

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## The predictive validity of A-level grades in medical school applicants in a time of COVID-19: A retrospective analysis of administrative data

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-047354
Article Type:	Original research
Date Submitted by the Author:	26-Nov-2020
Complete List of Authors:	McManus, Ian; UCL Medical School, Research Department of Medical Education Woolf, Katherine; UCL Medical School, Research Department for Medical Education Harrison, David; UCL Medical School, Research Department for Medical Education Tiffin, Paul; University of York, Department of Health Sciences; Hull York Medical School, Health Professions Education Unit Paton, Lewis; University of York, Cheung, Kevin; Cambridge Assessment Smith, Daniel; General Medical Council
Keywords:	COVID-19, MEDICAL EDUCATION & TRAINING, EDUCATION & TRAINING (see Medical Education & Training)

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

# The predictive validity of A-level grades in medical school applicants in a time of COVID-19: A retrospective analysis of administrative data

I C McManus <sup>a,b,e,\*</sup>

[i.mcmanus@ucl.ac.uk](mailto:i.mcmanus@ucl.ac.uk) ORCID 0000-0003-3510-4814

Katherine Woolf <sup>a,c,e</sup>

[k.woolf@ucl.ac.uk](mailto:k.woolf@ucl.ac.uk) ORCID 0000-0003-4915-0715

David Harrison <sup>a,e</sup>

[david.harrison@ucl.ac.uk](mailto:david.harrison@ucl.ac.uk) ORCID 0000-0002-6639-6752

Paul A Tiffin <sup>b,d,f</sup>

[paul.tiffin@york.ac.uk](mailto:paul.tiffin@york.ac.uk) ORCID 0000-0003-1770-5034

Lewis W Paton <sup>b,f</sup>

[lewis.paton@york.ac.uk](mailto:lewis.paton@york.ac.uk) ORCID 0000-0002-3328-5634

Kevin Yet Fong Cheung <sup>b,g</sup>

[Cheung.K@cambridgeenglish.org](mailto:Cheung.K@cambridgeenglish.org) ORCID 0000-0002-9548-2932

Daniel T. Smith <sup>a,b,h</sup>

[daniel.smith@gmc-uk.org](mailto:daniel.smith@gmc-uk.org) ORCID 0000-0003-1215-5811

<sup>a</sup> UKMEDP089 project

<sup>b</sup> UKMEDP051 project

<sup>c</sup> UKMEDP089 Principal Investigator

<sup>d</sup> UKMEDP051 Principal Investigator

<sup>e</sup> Research Dept of Medical Education, University College London, London WC1E 6BT, UK

<sup>f</sup> Department of Health Sciences, University of York, Heslington, York, YO10 5DD, UK

<sup>g</sup> Cambridge Assessment, Shaftesbury Road, Cambridge CB2 8EA, UK

<sup>h</sup> General Medical Council, Regent's Place, 350 Euston Road, London NW1 3JN, UK

\* Author for correspondence: [i.mcmanus@ucl.ac.uk](mailto:i.mcmanus@ucl.ac.uk)

## Abstract

### Objectives

Calculate the predictive validity and reliability of predicted A-level grades compared to attained A-level grades. Model potential consequences for UK medical schools of replacing public A-level examination grades with teacher-estimated calculated grades as a result of the COVID-19 pandemic.

### Design

Longitudinal observational study using UK Medical Education Database data.

### Setting

UK medical education and training.

### Participants

Dataset 1: 81,202 medical school applicants in 2010 to 2018 with predicted and attained A-level grades. Dataset 2: 22,150 18-year old medical school applicants in 2010 to 2014 with predicted and attained A-level grades, of whom 12,600 had medical school assessment outcomes and 1,340 had postgraduate outcomes available.

### Outcome measures

Correlations between predicted and attained A-level grades and undergraduate and postgraduate medical examination results. Estimated undergraduate medical school achievement.

### Results

Dataset 1: 48.8% of predicted grades were accurate, grades were over-predicted in 44.7% of cases and under-predicted in 6.5% of cases. Dataset 2: Undergraduate and postgraduate outcomes were predicted significantly better by attained than by predicted A-level grades. The lower predictive validity of teacher-estimated grades suggests medical students selected in 2020 are likely to under-attain, with 13% more gaining the equivalent of the current lowest decile of performance and 16% fewer gaining the equivalent of the current top decile, with follow-on effects for postgraduate training.

### Conclusions

The replacement of A-level examination grades with teacher-estimated calculated grades is likely to result in the medical students selected in 2020 having lower academic performance at medical school compared to previous years. Medical schools will probably also need to consider whether additional teaching is needed for entrants who are struggling, or might have missed out on important aspects of A-level teaching, with extra support being needed.

## Strengths and limitations of this study

- First comparison of the predicted validity of predicted and attained A-level grades using performance in subsequent undergraduate and postgraduate assessments.
- Large sample size of UK medical applicants from 2010 to 2018 provides adequate statistical power.
- Analysis of complete population data mean the results are unlikely to be biased.
- This study is with medical applicants only, meaning generalisability to students on other courses is uncertain.

## Background

“... the ... exam hall [is] a level playing field for all abilities, races and genders to get the grades they truly worked hard for and in true anonymity (as the examiners marking don't know you). [... Now we] are being given grades based on mere predictions.” Yasmin Hussein, letter to *The Guardian*, March 29<sup>th</sup> 2020 [1].

“[Let's] be honest, this year group will always be different...” Dave Thomson, blogpost on *FFT Educational Lab* [2]

“One headmistress commented that ‘entrance to university on teachers’ estimates may be fraught with unimagined difficulties’. ... If there is in the future considerable emphasis on school assessment, some work of calibration is imperatively called for.” James Petch, December 1964[3].

UK schools, in response to the COVID-19 pandemic, closed on 20<sup>th</sup> March 2020, and Key Stage 5 [Level 3] public examinations such as A-levels and SQA assessments were cancelled for summer 2020. On April 3<sup>rd</sup> *Ofqual* (Office of Qualifications and Examinations Regulation) in England announced that A-level, GCSE and other exams under its purview in England would be replaced by *Calculated Grades*, at the core of which are teachers’ estimates of the grades that their students would attain (called *Centre Assessment Grades, CAGs*), which would then be moderated by *Ofqual* using a computer algorithm which included the prior performance of the school attended by candidates – see the *Calculated Grades* subsection below for details. The Scottish Qualification Authority (SQA) and other national bodies also announced similar processes for their examinations. Inevitably the announcement of *Calculated Grades* resulted in confusion and uncertainty in examination candidates, particularly those needing A-levels or SQA Advanced Highers<sup>a</sup> to meet conditional offers for admission to university in autumn 2020. Universities also faced a major problem, having had A-levels taken away, which are, “the single most important bit of information [used in selection]” [4].

Some of the tensions implicit in *Calculated Grades* are well seen in the quotation above by Yasmin Hussein, a GCSE student in Birmingham, with its clear emphasis that a key strength of current examination systems such as GCSEs, A-levels and similar qualifications, is the *anonymity* and *externality* of the assessors, who know nothing of the students whose work they are marking. In contrast the replacement of actual grades attained in the exam hall with what Hussein describes as ‘mere predictions’ raises a host of questions, not the least being the possibility of bias when judgements are made by teachers<sup>b</sup>.

<sup>a</sup> SQA Highers are taken the year before (rather like AS-levels used to be) and therefore they will be available for 2020 applicants. Advanced Highers will not be available and will be estimated.

<sup>b</sup> A number of students, designated as ‘external’ or ‘private’, do not have regular teachers who can make such assessments and *Ofqual* has announced that they will not be able to receive *calculated grades*. Cases may

## Context of the current paper and the situation at the time of writing

Since the emergence of COVID-19 in Europe in early 2020, the situation has been constantly and rapidly changing. At the time of writing in November 2020, Europe is in the midst of a 'second wave' and England is in a second national lockdown. To help the reader situate the current paper in context, we explain briefly here what the exams situation was in the UK from April to August 2020, and provide more detailed in a postscript. The remainder of the paper should be read as it was published in pre-print in May 2020, when exams had been cancelled, the process for awarding Calculated Grades announced, but no grades yet released.

In Spring 2020, as universities entered into the final phases of the annual academic cycle of student selection, we considered in this paper the potential problems of using teacher-estimated grades such as the Calculated Grades proposed by Ofqual, rather than attained grades obtained in the usual way via examinations. We published the current paper in pre-print in May 2020 so that admissions tutors could see the paper within the admissions cycle. By August 2020 some immediate impacts were shown when the algorithms used by regulators resulted in many students, particularly those from historically poorly performing schools, having their expected results adjusted downwards. This forced the Scottish Government, followed by the English and Welsh Governments, to accept teacher-estimated Centre Assessment Grades (CAGs) without moderation by an algorithm or the Calculated Grade, whichever was the higher.

As predicted in this paper, this scrapping of the algorithm resulted in a very significant increase in grades compared to 2019 (see <https://ffteducationdatalab.org.uk/2020/08/gcse-and-a-level-results-2020-how-grades-have-changed-in-every-subject/>). One short-term impact of the increase in grades was to increase the number of students meeting university conditional offers. The longer-term impacts are still to be seen, but the current paper offers some clues, calculating the predictive validity of teacher-estimated predicted grades submitted to UCAS at application to university for applicants from 2010 to 2018 (i.e. before the pandemic) compared to attained grades, and modelling the potential effects on undergraduate and postgraduate performance in medical students and doctors of selecting medical students on teacher-estimated grades such as the Centre Assessment Grades submitted by schools to Ofqual during the pandemic in summer 2020.

## Medical school admissions

This paper will mainly concentrate on medical school applications. Medical education has a range of useful educational measures, including admissions tests during selection, and outcomes at the end of undergraduate training, which are linked together through UKMED (United Kingdom Medical Education Database; <https://www.ukmed.ac.uk/>). UKMED provides a sophisticated platform for assessing predictive validity in multiple entry cohorts in undergraduate and postgraduate training [5]. The paper should also be read in parallel with a second study from some members of the present team which assesses attitudes and perceptions to calculated grades and other changes in selection of current medical school applicants in the UKMACS (UK Medical Applicants Cohort Study) [6].

Fundamental questions about selection in 2020 concern the likely nature of Calculated Grades, and the extent to which they will predict outcomes to the same extent as currently do *actual or attained grades*. The discussion will involve actual grades, and then four types of teacher-estimated grades: predicted grades (sent to UCAS at application to university), centre assessment grades (CAGs –

---

include students taking a gap year and retaking A-levels, or a student studying GCSE or A-level German at a *Samstagsschule* who for administrative convenience takes the exam at a local school where the subject is not actually taught. There no doubt are many other A-level candidates affected as well.

submitted by schools to Ofqual in 2020) calculated grades (CAGs adjusted using an algorithm) and forecasted A-level grades (submitted by teachers to exam boards pre-2015 as a quality check for real exam grades). These are summarised in a box below, together with final grades, which were the grades eventually accepted by UCAS and were the higher of the calculated grade or centre assessed grade.

## Calculated grades

The status of calculated grades was made clear by *Ofqual* in April 2020:

“The grades awarded to students will have equal status to the grades awarded in other years and should be treated in this way by universities, colleges and employers. On the results slips and certificates, grades will be reported in the same way as in previous years”. [7], p.6.

The decisions of *Ofqual* are supported by Ministerial statement, and universities and other bodies have little choice therefore but to abide by them, although that does not mean that other factors may not need to be taken into account in some cases, as often occurs when applicants do not attain the grades in conditional offers.

None of the above means that calculated grades actually *will be* equivalent to conventional attained grades. Calculated grades will not actually *be* attained grades, they may well behave differently to attained grades, and in measurement terms they actually *are not* attained grades, even though in administrative and even in legal terms, by fiat, they have to be treated as equivalent<sup>c</sup>. From the perspective of educational research, although the key issue is the extent to which calculated grades actually will or can behave in an identical way to attained grades.

### Box 1: A-level grades: Actual, predicted, centre assessment, calculated, final, forecasted, and teacher-estimated grades

**Actual or attained grades.** The grades awarded by examination boards/awarding organisations based on written and other assessments which are set and marked externally. Typically sat in *May and June of year 13*, with results announced in *mid-August*.

**Predicted grades.** Teacher estimates of the likely attained grades of candidates, provided to UCAS in the *first term of year 13*, and by *October 15<sup>th</sup>* for medical and some other applicants.

**Centre assessment grades.** Used in the production of Calculated grades (see below). Provided by examination centres (typically schools) between 1st and 12<sup>th</sup> of June 2020, consisting of teacher-estimated grades and candidate rankings within examination centres.

**Calculated grades.** The final grades to be provided for candidates by exam boards for Summer 2020 assessments, in the absence of attained grades. Based on centre assessment grades, with final calculated grades involving standardisation/adjustment by exam boards using an algorithm. Calculated grades, “will have equal status to the grades awarded in other years and should be treated in this way by universities, colleges and employers” (*Ofqual*). These grades were often referred to as the ‘algorithm grades’ and were abandoned by the UK governments in August 2020

**Final grades.** The grades used by UCAS in the 2020 admissions cycle – the higher of the teacher estimated grade or the centre assessment grade

**Forecasted grades.** Prior to 2015, teachers, in *May of Year 13*, provided to exam boards a forecast of the likely grades of candidates along with rankings. Forecasted grades therefore take place later in the academic cycle than predicted grades, close to the time examinations are actually sat.

**Teacher-estimated grades.** Generic term used in this paper to refer to grades estimated by teachers. Includes predicted grades, centre assessment grades, calculated grades, and forecasted grades.



1 In April 2020 Ofqual issued guidance on how *calculated grades* would be provided for candidates for  
2 whom examinations have been cancelled. Essentially, teachers will be required, for individual  
3 candidates taking individual subjects within a *candidate assessment centre* (usually a school), to  
4 estimate *grades for* candidates, and then to *rank order* candidates within grades, to produce *centre*  
5 *assessment grades*. A statistical standardisation process will then be carried out centrally using a  
6 computer algorithm. Ranking is needed because standardisation, “will need more granular  
7 information than the grade alone” ([7] p.7), presumably to break ties at grade boundaries which  
8 occur because of standardisation. Standardisation, to produce *calculated grades*, will use an  
9 algorithm that takes into account the typical distribution of results from that centre for that subject  
10 in the three previous years, along with aggregated centre data on SATS and previous exam  
11 attainment as in GCSEs<sup>d</sup>. This approach is consistent with Ofqual’s approach to standard-setting.  
12 Following Cresswell [8], Ofqual has argued that during times of change in assessments, and perhaps  
13 more generally, there should be a shift away from “comparable performance” (i.e. criterion-  
14 referencing), and that there is an “ethical imperative” to use “comparable outcomes” (i.e. norm-  
15 referencing) to minimise advantages and disadvantages to the first cohort taking a new assessment,  
16 as perhaps also for later cohorts as teachers improve at teaching new assessments [9].

17 Ofqual says that centre assessment grades, the core of calculated grades, “are not the same as ...  
18 predicted grades provided to UCAS in support of university applications” [10], (p.7). Predicted  
19 grades in particular are provided by schools in October of year 13 and centre assessment grades in  
20 May/June of year 13, seven months later, when Ofqual says that teachers should also consider  
21 classwork, bookwork, assignments, mock exams and previous examinations such as AS-levels (taken  
22 only by a minority of candidates now), but should *not* include GCSE results, or any student work  
23 carried out after 20<sup>th</sup> March. Whether centre assessment grades, or calculated grades – centre  
24 assessment grades moderated by the algorithm - will be fundamentally different from predicted  
25 grades is ultimately an empirical question, which should be answerable when UCAS data for 2020  
26 are available for medical school applicants in UKMED. In the meantime, it seems a reasonable  
27 assumption, except for a small proportion of candidates who have improved dramatically from  
28 October 2019 to March 2020, that centre assessment grades and hence calculated grades will  
29 probably correlate highly with earlier predicted grades. Predicted grades, which have been collected  
30 for decades, should therefore act as a reasonable proxy in research terms for centre assessment  
31 grades and therefore calculated grades.

## 32 The rationale for using A-level grades in selection

33 Stepping back slightly it is worth revisiting the reasons that A-levels exist and why universities use  
34 them in selection. A-levels assess at least three things: subject knowledge, intellectual ability, and  
35 study habits such as conscientiousness [11]. Knowledge and understanding of, say, chemistry is  
36 probably necessary for high level study of medical science and medicine, to which it provides an  
37 underpinning, and experience suggests that students without such knowledge may have problems.  
38 A-levels also provide evidence for a student’s intellectual ability and capability of extended study at  
39 a high level. A-levels are regarded as a ‘gold standard’ qualification because of the rigour and  
40 objectivity of their setting and marking. Their measurement is therefore *reliable*, and the  
41 presumption is that they are also *valid*, in the many senses of that word [12-14], and as a result are  
42 *unbiased*. A crucial assumption is of *predictive validity*, that future outcomes at or after university  
43 are higher or better in those who have higher or better A-levels, as found both in predicting degree  
44 classes in general [15-17] and medical school performance in particular [18,19]. At the other

---

45 <sup>d</sup> It was this standardisation process that Governments reversed in August 2020 after the protests  
46 against calculated grades.

1 extreme, A-levels could be compared conceptually with, say, a mere assertion by a friend or  
2 colleague that, "Oh yes, they know lots of chemistry". That is likely neither to be reliable, valid nor  
3 unbiased, and hence is a base metal compared with the gold standard of A-levels. The empirical  
4 question therefore is where on the continuum from gold to base, lie calculated grades or predicted  
5 grades.  
6  
7

8 The issue of predictive validity has been little discussed in relation to calculated grades, but in a  
9 recent *TES (Times Educational Supplement)* survey of teachers, there were comments that,  
10 "predictions and staff assessments would never have the same validity as an exam", so that,  
11 "Predictions, past assessment data and mock data is not sufficient, and will never beat the real thing  
12 in terms of accuracy." [20]. The current changes in university selection inevitably mean that difficult  
13 policy decisions need to be made by universities and medical schools. Even in the absence of direct,  
14 high-quality, evidence, policy-makers still have an obligation to make decisions, and, therefore it is  
15 argued, must take theory, related evidence, and so on, into account [21]. This paper will provide  
16 both a review of other evidence, and also new results on the related issue of predicted grades,  
17 which it will be argued are likely to behave in a way that is similar to calculated grades.  
18  
19  
20

## 21 Review of literature on predicted and forecasted grades

### 22 Predicted grades in university selection

23 A notable feature of UK universities is that selection mostly takes place before A-levels or equivalent  
24 qualifications have been sat, so offers are largely conditional on later attained grades. As a result,  
25 UCAS application forms, since their inception in 1964, have included *predicted grades*, estimates by  
26 teachers of the A-level grades a student is likely to achieve. Admissions tutors also use other  
27 information in making conditional offers. A majority of applicants in England, applying in year 13 for  
28 university entry at age 18 will have taken GCSEs at age 16 in year 11, a few still take AS-levels in year  
29 12, some students submit an EPQ (Extended Project Qualification), and UCAS forms also contain  
30 candidate statements and school references. Medical school applicants mostly also take admissions  
31 tests such as U(K)CAT or BMAT at the beginning of year 13, and many will take part in interviews or  
32 MMIs (multiple mini-interviews) (see [https://www.medschools.ac.uk/studying-medicine/making-an-  
33 application/entry-requirements](https://www.medschools.ac.uk/studying-medicine/making-an-application/entry-requirements)).

34  
35  
36  
37  
38  
39 Predicted grades have always been controversial. A House of Commons Briefing Paper in 2019 noted  
40 that the UK was unusual among developed countries in using predicted grades<sup>e</sup>, and said that,

41  
42 "The use of predicted grades for university admissions has been questioned for a long time.  
43 Many critics argue that predicted grades should not be used for university entry because  
44 they are not sufficiently accurate and it has been suggested that disadvantaged students in  
45 particular lose out under this system." [22] p.4

46  
47 Others have suggested that as well as being "biased", "predicting A-level grades is clearly an  
48 imprecise science" [23] (p.418). There have been repeated suggestions over the years, none as yet  
49 successful, that predicted grades should be replaced with a PQA (Post-Qualification Applications)  
50 system. As Nick Hillman puts it,

51  
52 "The oddity of our system is not so much that people apply before receiving their results;  
53 the oddity is that huge weight is put on predicted grades, which are notoriously unreliable. ...  
54 PQA could tackle this..."<sup>f</sup>.

