for grading and possible appeals (Cambridge Assessment, 2015). However, from January 2015 one exam board (AQA) no longer required centres to submit estimated grades, and this approach was soon followed by the other UK examination boards. Secondly, estimated grades form part of the university application process via UCAS. Thirdly, estimated grades may be used to monitor student progress across the course, or as a motivational tool.

In the guidance documentation for one exam board (AQA, 2014) teachers were asked to make a "judgement of the grade a student is likely to gain in the exam" (p. 2). They are instructed that each grade should represent the student's attainment and not be based on unusual performance or effort. There are a variety of methods and sources of information that teachers can utilise when making estimated grading decisions about their students. The following section briefly outlines some of these methods below.

1.1. Methods for estimating grades at A level

There are several data sources that teachers can utilise in deciding estimated grades for students. These are broadly split into three areas: statistical methods; previous assessment performance; and in-class judgements.

Statistical information – a common statistical method is the Advanced Level Information System (ALIS), which was developed by the Centre for Evaluation and Monitoring (CEM, 2015). According to the CEM, this method is used by over a third of Sixth Form providers and makes predictions based on mean GCSE score and an adaptive baseline test. This system provides predictions for specific subjects related to national performance data. Other examples of statistical methods for estimating grades include ALPS, which sets predictions for a combination of A level grades rather than for specific subjects, the Cognitive Abilities Test (CAT) and Fischer Family Trust.

Other statistical information that might be utilised includes the AS level or GCSE performance of individual students. This could be related to the overall performance on the suite of qualifications taken by each student, or specifically related to performance in the equivalent subject at GCSE or AS level.

Assessment performance - it is unlikely that these statistical estimates are used in isolation. Teachers may use information gathered throughout each course to develop their grade estimates. This may include information gathered through formative assessments as part of the learning process, such as coursework or oral presentations.

In-class judgements – in making estimated grade decisions, teachers may also consider aspects related to individual students, such as motivation, student interest in the subject (e.g. as demonstrated by out of class reading) or day-to-day quality of work.

1.2. Accuracy of grade estimates

The range of methods that could be used suggests great potential variability amongst teachers in terms of how they estimate student A level grades. Thus far, little research has investigated how teachers reconcile this information when making grading decisions. However, there have been several studies that have investigated the accuracy of the estimated grading decisions made by teachers. A series of statistical reports produced by Cambridge Assessment (Table 1) have investigated the accuracy of teachers' predicted grades with actual grade outcomes (Gill & Rushton, 2011; Gill & Chang, 2013; Gill & Benton, 2015).

Table 1: Accuracy of estimated grades for final grade outcomes

Statistical report	% Accurate	% Optimistic	% Pessimistic	% One grade out
Gill & Rushton (2011)	54.65	33.13	12.22	93.04
Gill & Chang (2013)	48.29	38.70	13.00	91.89
Gill & Benton (2015)	43.14	43.17	13.69	88.10

Taken together, these studies suggest that teachers estimate the correct grade for their students approximately 50% of the time. When they estimate the incorrect grade, they typically overestimate student performance. This pattern of results was supported by research by the Department for Business, Innovation and Skills (2011) that investigated the accuracy of estimated grades submitted to UCAS. They found 51.7% accuracy overall, with 41.7% of grades being over-estimated and only 6.6% of grades being under-estimated. As the estimated grades submitted to UCAS are part of the application process for entry to university, it is a possibility that grade estimates were more based on aspirations for students rather than an appraisal of current (or future) attainment.

Interestingly, when the Cambridge Assessment studies are considered as a time series, there appears to be a general decrease in the accuracy of estimated grades for students. This may well relate to the reduction in the amount of information from individual units available to teachers since the removal of the January examination session (Gill & Benton, 2015). Teachers appear to increasingly overestimate students' grade outcomes at A level, whilst pessimistic estimated grading is reasonably steady. In the context of the current reform agenda, the reduction of summative information available to teachers may be a future hindrance to making accurate predicted grade decisions.

Furthermore, there were a number of trends observed across the three statistical reports: higher grades were easier to predict compared to lower grades; independent and grammar schools predicted grades more accurately compared to other school types; and comprehensive and FE colleges tended to over-estimate grades compared to other school types. Interestingly, Hopkin (2011) found that simply using the AS level grade was a slightly more accurate predictor of A2 performance compared to predictions made by teachers in Gill

and Chang (2013) and Gill and Benton (2015). Like these studies, the accuracy of predictions was increased higher up the grade scale.

1.3. Aims of the present study

Some of the perceptions related to marking quality rest, to an extent, on how estimated grades match up to student outcomes. Little research has attempted to investigate how teachers arrive at their estimated grading decisions. With the wealth of statistical and observational information available to them, teachers are likely to display considerable variability in how they synthesise this information.

The present study aimed to first explore the strategies teachers utilise to decide overall grade estimates at A level. We attempted to investigate this aim by giving A level teachers in two subjects (English literature and Chemistry) a questionnaire that was designed to allow teachers to (1) note down the methods they typically used, and (2) describe the interaction between these various methods.

The second aim of the research was to explore how predicted grades mapped onto outcome data for candidates, and how teachers reconciled these differences. To meet this aim, the questionnaire gathered information on the real time predictions made for a future live session (June 2015). This was followed up by the collection of interview data on how centres reconcile the differences (if any) between the predicted grade and the outcome grade for students.

2. Methods

2.1. Questionnaire

The first stage of data collection comprised a questionnaire that was sent to centres that offered either OCR A level English Literature or OCR A level Chemistry.

2.1.1. Participants

Participants were recruited via a snowball sampling method. Participants were initially invited to take part via three possible methods: direct email invites that were sent to a list of contacts provided by OCR colleagues, invites posted on social media via official OCR Twitter accounts; or by a general email invitation that was sent to the reception of each OCR centre. Furthermore, chemistry teachers were given paper-based questionnaire packs at several OCR-run events in June 2015. These packs contained the questionnaire, a self-paid envelope, the invitation letter, and a consent form.

As an incentive to participate, the invitation email/letter offered the opportunity for participants to enter a prize draw for £100 of Amazon vouchers. A further £20 was offered to participants if they agreed to consent to a follow-up interview at a later date.

At the end of each invitation, the teachers were encouraged to distribute the link to the questionnaire by forwarding on the invite to other teachers within their departments.

Overall, there were 16 (9 English Literature, seven Chemistry) responses to the questionnaire where some of the questions were answered. This resulted in 283 predicted overall grades across the two subjects.

2.1.2. Questionnaire design

There was a web version of the questionnaire, which could be accessed via an internet link, and a paper-based version. These questionnaires were identical in structure and in their response options.

There were four sections of the questionnaire. These sections had the titles and asked for the information described below:

- *About you* – asked for the participant's name, centre, teaching experience and subject.
- *Estimated grades for your students*- asked participants to give their grade estimates for students that were completing their A levels in 2015. They were also asked to note where students would reside within each grade (at the top, middle, and bottom of the grade).
- *How you decide estimated grades for your students* – participants were first asked to tick from list of options the information that they use to make estimated grading decisions. The list of options was developed in discussion with OCR colleagues and internet searches. Participants were also asked to describe in as much detail as possible their procedure for making estimated grading decisions.
- *Prize draw and follow-up* – participants were asked to write their email address so that the researchers could contact them for the follow-up phase of the research.

2.1.3. Piloting

A trial version of the complete questionnaire was given to two OCR subject specialists; one from each of the target subjects. The subject specialists were recent teachers, and so had an understanding of the potential clarity of the questionnaire from a teacher perspective. Any changes that were suggested at this stage were enacted before the final version of the questionnaire went live.

2.2. Interview

The questionnaire participants were asked to declare their interest in participating in a follow-up telephone interview. Overall five teachers offered to participate (four English literature teachers; one Chemistry teacher) and were informed by the research team that they would be contacted following A level results day.

The interview schedule comprised the following structure:

- *How you decide and use predicted grades* – this asked participants to detail how the relations between the different sources of information they said that they used in deciding predicted grades
- *Questions on specific candidates* – for each interview, the team identified three candidates that were of particular interest. This may have been if the student performed above predictions, below predictions, or if there was a particular movement in their ranking. They were also asked if they intended to make any EAR requests
- *Predicted grades in the future* – this set of questions asked participants if changes to the structure of examinations, or the change in requirements from examination boards, has changed how they will use predicted grades in future sessions.

The full version of the interview schedule can be found in Appendix A. Before the interview schedule was confirmed, a draft version was sent to one subject specialist in each of the English and Chemistry subject teams in OCR for their comments and suggestions.

Each interview took approximately 30 minutes to complete, and was conducted by telephone.

3. Results

3.1. Questionnaire analysis

3.1.1. Sample information

The 16 teachers had a mean of 19.5 years teaching experience, including teacher training.

For three of the chemistry teachers, no information was provided regarding their predicted student grades, although they provided information on the sources they use to make predictions. This left a sample of 239 grade predictions made by 14 teachers (*M* = 17.07 grade predictions per teacher).

3.1.2. Grade prediction information used

Table 2 shows the grade prediction information for all of the teachers that responded to the relevant questionnaire items. The most commonly used statistical information in predicting the outcome grade was the AS level grade in the relevant subject. This is perhaps unsurprising given the close links between overall grade and AS level grades. This appeared to usurp the importance of the GCSE grade in the same subject, which was used by less than half of the teachers surveyed. ALIS was used by over half of the teachers in both subjects.