55  
56  
57  
58  
59 <sup>e</sup> <https://www.bbc.co.uk/news/education-44525719>

60 <sup>f</sup> <https://www.hepi.ac.uk/2019/08/14/pqa-just-what-does-it-mean/>

The system of predicted grades is indeed odd, but also odd is the sparsity of academic research into predicted grades. The most important question that seems almost never to have been asked, and certainly not answered, is the fundamental one of whether it is predicted grades or actual grades which are better at predicting outcomes. Petch[3] considers that predicted and actual grades may be fundamentally different, perhaps being “complementary and not contradictory” (p.29), one being about scholarly attitude and the other about examination prowess, primarily because “the school knows the candidate as a pupil, knowledge not available to the examiners”. For Petch, either a zero correlation or a perfect correlation between predicted and actual grades would be problematic, the latter perhaps implying that actual grades might be seen as redundant (p.6).

The advent of Ofqual’s calculated grades, which are in effect predicted grades carried out in a slightly different way, means there is now an urgent need to know how effective predicted grades are likely to be as a substitute for attained A-level grades. Are they in fact ‘notoriously unreliable’, being ‘mere predictions’, or do they have equivalent predictive validity as attained grades?

### The research literature on predicted grades

As part of the Supplementary Information to this paper we have included a discursive overview of research studies on predicted grades. Here we will merely provide a brief set of comments.

Most studies look at predictions at the level of individual exam subjects, which are graded from E to either A or, from 2010 onwards, A\*. The most informative data show all combinations of predicted grades against attained grades, and Figure 1 gives an example for medical school applicants. Many commentators, though, look only at over-predictions (‘optimistic’) and under-predictions (‘pessimistic’). Figure 2 summarises data from five studies of university applicants. Accurate predictions occur in 52% of cases when A is the maximum grade and 17% when A\* is the maximum grade (and of course with more categories accuracy is likely to be lower). Grades are mostly over-predicted, in 42% of cases pre-2010 and 73% post-2010, with under-prediction rarer at 7% of cases pre-2010 and 10% post-2010. A number of studies have reported that under-prediction is more common in lower socio-economic groups, non-White applicants, and applicants from state school or further education [24-26]. A statistical issue means such differences are not easy to interpret, as a student predicted A\* cannot be under-estimated, and therefore under-estimation will inevitably be more frequent in groups with lower overall levels of attainment.

Some studies also consider grade-point predictions, the sum of grade scores for the three best attaining subjects, scored A\*=12, A=10, B=8, etc<sup>g</sup>. In particular a large study by UCAS [27] showed that applicants missing their predictions (i.e. they were over-predicted) tended to have lower predicted grades, lower GCSE attainment, were more likely to have taken physics, chemistry, biology and psychology, and were from disadvantaged areas. To some extent the same statistical problems of interpretation apply as with analysis at the level of individual exam subjects. For a number of years UCAS calculated grade-point predictions, and they are included in the P51 data described later.

### What are predicted grades and how are they made?

UCAS says that “A predicted grade is the grade of qualification an applicant’s school or college believes they’re likely to achieve in positive circumstances.”<sup>h</sup> Later though, the document says predicted grades should be, “**in the best interests of applicants** – fulfilment and success at college or

<sup>g</sup> In some studies a scoring of A\*=6, A=5, B=4 is used. The 12,10,8... scoring was introduced so that AS levels, weighted at half an A-level, could be scored as A=5, B=4 etc (there being no A\* grade at AS-level). For most purposes A\*=12, A=10 ... is equivalent in all respects to A\*=6, A=5, etc, apart from a scaling factor.

<sup>h</sup> <https://www.ucas.com/advisers/managing-applications/predicted-grades-what-you-need-know> [Accessed 13th April 2020].

1 university is the end goal “, and “**aspirational but achievable** – stretching predicted grades are  
2 motivational for students, unattainable predicted grades are not” (all emphases in original).  
3 Predicted grades should be professional judgements and be data-driven, including the use of, “past  
4 Level 2 and Level 3 performance, and/or internal examinations to inform ...predictions”.

5  
6  
7 Few empirical studies have asked how teachers estimate grades, with little major progress since  
8 1964 when Petch said, “Little seems to be known about measures taken by schools to standardize  
9 evaluations of pupils” [3] (p.7). Two important exceptions are the studies of Child and Wilson<sup>i</sup> in  
10 2015 and Gill [28] in May 2018, with only the latter published, Gill sending questionnaires to  
11 selected OCR Board exam centres concerning Chemistry, English Literature and Psychology exams.  
12 Teachers said the most important information used in predicting grades was performance in mock  
13 exams, observations of quality of work and commitment, oral presentation, the opinion of other  
14 teachers in the same subject, other subjects, and the head of department. Some teachers raised  
15 concerns about the lack of high stakes for mock exams which meant that students did not treat  
16 them seriously. AS-level grades were an important aid in making predictions. There were also  
17 concerns about the loss of AS-levels to help in prediction, as also mentioned elsewhere [29], and  
18 that is relevant to 2020 where most candidates will not have taken AS-levels.

### 22 Predicted grades in other Key Stage 5 qualifications than A-levels

23 Almost all studies on predicted grades have considered A-levels, with a few occasional exceptions  
24 looking at GCSEs. We know of no studies on the Extended Project Qualification (EPQ) in England, of  
25 Scottish Highers and Advanced Highers, or any other qualifications. The Supplementary Information  
26 to this report includes data on both EPQ and SQA examinations.

### 30 Forecasted grades

31 Until 2015, teachers in the May of Year 13 provided awarding organisations with *forecasted grade*,  
32 and those forecasts in part contributed to quality control of grades by the boards. Since forecasted  
33 grades are produced five to seven months after predicted grades, and closer to the exam date, they  
34 might be expected to be more accurate than predicted grades, being based on better and more  
35 recent information. Forecasted grades are important as they are more similar than predicted grades  
36 to the proposed calculated grades in the way they are calculated, and it is noted that “they may  
37 differ somewhat from the predicted grades sent to UCAS as part of the university application  
38 process” [30]. Three formal analyses are available, for candidates in 2009 [31], 2012 [32] and 2014  
39 [30], and four other studies from 1940 [33], 1963 [3], 1977 [35] and 2018 [28] are available, with  
40 one post-2000 study before A\* grades were introduced and three after (Figure 2). Petch [33] also  
41 provides a very early description of forecasted grades, looking at teachers’ predictions of pass or fail  
42 in School Certificate examinations in 1940, which also show clear over-prediction.

43 Forecasted A-level grades are similar in accuracy to predicted grades pre-2010 (42% vs 52%) but are  
44 more accurate post-2010 (47% vs 17%), in part due to a large drop in accuracy of later predicted  
45 grades. Despite there being *no aspirational or motivational reasons for over-prediction of forecasted*  
46 *grades*, particularly in the 1977 and 2018 studies, over-prediction nevertheless remains as frequent  
47 as in predicted grades (pre-2010: 39%; post-2010: 37%) and remains more common than under-  
48 prediction (pre-2010: 20%; post-2010 16%). Overall it is possible that calculated grades may be  
49 somewhat more accurate than predicted grades, if forecasted grades are taken as a model for  
50 calculated grades. Two set of forecasted grades are also available for GCSEs [35,36], and they show

---

51  
52  
53  
54  
55  
56  
57  
58  
59 <sup>i</sup> Child, S., & Wilson, F. (2015). An investigation of A level teachers’ methods when estimating student grades.  
60 Cambridge Assessment internal report. Cambridge, UK: Cambridge Assessment

1 similar proportion of over and under-prediction as do results for A-levels. Over-prediction seems to  
2 be a feature of all predictions by teachers.  
3

4 The three non-official studies of forecasted grades also asked teachers to rank-order candidates, a  
5 procedure which will also be included in calculated grades. The 1963 data[3] found a median  
6 correlation of rankings and exam marks within schools of 0.78, the 1977 data[35] a correlation of  
7 0.66 [35], and the recent 2018 data [28] a correlation of about .82. The three estimates, mean  $r =$   
8 0.75, are somewhat higher than a meta-analytic estimate of .63 (SE =.03) for teachers' ability to  
9 predict academic achievement [36].  
10  
11

12 The Gill study [28] is also of interest as one teacher commented on the difficulty of providing  
13 rankings with 260 students sitting one exam, and the author commented that, "it was easier for  
14 smaller centres to make predictions because they know individual students better" (p.42), with it  
15 also being the case that responses to the questionnaire were more likely to come from smaller  
16 centres. The 1963 study of Petch[3] , as well as commenting on "considerable divergencies ... in the  
17 methods by which estimates were produced" (p.27), as in the variable emphasis put on mock  
18 exams, also adds that, "some of the comments from schools suggested that at times there may be a  
19 moral ingredient lurking about some of the estimates"(p.28).  
20  
21

22 The studies of forecasted grades suggest that there is some possibility that calculated grades may be  
23 somewhat more accurate than predicted grades, but they also indicate the problems shown by  
24 teachers in trying to rank candidates. It also remains possible that examining boards have far more  
25 extensive and unpublished data on forecasted grades that they intend to use in assessing the likely  
26 effectiveness of calculated grades.  
27  
28

## 29 Applicants to medical school

30 So far, this paper has been entirely about university applicants across all subjects and the entire  
31 range of A-level grades. Only a handful of studies have looked at predicted grades in medical school  
32 applicants.  
33  
34

35 Lumb and Vail pointed out the importance of predicted grades since they determine in large part  
36 how shortlisting takes place [38]. In a study of 1995 applicants they found 52% of predictions were  
37 accurate, 41% were over-estimated and 7% under-estimated [38], values very similar to those  
38 reported in university selection in general (Figure 2).  
39  
40

41 A study by one of the present team used path modelling to assess the causal inter-relationships of  
42 GCSE grades, predicted grades, receipt of an offer, attained A-level grades, and acceptance at  
43 medical school [39]. Predicted grades were related to GCSE grades (beta=0.89), and attained A-level  
44 grades were predicted by both GCSE grades (beta=0.44) and predicted A-level grades (beta=0.74).  
45 The study supports claims that teachers may well be using GCSE grades in part to provide predicted  
46 grades, which is perhaps not unreasonable given the clear correlation.  
47  
48

49 Richardson et al [40] in a very important and seemingly unique study looked at the relative  
50 predictive validity of predicted as compared with attained A-level grades. Using a composite  
51 outcome of pre-clinical performance they found a minimal correlation with predicted grades ( $r=.024$ )  
52 compared with a correlation of 0.318 ( $p<.001$ ) with attained A-level grades. To our knowledge this is  
53 the only study of any sort assessing the predictive validity of predicted vs attained A-level grades.  
54  
55

## 56 The present study

57 Although calculated grades are novel and untested in their details, predicted grades have been  
58 around for half a century, and there is also a literature on forecasted grades. This paper will try to  
59 answer several empirical questions about predicted grades, for which data are now available in  
60

UKMED. Predicted grades will then be used, *faute de mieux*, to make inferences about the likely consequence of using calculated grades.

## Empirical questions to be addressed.

### *The relationship between predicted and attained grades in medical school applicants*

Few previous studies have looked in detail at this high-performing group of students. We will also provide brief results on Scottish Highers and Advanced Highers, and the EPQ (Extended Project Qualification), neither of which has been discussed anywhere else to our knowledge.

### *The predictive validity of predicted grades in comparison with attained grades*

A fundamental question concerning calculated grades is whether teacher estimated grades are better or worse at predicting outcomes than are actual A-level grades. The relationship between predicted grades and actual grades cannot answer that question. Instead what matters is how well predicted and actual grades predict subsequent outcomes at the end of undergraduate. The only relatively small study on this of which we are aware in medical students [40] found that only actual grades had predictive validity.

## Method

The method provided here is brief. A fuller description can be found in the Supplementary Information. Overall the project is **UKMEDP112**, approved by the UKMED Research Group in May 2020, with data coming from two separate but related UKMED projects, both of which include predicted grades.

Project **UKMEDP089**, “The UK Medical Applicant Cohort Study: Applications and Outcomes Study”, approved Dec 7<sup>th</sup>, 2018, with Dr Katherine Woolf as principal investigator, is an ongoing analysis of medical student selection as a part of UKMACS (UK Medical Applicant Cohort Study<sup>j</sup>). The data upload of 21<sup>st</sup> Jan 2020 included detailed information from UCAS and HESA on applicants for medicine from 2007 to 2018.

Project **UKMEDP051**, “A comparison of the properties of BMAT, GAMSAT and UKCAT”, approved Sept 25<sup>th</sup>, 2017, with Dr Paul Tiffin as principal investigator, is an ongoing analysis of the predictive validity of admissions tests and other selection methods such as A-levels and GCSEs in relation to undergraduate and postgraduate attainment. UCAS data are included, although at the time of the present analysis, 16<sup>th</sup> April 2020, this had not yet included the detailed subject level information available in UKMEDP089<sup>k</sup>. Outcome data for the P51 dataset are extensive, and in particular undergraduate progression data are included, such as UKFPO EPM and SJT, and PSA (Prescribing Safety Assessment).

Data from HESA are required to be reported using their rounding and suppression criteria (<https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics>) and those criteria have been used for all UKMED data. In particular the presence of a zero or a zero percentage may not always mean that there are no individuals in a cell of a table, and all integers are rounded to the nearest 5.

---

<sup>j</sup> <https://ukmacs.wordpress.com/>

<sup>k</sup> An upload for P51 dated 20<sup>th</sup> April 2020 is now available but pre-processing is required before the data can be analysed properly.

## Results

As with the Method section, a fuller description of the results can be found in the Supplementary Information.

### The relationships between predicted and actual grades in medical school applicants.

#### Predicted and actual A-level grades for individual A-level examinations

Figure 1 shows the relationship between predicted and attained A-level grades for 237,030 examinations from 2010 to 2018 (i.e. including A\* grades). 39.3% of predicted grades are A\* compared with 23.7% of attained grades. Figure 1.a shows predicted grades in relation to attained grades, with bold font for accurate predictions, green and blue shading for under-prediction and orange and red shading for over-prediction. Overall 48.8% of predicted grades are accurate, which is higher than for university applications in general (see Figure 2), reflecting the high proportion of A and A\* grades (69%). Over-prediction occurred in 44.7% of cases, and under-prediction in 6.5% of cases. Figure 1.b show the data as percentages. About a half of A\* predictions result in an attained A grade, and over a third of predicted A grades result in grade B or lower. Despite over-prediction by about half an A-level grade, predicted and attained grades have a Pearson correlation of  $r = 0.63$ .

#### Differences between A-level subjects

There is little in the literature on the extent to which different A-level subjects may differ in the accuracy of their predictions, perhaps with different degrees of bias or correlation. Detailed results are presented in the Supplementary Information. Overall, Biology, Chemistry, Maths and Physics are very similar in terms of over-prediction and correlation with actual grades. However General Studies is particularly over-estimated compared with other subjects.

#### Extended Project Qualification (EPQ) and SQA Advanced Highers

The Supplementary Information contains information on these qualifications. SQA Advanced Highers, as well as the EPQ, show similar proportions of over-estimation as other qualifications (see Figure 2).

#### Reliability of predicted and attained A-level grades

Considering the best three A-level grades, the reliability of an overall score can be calculated from the correlations of the individual subjects. For 66,006 candidates with at least three paired predicted and actual grades, Cronbach's alpha was 0.827 for actual grades and 0.786 for predicted grades, with a highly significant difference. The difference may in part reflect the higher proportion of A\* grades in predicted than actual grades, but may also reflect greater measurement precision in the marking of actual A-levels.

#### How reliable are attained A-level grades?

Attained A-level grades, like any behavioural measurement are not perfectly reliable, in the sense that if a candidate took a parallel test containing equivalent but different items it is highly unlikely that they would get exactly the same mark as on the first attempt. They may, for instance, have been lucky (or unlucky) at their first attempt, having questions which they happened to have studied

1 or revised more (or revised less), and so on. Reliability is a technical subject<sup>1</sup> with many different  
2 approaches [41,42]. For continuous measures of raw scores, the reliability can be expressed as a  
3 coefficient such as alpha (and in one A-level maths test in 2011, alpha for the full test was about 0.97  
4 [43], although it is suggested that value is unusually high). Boards though do not report raw scores,  
5 but instead award grades on a scale such as A\* to E. The 'classification accuracy' of grades is harder  
6 to estimate, and is greater with fewer grade points, wider grade intervals, and a wide spread of  
7 candidate ability [43]. There seem to be few published estimates of classification accuracy for A-  
8 levels (although they do exist for GCSEs and AS-levels [43]).

11 Estimating classification accuracy for the present high-attaining group of medical school applicants is  
12 not easy. A fundamental limit for any applicants is that predicted grades cannot possibly predict  
13 actual grades better than attained grades predict themselves (the reliability or classification  
14 accuracy). However, from considering the correlation of the three best predicted and actual grades it  
15 is unlikely that such a limit has currently been reached. The correlation of actual with predicted  
16 grades is .585 (see figure 3), and the alpha reliabilities of .827 for actual grades and .786 for  
17 predicted grades (see above). The disattenuated correlation between predicted and actual grades is  
18 therefore  $.585/(\sqrt{.827 \times .786}) = 0.726$ , which is substantially less than one, with predicted grades  
19 accounting for only about a half of the true variance present in actual grades. If the disattenuated  
20 correlation were close to one then it could be argued that predicted grades were doing as well as  
21 they could possibly do given that attained grades are not perfectly reliable, but that is clearly far  
22 from the case.

### 27 True scores and actual scores

28 From a theoretical, psychometric, point of view it could be argued that it is neither actual nor  
29 predicted grades which need to be estimated for applicants, but their 'true ability scores', or the  
30 'latent scores', to use the technical expressions, of which predicted and actual grades are but  
31 imperfect estimates. In an ideal world that would be the case, and a well-constructed exam tries to  
32 get as close as possible to true scores. However, it is not possible to know true scores (and if it were  
33 the boards would provide selectors with those scores). Selection itself does not work on true scores  
34 but on the actual grades that are written down, by teachers for predicted grades, and as grades on  
35 exam result certificates by boards. They are the currency in which transactions are conducted during  
36 selection, so that a predicted grade of less than a certain level means a candidate will not get a  
37 conditional offer, and likewise too low an actual grade means a candidate holding a conditional offer  
38 will be rejected. For that reason it is not strictly the correlation of predicted and actual grades which  
39 matters, the two measures being treated as symmetric, but the forward prediction of actual grades  
40 from predicted grades, i.e. the actual grades conditional on the predicted grades (as shown in figure  
41 1b).

## 47 Predictive validity of predicted and attained A-level grades in medical 48 students.

### 51 Predictive validity in UKMEDP051

52 The version of the P51 data used here consists entirely of applicants applying to medical schools, but  
53 there is also follow-up into undergraduate and postgraduate training. Predicted A-level grades were  
54 available only for the UCAS application cycles of 2010 to 2014, and consisted of a single score in the  
55 range 4 to 36 points, based on the three highest predictions scored as A\*=12, A=10, etc.. The modal  
56

---

58  
59 <sup>1</sup> See <https://www.gov.uk/government/publications/reliability-of-assessment-compedium> for a range of  
60 important papers commissioned and published by Ofqual.



1 score for 38,965 applicants was 30 (equivalent to AAA; mean=31.17; SD= 3.58; Median = 32; 5<sup>th</sup>, 25<sup>th</sup>,  
2 75<sup>th</sup> and 95<sup>th</sup> percentiles= 26, 30, 34 and 36). For simplicity the study was restricted to applicants  
3 aged 18 in the year of application, who had both predicted and attained A-levels, which also ensured  
4 the sample contained only first applications for non-graduate courses, from candidates who had not  
5 taken pre-2010 A-levels. Overall, 22,955 applicants were studied. Other selection measures included  
6 were GCSEs (mean grade for best eight grades), as well as U(K)CAT and BMAT scores, based on the  
7 most recent attempt which for most of these cases was also the first attempt. For simplicity we used  
8 the total of the four sub-scores of U(K)CAT, and the total of Section 1 and 2 scores for BMAT.  
9

10  
11 Follow-up is complicated as application cohorts enter medical school in different years, and spread  
12 out in time through medical school. For this UKMED data extract, applicants in 2010 entered the  
13 medical register from 2015-18, 2011 applicants in 2016-8, 2012 applicants in 2017-18 and 2013  
14 applicants in 2018. Applicants in 2014 would only qualify in 2019 or after, and the UKMED dataset  
15 did not include 2019 data. Some 2010 to 2013 applicants will also qualify after 2018. Again for  
16 simplicity, undergraduate outcome measures were restricted to the deciles of the UKFPO's  
17 Educational Performance Measure (EPM), the raw score of the UKFPO's Situational Judgement Test  
18 (SJT), and the score relative to the pass mark of the Prescribing Safety Assessment (PSA), all at first  
19 attempt. Relatively few doctors, mostly from the earlier cohorts, had progressed through to  
20 postgraduate assessments, but sufficient numbers for analysis were present for MRCP(UK) Part 1  
21 and MRCS Part A, scores being analysed at the first attempt.  
22

23  
24 EPM, is a complicated measure summarising academic progression through medical school, with  
25 individual schools deciding what measures to include [44], and expressed as deciles *within* each  
26 school and graduating cohort year. EPM is here used as the main undergraduate outcome measure.  
27 Deciles are confusing, as UKFPO scores them in the reverse of the usual order, the first decile being  
28 highest performance and the tenth the lowest<sup>m</sup> (see Figure 5). Here for ease of interpretation we  
29 reverse the scoring in what we call *revDecile*, so that higher *revDeciles* indicate higher performance.  
30 It should also be remembered that deciles are not an equal interval scale.  
31

32  
33 Correlations between the measures are summarised in Figure 3. Large differences in Ns reflect some  
34 measures being used in applicants during *selection*, and others being outcome measures that are  
35 only present in *entrants*, as well as the smaller numbers of doctors who had progressed to  
36 postgraduate assessments. The distinction is emphasised by dividing the correlation matrix into  
37 three separate parts. Correlations of selection and outcome measures necessarily show *range*  
38 *restriction* because candidates have been selected on the basis of the selection measures, and  
39 likewise doctors taking postgraduate examinations may be self-selected for earlier examination  
40 performance.  
41

42  
43 Figure 3 contains much of interest (see Supplementary Information), but the most important  
44 question for present purposes is the extent to which Predicted and Attained A-level grades (shown  
45 in pink and green in Figure 3) differ in their prediction of the five outcome measures, remembering  
46 that undergraduate outcomes are typically five or six years after selection, and postgraduate  
47 outcomes are seven or eight years after selection.  
48

49  
50 Attained A-levels predict EPM with  $r=0.297$  compared with a correlation of only 0.198 for predicted  
51 grades. N is large for these correlations and hence the difference, using a test for correlated  
52 correlations [45] is highly significant ( $Z=12.6$ ,  $p<10^{-33}$ ). These raw correlations are inevitably range  
53 restricted, and the construct level predictive validity accounting for range restriction and  
54  
55  
56  
57

58  
59 <sup>m</sup> <https://foundationprogramme.nhs.uk/wp-content/uploads/sites/2/2019/11/UKFP-2020-EPM-Framework-Final-1.pdf>,  
60

1 measurement error is likely to be much higher [46]. Multiple regression (see Supplementary  
 2 Information) suggests that predicted grades may have a small amount of predictive variance which is  
 3 not shared with attained A-levels. Figure 4 shows mean EPM *revDecile* scores in relation to actual  
 4 and predicted A-levels. The slope of the line is clearly less for predicted A-levels, showing a less good  
 5 prediction. It is also clear that attained grades predict well, A\*A\*A\* entrants scoring an average of  
 6 two deciles higher at the end of the course than those with AAA grades, each extra grade raising  
 7 average performance by about two-thirds of a decile. In contrast the slope is less for predicted  
 8 grades, being slightly less than half a decile per predicted A-level grade.  
 9

11 The broad pattern of results is similar for the other undergraduate outcomes, SJT and PSA, and is  
 12 available in the Supplementary Information.  
 13

14 The two postgraduate outcome measures, MRCP(UK) Part 1 (Membership of the Royal Colleges of  
 15 Physicians (UK) examination) Part 1, and MRCS (Membership of the Royal College of Surgeons) Part  
 16 A, although both based on smaller numbers of doctors, are still significant, actual grades correlating  
 17 more highly with MRCP(UK) Part 1 ( $r=.421$ ) than do predicted grades ( $r=.283$ ;  $Z= 4.54$ ,  $p=.000055$ ).  
 18 Likewise, actual grades correlate more highly with MRCS Part A ( $r=.421$ ) than do predicted grades  
 19 ( $r=.358$ ;  $Z= 3.67$ ,  $p=.000238$ ).  
 20  
 21

22 There are suggestions that predicted grades may not be equivalent in candidates from state schools  
 23 and private schools, grades being predicted more accurately in independent schools [24,25]. That is  
 24 looked at in the Supplementary Information, and while there is clear evidence, as found before in  
 25 the UKCAT-12 study [47], that private school entrants underperform relative to expectations based  
 26 on their A-levels, there is no evidence that predicted grades behave differently in candidates from  
 27 private schools.  
 28  
 29

30 A practical question relevant to calculated grades concerns the extent to which, in the absence of  
 31 attained A-level grades, other selection measures such as GCSEs, U(K)CAT and BMAT can replace the  
 32 predictive variance of attained A-level grades. That will be considered for EPM where the sample  
 33 sizes are large. Attained grades alone give  $R = 0.297$ , and predicted grades alone give  $R=.198$ ,  
 34 accounting for less than half as much outcome variance. Adding GCSEs to a regression model  
 35 including just predicted grades increases  $R$  to  $.225$ , and also including U(K)CAT and BMAT increases  $R$   
 36 to  $.231$ , which is though substantially less than the  $.297$  for attained A-levels alone. In the absence of  
 37 attained A-level grades, prediction is therefore improved by including GCSEs and U(K)CAT or BMAT,  
 38 although the prediction still falls short of that for actual A-levels alone.  
 39  
 40  
 41

### 42 Modelling the effect of only predicted grades being available for selection

43 In the context of the 2020 pandemic, an important question is the extent to which future outcomes  
 44 may change as a result of selection being in terms of calculated grades. Calculated grades  
 45 themselves cannot be known, but predicted grades are probably a reasonable surrogate for them in  
 46 the first instance. A modelling exercise was therefore carried out whereby the numbers of students  
 47 in the various EPM *revDeciles* were tabulated in relation to predicted grades at five grade levels, 36  
 48 pts  $\equiv$  A\*A\*A\*, 34 pts  $\equiv$  A\*A\*A, 32 pts  $\equiv$  A\*AA, 30 pts  $\equiv$  AAA and  $\leq 28$  pts  $\equiv$   $\leq$  AAB, the probability of  
 49 each decile found for each predicted A-level band. Assuming that selection results in the usual  
 50 numbers of entrants with grades of A\*A\*A\*, A\*A\*A, etc, but based on calculated grades rather than  
 51 actual grades, the expected numbers of students in the various EPM deciles can be found. Figure 5  
 52 shows deciles as standard UKFPO deciles (1 = highest), UKFPO scores (43 = highest), and *RevDeciles*,  
 53 (10 = highest). The blue column shows the actual proportions in the deciles based on attained A-  
 54 level grades. Note that for various reasons there are not exactly equal proportions in the ten  
 55  
 56  
 57  
 58  
 59  
 60

1 deciles<sup>n</sup>. Based on selection on attained A-level grades there are 7.2% of students in the lowest  
2 performing decile, compared with an expected proportion of 8.1% for selection on predicted grades,  
3 an increase of 0.9% percentage points, which is a relative increase of 13.0% in the proportion of the  
4 lowest decile, with an odds ratio of 1.141 of attaining the lowest decile. For the highest scoring  
5 decile, the proportion decreases from 10.1% with actual A-level grades to 8.8% if predicted A-level  
6 grades are used, an absolute decrease of 1.4%, and a relative decrease of 13.4% of top deciles, with  
7 an odds ratio of 0.853.  
8  
9

10 Of course the above calculations are based on the assumption that the 'deciles' for calculated grades  
11 are expressed at the same standard as currently. Were the outcomes to be re-standardised so that  
12 all deciles were equally represented then of course at finals no noticeable difference would be  
13 present<sup>o</sup>. However the academic backbone would still be present, and overall poorer performance  
14 on statistically equated postgraduate exams would be expected.  
15  
16

## 17 Discussion

18 The present data make clear that only slightly under half of predicted grades are accurate, with 45%  
19 being higher than attained grades, and only 17% being lower. The data also show that attained  
20 grades are far better predictors of medical school performance than are predicted grades, which  
21 account for only about a third as much outcome variance as attained grades. Attained grades are  
22 also more reliable than predicted grades.  
23  
24

25 Validation is the bottom line of all educational assessment, and in the present case it is validation  
26 against assessment five to eight years down the line from the original A-levels, in both  
27 undergraduate and postgraduate assessments. That is strong support for what we have called 'the  
28 academic backbone', prior attainment providing the underpinning for later attainment, and hence  
29 performance at assessments at all stages of training for GCSEs through to medical degrees (and  
30 probably on to postgraduate assessments) tend to be correlated [19]. The findings contradict  
31 suggestions that holistic judgments by teachers of predicted grades are better predictors of  
32 outcomes since teachers may know their students better than examiners. The immense efforts by  
33 exam boards and large numbers of trained markers to refine educational measurements is therefore  
34 gratifying and reassuring. Careful measurement does matter.  
35  
36

37 A remaining question is whether there is still some variance in predicted and actual grades which is  
38 complementary. A tiny amount, is the answer, predicted grades increasing R by about 0.05 when  
39 actual grades are already taken into account. What that variance might be is unclear, and it is  
40 possible that it is what Petch called 'scholarly attitude'. At present though it is worth remembering  
41 that *examination* grades at A-level are primarily predicting further *examination* grades at the end of  
42 medical school, although EPM scores do include some measures of course work, and practical and  
43 clinical skills. If other outcome measures, perhaps to do with communication, caring or other non-  
44 cognitive skills were available then predicted grades might show a greater predictive value.  
45  
46

47 The present data inevitably have some limitations, but they are few. There is little likelihood of bias  
48 since complete population samples have been considered, and there is good statistical power with  
49 large sample sizes. Inevitably not all outcomes can be considered, mainly because the cohorts  
50  
51

---

52 <sup>n</sup> In part this reflects the fact that some students, particularly weak ones, are given an EPM score, but then fail  
53 finals.

54 <sup>o</sup> This is based on deciles being calculated in a way that are equated to levels used in the present calculation.  
55 Of course if calculated strictly as deciles then of necessity 10% will still remain in the top decile, etc.. That  
56 difficulty of deciles will in large part be removed when the 2020 entrants graduate as UKMLA should be on  
57 stream by then.  
58  
59  
60

1 analysed have not yet progressed sufficiently through postgraduate training. However those who are  
2 included show effects which are highly significant statistically.  
3

4 Our questions about predicted grades have been asked in the practical context of the cancellation of  
5 A-level assessments and their replacement by calculated grades, as a result of the COVID-19  
6 pandemic. It seems reasonable to assume, given the literature on predicted grades, and particularly  
7 on forecasted grades, that calculated grades will probably have similar predictive ability to predicted  
8 grades, but perhaps be a little more effective due to occurring later in the academic cycle. Such a  
9 conclusion would be on firmer ground if exam boards had analysed the predictive validity of the data  
10 they had collected on forecasted grades, particularly in comparison with predicted and actual  
11 grades. Such data may exist, and if so then they need to be seen. In their absence, the present data  
12 may be the best available guesstimates of the likely predictive validity of calculated rather than  
13 actual grades.  
14  
15  
16

17 A further consideration is more general and asks what the broader effects of the COVID-19  
18 pandemic may be on education. Students at all levels of education have had teaching and learning  
19 disrupted, often extensively, and that is also true of all stages of medical education. The current  
20 cohort of applicants will not be assessed formally at A-level. As well as meaning that they may only  
21 have calculated grades, which are likely to be less accurate, they also will have missed out on several  
22 other things. Year 13 students may have missed normal schooling from, say, mid-March until mid-  
23 May, when they should have taken public examinations, perhaps 30 to 40 school days. Year 12  
24 students, from whom 2021 medical school entrants will be drawn, may miss school from mid-March  
25 until mid-July, totalling perhaps 80 school days. Burgess and Sievertsen [48], using data from two  
26 studies [50,51], estimate that 60 lost school days results in a reduction in performance of about 6%  
27 of a standard deviation, which they say is, “non-trivial”. These effects are likely to differ also by  
28 socio-economic background, particularly in effectiveness of home schooling. Applicants not taking A-  
29 levels will also suffer from the loss of the enhanced learning that occurs when learners are tested –  
30 the ‘testing effect’ – for which a meta-analysis found an effect size of 0.50 [51], which is also non-  
31 trivial. Taken overall, 2020 entrants to medical school, and perhaps those in 2021 as well, may  
32 perform less well as a result of missing out both on education and on its assessment.  
33  
34  
35  
36

37 What then are the specific implications of our findings for the use of calculated grades, for medical  
38 education and the selection of medical students? Whether our findings will generalise to other  
39 university courses is far from clear, not least because very few have the sophisticated undergraduate  
40 and postgraduate outcome measures found in medical education, which inevitably will make effects  
41 harder to demonstrate, although in principle the effects should still be there.  
42  
43

#### 44 For universities and medical schools.

45 Medical schools are skilled in understanding the various data they hold on applicants, and how  
46 results ought to look given prior attainment, and they recognise that some sets of results might be  
47 skewed, sometimes because of small group sizes. They also recognise that predicted grades often err  
48 on the side of generosity. No doubt they will be looking at calculated grades to check whether they  
49 are compatible with other data. Nevertheless, calculated grades are largely uncharted territory even  
50 for experienced university admissions officers, with many potential uncertainties, particularly in  
51 relation to predicting future medical school performance.  
52  
53  
54

55 A useful thought-experiment is to imagine what would happen were a computer-error accidentally  
56 to shuffle all A-level grades, but with the same numbers of A\*, A, B grades, etc still being awarded  
57 for individual subjects, but without any relation to the grades that candidates should have attained.  
58 Administratively, selection would proceed as normal. However instead of medical schools admitting  
59 high-achieving entrants with A\* and A grades, entrants would have what we might call nominal ‘A\*’  
60

1 and 'A' grades, but in reality would have had true grades anywhere from A\* to E, with a mode at  
2 about B. Since A-levels undoubtedly predict university outcomes, many entrants with lower true  
3 grades would therefore probably struggle with course content and exams. For medical schools using  
4 criterion-referencing, more students than usual would fail exams. If norm-referencing were used,  
5 the usual proportion of students would pass but pass marks would be lower, students would know  
6 less at graduation, and the knowledge deficit would then become clear when taking statistically  
7 equated postgraduate examinations such as MRCP(UK).  
8  
9

10 Attained grades within candidates in different A-level subjects normally tend to correlate, so that the  
11 reliability of a total points score can be calculated (see above). Grades allocated at random would  
12 not however correlate within candidates. Were a medical school to select on total A-level points  
13 (say, with a threshold equivalent to AAA) there would still be applicants with that number of points.  
14 However because random allocation results in zero correlations of grades within candidates, the  
15 variance of total points would be less and candidates with high point scores would be in shorter  
16 supply.  
17  
18

19 Although only a thought-experiment, calculated grades will probably have a partial similarity to  
20 randomly allocated grades (and the reliability of predicted grades is indeed less than that of attained  
21 grades, because correlations are lower) in that measurement error will have increased. Different  
22 subjects will correlate less than usual, and there may be a shortage of qualified candidates, and  
23 nominal grades will correlate less with outcomes than usual, regression to the mean making  
24 outcomes somewhat poorer than usual. The predictions therefore are qualitatively similar to the  
25 pattern of those with random allocation, but less extreme.  
26  
27  
28

29 A practical question is whether medical schools can do anything to ameliorate possible problems  
30 with calculated grades. For those holding conditional offers, a strong argument says there is a  
31 contractual agreement between a university and a candidate holding a conditional offer based on  
32 certain A-level grades (although a counter-argument might be that the contract implicitly said that  
33 payment would be in actual, gold-standard, attained grades, and not in the potentially baser  
34 currency of promissory notes based on calculated grades). If indeed there were to be too many  
35 candidates meeting their offers, then the precise nature of the contract could become crucial. For  
36 candidates not meeting the grades stated in conditional offers, so that there are unfilled places,  
37 there is no obvious contractual agreement, and medical schools may wish then to take into account  
38 GCSEs, U(K)CAT and BMAT scores, and perhaps other measures such as EPQ and MMI, although the  
39 latter two have not yet been validated as predictors of outcome. The use of some sorts of  
40 information, such as personal statements and work experience, may however favour more  
41 advantaged students [4], although medical schools may wish also to take educational disadvantage  
42 into account to widen participation and because applicants who do relatively better than their peers  
43 at poorly-performing schools tend to achieve higher grades in medical school than those with  
44 equivalent grades from high-performing schools [52]. Ultimately, universities will need to be,  
45 "flexible and responsive in their admissions processes" [53].  
46  
47  
48  
49

50 Whatever medical schools do choose to do (or not to do) the one certainty is that outcomes for the  
51 2020 entry cohort will be followed with interest through their careers, and UKMED will be a major  
52 tool in doing that.<sup>p</sup>  
53  
54  
55  
56

---

57  
58 <sup>p</sup> In the event, offers were honoured and the cap on medical school numbers was lifted to accommodate the  
59 extra students who met their conditional offers when the calculated grades algorithm was scrapped in August  
60 2020.

## For teachers

Centre assessment grades and rankings may well be problematic for teachers (and recent research suggests that rankings are hard to make for large numbers of candidates [28], and that different teachers use different sorts of information in making their judgements). Since the intention of calculated grades is that they will emulate the actual grades that students would have attained had they taken the exams, then a key emphasis probably has to be on what Petch called “anticipated examination prowess” rather than “scholarly attitude”. How easy it is to do that though is very unclear (and a way to help may be to ask teachers to make both assessments on students, so that differences become clearer). Many teachers will already have provided predicted grades for UCAS applications earlier in the academic year, and whether it is entirely possible to forget those earlier estimates seems unlikely, particularly when much but not all of the information will be the same as earlier. The possibility of parental challenge when centre assessment grades differ from earlier predicted grades may also make the two sets of grades more similar. Cognitive dissonance may additionally make it particularly difficult to forget earlier predictions, although it might be worth remembering the saying often attributed to John Maynard Keynes: “When events change, I change my mind. What do you do?”<sup>9</sup>. Finally, from a research point of view it would be nice to think that information will be collected about how teachers and schools in practice are producing their centre assessment grades, although it seems possible that may not occur due to fears of legal challenge.

## For medical school applicants

Cancellation of A-levels inevitably has produced uncertainty and anxiety for university applicants. Calculated grades have introduced what is to some extent an entirely unknown factor, although on balance it is likely that calculated grades will probably relate closely to predicted grades.