Popular assessment measures included coursework and mock examinations, both used by over 80% of teachers overall. Interestingly there were subject-related differences, with coursework marks being used by approximately half of Chemistry teachers but all of the English Literature teachers. This may reflect differences in the assessment models for the two subjects.

Finally, the teachers surveyed used in-class judgements consistently and regularly across the two subjects.

Table 2: Information used in deciding predicted grades by subject

		Chemistry (*N*=7)		English Literature (*N*=9)		Combined (*N*=16)	
Prediction information		Sum	Proportion	Sum	Proportion	Sum	Proportion
Statistical	ALIS	4	.57	6	.67	10	.62
	ALPS	1	.14	2	.22	3	.19
	CAT	0	.00	0	.00	0	.00
	FFT	3	.43	1	.11	4	.25
	AS (mean)	2	.29	4	.44	6	.38
	AS (subject)	6	.86	9	1.00	15	.94
	GCSE (mean)	4	.57	2	.22	6	.38
	GCSE (subject)	2	.29	4	.44	6	.38
Assessment performance	Coursework	4	.57	9	1.00	13	.81
	Mock exam	6	.86	7	.78	13	.81
	Oral	1	.14	2	.22	3	.19
	Other formative	2	.29	6	.67	8	.50
In-class judgement	Obs of quality of work	7	1.00	9	1.00	16	1.00
	Obs of student commitment	6	.86	9	1.00	15	.94
	Obs of student interest	4	.57	8	.89	12	0.75

3.1.3. Grade predictions

Figure 1 shows the pattern of grade predictions for the two subjects. For Chemistry, there was a reasonably even distribution of the proportions of students predicted each grade. The most commonly predicted grade was an A, whilst an A* was predicted rarely. In English literature, the predicted percentage of A* was much higher, and the most commonly predicted grade was A. No E grades were predicted in English literature

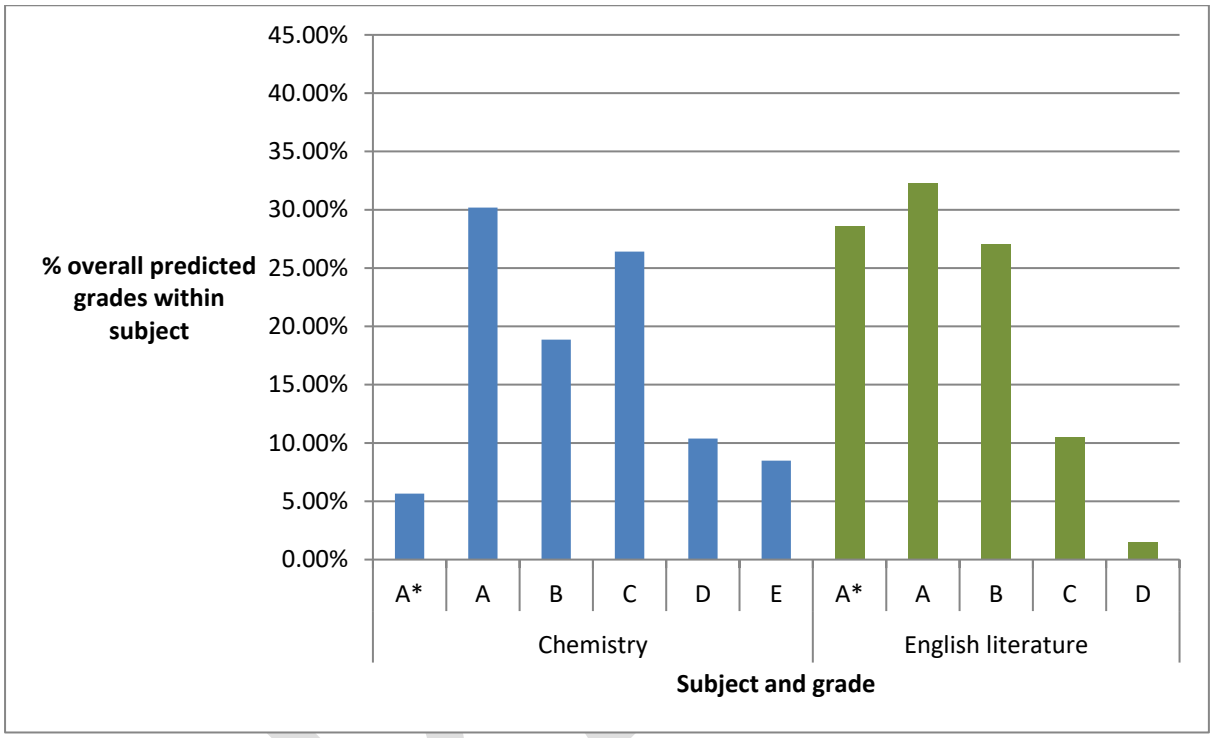


Figure 1: Predicted grades by subject

This difference in the distribution of predicted grades is shown in greater detail in Figure 3, which shows grade predictions by teacher. Teacher CC predicted the most C grades and this contributed to the distribution characteristics in Figure 2. The relatively high proportion of A* predictions in English Literature is due to the high number by teachers EG, EH and EJ relative to their predictions at other grades.

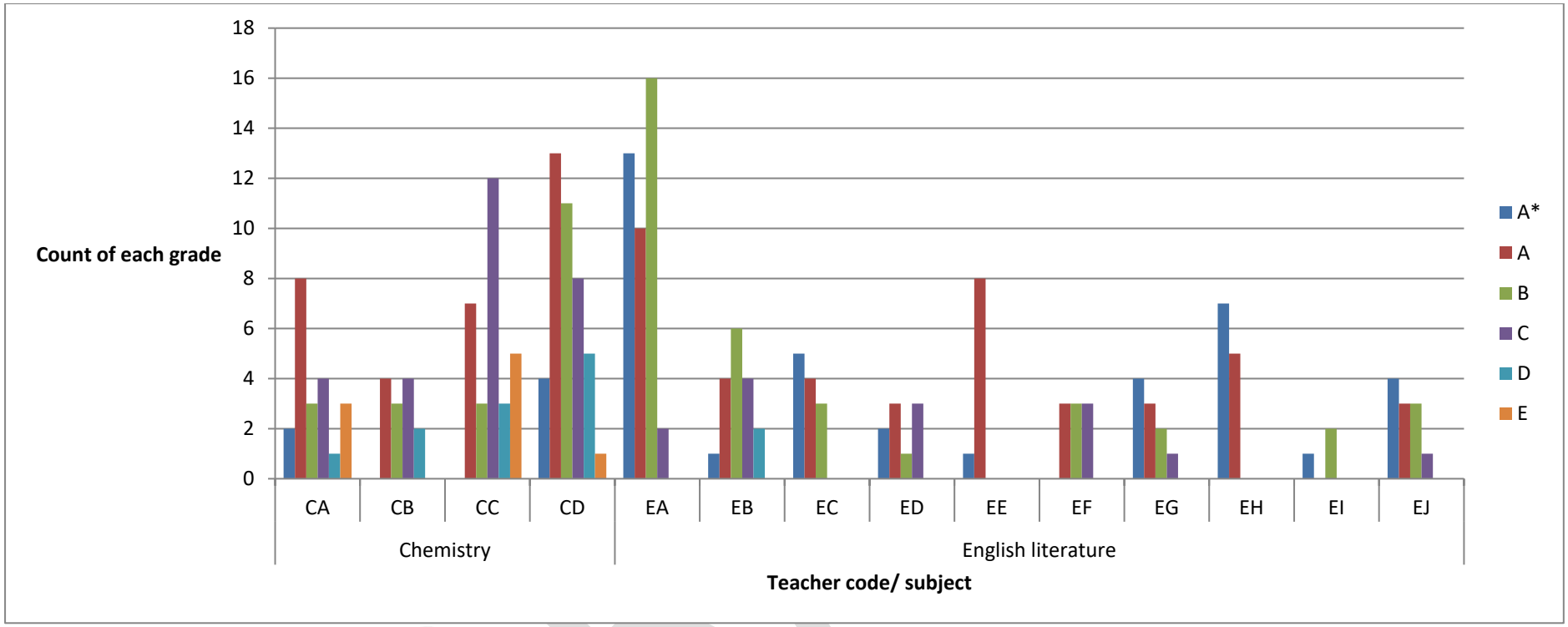


Figure 2: Grade prediction counts by teacher

3.1.4. Grade outcomes

Figure 3 shows the outcome grade distribution by subject. For both subjects, the percentage of A* and A grades was lower compared to the predicted grades (Figure 1). For chemistry, there was an even distribution of grades, with grade B the most common. Grade B was also the most common grade outcome for English literature, with over 40% of the student sample achieving this grade.

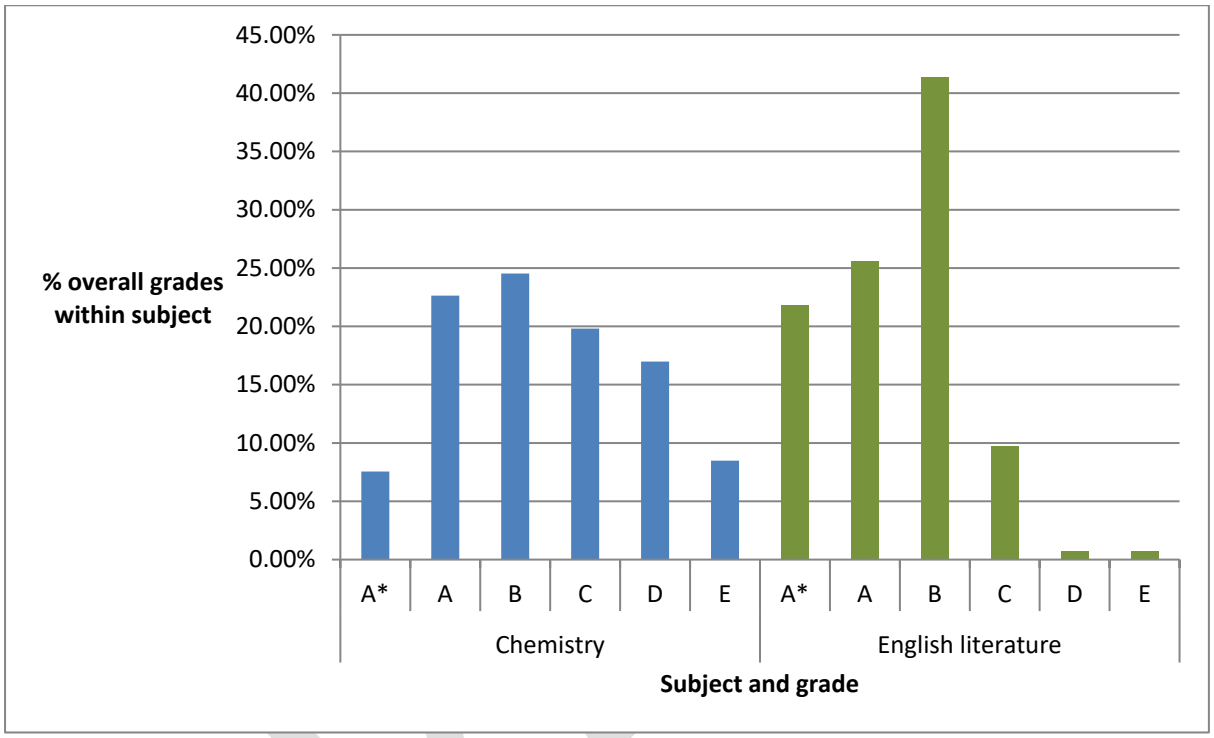


Figure 3: Outcome grades by subject

The data from Figures 1 and 3 suggests that overall the teachers tended to overestimate their grade predictions for their students. To investigate this further, for each student we calculated the difference between the teachers' estimated grades and the outcome grade (Figure 4). Figure 4 shows that teachers were accurate in their grade predictions over 50% of the time for both subjects. This is slightly higher that recent statistical analyses (Gill & Benton, 2015). Teachers in both subjects were overall more likely to be optimistic in their grade predictions rather than pessimistic.

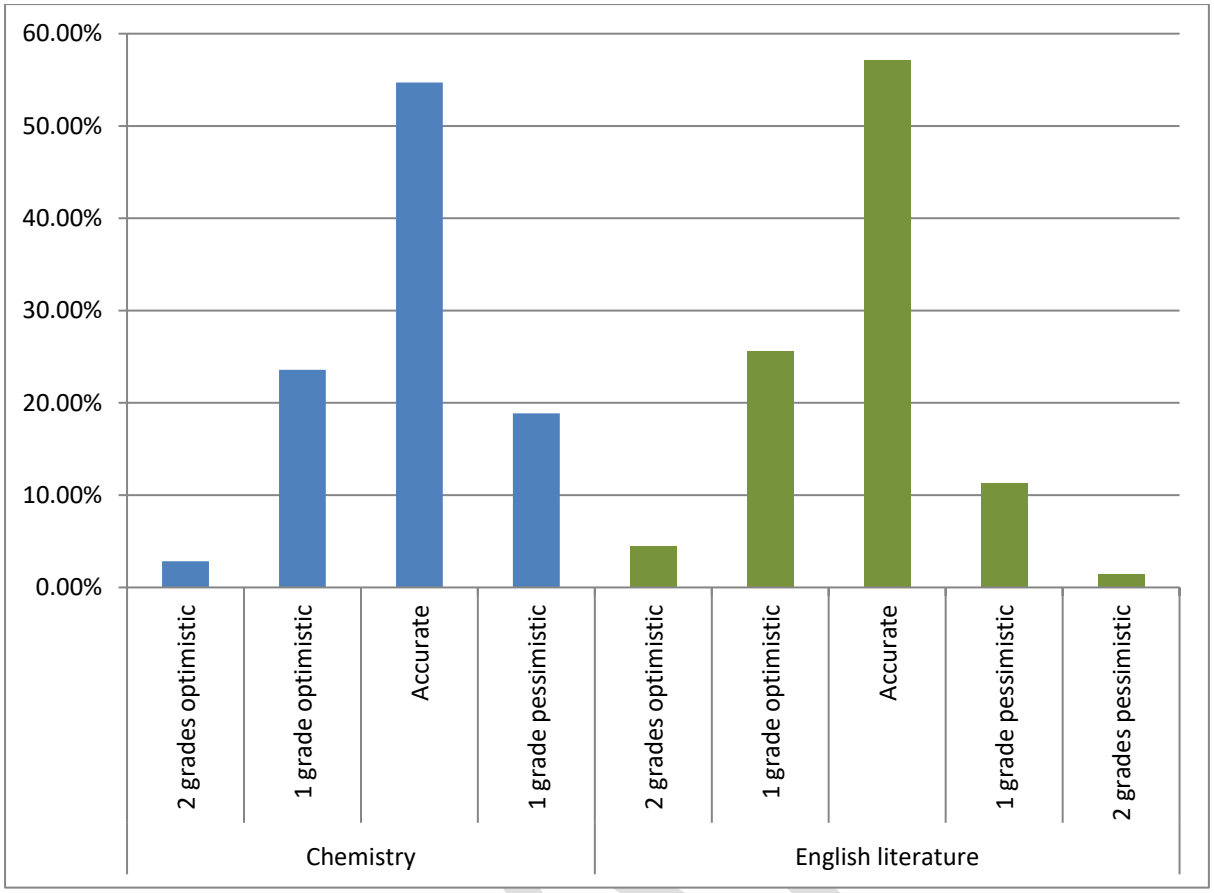


Figure 4: comparison of predicted grades to outcome grades by subject

When absolute differences in grade were compared (Appendix B), it was found that there was a comparable degree of inaccuracy between both subjects. The overall mean absolute grade difference was slightly below .50 for both subjects. When this was split by outcome grade, grade B was found to have the largest average absolute grade difference for both English literature (mean = .56) and chemistry (mean = .69). This finding is illustrated further in Table 3, which shows the percentage of candidates for each estimated grade and outcome grade combination, collapsed across both subjects (see Appendix C for this table split by subject). Over 10% of the candidates who were predicted an A grade achieved a grade B, whilst nearly 5% of candidates who were predicted a grade C achieved a grade B.

Table 3: Percentage of candidates by outcome and estimated grades

	Outcome grade						
Estimated grade	**A***	**A**	**B**	**C**	**D**	**E**	**Total**
A*	9.62%	6.28%	2.51%	0.00%	0.00%	0.00%	18.41%
A	5.02%	15.06%	10.88%	0.42%	0.00%	0.00%	31.38%
B	0.84%	2.93%	15.90%	3.35%	0.42%	0.00%	23.43%
C	0.00%	0.00%	4.60%	9.21%	3.35%	0.42%	17.57%
D	0.00%	0.00%	0.00%	1.26%	3.35%	0.84%	5.44%
E	0.00%	0.00%	0.00%	0.00%	0.84%	2.93%	3.77%
Total	**15.48%**	**24.27%**	**33.89%**	**14.23%**	**7.95%**	**4.18%**	**100.00%**

In terms of rank differences, English literature teachers appeared to be more accurate in placing students relative to other students, with an overall absolute rank difference of 2.21, compared to 3.42 for chemistry (Appendix B).

3.1.5. Student ranking analysis

A further analysis was conducted that investigated the rankings of the student cohort. Figure 5 is a scatter plot that shows the degree of accuracy that teachers had in their ranking predictions (see Appendix D for plots separated by teacher). Overall, teachers appeared to be accurate in their ranking predictions with a Spearman correlation coefficient of .91 (p <.0001).