If calculated grades are similar to predicted grades, and given that much selection has in fact already taken place during shortlisting largely, but not entirely, on the basis of predicted grades (and other information such as admissions tests and interviews), then selection this year will mainly be on predicted grades. A particular problem normally is applicants for whom there is under-prediction of grades, as calculated grades are likely also to be under-predicted, but without any actual grades to assess whether that is the case. For those whose grades in normal circumstance are under-predicted, being predicted, say, BBB, but achieving A\*AA, there is usually a safety net for a second chance at a place, formally through Clearing or sometimes through other more informal processes such as in medical schools which maintain waiting lists. Quite how, beyond random chance, calculated grades might differ substantially from predicted grades, is unclear. That also means it is difficult to see how Clearing and similar processes will work in 2020. Under-prediction is a particular risk in cases where teachers do not know, or in some cases do not like, their students, perhaps because of attitude, personal characteristics, or other factors, the externality and objectivity of examinations traditionally solving the problem. Petch, once again, put it well, describing,

“instances, where, in the examination room, candidates have convinced the examiners that they are capable of more than their schools said that they were ... Paradoxical as it will seem, examiners are not always on the side of authority; an able rebel can find his wider scope within the so-called cramping confines of an examination.” [3](p.29).

There is a clear echo here of the quote by Yasmin Hussein with which this paper began. Hussein’s concerns are not alone, and the UKMACS study in April 2020 found concerns about fairness were particularly present in medical school applicants from non-selective schools, from Black Asian and

<sup>9</sup> The phrase was probably said in 1970 by Paul Samuelson, a Nobel-prize winning economist, although in 1978 he himself attributed it to Keynes. <https://quoteinvestigator.com/2011/07/22/keynes-change-mind/>

1 Minority Ethnic (BAME) applicants, from female applicants, and from those living in more deprived  
2 areas [6].  
3

4 The question of whether predicted grades are more accurate in private sector schools than public  
5 sector, state schools is complex and is discussed in an appendix to the Supplementary Information.  
6 There are subtle ceiling effects and non-linearities, but overall the best conclusion seems to be that  
7 private sector predictions may seem to be more accurate but that mainly reflects the higher actual  
8 grades of students at selective schools – if a student is predicted an A\* they cannot over-achieve and  
9 private school pupils tend to be less likely to have low grades (not least because they have been  
10 selected to have higher than average ability levels) meaning that teachers predicting high grades are  
11 more likely to be correct. That conclusion is also important as it means that although there is an  
12 argument that calculated grades may predict outcomes better than predicted grades because they  
13 will have been standardised, predicted grades in effect are already standardised in so far as teachers  
14 at selective schools provide higher predicted grades because those candidates in actually do gain  
15 higher attained A-level grades. Standardisation by previous school performance is unlikely to change  
16 that conclusion.  
17  
18  
19  
20

## 21 Conclusions

22 The events of 2020 as a result of the COVID-19 pandemic are extra-ordinary, and unprecedented  
23 situations have occurred, of which the cancellations of GCSE and A-level exam cancellations are but  
24 one example. The current study should not be seen as criticism of the response of Ofqual to that  
25 situation; given the circumstances in which it found itself, with examinations cancelled (when the  
26 Chair of Ofqual Roger Taylor had recommended socially-distanced or delayed exams), Ofqual's  
27 solution to the problems has many obvious virtues.  
28  
29  
30

31 For most university applicants there already existed predicted grades from the previous autumn  
32 when UCAS applications were submitted, but they would have been on average half a grade or so  
33 too high, being aspirational as much as realistic, and also for medical students would have been  
34 made by October 2019, whereas calculated grades would be based on teacher predictions in May  
35 2020, albeit with several months of courses missing since March 2020.  
36  
37

38 Raw predicted grades would have wrecked much university planning as numbers of acceptances  
39 would inevitably have been far too high. There was also a risk that predicted grades could have been  
40 systematically higher from some schools than others – the ones with a tendency to call all of their  
41 geese swans -- and that probably applies also to the centre assessment grades sent to examination  
42 boards and mostly eventually accepted without central standardisation in August 2020.  
43  
44

45 This paper has provided evidence that the grades awarded to medical applicants in summer 2020  
46 will probably not predict future outcomes with the same effectiveness as actual, attained grades,  
47 and that is a problem that universities and medical schools and postgraduate deaneries will have to  
48 work with, probably for many years as the 2020 cohort works through the system, some eventually  
49 being part of the first cohort to take the new national Medical Licensing Examination in 2024/2025  
50 (<https://www.gmc-uk.org/education/medical-licensing-assessment/uk-students-guide-to-the-mla>).  
51 It seems likely therefore, as Thomson has said, "... this year group will always be different..."[2].  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Postscript 25<sup>th</sup> November 2020

### Introduction

The research for the present paper was carried out in April and May 2020, in parallel with the study of attitudes and responses of medical school applicants to the cancellation of A-level examinations and their replacement by ‘Calculated Grades’<sup>r</sup>. These Calculated Grades consisted of teacher-estimated Centre Assessment Grades (CAGs) which were submitted to *Ofqual* and moderated using an algorithm. Calculated grades were released on 13<sup>th</sup> August, but then fairly rapidly the algorithm was abandoned and grades reverted to the unmoderated, unadjusted, teacher-estimated CAGs (or the Calculated Grade, if higher). This process, and the fallout from it, is described below.

Key findings from this paper and the accompanying applicant attitudes paper were presented to UK medical admissions tutors at a meeting of the MSC-SA (Medical Schools Council Selection Alliance) on 6<sup>th</sup> May 2020, and drafts of the two papers distributed. Both papers were finalized in May 2020 and uploaded to *medRxiv* on June 5<sup>th</sup> 2020. Little further happened over the next few weeks, and the present paper is therefore a statement of how we understood the situation at that time, with a few amendments for clarity.

In July and August 2020 however, things moved rapidly, with dramatic changes taking place. It would have been extremely confusing and probably misleading to have tried to incorporate those changes into the text of the main paper. Instead we hope this postscript will give readers a sense of what happened, to what extent events were correctly or incorrectly predicted by us, what impact the present paper may have had, and what may be the implications for the future.

The story is best told chronologically and we mostly use reports from newspapers, and refer interested readers to a brief summary on Wikipedia<sup>s</sup> and a journalistic review on the BBC website<sup>t</sup>.

### The awarding of A-level grades in 2020: the story from March to November 2020

As described in our main paper, on 20<sup>th</sup> March 2020 public examinations in the UK including A-levels were cancelled. On 3<sup>rd</sup> April *Ofqual* announced that exam grades in England would be replaced with Calculated Grades. Calculated Grades were to consist of Centre Assessment Grades (also called Centre Assessed Grades or CAGs), estimated by teachers that centres (mostly schools and colleges) would submit to *Ofqual*. *Ofqual* would moderate these CAGs using an algorithm – the details of which had not yet been published but, it was stated, would be based on the prior performance of pupils within schools. The Scottish Qualification Authority (SQA), Qualifications Wales, and the Northern Irish Council for Curriculum, Examination and Assessment (CCEA) also announced that they would use a broadly similar approach to that of *Ofqual*.

Schools (centres) had to return their teacher-estimated CAGs to *Ofqual* in June 2020. On June 16<sup>th</sup> a report in *The Times* said that “Teachers have marked too generously in allocating GCSE and A-level grades this year, research suggests” (*The Times*, 16<sup>th</sup> June), the article being based on a report from

---

<sup>r</sup> Woolf, K., Harrison, D., McManus, I.C. “The attitudes, perceptions and experiences of medical school applicants following the closure of schools and cancellation of public examinations in 2020 due to the COVID-19 pandemic”, *medRxiv*, <https://doi.org/10.1101/2020.06.02.20116855>

<sup>s</sup> [https://en.wikipedia.org/wiki/2020\\_UK\\_GCSE\\_and\\_A-Level\\_grading\\_controversy](https://en.wikipedia.org/wiki/2020_UK_GCSE_and_A-Level_grading_controversy)

<sup>t</sup> Coughlan, S. “Coronavirus: The story of the big U-turn of the summer”, <https://www.bbc.co.uk/news/education-54103612>



1 FFT Education Datalab, which actually had only asked about GCSEs, and had no data on A-levels<sup>u</sup>. In  
2 July a Guardian article reported a statement from *Ofqual* that “a substantial number of students  
3 would receive at least one adjusted grade – usually downwards – as a result of a standardisation  
4 process” although they “sought to allay fears that certain groups of pupils, ... could be disadvantaged  
5 by calculated grades. *Ofqual* said their analysis had found no evidence of widening of gaps in  
6 attainment”. (*The Guardian (G)*, 21<sup>st</sup> July).  
7  
8

9 SQA results in Scotland are announced a week before those in England, so the SQA results  
10 announced on August 5<sup>th</sup> 2020 gave a preview of what was to come the following week in the rest of  
11 the UK. The Scottish results were immediately controversial when it emerged that the moderation of  
12 teacher-estimated grades (CAGs) by an algorithm had resulted in a quarter of grades being adjusted  
13 downwards. The Scottish Education Secretary, John Swinney, said that without those adjustments  
14 the pass rates would be up on the previous year by 14% for Highers and 13.4% for Advanced  
15 Highers. He added that, “... these robust processes mean we have upheld standards... All exam  
16 systems rely on an essential process known as moderation to uphold standards. This ensures an A  
17 grade is the same in every part of the country, making the system fair for everyone, and across all  
18 years.” (*G*, 4<sup>th</sup> August).  
19  
20  
21

22 Teachers, students, parents and the media were unhappy with the moderation. A *Daily Telegraph*  
23 editorial entitled “Exam moderation is a gross injustice” attacked statistical modelling in general, and  
24 the SQA process in particular, which “gives poorer marks to children living in deprived areas ...  
25 [without] recognition of individuals who buck the general trend” (*Daily Telegraph (DT)*, 5<sup>th</sup> Aug), and  
26 asked “Is the same fiasco about to be inflicted on A-level students in England and Wales?”. By 11<sup>th</sup>  
27 August, students in Scotland were protesting on the streets, Nicola Sturgeon, the First Minister, was  
28 apologising for the exam results debacle, the Scottish government was facing a vote of no  
29 confidence (*G*, 11<sup>th</sup> Aug). Expectations in the media of problems with A-levels were also growing. On  
30 12<sup>th</sup> August it was announced in Scotland that the teacher-estimated CAGs downgraded by the SQA  
31 algorithm during moderation “would be reinstated” (*G*, 12<sup>th</sup> Aug).  
32  
33  
34

35 In an attempt to prevent problems with A-levels, on August 12<sup>th</sup> 2020 the English education  
36 secretary, Gavin Williamson announced “a triple lock” for A-level students, whereby students could  
37 accept their Calculated Grade results, use the results of mock exams (practice exams which students  
38 take in schools), or use the results of real exams due to take place in Autumn 2020 after the start of  
39 the university academic year (*The Times, (T)*, 12<sup>th</sup> August; *T*, 13<sup>th</sup> August). Protests were immediate  
40 as mock exams vary immensely, and many schools had been encouraged to cancel mock exams as a  
41 part of the Covid lockdown in March 2020.  
42  
43

44 A-level Calculated Grades (i.e. the teacher-estimated CAGs adjusted by the algorithm during  
45 moderation) were announced on 13<sup>th</sup> August 2020, and UCAS announced which students had  
46 obtained places at their chosen university based on these Calculated Grades. An immediate problem  
47 arose: following the Scottish government’s reversal students in Scotland now had unadjusted SQA  
48 grades, which were higher on average and gave them an advantage over applicants with A-level  
49 Calculated Grades, which had been adjusted (*G*, 13<sup>th</sup> Aug). University admission processes were also  
50 becoming embroiled in confusion, and although universities had, “reassured ministers that they will  
51 ‘soften’ the grades they normally require” (*T*, 13<sup>th</sup> August), by the next day universities were accused  
52 of being inflexible (*G*, 14<sup>th</sup> Aug).  
53  
54  
55  
56  
57

58  
59 <sup>u</sup> <https://ffteducationdatalab.org.uk/2020/06/gcse-results-2020-a-look-at-the-grades-proposed-by-schools/>  
60

1 It soon became apparent that schools in the private, fee-paying sector, had probably benefited from  
2 the algorithm, primarily because statistical predictions were less accurate for the small class sizes  
3 more likely to be found in private schools, and in those cases the teacher-estimated CAGs had been  
4 allowed to stand unadjusted. Although the Prime Minister Boris Johnson defended the system  
5 saying, "Let's be in no doubt about it: the exam results that we've got today are robust, they're  
6 good, they're dependable for employers" (G, 14<sup>th</sup> August), many backbench MPs were in revolt,  
7 having been deluged with complaints from constituents (T, 14<sup>th</sup> August). The next day Gavin  
8 Williamson said, "No U-turn. No change" (T, 15<sup>th</sup> August), and although he did agree to waive fees  
9 for appeals against Calculated Grades, he insisted that the grades themselves would not change in  
10 order to avoid the grade inflation that had occurred in Scotland (T, 15<sup>th</sup> August).  
11  
12  
13

14 Meanwhile the *Ofqual* algorithm, published in a document over 300 pages in length, was being  
15 dissected carefully, and when one headteacher anonymously shared their school's results, the  
16 problems became particularly apparent<sup>v</sup>. In the previous three years at their school, 12.5% of pupils  
17 had achieved A\* and none had got a U; however the algorithm meant only 3.7% of students  
18 (equivalent to just one student) received an A\* - much below the historic 12.5%. The algorithm also  
19 resulted in one student being awarded a U, despite no students at that school having received a U  
20 previously. The weekend newspapers attacked the government, "which deserved a U grade for this  
21 debacle" (Sunday Times (ST), 16<sup>th</sup> August). GCSE results, due on Aug 20<sup>th</sup>, were also on the horizon,  
22 with similar problems predicted (T, 15<sup>th</sup> Aug, p.14; *Observer* (O), 16<sup>th</sup> August). Students in England  
23 demonstrated outside the Department for Education in London. *Ofqual* also announced guidance on  
24 the role of mock exams in appeals only to withdraw it a few hours later (T, 17<sup>th</sup> August).  
25  
26  
27

28 On August 18<sup>th</sup> the government scrapped *Ofqual's* algorithm and reverted to unadjusted teacher-  
29 estimated CAGs (G, 18<sup>th</sup> August). *The Times*, notionally a supporter of the Conservative Party, simply  
30 called its main editorial, "Another Fine Mess" (T, 18<sup>th</sup> August). The chairman and chief executive of  
31 *Ofqual* were criticised for having little experience of education (T, 19<sup>th</sup> August), and the Chief  
32 Executive eventually resigned on August 25<sup>th</sup>.  
33  
34

35 Several other problems now emerged, not the least being that universities would not know the (now  
36 unadjusted CAG) grades for several days. Once universities did receive these grades, they found that  
37 they did not have enough places to honour all the offers they had made students months earlier.  
38 This was because universities typically make more offers than they have places, knowing that a  
39 significant number of students will not meet those offers when they achieve lower exam grades than  
40 the Predicted Grades estimated by their teachers and submitted to UCAS when they apply to  
41 university. But now with the teacher-estimated CAGs replacing exam grades, and the resulting  
42 increase in the percentage of A and A\* grades from 28% to 38%, many more students than expected  
43 did in fact meet their university offers (T, 18<sup>th</sup> August). This caused problems for some students  
44 whose Calculated Grades (the grades adjusted by the algorithm) had been too low for their first  
45 choice university and so had accepted offers from their second choice, but now they found their  
46 unadjusted teacher-estimated CAGs enabled them to meet their first choice offer they inevitably  
47 wanted to go there. Other students seemed to have been left in limbo, needing to delay entry until  
48 the next year, with a potential knock-on effect for students taking A-levels in 2021 (T, 19<sup>th</sup> August).  
49 Some groups of A-level students had clearly fallen through the net and were in limbo, such as who  
50 were home-schooled or private A-level students, having no teachers to estimate their grades (G, 21<sup>st</sup>  
51 August).  
52  
53  
54  
55

---

56  
57 <sup>v</sup> Hern,Alex. (2020) "Do the maths: analysis shows why England's grading system is both imprecise and unfair", *Guardian*,  
58 15<sup>th</sup> August, 2020, p.13; the analyses are based on Thomson, Dave, "A-Level results 2020: How have grades been  
59 calculated?", 13<sup>th</sup> August 2020, <https://ffteducationdatalab.org.uk/2020/08/a-level-results-2020-how-have-grades-been-calculated/>.  
60

1 Regarding medical schools specifically, *The Guardian* reported that the Norwich Medical School at  
2 the University of East Anglia had 185 places and a possible overshoot of 50, a 27% increase,  
3 emphasising that with medical school numbers more strictly limited than other university places and  
4 costing £50,000 a year this would have clear financial implications (G, 19<sup>th</sup> August). A news story  
5 based in part on the current research as published in pre-print on *medRxiv*, suggested that more  
6 medical students may be liable to drop out, a co-chairman of the Medical Schools Council  
7 suggesting that “we are going to have, on average, students with lower grades than in previous  
8 years” (DT, 20<sup>th</sup> August). A similar concern was also raised by headteachers in relation to GCSE  
9 grades, which it had now been announced would also be based on unadjusted teacher estimates,  
10 with fears that students, “could end up on unsuitable courses post-16 which could set them up for  
11 failure” (G, 21<sup>st</sup> August).  
12  
13  
14

15 The cap on university student numbers was released on 17<sup>th</sup> August, the medical school cap  
16 following on 20<sup>th</sup> August, resulting in a bulge of new undergraduates. There was also a bulge in  
17 admissions to sixth form colleges as a result of unadjusted teacher-estimated GCSE grades being  
18 higher than the expected exam grades (G, 21<sup>st</sup> August). There are financial implications for  
19 educational institutions but until UCAS releases its entry statistics for October 2020, the size of the  
20 impacts will be unclear.  
21  
22

23 The controversies rumbled on into September 2020 as the House of Commons Select Committee on  
24 Education heard evidence from *Ofqual* and other bodies. *Ofqual* put out a lengthy statement on 2<sup>nd</sup>  
25 September in evidence to the Committee<sup>w</sup>. It set out the history according to the regulator, stating  
26 that in March its advice to the Secretary of State had been that, “the best option in terms of valid  
27 qualifications would be to hold exams in a socially distanced manner”; however, “The decision to use  
28 a system of statistical standardised teacher assessments was taken by the Secretary of State and  
29 issued as a direction to *Ofqual*”. In reviewing the failure of the system, the conclusion was reached  
30 that, “a ‘better’ algorithm would not have made the outcomes significantly more acceptable. The  
31 inherent limitations of the data and the nature of the process were what made it unacceptable.... *To*  
32 *try to deliver comparable qualification results in the absence of students having taken any*  
33 *assessments (examinations) proved to be an impossible task*” (our emphasis). Cambridge  
34 Assessment’s submission to Select Committee provides a detailed timeline of collaborative efforts to  
35 inform decision making.<sup>x</sup> With the model running and results being calculated, from late July through  
36 August, Cambridge Assessment worked with *Ofqual* and DfE to understand possible unfairness in the  
37 outcomes, and to put in place adequate remedy. No doubt the post-mortem will continue for a long  
38 while.  
39  
40  
41  
42

43 In autumn 2020, the needs of the next year began to be considered. One group, including several  
44 university Vice-Chancellors, argued that A-levels should be cancelled once more and replaced by  
45 teacher-estimated grades (T, 2<sup>nd</sup> October). Others argued more radically that teacher-estimated  
46 grades should permanently replace A-level examinations. On 12<sup>th</sup> October 2020 the Secretary of  
47 State announced that A-level exams would go ahead in England in summer 2021 with some minor  
48 changes, including being three weeks later than usual, and with results announced a few days later  
49 than typical<sup>y</sup>. Contingency measures would be in place for possible disruption, but were yet to be  
50 described, although a leaked newspaper report suggested that they might include more formal mock  
51 exams as a back-up when following earlier leaked reports (G, 10<sup>th</sup> October). Perhaps the most  
52  
53  
54

55 <sup>w</sup> <https://www.gov.uk/government/news/written-statement-from-chair-of-ofqual-to-the-education-select-committee>,  
56 “Written statement from Chair of *Ofqual* to the Education Select Committee”, 2<sup>nd</sup> Sept 2020.