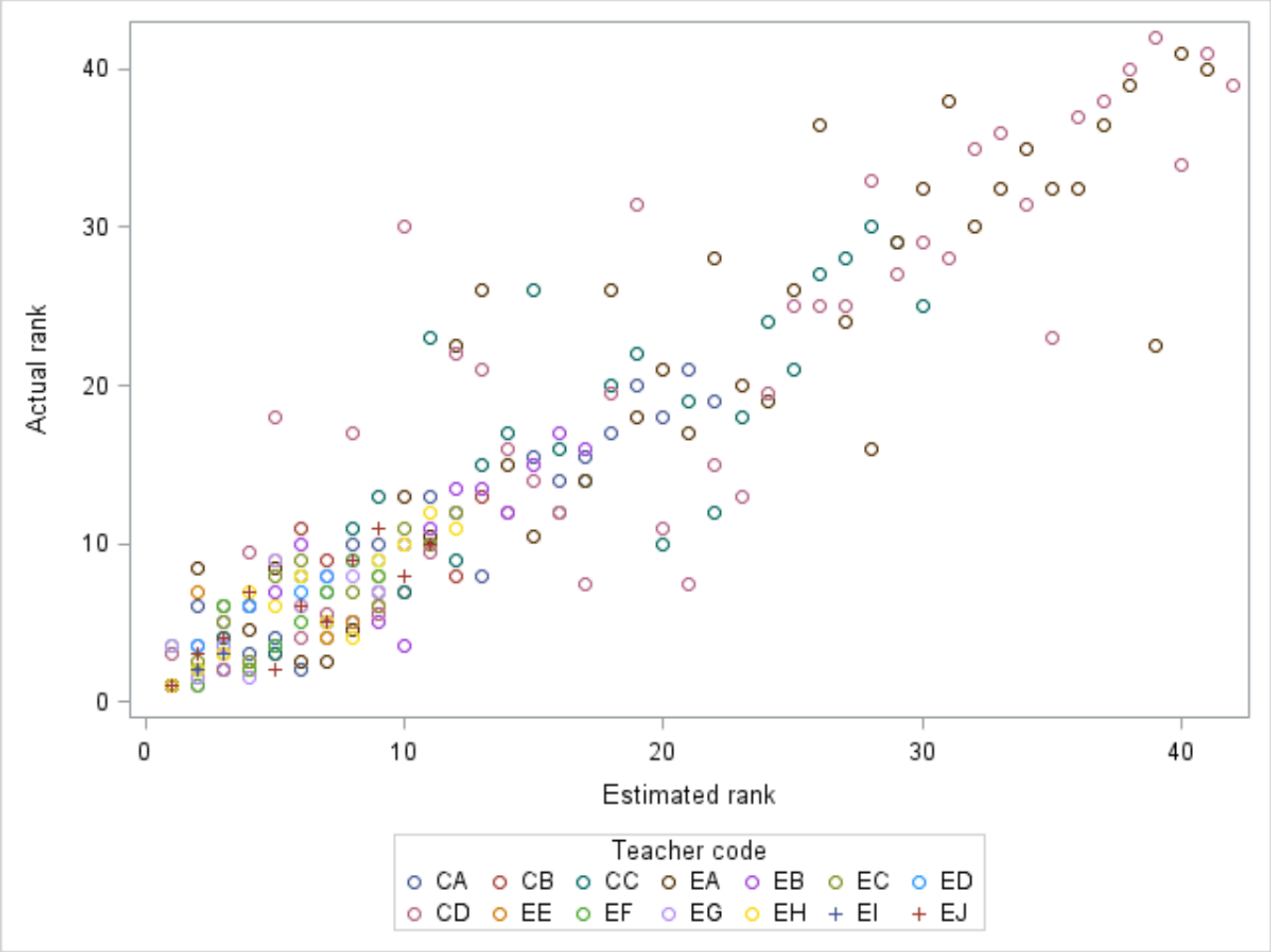


Figure 5: Scatter plot of estimated rankings vs actual rankings

Table 4 shows the Spearman correlations for individual teachers. The majority of the correlations were significant. For English literature teachers, the strength of the correlations was typically weaker, and for one teacher (EE) the correlation was non-significant. This particular teacher predicted that all of their students (*N* = 9) would get at least a grade A in English Literature, with one student predicted an A*. The students overall performed above expectations, with five achieving and A* grade, and only two students receiving a B grade. This mismatch between estimated and actual grades resulted in large ranking differences when UMS scores were considered, with one student moving down five places in the rankings, even after achieving a higher grade than expected.

Table 4: Correlation coefficients for estimated vs actual student rankings

Teacher code	*N*	Spearman correlation	*p*
CA	21	.94	<.0001
CB	13	.79	.001
CC	30	.86	<.0001
CD	42	.85	<.0001
EA	41	.89	<.0001
EB	17	.88	<.0001
EC	12	.84	.0006
ED	9	.83	.0058
EE	9	.47	.2054
EF	9	.79	.0119
EG	10	.77	.0085
EH	12	.87	.0002
EI	3	1.00	<.0001
EJ	11	.85	.0010

Table 5 shows the mean absolute difference between the predicted grade and the actual grade by prediction information used. For this calculation, it was assumed that for every student in a specific class the teacher used the same information. The absolute grade difference was lowest for AS level (mean) and oral assessment performance, whilst for the ranking difference the lowest difference was seen in GCSE (subject) and the Fischer Family Trust.

Table 5: Absolute grade and ranking difference by prediction source type

Prediction information		*N* students	Mean absolute grade difference	*SD*	Mean absolute rank difference	*SD*
Statistical	ALIS	127	.54	.57	2.74	3.22
	ALPS	71	.54	.58	3.46	4.01
	CAT	0	N/A	N/A	N/A	N/A
	FFT	33	.67	.60	1.42	1.40
	AS (mean)	53	.47	.58	1.58	1.39
	AS (subject)	239	.49	.59	2.74	3.20
	GCSE (mean)	49	.49	.54	2.51	2.83
	GCSE (subject)	42	.69	.56	1.31	1.16
Assessment performance	Coursework	209	.50	.60	2.67	3.19
	Mock exam	218	.45	.58	2.89	3.29
	Oral	35	.37	.60	1.74	1.35
	Other formative	115	.43	.56	2.03	2.20
In-class judgement	Obs of quality of work	239	.49	.59	2.74	3.20
	Obs of student commitment	239	.49	.59	2.74	3.20
	Obs of student interest	198	.49	.57	2.51	3.00

3.2. Interview and qualitative data

The interview and questionnaire qualitative data are presented according to themes that were identified by the research team. There were four themes identified: unease of using statistical modelling; the status of AS level; support for future predictions; and student characteristics for non-predicted results.

3.2.1. Unease of using statistical modelling alone

The questionnaire asked teachers to write, in as much detail as possible, how they utilise the information from statistics, assessment performance and in-class judgements to predict grades for their students. The qualitative data revealed nuanced and variable weightings to different data sources. For example, some of the teachers used ALIS data as a framework from which all predictions are made.

> *The school uses ALIS predictions to set a baseline target plus 1 grade.*
>
> Teacher EJ

However, this was often supplemented by some form of on-going assessment of work. Although statistical information provided by external sources (e.g. ALIS) was used there was typically mistrust in using the outcome for each student as given. Participant CA suggested that in his opinion that the FFT tends to overinflate results at the top end and underinflate at the bottom end of the grade range. Teacher EJ noted that for English literature ALIS is not able to consider wider reading in the subject, which is an important aspect in later success. ALIS is often considered alongside AS level performance, as the following quotations illustrate:

> *We would begin with target grade (calculated by ALIS GCSE score and test). We would relate this to performance in Module tests and AS exams.*
>
> Teacher CD

> *Predictions are based on a combination of their ALIS scores, their performance at AS and their coursework - as it emerges alongside their final mark.*
>
> Teacher ED

It appears then that teachers use information measured external in the first instance and then use this information in connection with in-class performance and other more qualitative measures to come to a final prediction. Participant EE suggested that whilst ALIS was a realistic predictor her centre wanted to 'add value' whenever possible, and so used ALIS as a baseline and added one grade to predictions. Several of the interviewees referred to ALIS as a 'benchmark' grade, providing the minimum target for students.

3.2.2. The status of AS level

The AS level grade was widely considered to be the strongest predictor of future A level performance. This status, coupled with some issues raised about the accuracy of other statistical predictions, meant that the interviewees valued AS grades over and above other sources of information. As participant EJ stated:

> *AS level grades inform the A2 prediction closely. There is quite a formal relation between the two.*

This view appears to be confirmed by the questionnaire findings. AS level results in the target subject was the only statistical information source that was used by all of the questionnaire participants. Performance at AS, as indicated by overall UMS score, was claimed to provide useful information about the likelihood for a student to move up/down a grade at A2.

> *My decision is based on an overview of factors as listed with AS performance and the UMS score a key indicator. The actual position of a candidate within the grade UMS score range is also helpful. For example a candidate scoring nearer to 240 UMS on AS will more likely achieve a B or a low A, whereas a candidate scoring 270 -300 would be destined for a good A grade at A2.*
>
> Teacher CC

> *At my centre we have a theory that with OCR the AS grade is a good predictor of A2 grade. We tell our students that they are unlikely to exceed their AS level grade unless they re-sit.*
>
> Teacher EB

There were two main reasons given for the AS level grade being used as a strong predictor of A2 level performance. First, the interviewees stated that they were able to calculate the required number of UMS points required to achieve particular target grades. Unrealistic targets were often adjusted at this stage:

> *Obviously there are adjustments as well to any initially predicted grade. It can change in some circumstances to be realistic in relation to the student. What they got at AS level is crucial, based on UMS calculations.*
>
> Participant EH

In this sense, the initial predictions made by ALIS and similar are usurped by actual performance in live examinations. Secondly, AS levels appeared to have some currency value in terms of university applications:

> *Universities like the AS level as it provides information more relevant to A level performance.*
>
> Participant CA

When interviewees were asked about how they intend to make predicted grades in the context of reformed qualifications, several stated that their centres intend to co-teach the AS level in the first year to help inform progress tracking.

> *Our school will do AS level this year to provide an end of Yr 12 grade. This makes applications to university stronger.*
>
> Participant EH

3.2.3. Support for future predictions

The interviewees stated that predicting A level grades for the new qualifications will be more difficult and will rely on close tracking of formative work.

> *Two years hence is difficult - regular marking/assessment to take the temperature at different times of the year.*
>
> Participant EJ

They also said that in future they may attach more weighting to coursework-based assessments. One teacher mentioned that indeed, coursework marks are used to help inform predicted grades because they represent a ‘best case scenario’ for the student, in the sense that it provides an indication of what the student can produce in a nurtured environment.

Several of the teachers noted that it was important for the awarding bodies to provide support on this aspect of qualifications, beyond offering the AS level. They suggested that awarding bodies could support the development of effective mock examinations. Participant CA suggested that he will look at Salters papers to help mock examination development. Participant EB stated that she expected her centre would offer AS level for the first year of students in the reformed qualification, but would use internally developed examinations in future years. She expected that students might be aware of the lowered ‘status’ of AS levels in determining their overall A level grade and so might strategically focus on qualifications where the AS and A2 are coupled together.

3.2.4. Student characteristics for non-predicted results

The majority of the teachers suggested that they used some additional criteria to sculpt grade predictions over the final year as student progress is established. This can be related to aspects of coursework performance, or student attitude to work:

> *The grade may then be revised depending on student's performance in two mock examinations in the course of Year 13, their performance in the coursework and their general attitude to learning and work throughout the year.*
>
> Teacher EB

> *We would begin with target grade (calculated by ALIS GCSE score and test). We would relate this to performance in Module tests and AS exams. We would then look at progress during Year 13 especially improvements in attitude and work rate and strategies for which papers they might re-sit.*
>
> Teacher CD

Following the publication of results, interview participants were each asked about three students that had either performed as predicted, or had moved up or down in grade or rankings. They were asked to offer some reflections on why these final outcomes had occurred and whether any EARs requests were submitted. The teachers often attributed accurate predictions to consistency of the student in whichever grade they ended up with. For A grade students, it was often understood that they were unlikely to perform well enough for the A*, but were solidly within the A grade.

For students that performed beneath expectations, interviewees typically put this down to student motivation, an overly optimistic prediction, or issues with performance under exam conditions. For example participant EA explained for one student that was predicted an A* but received a B lost motivation after they had received an unconditional offer from a Russell Group university. Interestingly, the interviewees rarely attributed underperformance to concerns about reliability of marking. Although all of the participants said that their centres were submitting some EAR requests, this was typically motivated by students being close to grade boundaries rather than a concern with the overall quality of marking. As one participant noted:

> *[An EAR is requested] when the UMS is close to the boundary, or the exam performance appears out of kilter.*
>
> Participant EJ

Interviewees were also asked about students that performed above expectations, although this was rarer than both accurate and optimistic predictions (see Figure 4). This was often attributed to particular teaching approaches that had worked well for students.

4. Discussion

The teachers in the present study tended to be optimistic in their predictions, a finding that is in line with previous research (e.g. Gill & Benton, 2015). The participants stated that these predictions were often *deliberately* optimistic. This was to provide motivation for students and to reflect the reality that students are competing for university places. The pattern of the accuracy of predictions was consistent between Chemistry and English literature, which was also in line with previous studies. The slightly higher degree of accuracy observed in the present study may reflect the high-performing cohort in the sample. Gill and Benton (2015) found that grade predictions are more accurate for higher grades. Interestingly, when students sampled in their research were predicted the A* grade, accuracy was higher than for other grades, and there was a difference between language subjects (such as English – 74% accurate) compared to science subjects (67% accurate). This is perhaps surprising given the subjective nature of English as a subject discipline and the generally lower marker reliability compared to science subjects.

The correlational analysis showed that teachers were reasonably effective at ranking their students, although they appeared to be slightly more accurate at the top end of the ranking scale. This suggests that teachers are able to make sound relative judgements about candidates' abilities (Laming, 2004) even if their representation of how these judgements convert into a grade is influenced by factors related to student motivation and competition for university places. The teachers stated that they did not usually rank students in making their predictions. This may be a reason why the questionnaire had only a small number of responses, as the request to rank students was potentially incongruent to teachers' normal estimating activities. This incongruence may also be related to the criterion-referenced nature of qualifications and grading in England and Wales. In countries where the use of norm-referenced tests is more common (see Ryan & Keir, 2008 for examples) student rankings may find more currency, as this can indicate a student's final placement on the grading scale. It is possible that future changes to the grading system at GCSE may influence teachers' approaches to predicting grades at Level 2, and encourage different prediction methods. Although the focus of the present study was on A level predictions, it revealed that teachers were uncertain as to where standards would be set in the context of new qualifications. It is possible that teachers will respond by attaching more value to relative judgements (both within and between years) in making grade predictions. This may also be influenced by the norm-referenced element to the awarding of grades 8 and 9 in the new GCSE.

Historically, the accuracy of grade predictions was potentially problematic in cases where the predicted grades influenced awarding decisions, and in exceptional circumstances (such as candidate illness). However, the recent removal of the requirement to provide estimated grades removes the possibility for these optimistic predictions to influence final grade outcomes. Teachers' judgements on prospective academic performance of their students, however, can have consequences for pedagogical practice, later evaluations of student performance, and individual students' own concepts of their abilities and potential (Sudkamp & Kaiser & Moller, 2012). Although more 'objective' statistical predictors are available, teachers in the present study were wary of taking the statistical information that they received as given. There were numerous references to professional judgement when making grading decisions, which allowed for a more iterative process.

The questionnaire data revealed that all of the teachers considered student commitment and judgements on in-class work in making grade predictions. When teachers were asked about students that underperformed relative to predictions, this was often attributed to student motivation or issues related to student performance in an examined unit. This was slightly surprising given the recent rhetoric that has raised concerns with quality of marking, and the year on year increase of EARs (Ofqual, 2014). Thus far there has been no research conducted on how student motivation influences the accuracy of grade predictions by teachers (Sudkamp et al., 2012). Student characteristics, and elements related to in-class motivation, came out as strong influencers of the grade teachers predicted. An interesting avenue for future research would be to establish whether there was a relation between *consistency* of student motivation and in-class performance, and the accuracy of teachers' predicted grades.

In making their predictions, teachers stated that they synthesised information from several different sources. A common process was for teachers to initially use statistical information from an external modelling service such as ALIS. This information was then manipulated to develop either an expected, target, or aspirational grade. However, the AS level was typically considered the strongest predictor of A2 level performance by the teachers, and was used by all of the teachers in the second year of the course. In the present study, there was some evidence to suggest that AS level grades provided a slightly stronger predictor of final grades, compared to other statistical measures. Hopkin (2011) found that AS level grades matched A2 level grades 54% of the time, and when they were not matched, there was an even distribution of grades that were below and above the AS level grade. This suggests that AS level grade may indeed be a better predictor of final outcomes than teacher judgement (as indicated by their grade predictions), at least in the sense that they are less prone to bias. This status is reflected in how closely teachers claim to use the AS level grade as a barometer for later performance.

Related to the point above, the participants in the present study also studied the AS level UMS score closely when deciding a final predicted grade for their students. Participants suggested that the UMS mark provided important information about the probability that students would be able to enter a higher grade at A2 level or go down a grade (if they were close to a lower grade boundary). Searle (2013) in a study analysing 53 subjects, found that UMS scores between AS and A2 level were closely correlated (at approximately 0.8). This suggests that a candidate's predicted A level grade is likely to be more reliable if it is based on the candidate's AS level UMS score. In the reformed suite of A level qualifications, it is possible for AS levels to be co-taught. Some of the teachers anticipated they would offer the reformed AS level as a staging post for to track the progress of students. Several of the teachers stated that they would offer AS levels for this reason. Otherwise they would be reliant on effective mock examinations and assessment materials to be provided by awarding bodies.

4.1. Future work

The reformed A level qualifications introduce an interesting divide in the cohort at AS level (i.e. students that take the AS level and those that do not). In a recent Ofqual blog, Oposs (2015) stated, 'we don't know what the impact will be of removing that formal feedback [that AS level provides]'. There are potentially two avenues for future research emerging from the present research that may boost the utility of AS levels to teachers. First, it would be

interesting to investigate whether teachers that have co-taught the AS level with the A level provided significantly more accurate predictions of overall A level performance. Secondly, it is important to investigate whether students that take AS levels perform over and above their predictions, compared to those that did not take it. This research could potentially inform the strategy for engaging teachers with the reformed AS level qualifications.

5. References

AQA. (2014). Instructions for submitting estimated grades. Retrieved 15th April, 2015 from http://store.aqa.org.uk/admin/library/AQA-EST-GRADES.PDF

Department for Business, Innovation & Skills (2011). Investigating the accuracy of predicted A level grades as part of the 2009 UCAS admission process. Retrieved 15th April, 2015 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32412/11-1043-investigating-accuracy-predicted-a-level-grades.pdf

Cambridge Assessment. (2013). *A simple guide: Assuring OCR's marking accuracy.* Retrieved 5th May, 2015 from http://www.cambridgeassessment.org.uk/Images/143058-a-simple-guide-to-marking-and-grading.pdf.

CEM. (2015). Centre for Evaluation and Monitoring. Retrieved 5th May, 2015 from http://www.cem.org/post-16

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, Thomson.

Ofqual. (2014). Commentary of enquiries about results for GCSE and A level. Retrieved 15th October, 2015 from //www.gov.uk/government/uploads/system/uploads/attachment_data/file/378156/2014-10-21-enquiries-about-results-provisional-statistics-summer-2014-commentary.pdf

Oposs, D. (2015). A level awarding in 2016 and 2017. Retrieved 26th October, 2015 from https://ofqual.blog.gov.uk/2015/08/14/a-level-awarding-in-2016-and-2017/

Ryan, J., & Keir, S. (2008). The status of large-scale assessment in the Pacific Region. Retrieved 15th October, 2015 from http://ies.ed.gov/ncee/edlabs/regions/pacific/pdf/techbrief/tr_00308.pdf

Searle, J. (2013). *Predicting A level performance*. Retrieved 22nd October, 2015 from https://cerp.aqa.org.uk/perspectives/predicting-level-performance.