57 <sup>x</sup> <https://committees.parliament.uk/writtenevidence/11358/default/>, “Written evidence submitted by  
58 Cambridge Assessment”, 2<sup>nd</sup> Sept 2020.

59 <sup>y</sup> <https://www.gov.uk/government/news/students-to-be-given-more-time-to-prepare-for-2021-exams>, “Students to be  
60 given more time to prepare for 2021 exams”, 12<sup>th</sup> October 2020.

1 interesting comment by Williamson was that, “Exams are the fairest way of judging a student’s  
2 performance ...”, with its tacit acceptance that teacher-estimated grades perhaps actually were not  
3 fair to many students.  
4

5 To date other UK countries have taken different approaches. On 7<sup>th</sup> October the Scottish  
6 Government announced the cancellation of National 5 exams in 2021 although Higher and Advanced  
7 Higher examinations would be taken (BBC, 7<sup>th</sup> October). A month later on 10<sup>th</sup> November the Welsh  
8 Government announced that GCSE and A-level examinations in Wales would be cancelled, with  
9 grades being based on classroom assessments instead (BBC, 10<sup>th</sup> November). On the same day  
10 Education Minister Peter Weir announced that GCSE and A-levels examinations would be taken in  
11 Northern Ireland (BBC, 10<sup>th</sup> November). In November 2020 the Secretary of State announced his  
12 intention to consider post-qualification university admissions  
13 (<http://researchbriefings.files.parliament.uk/documents/CBP-8538/CBP-8538.pdf>).  
14

15 This postscript has been relatively brief, given the complexity of the events, but may help those new  
16 to the issues to navigate through the major changes that occurred. In research terms, in medical  
17 education, higher education more generally, and in secondary education, there seems little doubt  
18 that researchers will be following in detail the outcomes for the cohorts affected by the dramatic  
19 changes which resulted in a giant, unplanned, experiment, where notional grades awarded were  
20 probably different in many cases to what students would have been awarded in normal  
21 circumstances.  
22

23 We began our main paper by quoting a letter to a newspaper at the beginning of lockdown by a  
24 student taking GCSEs, and so it is probably appropriate to finish with a letter to a different  
25 newspaper by an A-level student, written at the height of the A-levels crisis, which raises many  
26 subtle, important and mostly neglected questions, one which researchers will need to grapple with:  
27

28 “Ofqual’s grading system appears to be lacking in advocates. Blinded by rhetoric about what  
29 protesters call a ‘classist’ algorithm, key facts have been overlooked. It is very clear that  
30 teachers are shockingly bad at predicting grades; using teacher predictions there will be a  
31 12% inflation in higher grades compared with last year. While some centres predicted  
32 accurately, some centres predicted only the highest grades for their students. This U-turn  
33 from the government entails a huge injustice for the pupils who had fair and accurate  
34 predictions, as well as for those taking exams next year. In the zero-sum game of university  
35 applications, the results of these pupils make them appear weaker than they are.  
36 Irresponsible teachers who over-predicted their pupils’ results ought to be ashamed that  
37 they too have thereby ‘dashed the dreams’ of many young people across the country. That it  
38 is less obvious does not make it any less true.” (Letter to *The Times*, 19<sup>th</sup> August 2020, by Seb  
39 Bird, A-level student, Bristol).  
40  
41  
42  
43  
44  
45  
46  
47

## 48 Acknowledgements

49 We are grateful to Paul Garrud, Gill Wyness, Paul Newton, Colin Melville and Christian Woodward  
50 for their comments on earlier versions of this manuscript, and to Jon Dowell, Peter Tang, Rachel  
51 Greatrix and other members of the UKMED Research Group and Advisory Board for their assistance  
52 in fast-tracking the paper, and for their comments on it. We also thank Tim Gill for providing us with  
53 an unpublished manuscript.  
54  
55

## 56 Contributors

57 DTS prepared the data extracts, provided details on data sources and variable definitions where  
58 required and commented on manuscript drafts. ICM originated the idea for the study, and discussed  
59  
60

1 it with other authors throughout the project. ICM wrote the first draft of the manuscript, and KW,  
2 DH, PAT, LP, KYFC and DTS have read, reviewed and commented on earlier drafts and contributed  
3 ideas, as well as approving the final draft.  
4

## 5 6 Funding

7  
8 KW is a National Institute for Health Research (NIHR) Career Development Fellow (NIHR CDF-2017-  
9 10-008) and is principal investigator for the UKMACS and UKMEDP089 projects supported by the  
10 NIHR funding.  
11

12 DH is funded by NIHR grant CDF-2017-10-008 to KW.

13  
14 PAT's research time is supported by an NIHR Career Development Fellowship (CDF 2015-08-11), and  
15 PAT is also principal investigator for the UKMEDP051 project.  
16

17 LWP is partly supported by NIHR grant CDF 2015-08-11 to PAT, and a portion of his research time is  
18 funded by the UCAT board.  
19

20 ICM, KYFC and DTS have received no specific funding for this project.  
21

## 22 23 Disclaimers

24  
25 KW and DH state that this publication presents independent research funded by the National  
26 Institute for Health Research (NIHR). The views expressed are those of the authors and not  
27 necessarily those of the NHS, the NIHR or the Department of Health and Social Care.  
28

29 PAT states this research is supported by an NIHR Career Development Fellowship (CDF 2015-08-11).  
30 This paper presents independent research partly funded by the National Institute for Health  
31 Research. The views expressed are those of the authors and not necessarily those of the NHS, the  
32 NIHR or the Department of Health and Social Care.  
33

34  
35 KYFC is employed as the Head of Marking and Results at Cambridge Assessment English. The views  
36 expressed are those of the authors and do not represent the views of Cambridge Assessment.  
37

38 DTS is employed by the GMC as a data analyst working on the UKMED project. The views expressed  
39 here are his views and not the views of the GMC.  
40

41 Data sources. UK Medical Education Database (UKMED) UKMEDP051 data extract generated on 13<sup>th</sup>  
42 May 2019. UKMEDP089 extract generated 21<sup>st</sup> January 2020. UKMEDP112 project, using  
43 UKMEDP051 and UKMEDP089 data, approved for publication on 29<sup>th</sup> May 2020. We are grateful to  
44 UKMED for the use of these data. However, UKMED bears no responsibility for their analysis or  
45 interpretation.  
46  
47

48 UKMEDP051 data includes information derived from that collected by the Higher Education Statistics  
49 Agency Limited (HESA) and provided to the GMC (HESA Data). Source: HESA Student Record  
50 2002/2003 to 2014/2015. Copyright Higher Education Statistics Agency Limited. The Higher  
51 Education Statistics Agency Limited makes no warranty as to the accuracy of the HESA Data, cannot  
52 accept responsibility for any inferences or conclusions derived by third parties from data or other  
53 information supplied by it.  
54  
55

56 UKMEDP051 and UKMEDP089 include Universities and Colleges Admissions Service (UCAS) data  
57 provided to the GMC (UCAS data). Source: UCAS (application cycles 2007 to 2018). Copyright  
58 Universities and Colleges Admissions Service (UCAS). UCAS makes no warranty as to the accuracy of  
59 the UCAS Data and cannot accept responsibility for any inferences or conclusions derived by third  
60 parties from data or other information supplied by it.

All data from HESA are required to be reported using their rounding and suppression criteria (<https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics>) and we have applied those criteria to all UKMED-based tables and values reported here.

## Competing interests

ICM is a member of the UKMED Research Group and the UKMED Advisory Board, and is also on the UKMACS advisory group.

PAT is a member of the UKMED Research Group. PAT has previously received research funding from the ESRC, the EPSRC, the Department of Health for England, the UCAT Board, and the GMC. In addition, PAT has previously performed consultancy work on behalf of his employing University for the UCAT Board and Work Psychology Group and has received travel and subsistence expenses for attendance at the UCAT Research Group.

KYFC is a member of the UKMED Research Group, and is an employee of Cambridge Assessment - a group of exam boards that owns and administers the BioMedical Admissions Test (BMAT); UK GCSEs and A-levels; and International GCSEs and A-levels.

DTS is a member of the UKMED Research Group and the UKMED Advisory Board and is employed by the GMC as a data analyst working on the UKMED project.

KW, DH and LP declare no competing interests.

## Authorship

ICM conceived the idea for the study, conducted the statistical analysis and wrote the first draft of the paper. All authors contributed to subsequent drafts and approved the final version.

## Ethical approval

Queen Mary Research Ethics Committee, University of London, agreed on 11 November 2015 that there was no need for ethical review of UK Medical Education Database research studies.

[https://www.ukmed.ac.uk/documents/UKMED\\_research\\_projects\\_ethics\\_exemption.pdf](https://www.ukmed.ac.uk/documents/UKMED_research_projects_ethics_exemption.pdf)

All data from HESA are required to be reported using their rounding and suppression criteria (<https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics>) and we have applied those criteria to all UKMED-based tables and values reported here.

## Provenance and peer review

Not commissioned; reviewed by the UKMED Research Subgroup and Advisory board, and to submitted to an externally peer reviewed journal. This paper was fast-tracked through the UKMED governance processes as with other COVID-19 related research projects elsewhere.

## Patient and public involvement.

No patient involvement.

## Data sharing statement

Researchers wishing to re-analyse the data used for this study can apply for access to the same datasets via UKMED ([www.ukmed.ac.uk](http://www.ukmed.ac.uk)).

## References

1. Hussein Y: **Cancellation of GCSE is unfair to some students.** *The Guardian* 2020, <https://www.theguardian.com/world/2020/mar/29/cancellation-of-gcse-exams-unfair-to-some-students>.
2. Thomson D: *Moderating teaching judgments in 2020 [Blog post, 25th March 2020]*. London: FFT Educational Lab: <https://ffteducationdatalab.org.uk/2020/03/moderating-teacher-judgments-in-2020/> (accessed 16th April 2020); 2020.
3. Petch JA: *School estimates and examination results compared*. Manchester: Joint Matriculation Board; 1964.
4. McKie A: **Scapped exams may spark UK admissions 'scramble'.** *Times Higher Education* 2020, **26th March 2020**: 9.
5. Dowell J, Cleland J, Fitzpatrick S, McManus IC, Nicholson S, Oppé T *et al.*: **The UK medical education database (UKMED): What is it? Why and how might you use it?** *BMC Medical Education* 2018, **18 (DOI 10.1186/s12909-017-1115-9)**: 1-8.
6. Woolf K, Harrison D, McManus IC: **The attitudes, perceptions and experiences of medical school applicants following the closure of schools and cancellation of public examinations due to the COVID-19 pandemic in 2020 .** *medRxiv* 2020, **submitted**.
7. Ofqual: *Summer 2020 grades for GCSE, AS and A level, Extended Project Qualification and Advanced Extension Award in maths: Guidance for teachers, students, parents and carers*. Coventry: Ofqual: Ofqual/20/6607/2 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/877842/Summer\\_2020\\_grades\\_for\\_GCSE\\_AS\\_A\\_level\\_EPQ\\_AEA\\_in\\_maths\\_-\\_guidance\\_for\\_teachers\\_students\\_parents.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/877842/Summer_2020_grades_for_GCSE_AS_A_level_EPQ_AEA_in_maths_-_guidance_for_teachers_students_parents.pdf) (accessed 3rd April 2020); 2020.
8. Cresswell M: *Heaps, prototypes and ethics: The consequence of using judgments of student performance to set examination standards in a time of change*. London: Institute of Education; 2003.
9. Ofqual: *Setting GCSE, AS and A Level Grade Standards in Summer 2014 and 2015*. London: <https://www.gov.uk/government/publications/setting-gcse-and-a-level-grade-standards-in-summer-2014-and-2015> [accessed 18th April 2020]; 2020.
10. Ofqual: *Summer 2020 grades for GCSE, AS and A level, Extended Project Qualification and Advanced Extension Award in maths: Information for Heads of Centre, Heads of Department and teachers on the submission of Centre assessment grades*. Coventry: Ofqual: Ofqual/20/6607/1 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/877930/Summer\\_2020\\_grades\\_for\\_GCSE\\_AS\\_A\\_level\\_EPQ\\_AEA\\_in\\_maths\\_-\\_guidance\\_for\\_heads\\_of\\_centres.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/877930/Summer_2020_grades_for_GCSE_AS_A_level_EPQ_AEA_in_maths_-_guidance_for_heads_of_centres.pdf) (accessed 3rd April 2020); 2020.
11. McManus IC, Powis DA, Wakeford R, Ferguson E, James D, Richards P: **Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school.** *Brit Med J* 2005, **331**: 555-559.
12. Downing SM: **Validity: On the meaningful interpretation of assessment data.** *Med Educ* 2003, **37**: 830-837.

13. American Educational Research Association, American Psychological Association, National Council on Measurement in Education: *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 2014.
14. Kane MT: **Validating the interpretations and uses of test scores**. *JEM* 2014, **50**: 1-73.
15. Bekhradnia B, Thompson J: *Who does best at University?* London: Higher Education Funding Council England (<http://webarchive.nationalarchives.gov.uk/20081202000732/http://hefce.ac.uk/Learning/whodoes/>); 2002.
16. Higher Education Funding Council for England [HEFCE]: *Differences in degree outcomes: The effect of subject and student characteristics. Issues Paper 2015/21*. Bristol: HEFCE; <https://webarchive.nationalarchives.gov.uk/20180319114800/http://www.hefce.ac.uk/pubs/Year/2015/201521/> 2015.
17. Higher Education Funding Council for England [HEFCE]: *Differences in student outcomes: The effect of student characteristics. Data Analysis March 2018/05*. Bristol: HEFCE; 2018.
18. McManus IC, Smithers E, Partridge P, Keeling A, Fleming PR: **A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study**. *Brit Med J* 2003, **327**: 139-142.
19. McManus IC, Woolf K, Dacre J, Paice E, Dewberry C: **The academic backbone: Longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the Specialist Register in UK medical students and doctors**. *BMC Medicine* 2013, **11**:242: doi:10.1186/1741-7015-11-242.
20. Lough C: **GCSEs: Only 39% teachers think 2020 grades fair for all: Plan for teacher-assessed GCSE and A-level grades prompts concerns about potential teacher bias, TES survey of 19,000 finds**. *TES (Times Educational Supplement)* 2020, **13th May 2020**: <https://www.tes.com/news/coronavirus-gcse-only-39-teachers-think-2020-grades-fair-all>.
21. Lilford R: **Policy makers should use evidence, but what should they do in an evidence vacuum**. *ARC West Midlands News Blog [NIHR Applied Research Collaboration, West Midlands]* 2020, **2**: 1-2 (<https://arcwm.files.wordpress.com/2020/04/arc-wm-newsblog-2020-04-24.pdf>).
22. Hubbles S, Bolton P: *The review of university admissions [Briefing Paper Number 8538, 10 April 2019]*. London: House of Commons (<https://researchbriefings.files.parliament.uk/documents/CBP-8538/CBP-8538.pdf>); 2019.
23. Snell M, Thorpe A, Hoskins S, Chevalier A: **Teachers' Perceptions and A-Level Performance: Is There Any Evidence of Systematic Bias?** *ORE* 2008, **34**: 403-423.
24. Everett N, Papageorgiou J: *Investigating the Accuracy of Predicted A Level Grades as part of 2009 UCAS Admission Process*. London: Department for Business, Innovation and Skills; 2011.
25. Wyness G: *Predicted grades: Accuracy and impact. A report of University and College Union*. London: University and College Union ([https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted\\_grades\\_report\\_Dec2016.pdf](https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted_grades_report_Dec2016.pdf)); 2016.



- 1 26. UCAS: *End of cycle report 2017: Qualifications and competition*. Cheltenham: UCAS  
2 [https://www.ucas.com/data-and-analysis/ucas-undergraduate-releases/ucas-  
3 [undergraduate-analysis-reports/2017-end-cycle-report](https://www.ucas.com/data-and-analysis/ucas-undergraduate-releases/ucas-undergraduate-analysis-reports/2017-end-cycle-report)]; 2017.
- 4  
5 27. UCAS: *Factors associated with predicted and achieved A level attainment, August 2016*.  
6 Cheltenham: UCAS: <https://www.ucas.com/file/71796/download?token=D4uuSzur>; 2016.
- 7  
8 28. Gill T: **Methods used by teachers to predict final Alevel grades for their students**. *Research*  
9 *Matters (UCLES)* 2019, 33-42.
- 10  
11 29. Walland E, Darlington E: *Insights on trends in AS Levels, the EPQ and Core Maths: summary*  
12 *report*. Cambridge: 35859 /id}. Cambridge Assessment:  
13 <https://www.cambridgeassessment.org.uk/.../527125-insights-on-trends-in-as-levels-the-epq-and-core-maths-summary-report.pdf>; 2019.
- 14  
15 30. Gill T, Benton T: *The accuracy of forecast grades for OCR Alevels in June 2014: Statistics*  
16 *Report Series No 90*. Cambridge: Cambridge Assessment  
17 [https://www.cambridgeassessment.org.uk/Images/241261-the-accuracy-of-forecast-  
18 [grades-for-ocr-a-levels-in-june-2014.pdf](https://www.cambridgeassessment.org.uk/Images/241261-the-accuracy-of-forecast-grades-for-ocr-a-levels-in-june-2014.pdf)]; 2015.
- 19  
20 31. Gill T, Rushton N: *The accuracy of forecast grades for OCR Alevels: Statistics Report Series No*  
21 *26*. Cambridge: Cambridge Assessment [https://www.cambridgeassessment.org.uk/our-  
22 [research/all-published-resources/statistical-reports/150215-the-accuracy-of-forecast%20grades-for-ocr-a-levels-in-june-2012.pdf/](https://www.cambridgeassessment.org.uk/our-research/all-published-resources/statistical-reports/150215-the-accuracy-of-forecast%20grades-for-ocr-a-levels-in-june-2012.pdf/)]; 2011.
- 23  
24 32. Gill T, Chang Y: *The accuracy of forecast grades for OCR A levels in June 2012*.  
25 Cambridge: Cambridge Assessment Statistics Report Series No.64; 2013.
- 26  
27 33. Petch JA: *Fifty years of examining: The Joint Matriculation Board, 1903-1953*. 1953.
- 28  
29 34. Murphy RJL: **Teachers' assessments and GCE results compared**. *Educational Research* 1979,  
30 **22**: 54-59.
- 31  
32 35. Gill T, Chang Y: *The accuracy of forecast grades for OCR Alevels in June 2013: Statistics*  
33 *Report Series No 89*. Cambridge: Cambridge Assessment  
34 [https://www.cambridgeassessment.org.uk/Images/241260-the-accuracy-of-forecast-  
35 [grades-for-ocr-gcses-in-june-2013.pdf](https://www.cambridgeassessment.org.uk/Images/241260-the-accuracy-of-forecast-grades-for-ocr-gcses-in-june-2013.pdf)]; 2015.
- 36  
37 36. Gill T, Benton T: *The accuracy of forecast grades for OCR GCSEs in June 2014: Statistics*  
38 *Report Series No 91*. Cambridge: Cambridge Assessment  
39 [https://www.cambridgeassessment.org.uk/Images/241265-the-accuracy-of-forecast-  
40 [grades-for-ocr-gcses-in-june-2014.pdf](https://www.cambridgeassessment.org.uk/Images/241265-the-accuracy-of-forecast-grades-for-ocr-gcses-in-june-2014.pdf)]; 2015.
- 41  
42 37. Südkamp A, Kaiser J, Möller J: **Accuracy of Teachers' Judgments of Students' Academic**  
43 **Achievement: A Meta-Analysis**. *Journal of Educational Research* 2012, **104**: 743-762.
- 44  
45 38. Lumb AB, Vail A: **Applicants to medical school: the value of predicted school leaving**  
46 **grades**. *Med Educ* 1997, **31**: 307-311.
- 47  
48 39. McManus IC, Richards P, Winder BC, Sproston KA, Styles V: **Medical school applicants from**  
49 **ethnic minorities: identifying if and when they are disadvantaged**. *Brit Med J* 1995, **310**:  
50 496-500.
- 51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
40. Richardson PH, Winder B, Briggs K, Tydeman C: **Grade predictions for school-leaving examinations: do they predict anything?** *Med Educ* 1998, **32**: 294-297.
  41. Wilmut J, Wood R, Murphy R: *A review of research into the reliability of examinations: A discussion paper prepared for the School Curriculum and Assessment Authority*. Nottingham: School of Education [available at [www.gov.uk/systems/uploads](http://www.gov.uk/systems/uploads)]; 1996.
  42. Bramley T, Dhawan V: *Estimates of Reliability of Qualifications*. Cambridge: Cambridge Assessment:[[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/578868/2011-03-16-estimates-of-reliability-of-qualifications.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/578868/2011-03-16-estimates-of-reliability-of-qualifications.pdf)]; 2010.
  43. Wheadon C, Stockford I: *Classification accuracy and consistency in GCSE and A level examinations offered by the Assessment and Qualifications Alliance (AQA) November 2008 to June 2009*. Office of Qualifications and Examinations Regulation: Coventry (<http://www.ofqual.gov.uk/files/reliability/11-03-16-AQA-Classification-Accuracy-and-Consistency-in-GCSE-and-A-levels.pdf>); 2011.
  44. Curtis S, Smith D: **A comparison of undergraduate outcomes for students from gateway courses and standard entry medicine courses.** *BMC Medical Education* 2020, **20**: <https://doi.org/10.1186/s12909-019-1918-y>.
  45. Meng X-L, Rosenthal R, Rubin DB: **Comparing correlated correlation coefficients.** *Psychol Bull* 1992, **111**: 172-175.
  46. McManus IC, Dewberry C, Nicholson S, Dowell JS, Woolf K, Potts HWW: **Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: Meta-regression of six UK longitudinal studies.** *BMC Medicine* 2013, **11**:243: doi:10.1186/1741-7015-11-243.
  47. McManus IC, Dewberry C, Nicholson S, Dowell J: **The UKCAT-12 study: Educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a collaborative study of twelve UK medical schools.** *BMC Medicine* 2013, **11** :244: doi:10.1186/1741-7015-11-244.
  48. Burgess S, Sievertsen HH: *Schools, skills, and learning: The impact of COVID-19 on education.* <https://voxeu.org/article/impact-COVID-19-education> (1st April 2020; accessed 31st May 2020): 2020.
  49. Carlsson M, Dahl GB, Öckert B, Roth D: **The effect of schooling on cognitive skills.** *Review of Economics and Statistics* 2015, **97**: 533-547.
  50. Lavy V: **Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries.** *The Economic Journal* 2015, **125**: 397-424.
  51. Rowland CA: **The effect of testing versus restudy on retention: A meta-analytic review of the testing effect.** *Psychol Bull* 2014, **140**: 1432-1463.
  52. Mwandigha LM, Tiffin PA, Paton LW, *et al* What is the effect of secondary (high) schooling on subsequent medical school performance? A national, UK-based, cohort study *BMJ Open* 2018;**8**:e020291. doi: 10.1136/bmjopen-2017-020291