Sudkamp, A., Kaiser, J., & Moller, J. (2012). Accuracy of teachers' judgements of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743-762.

Appendix A: Interview schedule for teachers

Investigating how teachers make estimated grading decisions at A level- follow up interview

Documentation required: interview schedule, predicted grade list, specific candidate number/names to enquire about

Teacher name:

Subject:

Date of interview:

Introduction: *Thank you for logging into the conference call today so that I could interview you. As you will remember from our email/discussions, the aim of the interview is to get some of your reflections on the grade outcomes for the students that you sent us information about in the previous questionnaire.*

I expect that the interview will last approximately X minutes. Please note that the interview is to be recorded for the purpose of data collection, but anything you say will be fully anonymised in any report and will not be traceable to you. Is this ok with you before we begin?

As you may remember in the email that I sent to you, I will ask you about some specific examples of students in the sample you sent in the questionnaire. Do you have a list to hand?

Ok, if you are ready, we can get started!

How you decide and use predicted grades

Note: hollow bullet-points to act as prompts

In the first part of the interview I would like to ask you about the reasons for making predicted grades, and the information you utilise in doing so.

- What are the main reasons for you making predicted grades for your students?
    - Information for students, teacher accountability, university applications, motivation?

- You mentioned in the questionnaire that you used a variety of sources of information in deciding the predicted grades for your students. Could you outline for me how you use this information? *Note: may be useful to have the list of sources they use for reference.*
    - Is it different for different types of student?
    - Do you change your predicted grade over time, or for other reasons such as UCAS?

If they use statistical information (e.g. ALIS)

- When do you collect the data?
    - Beginning of the course, towards the end?

If they use ALIS, ALPS, CAT or FFT data

- What is the typical relationship between the statistical prediction this method uses and the predicted grade you provide to students?

- How do you use the statistical information that you receive?
    - Do you often adjust it based on other observations about the student?
    - Does it match what your initial view is of the student?
    - Are there any pressures to increase predicted grades?

- You may be aware that from 2015 the examination boards no longer required you to send predicted grades to them before the final examinations took place. Has this changed anything about how you *decide* predicted grades?
    - Information you utilise, decided earlier or later?

- Has this changed anything about how you *use* predicted grades?
    - Timing of discussion with students/parents, submissions to UCAS

If they use assessment information (e.g. mock exams)

- When is this information gathered during the course?

- Do you use any information from other subjects in decided predicted grades for your subject (e.g. English Language or other Sciences)

- Do you value any type of information (statistical, assessment-based or observational) over others?
    - If so, why?

Questions on specific candidates

I would now like to ask you about the grade outcomes for your students, using some examples from the sample you sent to me. First of all, I would like to ask you about your thoughts of how the class did overall.

- Overall, what are your reflections on the grade outcomes for your students.
    - Did they perform to expectations? Were there many surprises?

From the list you sent to me it was possible to both understand the grade that you predicted and also the position of each student overall in the class (i.e. their predicted rank). I would like to ask you about some interesting examples in your class. I will give you the candidate number so that you can remind yourself of the name of the candidate.

Note: Look at sample to check for the following scenarios before the interview commences

- When the predicted grade matched the outcome grade
- When a student performed above the predicted grade
- When a student performed below the predicted grade
- When the student moved up/down the rankings (i.e. relative to the rest of the sample)

List candidate numbers, grades and why interesting here.

- Were there any particular reasons why you think the predicted grade matched/mismatched with the outcome grade?
    - Personal circumstances, particular performance on one unit, perceived marker error

- Have you/will you be putting in an EAR for any of the students in this class?

If yes to the previous question ask question below

- How do you decide whether to put in an EAR or not?
    - When UMS close to grade boundary, university place, student request, unit-related performance.

Future predicted grades

Finally, I would like to ask you a couple of questions about how you intend to use predicted grades in the future.

- Will the results from 2015 influence your approach to making predicted grades in the next session?

- Do you think there will be any changes in how you use predicted grades when you begin taking the reformed qualifications?
    - Since introduction of changes to the assessments (e.g. endorsement model; speaking and listening changes)

End of interview

That is the end of the interview. Once again thank you for taking the time to speak to me today. I hope you have found it an interesting experience. As a thank you for your participation, we would like to offer you a £20 voucher. I can send this through to you via email and will do so as soon as possible. I would also like to send you some information about the results of the study once it is completed.

Appendix B: Mean absolute grade and ranking differences by subject and outcome grade

Subject/outcome grade	*N*	Mean absolute grade difference	Mean absolute rank difference
A*	8	.50	1.38
A	24	.21	4.29
B	26	.69	3.31
C	21	.48	4.52
D	18	.67	3.00
E	9	.22	1.44
Chemistry	**106**	**.48**	**3.42**
A*	29	.41	2.26
A	34	.50	2.04
B	55	.56	2.64
C	13	.23	.92
D	1	.00	1.00
E	1	2.00	1.00
English literature	**133**	**.49**	**2.21**
Total	**239**	**.49**	**2.74**

Appendix C: Percentage of candidates by outcome and estimated grades, split by subject