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
53. Russell Group, GuildHE, Million Plus, University Alliance: *Joint statement on university admissions*. London: <https://russellgroup.ac.uk/news/joint-statement-on-university-admissions/> (accessed 23rd April 2020); 2020.

For peer review only

## Figure captions

Figure 1: Predicted vs attained A-level grades for individual subjects in applicants to UK medical schools. Accurate predictions are in bold; yellow – over-estimates by 1 grade; orange – over-estimates by 2+ grades; green – under-estimates by 1 grade; blue – under-estimates by 2+ grades. a) Counts; b) attained grades as percentages within predicted grades.

Figure 2: Over-estimates, under-estimate and accurate predicted grades in various studies. Black font: predicted grades; red font: forecasted grades; yellow background: pre-2000; blue background: pre-2010; bold, underlined: averaged results post-2000.

Figure 3: Correlation matrix of selection measures, undergraduate outcome measures, and postgraduate outcome measures (separated by grey lines for clarity). Cells indicate Pearson correlation and N.

Figure 4: Mean Educational Performance Measure (EPM) deciles (revDeciles; 95% CI) in relation to actual A-level grades (green) and predicted A-level grades (red)

Figure 5: Predicted decile outcomes if selection were on Predicted A-level grades (blue) rather than actual A-level grades (orange).

a) Counts of number of cases

		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades (points)	E (2 pts)	<b>200</b>	35	10	5	0	0	255 (0%)
	D (4 pts)	235	<b>610</b>	155	35	10	0	1045 (0%)
	C (6 pts)	635	1220	<b>2110</b>	505	95	5	4570 (2%)
	B (8 pts)	635	2095	4755	<b>7355</b>	1695	175	16715 (7%)
	A (10 pts)	430	1925	8785	35640	<b>61950</b>	12655	121390 (51%)
	A* (12 pts)	50	135	635	6025	42815	<b>43395</b>	93060 (39%)
	Total	2185	6020	16450	49570	106570	56235	237030
		(1%)	(3%)	(7%)	(21%)	(45%)	(24%)	

b) Percentages within predicted grades

		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades (points)	E (2 pts)	<b>79%</b>	14%	..	..	..	..	100%
	D (4 pts)	23%	<b>58%</b>	15%	3%	..	..	100%
	C (6 pts)	14%	27%	<b>46%</b>	11%	2%	..	100%
	B (8 pts)	4%	13%	28%	<b>44%</b>	10%	1%	100%
	A (10 pts)	0%	2%	7%	29%	<b>51%</b>	10%	100%
	A* (12 pts)	0%	0%	1%	7%	46%	<b>47%</b>	100%
	Total	1%	3%	7%	21%	45%	24%	100%

Figure 1: Predicted vs attained A-level grades for individual subjects in applicants to UK medical schools. Accurate predictions are in bold; yellow – over-estimates by 1 grade; orange – over-estimates by 2+ grades; green – under-estimates by 1 grade; blue – under-estimates by 2+ grades. a) Counts; b) attained grades as percentages within predicted grades.

108x60mm (300 x 300 DPI)

Study	Context	Year	A-level range	Over-estimated "Pessimistic"	Accurate	Under-estimated "Optimistic"
<b>University applicants overall: A-levels, etc</b>						
Everett & Papageorgiou (2011) [24]	Predicted Grades	2009	A-E	7%	52%	42%
UCAS (27)	Predicted Grades	2012	A*-E	12%	20%	68%
Wyness (2016) [25]	Predicted Grades	2013-15	A*-E	9%	16%	75%
UCAS (27)	Predicted Grades	2016	A*-E	9%	16%	74%
UCAS (27)	Predicted Grades	2017	A*-E	10%	16%	73%
Petch (1953) [32]	Forecasted Grades	1940	School Cert Pass/Fail	2%	89%	9%
Petch (1964) [4]	Non-official forecasted Grades	1963	A+B/C+D/E/O/F	18%	43%	39%
Murphy (1979) [33]	Non-official forecasted Grades	1977	A-E	29%	27%	44%
Gill & Rushton (2011) [30]	Forecasted Grades	2009	A-E	12%	55%	33%
Gill & Chang (2013) [31]	Forecasted Grades	2012	A*-E	13%	48%	39%
Gill & Benton (2015) [29]	Forecasted Grades	2014	A*-E	14%	43%	43%
Gill (2019) [28]	Non-official forecasted Grades	2018	A*-E	20%	45%	35%
	<b>Mean Predicted Grades</b>	Pre-2010	A-E	7%	52%	42%
	<b>Mean Forecasted Grades</b>	Pre-2010	A-E	20%	42%	39%
	<b>Mean Predicted Grades</b>	Post-2010	A*-E	<b>19%</b>	<b>17%</b>	<b>73%</b>
	<b>Mean Forecasted Grades</b>	Post-2010	A*-E	<b>15%</b>	<b>46%</b>	<b>39%</b>
<b>Medical school applicants and students: A-levels and other qualifications</b>						
Students: Lumb & Vail [37]	Predicted Grades	1995	A-E	7%	52%	41%
Applicants: A-levels (this study)	Predicted Grades	2010-18	A*-E	7%	49%	45%
Applicants: EPQ (this study)	Predicted Grades	2010-18	A*-E	14%	52%	34%
Applicants: SQA Adv Highers (this study)	Predicted Grades	2010-18	A*-E	3%	60%	38%
	<b>Mean Predicted Grades (Medics)</b>	2010-18	A*-E	<b>8%</b>	<b>53%</b>	<b>39%</b>
<b>GCSE grades: All candidates</b>						
Gill & Chang (2015) [35]	Forecasted GCSE Grades	2013	A*-G,U	12%	47%	41%
Gill & Benton (2015) [34]	Forecasted GCSE Grades	2014	A*-G,U	14%	44%	42%
	<b>Mean Forecasted GCSE Grades</b>			<b>13%</b>	<b>45%</b>	<b>42%</b>

Figure 2: Over-estimates, under-estimate and accurate predicted grades in various studies. Black font: predicted grades; red font: forecasted grades; yellow background: pre-2000; blue background: pre-2010; bold, underlined: averaged results post-2000.

108x60mm (300 x 300 DPI)

		Selection measures applicants					Undergraduate outcome measures			Postgraduate outcome measures	
		GCSE grades	Predicted Alevels	Alevel grades	UKCAT	BMAT	EPM	SJT	PSA	MRCP(UK) Part 1	MRCS Part A
	GCSE grades	<b>1</b>	<b>0.452</b>	<b>0.421</b>	<b>0.265</b>	<b>0.223</b>	<b>0.180</b>	<b>0.190</b>	<b>0.201</b>	<b>0.212</b>	<b>0.173</b>
	Predicted A-level grades	<b>0.452</b>	<b>1</b>	<b>0.585</b>	<b>0.272</b>	<b>0.326</b>	<b>0.198</b>	<b>0.160</b>	<b>0.226</b>	<b>0.283</b>	<b>0.181</b>
<b>Selection measures in all applicants</b>	Attained A-level grades	<b>0.421</b>	<b>0.585</b>	<b>1</b>	<b>0.326</b>	<b>0.416</b>	<b>0.297</b>	<b>0.195</b>	<b>0.306</b>	<b>0.421</b>	<b>0.358</b>
	UKCAT total	<b>0.265</b>	<b>0.272</b>	<b>0.326</b>	<b>1</b>	<b>0.483</b>	<b>0.115</b>	<b>0.243</b>	<b>0.238</b>	<b>0.200</b>	<b>0.181</b>
	BMAT sections 1 and 2	<b>0.223</b>	<b>0.326</b>	<b>0.416</b>	<b>0.483</b>	<b>1</b>	<b>0.089</b>	<b>0.239</b>	<b>0.321</b>	<b>0.378</b>	<b>0.319</b>
	UKFPO EPM decile	<b>0.180</b>	<b>0.198</b>	<b>0.297</b>	<b>0.115</b>	<b>0.089</b>	<b>1</b>	<b>0.319</b>	<b>0.470</b>	<b>0.509</b>	<b>0.535</b>
<b>Undergraduate outcome measures</b>	UKFPO SJT score	<b>0.190</b>	<b>0.160</b>	<b>0.195</b>	<b>0.243</b>	<b>0.239</b>	<b>0.319</b>	<b>1</b>	<b>0.346</b>	<b>0.351</b>	<b>0.274</b>
	PSA score	<b>0.201</b>	<b>0.226</b>	<b>0.306</b>	<b>0.238</b>	<b>0.321</b>	<b>0.470</b>	<b>0.346</b>	<b>1</b>	<b>0.500</b>	<b>0.483</b>
	MRCP(UK) Part 1	<b>0.212</b>	<b>0.283</b>	<b>0.421</b>	<b>0.200</b>	<b>0.378</b>	<b>0.509</b>	<b>0.351</b>	<b>0.500</b>	<b>1</b>	<b>...</b>
<b>Postgraduate outcome measures</b>	MRCS Part A	<b>0.173</b>	<b>0.181</b>	<b>0.358</b>	<b>0.181</b>	<b>0.319</b>	<b>0.535</b>	<b>0.274</b>	<b>0.483</b>	<b>...</b>	<b>1</b>

Figure 3: Correlation matrix of selection measures, undergraduate outcome measures, and postgraduate outcome measures (separated by grey lines for clarity). Cells indicate Pearson correlation and N.

108x60mm (300 x 300 DPI)

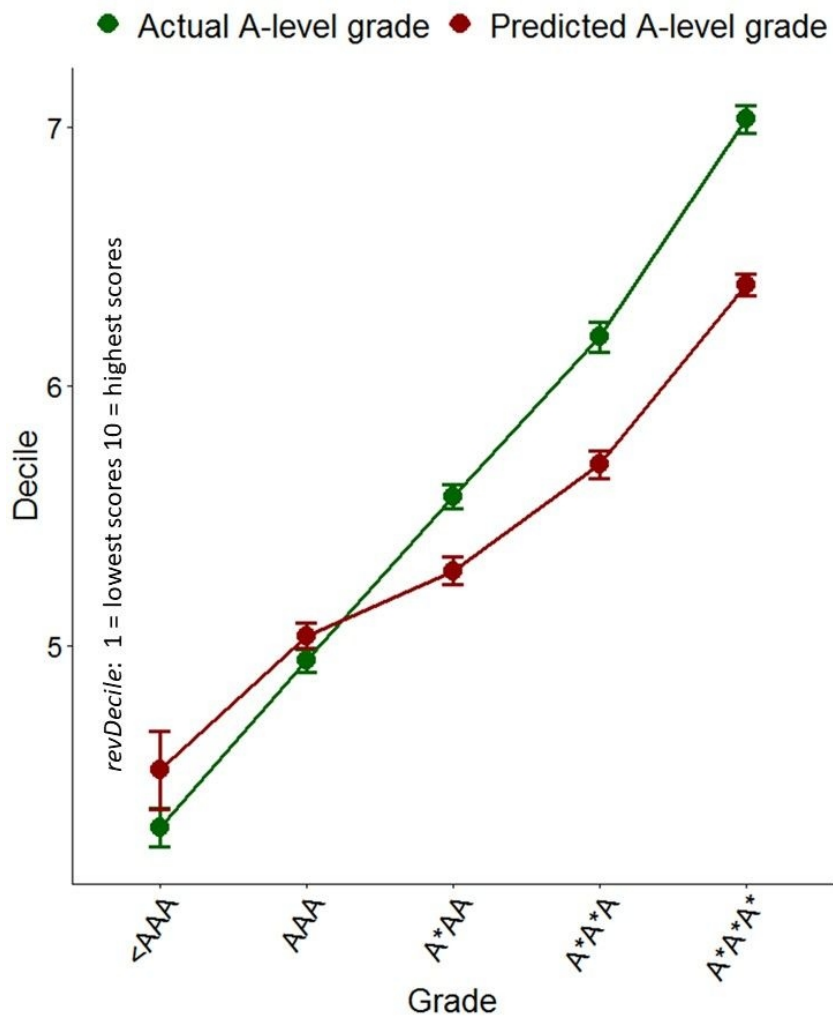


Figure 4: Mean Educational Performance Measure (EPM) deciles (revDeciles; 95% CI) in relation to actual A-level grades (green) and predicted A-level grades (red)

60x88mm (300 x 300 DPI)



	Decile	UKFPO score	RevDecile	Selection grades:		Odds Ratio	Absolute difference	Relative increase
				Attained	Predicted			
Worst	10	34	1	7.2%	8.1%	1.141	0.9%	13.0%
	9	35	2	9.4%	10.6%	1.135	1.1%	12.0%
	8	36	3	10.1%	11.1%	1.107	1.0%	9.5%
	7	37	4	10.7%	11.2%	1.052	0.5%	4.6%
	6	38	5	10.7%	10.8%	1.003	0.0%	0.3%
	5	39	6	10.6%	10.4%	0.978	-0.2%	-2.0%
	4	40	7	10.7%	10.4%	0.970	-0.3%	-2.7%
	3	41	8	10.3%	9.7%	0.935	-0.6%	-5.8%
	2	42	9	10.2%	9.1%	0.882	-1.1%	-10.7%
Best	1	43	10	10.1%	8.8%	0.853	-1.4%	-13.4%

Figure 5: Predicted decile outcomes if selection were on Predicted A-level grades (blue) rather than actual A-level grades (orange).

108x60mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

*Calculated grades, predicted grades,  
forecasted grades and actual A-level grades:  
Reliability, correlations and predictive validity in  
medical school applicants, and undergraduates*

## Supplementary information

- Supplementary literature review
- Supplementary Methods
- Supplementary Results
- *Appendix*: Are independent (private sector) schools more accurate in their A-level predictions?

### Supplementary literature review

*Overview of literature on predicted, forecasted and attained A-level grades.* The majority of studies reported here are also discussed in the main paper, in much more abbreviated form, but here are described more discursively.

#### *University applications in general*

Petch in 1964 [1] did what Wilmut has described as “one of the earliest and most celebrated studies of teacher estimates of examination result” [2] (p.60), describing how Petch found, “grade agreement in about 43% of cases, but the examination grade was higher than the teacher estimate in 18% of cases, but lower in 39% of cases, sometimes heavily so”.

Two other early studies were by Murphy in the first of which in 1979 he compared actual and predicted grades both for A-levels and, unusually, for O-levels (the predecessor of GCSEs) [3], including two-way tables of predicted vs actual grades. Of 291 results the predicted grades were accurate in 27% of cases, over-predictions in 44% and under-prediction in 29% of cases. Teachers were also asked to provide a rank order of students, and overall these correlated 0.6 with rank order in the examination, although individual teachers showed a range of correlations from just less than zero through to more than 0.9. Murphy’s 1981 study drew on application forms submitted to UCCA (now UCAS) by 15,109 candidates, of which “a large number included teachers’ pre-examination estimates of A-level grades” (with predicted grades being A, A/B, B, B/C, C etc). Results were broken down by exam board and also by subject. The overall correlation of predicted and actual grades was 0.66, with Physics, Chemistry and French showing the highest correlations. The study also looked at A-level – O-level correlations [4]. Although described as predicted grades, these data are actually best described as being *forecasted grades*.