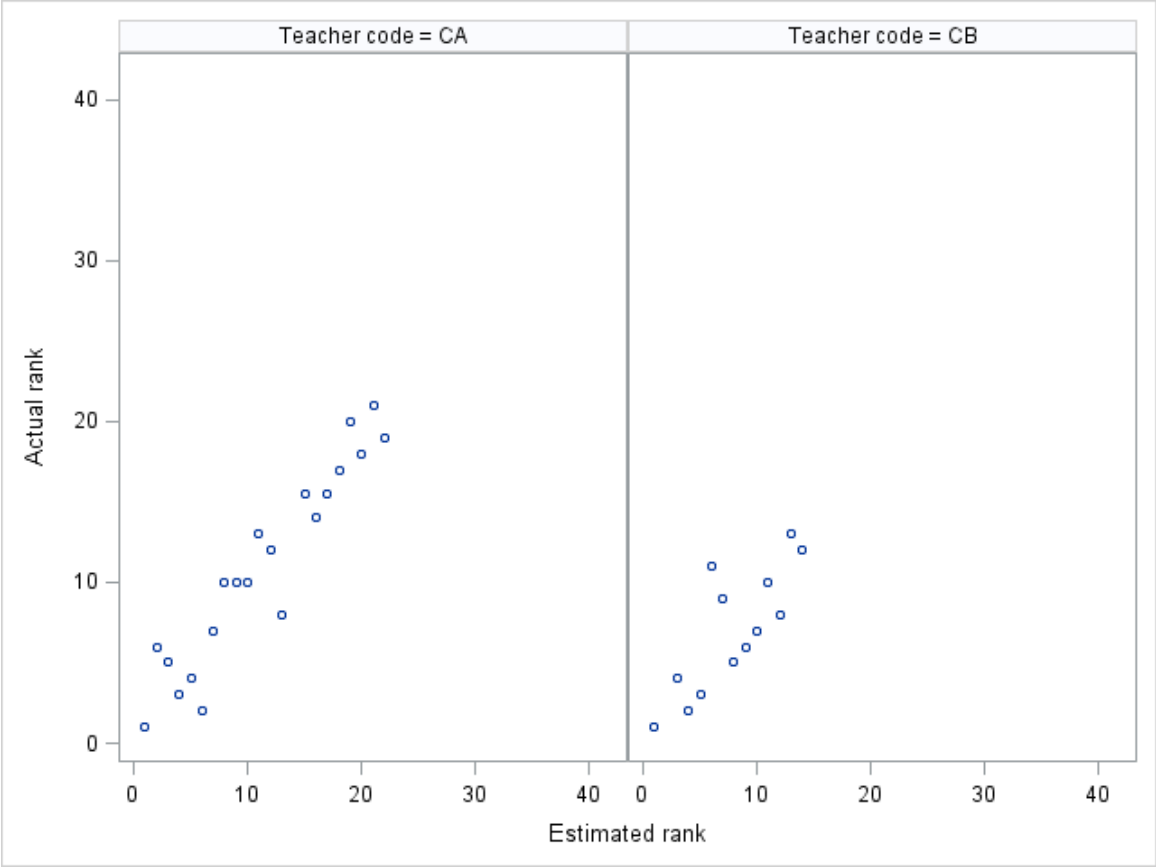
Chemistry

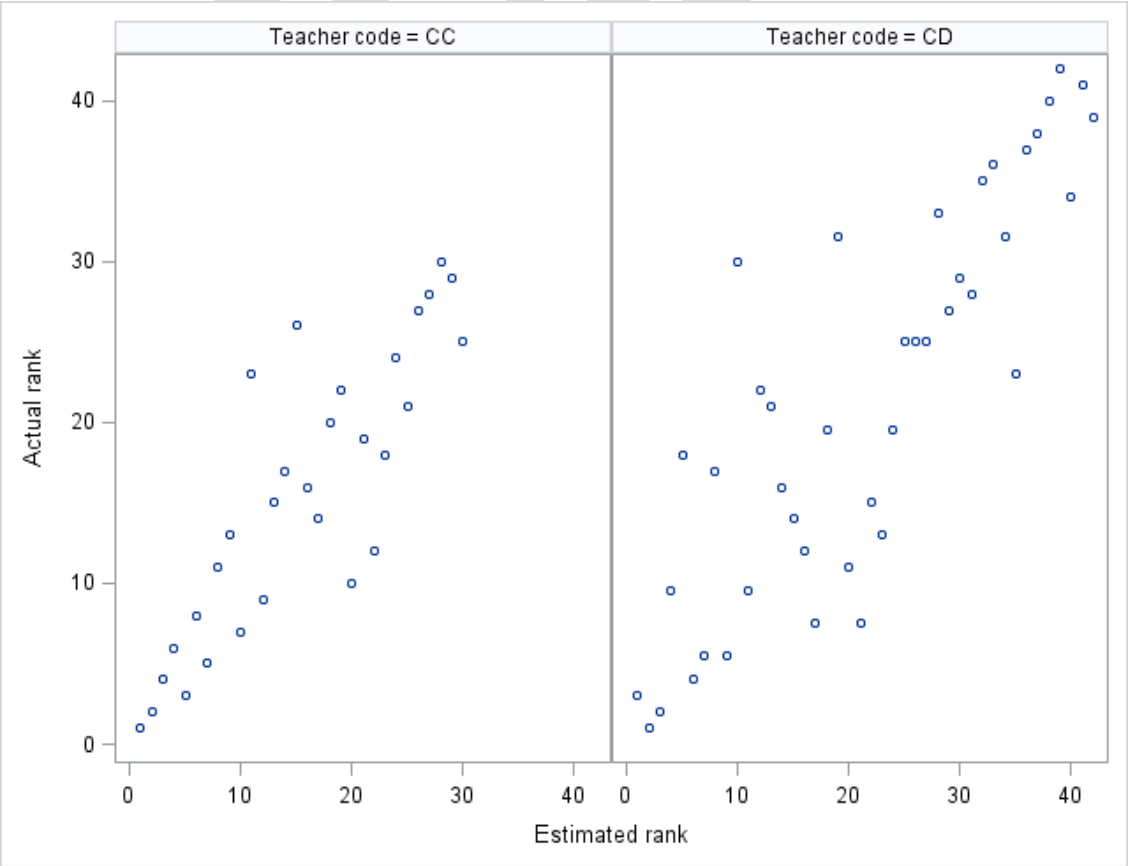
	Outcome grade						
Estimated grade	**A***	**A**	**B**	**C**	**D**	**E**	**Total**
A*	3.77%	0.94%	0.94%	0.00%	0.00%	0.00%	5.66%
A	3.77%	17.92%	7.55%	0.94%	0.00%	0.00%	30.19%
B	0.00%	3.77%	8.49%	5.66%	0.94%	0.00%	18.87%
C	0.00%	0.00%	7.55%	11.32%	7.55%	0.00%	26.42%
D	0.00%	0.00%	0.00%	1.89%	6.60%	1.89%	10.38%
E	0.00%	0.00%	0.00%	0.00%	1.89%	6.60%	8.49%
Total	**7.55%**	**22.64%**	**24.53%**	**19.81%**	**16.98%**	**8.49%**	**100.00%**

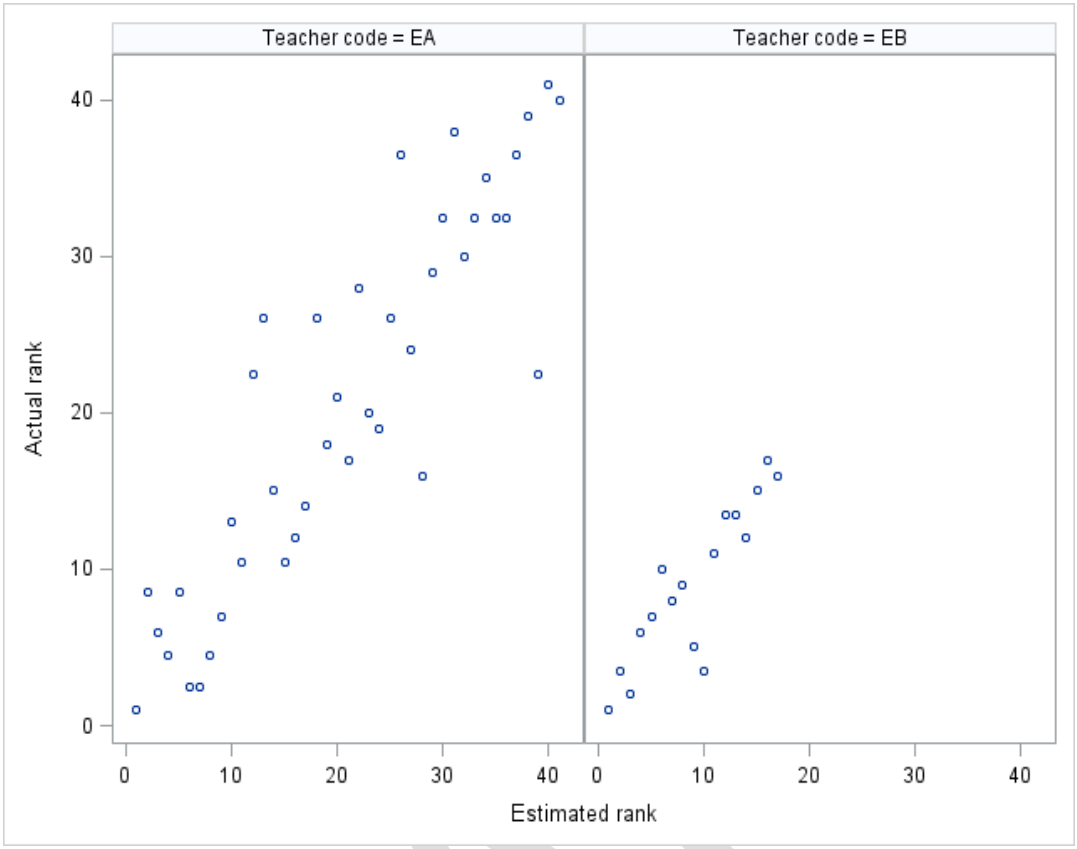
English literature

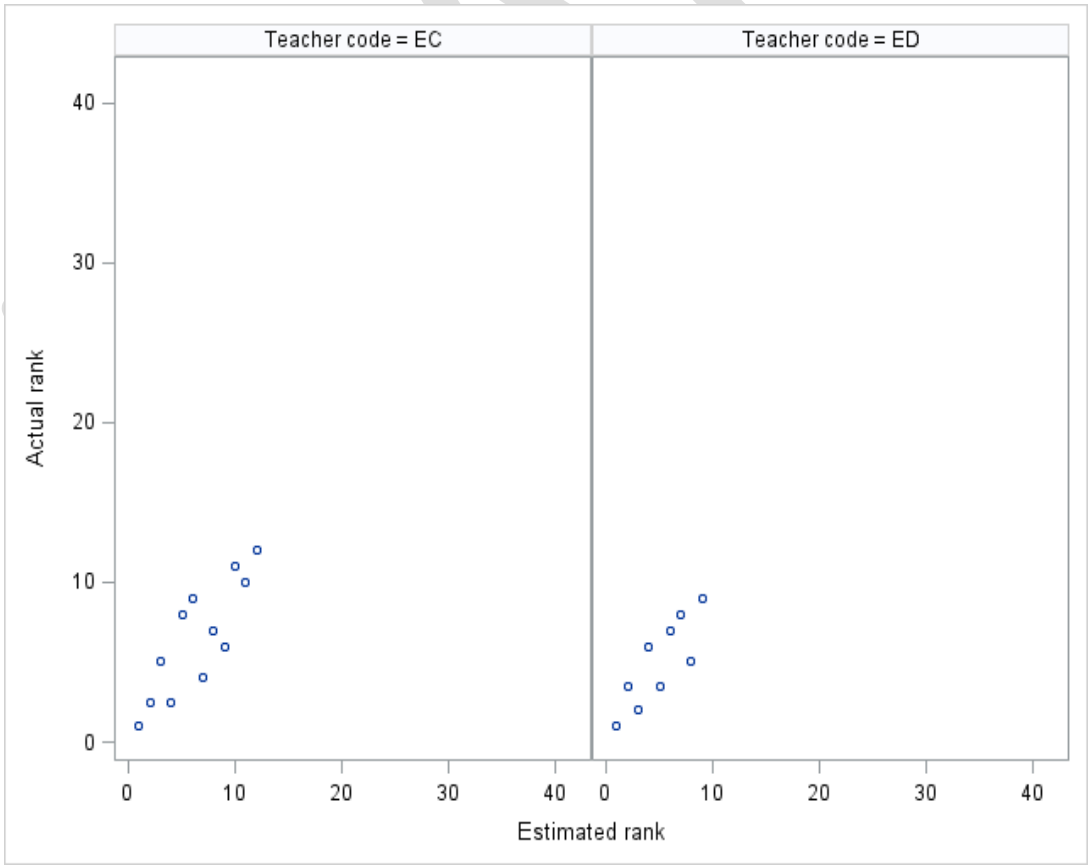
	Outcome grade						
Estimated grade	**A***	**A**	**B**	**C**	**D**	**E**	**Grand Total**
A*	14.29%	10.53%	3.76%	0.00%	0.00%	0.00%	28.57%
A	6.02%	12.78%	13.53%	0.00%	0.00%	0.00%	32.33%
B	1.50%	2.26%	21.80%	1.50%	0.00%	0.00%	27.07%
C	0.00%	0.00%	2.26%	7.52%	0.00%	0.75%	10.53%
D	0.00%	0.00%	0.00%	0.75%	0.75%	0.00%	1.50%
E	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Grand Total	**21.80%**	**25.56%**	**41.35%**	**9.77%**	**0.75%**	**0.75%**	**100.00%**