More recent studies have mostly been concerned with the relationship of attained A-level grades and the predicted A-level grades entered on UCAS application forms by teachers. UCAS changed the way it collected such data in 2009, so that for UK-domiciled applicants subject-level predicted grades

1  
2  
3 were available, rather than as earlier when predicted grades were only available as total point scores  
4 [5]. For various reasons, not all A-levels have predicted grades. Most analyses are for candidates  
5 across all ability levels. Note that A\* grades were only introduced in 2010.  
6

7 In a study of 2009 applicants [5], overall *accuracy* at the subject level for A-levels for 219,744 A-  
8 levels was 52%, with predicted and attained grade being the same. In 42% of cases predicted grades  
9 were over-estimates, and in only 7% were they under-estimates. A grades tended to be predicted  
10 more accurately but that in part reflects that A grades cannot be under-predicted (or E grades over-  
11 predicted).  
12

13 Female candidates showed a slight tendency for grades to be more accurately predicted (52.3% vs  
14 51.1% in males). Socio-economic group showed strong relationships to accuracy, with 58% accurate  
15 predictions in the Higher Managerial group and 43% in the Routine group, but that in part reflects  
16 different actual A-level achievement (58% of Managerial candidates receiving an A grade compared  
17 to 33% of Routine candidates). The Higher Managerial group had the greatest over-prediction and  
18 the Routine group the highest under-prediction. Considering ethnicity, 53% of White applicants had  
19 accurate predictions compared with 47% of Asian ethnicity, and 39% of those of Black ethnicity.  
20 Centre (school) was related to accuracy, with 64% accuracy in Independent schools, 47% in state  
21 schools, and 40% in those in Further or Higher education. The authors note that multivariate  
22 analyses are probably needed to tease apart the relationships between the various correlates of  
23 accuracy. Other analyses looked at disability, region, and nation within the UK. Number of choices  
24 also related to accuracy, applicants making four choices being more accurate than those making five  
25 choices, but it was suggested that was because of the majority of the former being higher attainers  
26 applying to Medicine, Dentistry or Veterinary Medicine. The paper concluded that it is difficult to  
27 separate out the various factors involved in accuracy, not least because of the ceiling and floor  
28 effects for high and low attainers [5].  
29  
30

31 Wyness [6] analysed aggregated data provided by UCAS for the applicants from 2013-15, and hence  
32 A\* grades were included in the analysis. Overall only 16.1% of grades were accurately predicted, a  
33 much lower figure than the earlier study using 2009 data [5], perhaps because of the inclusion of the  
34 new A\* grades. 8.54% of grades were under-predicted, while 75.4% of grades were over-predicted.  
35 As with the 2009 data, there was a clear relationship between over-prediction and attained grade,  
36 although it is noted that there are strong ceiling effects at work. As with the 2009 study,  
37 independent schools provided the most accurate predictions. Applicants from disadvantaged  
38 backgrounds showed moderate to severe over-prediction. Asian and Black applicants were also  
39 more likely to be severely over-predicted. There were no differences between male and female  
40 applicants. The report is particularly interesting as it looks at prediction in high ability students,  
41 defined as AAB or above. The difference between the most and least disadvantaged in this group is  
42 much smaller, with 44.0% overpredicted in the most disadvantaged and 47.4% in the least  
43 disadvantaged. There was some evidence that under-predicted applicants tended to show under-  
44 matching (i.e. entering less competitive universities than their actual grades might predict). Further  
45 analyse and discussion of these data are provided elsewhere [7,8].  
46  
47  
48

49 UCAS in 2017 provided some limited data on over-prediction and under-prediction of A-levels since  
50 the introduction of A\* grades, with data for 2012, 2016 and 2017 [9]. Overall 19.5%, 16.3% and  
51 16.0% of predictions were accurate, with over-prediction in 68.4%, 74.3% and 73.3% of cases, and  
52 under-prediction in 11.8%, 9.0% and 10.4% (figures from EoC17\_Figure7\_9\_database.csv<sup>1</sup>). UCAS  
53 commented that, the gap between achieved and predicted A-level grades, "continues to widen"  
54 (p.23), although a comparison of 2016 and 2017 results concluded that there was little effect  
55 due to the reforms in A-levels that took place in 2017.  
56  
57  
58  
59  
60

---

<sup>1</sup> <https://www.ucas.com/file/140426/download?token=tUxAGXtt>

1  
2  
3 Not all studies have used the *predicted grades* provided to UCAS for use by universities in selection,  
4 which for medical school applicants would have been by mid-October). Until 2015 teachers were  
5 also asked, by the end of the following May, just before A-levels were sat, to provide *forecasted*  
6 *grades* to Awarding Organisations, and those grades then contributed in part to decisions on  
7 grading. Forecasted grades are clearly of particular interest given proposals for calculated grades to  
8 be based on estimates of performance by schools during May. Three analyses are available, for  
9 candidates taking A-levels in 2009 [10], 2012 [11] and 2014 [12] which are before and after A\*  
10 grades were introduced. A primary interest must be the comparison of these forecasted grades with  
11 the more usually studied predicted grades, described earlier for 2009 [5] and 2012 [9]. Note that the  
12 studies of forecasted grades are only for OCR (Oxford, Cambridge and Royal Society of Arts  
13 Examination Board) and hence include all A-level candidates, whereas the studies of predicted  
14 grades are for university applicants. Supplementary table 1 compares the two sets of predictions. In  
15 2009 there is little difference between predicted and forecasted grades in accuracy, with a small  
16 diminution of over-predictions. The picture three years later, in 2012 after A\* grades have been  
17 introduced, is rather different. Forecasted grades have an accuracy of 48% compared with only 20%  
18 for predicted grades. Taken overall it is difficult to reconcile the two studies which are only three  
19 years apart. Based on the 2009 data it would seem that predictions in May are no more accurate  
20 than those in October, whereas the 2012 data suggest that May predictions are much more accurate  
21 than October predictions. Having said that, even in May 2012, slightly less than a half of forecasted  
22 grades are accurate, with the same grade as in October.

23  
24  
25  
26 It should be noted, as pointed out earlier, that the early studies by Murphy should probably be  
27 regarded as being of forecasted and not predicted grades.

28  
29 *Grade point predictions.* The analyses described so far have been at the level of A-level subjects.  
30 Students mostly take three or sometimes more A-levels, and universities usually look at the three  
31 best grades attained. Scoring grades as A\*=12, A=10, B=8, C=6, D=4 and E=2 then a candidate  
32 passing three A-levels will score between 6 and 36 points for their three best grades<sup>2</sup>. Two studies  
33 [10,11] have pointed out the difficulty of using totalled points. As an example, a candidate predicted  
34 AAA will be predicted 30 points but may attain grades AAA or grades A\*A\*D; both are equally  
35 accurate in point terms but not in grade terms. Total predicted points are important in that UCAS for  
36 a number of years only provided total predicted points for the best three A-levels, without subjects  
37 or individual grades being specified<sup>3</sup>.

38  
39 UCAS in 2016 reviewed predicted and actual A-level grade points in applicants from 2010 to 2015  
40 [13] considering the best three grades attained. Achieved grades were one or two grades in total  
41 lower for attained than predicted grades. About a half of applicants in 2015 missed predicted total  
42 grades by two or more grades (e.g. ABB rather than AAA), a proportion that had increased by a third  
43 since 2010. Simple analyses in particular showed that missing predicted grades was associated with  
44 having *lower* predicted grades overall (as in the earlier analyses at the subject level). Multivariate  
45 analyses i.e. taking other factors into account, found missing predicted grades was associated with  
46 having *higher* predicted grades, lower GCSE attainment, taking biology, chemistry and maths, having  
47 Asian, Black, Mixed and Other ethnicity, coming from disadvantaged areas, being female, and having  
48 '[pre-A-level]unconditional offers'. Of particular interest is the relationship to GCSE grades, which  
49 have a strong relationship to A-level attainment [14] which is clearly seen in the UCAS data (see their  
50 figures 5 and 6).

51  
52  
53  
54  
55 <sup>2</sup> Some studies, including my own earlier ones, score A\*=6, B=5, etc.. Such schemes became less popular with  
56 the advent of AS-grades, which were scored as half of an A-level, and hence it made sense to double the points  
57 available for a full A-level so that totals remained integer. With the near disappearance now of AS-levels that  
58 rationale makes less sense.

59 <sup>3</sup> Earlier studies, such my 1991 cohort, had to extract predicted grades from UCAS references, and hence they  
60 are often embedded in free text, making it difficult to match them up with specific A-level subjects.

### *What are predicted grades and how are they made?*

UCAS, in its document, "Predicted grades – what you need to know"<sup>4</sup> says that "A predicted grade is the grade of qualification an applicant's school or college believes they're likely to achieve in positive circumstances." Later the document says predicted grades should be, "**in the best interests of applicants** – fulfilment and success at college or university is the end goal", and "**aspirational but achievable** – stretching predicted grades are motivational for students, unattainable predicted grades are not" (all emphases in original). It also says that grades should be "determined by professional judgement" and be data-driven, including "past Level 2 and Level 3 performance, and/or internal examinations to inform your predictions".

Gill [15] has described the relatively sparse literature on how teachers estimate grades. Gill's own study followed the methodology of Child and Wilson [16] although that study is not in the public domain. Gill sent questionnaires in May to selected OCR exam centres concerning Chemistry, English Literature and Psychology, and as well as estimating grades teachers were also asked to rank within grades, the method currently being adopted by Ofqual for calculated grades<sup>5</sup>. Teachers also indicated the evidence they had used for each decision. The response rate was extremely low (2.8%). About 45% of forecasted grades were accurate (which is similar to the 48% in supplementary table 1). Detailed A-level raw marks were also available and could be correlated with rankings, giving correlations of .87, .76 and .83 for the three subjects. Those correlations are high, and certainly are higher than a meta-analytic estimate of the effect size for teachers predicting academic achievement in pupils of 0.63 (SE=.03), although there was substantial heterogeneity. They are also higher than Murphy's 1979 estimate of 0.66 for the correlation of rankings and exam marks [3]. The most important information said by teachers to be used when predicting grades was performance in mock exams, and observations of quality of work and commitment, with oral presentation also important. Amongst other topics written in, the most important was the opinion of other teachers both in the same subject and other subjects, including the head of department. Other teachers raised concerns about the lack of high stakes for mock exams which meant that students did not treat them seriously. There were also concerns about the loss of AS-levels to help in prediction.

*Other examinations.* We know of no studies that have looked at accuracy of prediction of Scottish Highers or Advanced Highers, of the EPQ (Extended Project Question) used in England, or of other examinations carried out in the UK.

### *Applications to medical school*

Relatively few studies have looked at predicted grades in medical school applicants, although those studies do show a tendency to ask rather more stretching questions, perhaps because of the different interests of the researchers, and the specificity of the course and its outcomes.

Lumb and Vail pointed out that predicted grades are particularly important in the shortlisting phase of medical student selection [17]. They studied 1661 applications in 1995 to a single medical school who had estimated grades for 5053 A-levels, 52% of predictions being accurate, 41% were over-estimated and 7% under-estimated [17]. The authors presented an ROC curve (but not the area under the curve), and concluded that, "... selectors for medical schools can have some confidence in

---

<sup>4</sup> <https://www.ucas.com/advisers/managing-applications/predicted-grades-what-you-need-know> [Accessed 13th April 2020].

<sup>5</sup> One teacher refused to take part because of the difficulty of ranking 260 students sitting one exam. Another teacher commented, "it was easier for smaller centres to make predictions because they know individual students better" (p.42). The paper in fact comments that, "Responses to the questionnaire were more likely to come from smaller centres. ... [T]he 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)."

1  
2  
3 the accuracy of predictions and we should therefore continue to use them ... [for] selecting the  
4 doctors of the future.” (p.311).  
5

6 Richardson et al, studied 721 entrants from 1991 to 1994 to a single medical school [18]. Unusually  
7 they looked at predictive validity, assessing how well predicted and actual A-level grades related to a  
8 composite outcome on the pre-clinical course. Predicted and actual A-level grades showed a minimal  
9 correlation ( $r=0.024$ ), but selection would have imposed range restriction. Pre-clinical exam  
10 performance correlated 0.318 ( $p<.001$ ) with attained A-level grades, but only 0.041 (NS) with  
11 predicted A-level grades. This is a rare study in which predictive validity was assessed and it implied  
12 that selection should be on actual grades rather than predicted grades, concluding in contra-  
13 distinction to Lumb and Vail that, “medical school admissions panels would be well advised to take  
14 the predicted grade with a sizeable pinch of salt” (p.296).  
15  
16

17  
18 A third study, by one of the present team, took a different approach, using path modelling to assess  
19 the causal inter-relationships between GCSE grades, predicted A-level grades, receipt of an offer,  
20 actual A-level grades, and acceptance at medical school in an original sample size of 6901 applicants  
21 to five English medical schools [19]. A-level estimates were predicted by GCSE grades ( $\beta=0.89$ ),  
22 with attained A-level grades predicted by both GCSE grades ( $\beta=0.44$ ) and predicted A-level grades  
23 ( $\beta=0.74$ ). A substantive question of interest was whether the paths in the model differed  
24 between White and non-White candidates, with it being shown that none of the relationships  
25 described showed ethnic differences (although non-white candidates were significantly less likely  
26 than White candidates to receive an offer based on predicted A-level grades). Although the study  
27 reported no follow-up into the medical course, this dataset is analysed further below to assess  
28 predictive validity for postgraduate examination performance.  
29

30  
31 *A comment on issues in studying predicted A-level grades.*

32 Although predicted A-level grades have been an integral part of university application and selection  
33 in the UK for four decades, obtaining data on them is less than easy. Early studies, including my own,  
34 as well as those of other medical researchers, simply resorted to having researchers transcribe  
35 grades from paper UCCA and UCAS application forms, although often that was not easy in earlier  
36 forms as the predictions were often embedded in the free text of the Referee’s Statement. Until  
37 2009 UCAS only recorded the summed score of the best three A-levels, so that study of specific  
38 subjects was not possible. Even now obtaining UCAS data on predicted grades is less than easy, and  
39 Boliver in 2013 comments, “It would have been desirable to include predicted A-level grades... .  
40 Unfortunately UCAS are unable to provide this information in microdata form because of uncertainty  
41 about its validity in the case of applicants whose application is not linked to a school or college ...  
42 (personal communication from UCAS).” [20]. Similarly Wyness in 2016 in her study of three years of  
43 UCAS data comments that, “The data are aggregate (for reasons of privacy)” [6], which means of  
44 course that proper analyses at the level of individual participants are not possible. There is an irony  
45 here in that of course all universities have access to predicted grades provided by UCAS as a part of  
46 the admissions process, but subsequently obtaining those data for research is often very difficult.  
47 The data for the present study are the result of an important collaboration between UKMED and  
48 UCAS, with UCAS providing detailed information on applicants to UK medical schools for inclusion in  
49 the database, which is hosted in a safe haven to ensure strict controls on access; we are very  
50 grateful to UCAS for that collaboration without which the present study would not be possible.  
51  
52  
53  
54  
55

## 56 **Supplementary methods**

57 Data for the present study comes from two separate primary sources:  
58  
59  
60

1  
2  
3 “P89”. UKMED project UKMEDP089, “The UK Medical Applicant Cohort Study: Applications and  
4 Outcomes Study”, approved Dec 7<sup>th</sup>, 2018, with Dr Katherine Woolf as principal investigator, is an  
5 ongoing analysis as a part of UKMACS (UK Medical Applicant Cohort Study). Data are primarily  
6 concerned with the process of selection. In particular in the upload of 21<sup>st</sup> Jan 2020<sup>6</sup> there is detailed  
7 information from UCAS on all applicants to medical schools from 2007 to 2018, including all attained  
8 Key Stage 5 (Level 3) qualifications (e.g. A-levels and SQA) as well as teacher predicted grades for  
9 individual Key stage 5 qualifications.

10  
11 “P51”. UKMED project UKMEDP051, “A comparison of the properties of BMAT, GAMSAT and  
12 UKCAT”, approved Sept 25<sup>th</sup>, 2017, with Dr Paul Tiffin as principal investigator, is an ongoing analysis  
13 of the predictive validity of admissions tests and other selection methods such as A-levels and GCSEs  
14 in relation to undergraduate and postgraduate attainment. A major feature of the study is the  
15 inclusion of data from UCAS, although in the 13<sup>th</sup> May 2019 data upload, which was used here<sup>7</sup>,  
16 UCAS predicted grades were only available as a composite, 18-point score, for application years  
17 2010 to 2014. A new upload of the data in late April 2020 will provide more detailed information,  
18 but that will require quite extensive coding, etc., making it similar to the qualifications data for  
19 applicants in P89. For the present data upload, predicted A-level grades are in the old UCAS format  
20 consisting of a single number from 6 to 18 (i.e. 3 Es to 3 A\*s using A\*=6 coding). Outcome data for  
21 the P51 dataset are more extensive, and in particular include data for end of undergraduate training,  
22 including the UKFPO EPM measures, the UKFPO SJT as well as PSA (Prescribing Safety Assessment).  
23 Some data are available for later postgraduate examinations, but numbers inevitably are small for  
24 cohorts entering medical school in 2011 onwards.

25  
26 *A-level grade scoring.* In both P89 and P51, A-level grades are expressed numerically on a standard  
27 scale of A\*=12 points, A=10, B=8, C=6, D=4 and E=2, or have been rescaled to that score.

28  
29 *Rounding and suppression criteria.* All data from HESA are required to be reported using their  
30 rounding and suppression criteria ([https://www.hesa.ac.uk/about/regulation/data-  
31 protection/rounding-and-suppression-anonymise-statistics](https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics)) and although not all data in the current  
32 study use HESA measures we have nevertheless applied the HESA criteria to all UKMED-based tables  
33 and values reported in this study. It should be noted in particular that the presence of a zero or a  
34 zero percentage may not always mean that there are no individuals in a cell of a table. All Ns are  
35 rounded to the nearest 5 which should easily flag up that rounding has been applied, all counts  
36 ending in 0 or 5. Percentages are only reported when the number of participants is greater than  
37 22.5.  
38  
39  
40  
41  
42

## 43 **Supplementary Results**

44 *Predicted and actual grades for Key Stage 5 qualifications.*

45  
46 *Predicted and actual grades for individual A-levels.* Supplementary table 2 shows the relationship  
47 between predicted and attained A-level grades for 237,030 individual examinations from 2010 to  
48 2018. Supplementary table 2.a shows frequencies in the various combinations, with bold values in  
49 grey boxes on the diagonal indicating accurate prediction of grades, green and blue indicating under-  
50 prediction by 1 or 2 grades, and orange and red indicating over-prediction by 1 or 2 grades. Overall  
51 48.8% of predicted grades are accurate. Under-prediction occurs by one grade in 35.7% of cases, and  
52 by two or more grades in 9.0% of cases. Over-prediction is by one grade for 6.3% of A-levels, and  
53 0.1% by two or more grades. It should be remembered that since the median grade for actual A-  
54  
55

56  
57 <sup>6</sup> OUTPUT\_UCAS\_QUALS\_DEC\_20200121\_1.TXT and OUTPUT\_UCAS\_QUALS\_DEC\_20200121\_2.TXT (both  
58 dated 21/1/2020), and OUTPUT\_UCAS\_QUALS\_VER.TXT (dated 5/12/2019).

59 <sup>7</sup> UKCAT51\_APP\_ALL\_DATA\_13052019\_FILE1.SAV and UKCAT51\_APP\_ALL\_DATA\_13052019\_FILE2.SAV (both  
60 dated 13/5/2019).

level grades is A, then over-prediction in such cases can only be by a maximum of one grade, since A\* is the highest grade.

Supplementary tables 2.b and 2.c show the data of supplementary table 2.a as percentages. As has been pointed out [5] percentages within predicted grades and percentages within actual grades have different interpretations and uses. Both are presented here, but from the perspective of admissions tutors perhaps the most useful are those in supplementary table 2.b of percentages within predicted grades in relation to actual grades, as they show the likelihood that a predicted grade will actually manifest as particular actual grades. About a half of A\* predictions actually gain an A grade, and over a third of predicted A grades result in a grade B or lower.

Allocating points on the basis of A\*=12, A=10, B=8, C=6, D=4 and E=2, predicted grades show systematic *bias*, the mean prediction of 10.53 points being systematically higher than the mean actual grade of 9.55 points, the difference of 0.98 points being about half of an A-level grade, and can be seen in the greater numbers in red and orange cells in supplementary table 2.a (over-prediction, 45%) than in the blue and green cells (under-prediction, 6%).

Despite the bias, predicted grades overall show a reasonable *correlation* with actual grades, with a Pearson  $r_p$  of 0.624 and a Spearman correlation  $r_s$  of 0.581. Both predicted and actual grades are skewed because of censorship, values above A\* not being possible. A tetrachoric or polychoric correlation fits an underlying latent normal distribution into account, accepting that row and column totals may not be equally spaced, being ordinary in nature [21]. Using the *polychor()* function in R the polychoric correlation,  $r_t$  is somewhat higher at 0.716 (SE 0.002), and is probably the best estimate of the true extent of correlation.

*Differences between A-level subjects.* A-levels in different subjects may show differences in their degrees of bias or correlation. Subjects were divided into 26 broad groups (see supplementary table 3), with the Modern Languages group including 21 languages.

Supplementary table 3 shows the mean predicted points, the mean actual points, actual minus predicted points, and the Pearson correlation of predicted and actual points. Subjects are sorted by the number of examination entries, and values are colour coded on a green-yellow-red scale, green indicating higher predicted and actual grades, a smaller difference between predicted and actual grades (i.e. less bias), and higher correlation of predicted and actual grades.