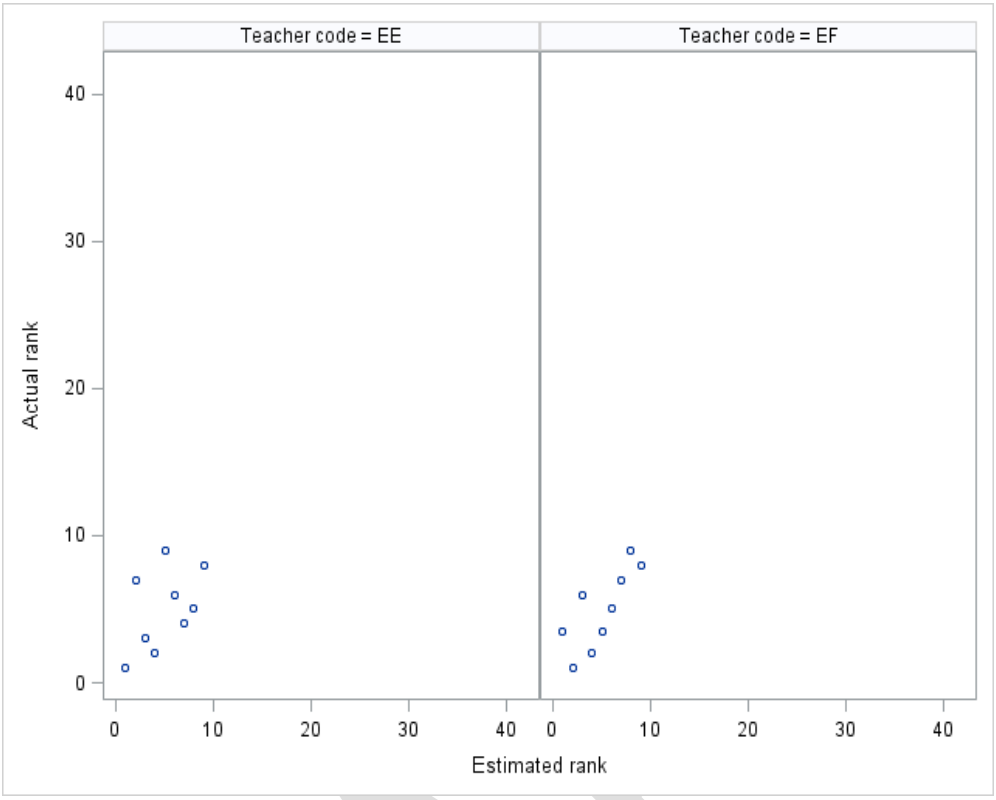
Appendix D: Ranking correlations by teacher

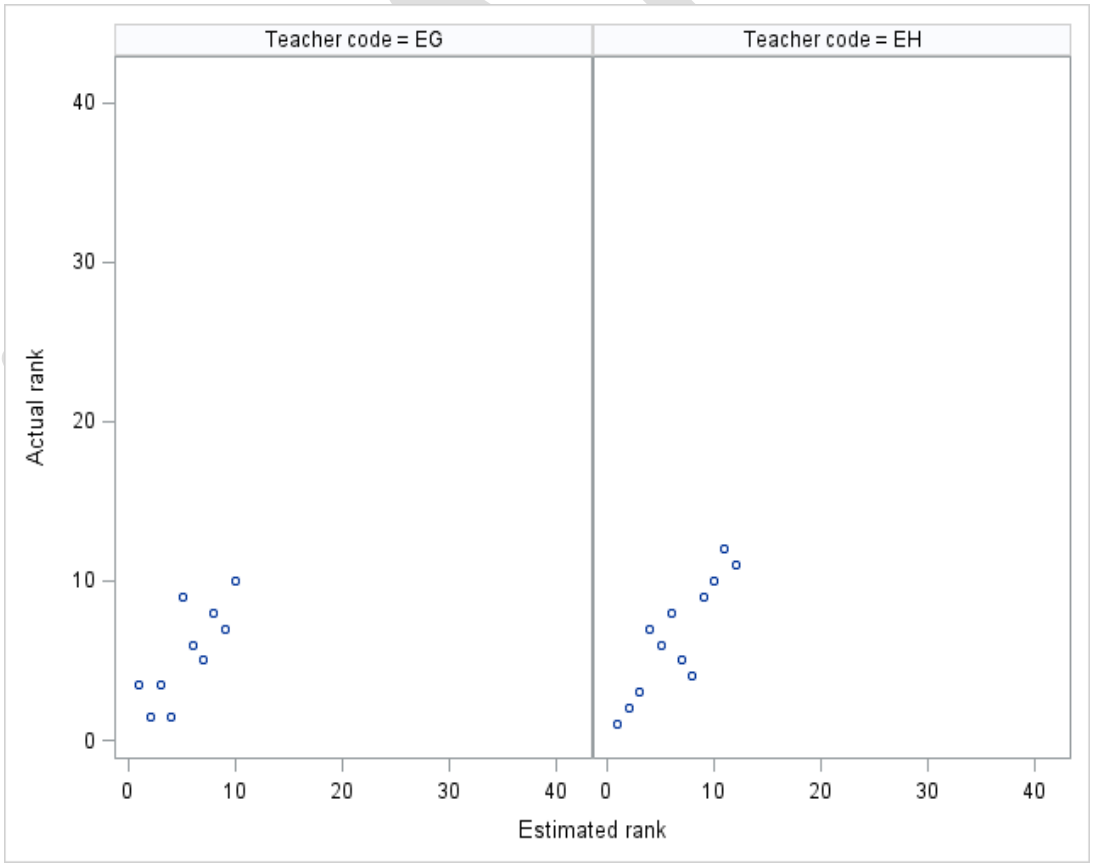


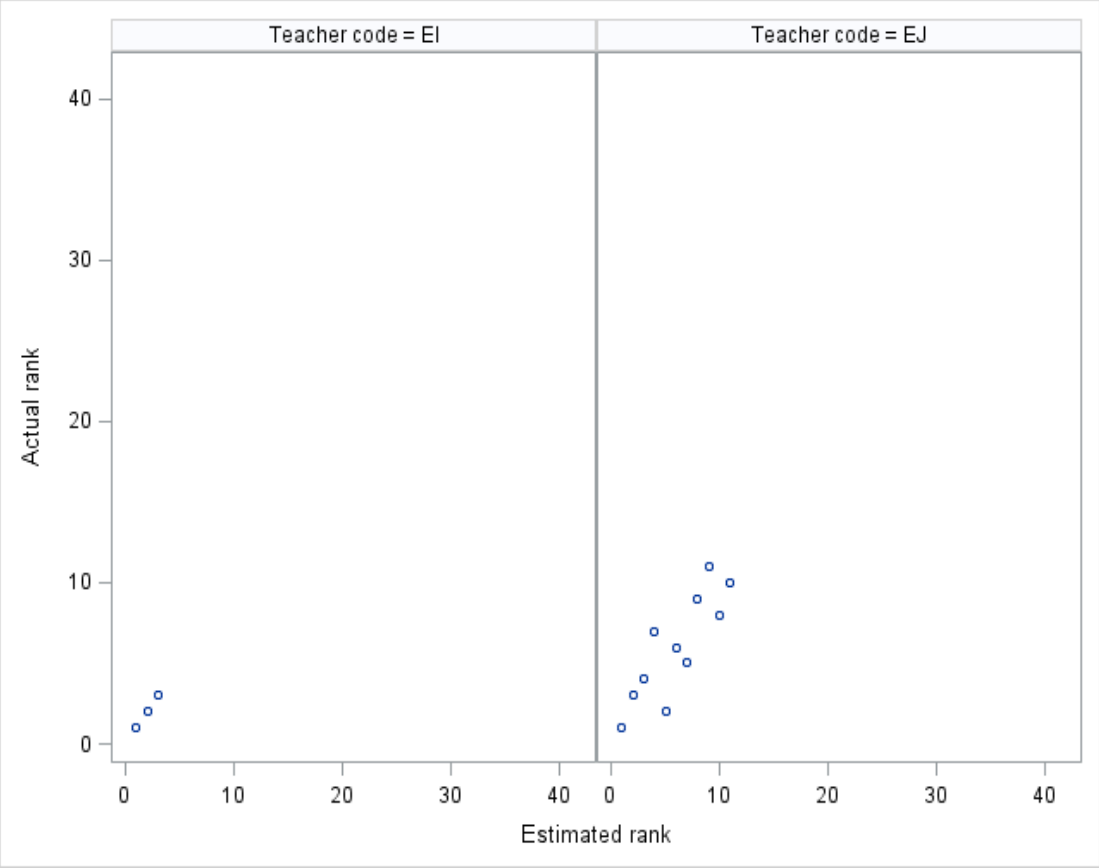



Teacher code = EA
Teacher code = EB
Actual rank
Estimated rank
0
10
20
30
40


Teacher code = EC
Teacher code = ED
Actual rank
Estimated rank
0
10
20
30
40


Teacher code = EE
Teacher code = EF
Actual rank
Estimated rank
0
10
20
30
40


Teacher code = EG
Teacher code = EH
Actual rank
Estimated rank
0
10
20
30
40


Teacher code = EI
Teacher code = EJ
Actual rank
Estimated rank
0
10
20
30
40