Considering the four major subjects of chemistry, biology, maths and physics, differences between actual and predicted grades are very similar (-1.15 to -0.98) indicating a bias of about 1 point (i.e. half of a grade) and very similar correlations of 0.600 to 0.635. Amongst other subjects there is inevitably greater variation in those subjects taken less frequently. Of particular interest, given that some medical schools use it for selection, is General Studies, which has the largest difference of predicted and actual grades of -1.96 points, equivalent to a whole grade. The smallest bias is for art and design subjects at -.57 points, perhaps indicating the role of an in-course portfolio in these subjects giving teachers a better sense of how students are performing. Correlations of predicted and actual grades are mostly very similar, although the lower correlations are for general studies, modern languages, geography, history, economics, music and classics, and, as mentioned, for general studies.

*Total predicted and actual points, correlations between grades and reliability of measures*

*Reliability of actual and predicted A-levels.* The reliability of total points from the three best actual and predicted A-levels was calculated by randomly sampling a pair of grades from the best three and finding the correlation. Cronbach's alpha for the three totalled grades could then be calculated from the standard formula,  $\text{Alpha} = 3.r/(1+2.r)$  where  $r$  is the mean correlation, and is equivalent to a single randomly sampled correlation between a pair of grades since any pair should give similar results. Analysis was restricted to the 66,006 candidates who had at least three paired predicted and actual grades. For actual grades  $r=0.615$  (SE .003) giving  $\text{alpha}=0.827$ , while for predicted grades



1  
2  
3  $r=0.550$  (SE = .004) and hence  $\alpha=0.786$ . Given the standard errors, the correlation between  
4 grades is clearly substantially lower for predicted than actual grades, and the same must be true of  
5  $\alpha$ . Interpreting the difference is not entirely straightforward, since on the one hand more  
6 predicted grades are at A\*, meaning that there should be fewer non-identical grades, but range  
7 restriction might also result in a lower correlation. In terms of mechanism, teachers may collaborate  
8 in producing predicted grades [15], and such non-independence would increase correlations and  
9 increase  $\alpha$ . However teachers may also spend less time making judgements than do A-level  
10 examiners, and hence there should be lower correlations. On balance it seems that the most likely  
11 conclusion is that estimated grades are somewhat less reliable than actual grades, but there is  
12 clearly a need for more complex modelling of the reliability of actual and estimated grades.  
13  
14

15 *Predicted and actual grades for Extended Project Qualification (EPQ).* The English EPQ has become  
16 popular qualification for medical school applicants, being taken by 18616 applicants over the years  
17 2018 to 2018, about 2100 applicants a year (perhaps 10% of all applicants). There is evidence that it  
18 has predictive validity for degree outcomes [21]. At present it is not known if it predicts outcomes in  
19 application or at medical school. Supplementary table 4 shows the relationship between actual and  
20 predicted grades. Grades are over-estimated in 33.7% of cases, under-estimated in 14.0% and  
21 accurate in 52.3% of cases, the mean score difference, the bias, being 0.805, which is a little under  
22 half a grade. Pearson's correlation is  $r_p=.459$ , Spearman's correlation is  $r_s=.457$ , whereas the  
23 polychoric correlation is somewhat higher at  $r_t=.569$ .  
24

25 *Predicted and actual grades for SQA Advanced Highers.* SQA Advanced Highers, as with SQA Highers,  
26 are scored both as simple literals (A, B, C D) and as a more extended scoring (A1, A2, B3, B4, C5, C6,  
27 D7), although predicted grades are only in terms of literals. Supplementary tables 5.a and 5.b show,  
28 that A grades are more frequent in predicted than in attained grades. Using literals, 59.8% of  
29 predictions are accurate, 37.7% are over-estimated, and 2.6% are under-estimated, and for literal  
30 grades the bias was 0.976 points, equivalent to half a grade. Correlations of predicted grades with  
31 literal attained grades were  $r_p=.407$  and  $r_s=.357$ , whereas with extended grades were  $r_p=.409$  and  
32  $r_s=.355$ . Polychoric correlations were  $r_t=.575$  for literal grades and  $r_t=.587$  for extended grades,  
33 again showing the similarity across the two grading schemes.  
34  
35

36 *Summary.* Taking all the exam types together, A-Levels, EPQ and SQA Advanced Highers, it is  
37 generally clear that predicted grades are usually about a half-grade higher than actual grades. Where  
38 grades are not accurate there are about four times as many grades over-estimated as under-  
39 estimated.  
40

41 *Predictive validity of predicted and attained A-level grades.*

42 A key question throughout discussions of calculated grades is whether grades estimated by teachers  
43 are better or worse at predicting outcomes than are actual A-level grades. That question is answered  
44 not in terms of how well predicted grades relate to actual A-level grades, but by assessing how well  
45 predicted and actual grades predict subsequent outcomes during undergraduate and postgraduate  
46 training. It should also be said that it is not entirely self-evident that teachers' grades will be less  
47 good, and in the context of GCSEs rather than A-levels, Thomson said, "It is possible, in theory at  
48 least, that teacher judgements may be more reliable than exam grades, particularly in those subjects  
49 where exam reliability is lower" [22], with "more reliable" being somewhat ambiguous and perhaps  
50 also meaning more valid as well as more reliable in the narrow statistical sense. Questions about  
51 predictive validity can be answered by the P51 dataset.  
52  
53

54 *Predictive validity in P51.* The P51 UKMED data includes only applicants applying for medical  
55 schools. Predicted A-level grades were available only for the UCAS application cycles of 2010 to  
56 2014, and consisted of a single score in the range 2 to 18 points, based on the three highest  
57 predictions scored as A\*=6, A=5, etc.. The modal score for 38964 applicants was 15 (equivalent to  
58 AAA; mean=15.88; SD= 1.79; Median = 16; 5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup> and 95<sup>th</sup> percentiles= 13, 15, 17 and 18).  
59 Some older applicants had only pre-A\* A-levels, and it was also desirable to restrict the analysis to  
60

1  
2  
3 standard applicants in their first year of application, and so only those aged 18 in the UCAS year  
4 were included. For multiple reasons not all applicants had both predicted grades and attained A-  
5 level grades, and analysis was restricted to the 22954 applicants with both predicted and attained  
6 grades. Other selection measures which were included in the analysis are GCSEs (mean grade for  
7 best eight grades), as well as U(K)CAT and BMAT scores, which are based on the most recent  
8 attempt which in most of the present cases is also the first attempt. For simplicity we used the total  
9 of the four sub-scores for U(K)CAT, and for BMAT the total of the Section 1 and 2 scores. No  
10 GAMSAT scores were available for this age-group.  
11

12 Outcome measures are complicated as different application cohorts enter medical school and  
13 graduate at different times, and lags within the system mean that not all outcome measures are  
14 available. In this UKMED data extract, applicants to UCAS in 2010 entered the medical register from  
15 2015-18, 2011 applicants in 2016-8, 2012 applicants in 2017-18 and 2013 applicants in 2018.  
16 Applicants for 2014 would only have qualified in 2019 but the UKMED dataset did not yet include  
17 that years, and some earlier entrants would also be expected to qualify after 2018. For simplicity,  
18 outcome measures were restricted to the deciles of the UKFPO's Educational Performance Measure  
19 (EPM), the raw score of the UKFPO's Situational Judgement Test (SJT), and the score relative to the  
20 pass mark of the Prescribing Safety Assessment (PSA), all at first attempt, as these are the main  
21 outcomes from undergraduate training. Insufficient numbers of doctors had progressed further in  
22 postgraduate training to make analysis meaningful in this data extract.  
23

24 Supplementary table 6 (presented also in the main paper) summarises the correlation matrix of the  
25 various measures. It is important to note that the large differences in Ns are primarily because some  
26 measures are present in applicants and used during *selection*, and others are undergraduate  
27 outcome measures from medical school, which of necessity are only present in *entrants*, and some  
28 are postgraduate outcome measures, only present in *graduates*, not all cohorts yet having reached  
29 that stage. The three parts of the correlation matrix are separated to clarify the distinction.  
30 Correlations of selection and outcome measures necessarily show range restriction because  
31 candidates have been selected on the basis of these measures, and in the case of graduates,  
32 selected and self-selected, so that they are less variable than would be the case in an unrestricted  
33 population of applicants. The most important question for these data is the extent to which  
34 Predicted and Attained A-level grades (shown in pink and green in Supplementary table 6) differ in  
35 how much they predict the three outcome measures, which typically are taken five or six years later.  
36

37  
38  
39 *Prediction of Educational Performance Measure (EPM)*. EPM is probably the most important  
40 outcome measure since it integrates educational performance across assessments for all but the  
41 final year of the undergraduate course<sup>8</sup>. Note that deciles are confusing, as UKFPO scores them in  
42 the reverse of the usual order, the first decile being highest performance and the tenth the lowest.  
43 Here for ease of interpretation we reverse the scoring in what we call *revDecile*, so that higher  
44 *revDeciles* indicate higher performance. EPM is a summary of outcome across assessments within a  
45 medical school, expressed as deciles of achievement within each school. EPM is predicted  $r=0.297$  by  
46 attained A-level grades but only  $r=0.198$  by predicted grades. Although in absolute terms those  
47 correlations may seem small it must be remembered that they are range restricted, and the  
48 construct level predictive validity, taking into account range restriction and measurement error is  
49 likely to be much higher [23]. N is large for these correlations and hence the differences are highly  
50 significant using Meng and Rosenthal's test for correlated correlations [24],  $Z = 12.6$ , with  $p < 10^{-33}$ .  
51 Although predicted grades predict less well than attained grades, they may predict differently, and  
52 hence contribute something over and above attained grades in predicting outcome? Entering  
53 predicted grades after attained grades in a multiple regression shows a highly significant but small  
54 additional prediction of predicted grades ( $\beta = .052$ , compared with  $\beta = .269$  for attained grades).  
55  
56  
57

58  
59 <sup>8</sup> <https://foundationprogramme.nhs.uk/wp-content/uploads/sites/2/2019/11/UKFP-2020-EPM-Framework-Final-1.pdf>  
60

1  
2  
3 Attained grades are therefore substantially better at predicting undergraduate outcome, but  
4 predicted grades may have a small amount of variance which is not shared with attained A-levels.  
5

6 *Can other measures replaced attained A-level grades for predicting EPM?* In the absence of attained  
7 grades, to what extent can other selection measures such as GCSE grades, U(K)CAT and BMAT  
8 replace the predictive variance in attained A-level grades? Regressing EPM on just predicted grades  
9 gives multiple R = .198, compared with an R of 0.297 when regressed on just actual grades. Adding  
10 GCSEs to Predicted grades increases R to .225, while also including U(K)CAT and BMAT increases R to  
11 .231, although that is still far short of the .297 from A-levels alone. Interestingly if Actual Grades are  
12 now added in to the equation as well, R increases to .308, which is higher than the R for just actual  
13 grades. Exploration suggests that the effect is due to the additional effect of GCSEs grades compared  
14 with just having attained A-level grades in the model (R=.306; Beta(attained grades)=.268,  
15 beta(GCSES)=.077). Overall therefore if only Predicted Grades are available, an improved prediction  
16 is obtain by including GCSEs and U(K)CAT/BMAT, although the model still falls short of that of actual  
17 A-levels in terms of prediction.  
18  
19

20 *Private and State Sector schooling and EPM.* The UKCAT-12 study [25] found that medical students  
21 educated in the private sector performed less well at medical school than those educated in the  
22 state sector with equivalent A-level grades. It is important to replicate that finding in the present  
23 data, and to explore the extent to which there are effects related to predicted as opposed to  
24 attained grades. Overall 6149 (26.8%) of students were educated in the private sector, compared  
25 with 16805 (73.2%) in the state sector. Supplementary figure 1 plots *revDecile* in relation to attained  
26 and predicted grades, separately by private and state education. Visually it is immediately clear that  
27 there is an overall main effect of schooling, the lines for private sector schools (pale green and pale  
28 red) being below those for state schools. Note that the point for private schools with predicted  
29 grades <AAA is missing, as N was very small, because of few private schools predicting grades below  
30 AAA. Considering just attained grades, regression showed effects of both A-level grade (b=.299 (SE  
31 .008)<sup>9</sup>, beta=.301, t=35.24, p<10<sup>-100</sup>) and private schooling (b= -.292 (SE=.053), beta= -.047, t=-5.478,  
32 p=4x10<sup>-8</sup>), but the addition of an interaction was not significant (t=0.746, p=.455) meaning that the  
33 slopes in supplementary figure 1.a 1.b are the same. A similar analysis for predicted grades found  
34 effects of predicted grade (b=.213 (SE .009), beta=.201, t=22.94, p<10<sup>-100</sup>) and private schooling (b= -  
35 .256 (SE .055), beta= -.041, t=-4.679, p=0.000003), but the addition of an interaction was not  
36 significant (t=0.680746, p=.455), again meaning that the slopes are similar in the two types of school  
37 in supplementary figure 1.b. The standard errors for the effects of private schooling suggest that the  
38 difference between the slope is similar for actual and predicted grades.  
39  
40

41 Supplementary table 6 contains a number of other interesting features.  
42

43 *Other outcome measures in relation to actual and predicted A-levels.* There are four other outcome  
44 variables, two undergraduate and two postgraduate. For the undergraduate measures, PSA mark  
45 (supplementary figure 2) and SJT score (supplementary figure 3), both correlate more strongly with  
46 attained A-level grades than predicted A-levels (PSA: Z= 10.31, p<10<sup>-23</sup>; SJT Z= 4.38, p=0.000012).  
47 The two postgraduate outcome measures, are based on smaller, but still substantial, numbers of  
48 doctors, MRCP(UK) Part 1 being taken by 910 doctors, and MRCS Part A by 440 doctors. Both  
49 outcomes have higher correlations with attained A-level grades than predicted grades, MRCP(UK)  
50 Part 1 correlating 0.421 with actual A-level grades (supplementary figure 4), and 0.283 with  
51 predicted grades (Z= 4.54, p=.000055), and MRCS Part A correlating 0.421 with actual grades  
52 (supplementary figure 5) compared with 0.358 with predicted grades (Z= 3.67, p=.000238). The five  
53 outcome measures therefore show the same broad pattern of results.  
54  
55  
56  
57

58 <sup>9</sup> Actual and Predicted grades are scored on the basis of A\*=12, A=10 etc so are in the range 6 to 36 for three  
59 best grades. b=.299 therefore means an increase of 0.3 deciles per step on the A-level grade score, and  
60 therefore a full A-level grade (e.g A\*AA compared with AAA is 0.6 EPM deciles higher).

*Correlations of outcome measures and the status of the SJT.* The five outcome measures correlate well with each other (mean  $r = .420$ )<sup>10</sup>, as might be expected given the academic backbone [27]. Noteworthy is the relatively low correlation of SJT with EPM (.319) and PSA (.346), compared with the correlation of EPM and PSA (.470). That pattern is repeated when postgraduate exams are included, the four non-SJT assessments showing a higher correlation (mean  $r = .499$ ) than the correlations of the four non-SJT assessments with SJT (mean  $r = .322$ ). Overall that suggests that SJT may be measuring a construct that is different in part from the other more academic assessments, and that will need investigating more closely in the future. It is also of interest when considering predicted grades that SJT correlates only slightly better with actual grades than predicted grades (.195 vs .160), compared with the other four outcomes (.297 vs .198; .306 vs .226; .421 vs .283; and .358 vs .181; mean  $r = .346$  vs .222) raising the possibility that predicted grades may include some non-academic variance which then is predictive for SJT. That can be tested by regressing SJT on actual and predicted grades, when including predicted grades increases R from .195 to .206. The model including both grade types, shows an effect of actual grades (beta=.153,  $t=14.8$ ,  $p=10^{-49}$ ) and an effect of predicted grades (beta=.077,  $t=7.42$ ,  $p=1.2 \times 10^{-13}$ ), so that the beta effect of predicted grades is 50% of that for actual grades, compared with the earlier regression for deciles, where the beta of .052 for predicted grades is only 19% of the beta of .269 for attained grades.

The present SJT test is administered at the time of graduation. There is also a separate SJT administered as a part of the U(K)CAT tests, which was only introduced in 2014, and none of that cohort have outcome variables in the present data set. However it is of interest that, for the 4286 applicants in 2014 with U(K)CAT SJT, there is a correlation of .145 with Actual A-levels and .127 with predicted A-levels ( $Z=1.28$ ,  $p=0.192$ ). Overall it is possible that SJT tests are behaving differently to academic outcomes, despite moderately strong correlations of SJT with other academic outcomes. SJT tests are, “designed to assess for key attributes ... including commitment to professionalism, coping with pressure, effective communication, patient focus, and working effectively as part of a team” [28] [29].

*Correlations of A-levels with GCSEs, U(K)CAT and BMAT.* Without going into details, attained A-levels correlate more strongly with U(K)CAT and BMAT ( $r=.326$  and .416) than do predicted A-levels ( $r=.272$  and .326), suggesting that admissions tests are particularly assessing academic attainment. However GCSE grades show the reversed pattern and correlated *more strongly with predicted A-levels* (0.452) than with attained A-level grades (0.421), perhaps implying that teachers in part use GCSE grades to make predictions (as has been found in a previous study[19]).

*Correlations of admissions tests with outcome measures.* Neither of the two admissions tests, U(K)CAT and BMAT, has a strong prediction of EPM ( $r=.115$  and .089 respectively), and both clearly correlate less with EPM than does attained A-levels,  $r=.297$ , despite A-levels showing range restriction due to a ceiling effect at A\*. PSA and SJT though show a somewhat different picture. PSA correlates more highly with BMAT ( $r=.321$ ) than with U(K)CAT ( $r=.238$ ), and the correlation with BMAT is higher than that with attained A-levels ( $r=.306$ ). In contrast U(K)CAT and BMAT both correlate similarly with SJT ( $r=.243$  and .249), and both correlations are higher than with attained A-levels ( $r=.195$ ). BMAT and U(K)CAT both show correlations with the two postgraduate outcomes (0.200 and 0.378 for MRCP(UK) Part 1 and 0.181 and 0.319 for MRCS Part A, but both are lower than the correlations with A-levels (0.421 and 0.358). Taken overall, BMAT has somewhat higher correlations with the five outcome measures (mean  $r = .269$ ) than does U(K)CAT (mean  $r = .195$ ) but both correlate less with outcomes than do attained A-levels (mean  $r=.315$ ). U(K)CAT correlates at a similar level to predicted A-levels (mean  $r=.209$ ) but BMAT at a somewhat higher level.

<sup>10</sup> Note that there are too few doctors who took both MRCP(UK) Part 1 and MRCS Part A to be able to calculate a correlation. Elsewhere we have looked at the relatively rare groups of doctors taking both MRCP(UK) and MRCP, and shown high correlations between performance on the two assessments [26], making it likely that the same would also apply to MRCP(UK) Part 1 and MRCS Part A.

## Reference List

1. Petch JA: *School estimates and examination results compared*. Manchester: Joint Matriculation Board; 1964.
2. Wilmot J: *Experiences of summative teacher assessments in the UK*. London: Qualifications and Curriculum Authority; 2011.
3. Murphy RJL: **Teachers' assessments and GCE results compared**. *Educational Research* 1979, **22**: 54-59.
4. Murphy RJL: **O-level grades and teachers' estimates as predictors of the A-level results of UCCA applicants**. *BJEP* 1981, **51**: 1-9.
5. Everett N, Papageorgiou J: *Investigating the Accuracy of Predicted A Level Grades as part of 2009 UCAS Admission Process*. London: Department for Business, Innovation and Skills; 2011.
6. Wyness G: *Predicted grades: Accuracy and impact. A report of University and College Union*. London: University and College Union ([https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted\\_grades\\_report\\_Dec2016.pdf](https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted_grades_report_Dec2016.pdf)); 2016.
7. Wyness G: *Rules of the game: Disadvantaged students and the university admissions process*. London: The Sutton Trust; 2017.
8. Murphy R, Wyness G: *Minority Report: the impact of predicted grades on university admissions of disadvantaged groups*. London: **Centre for Education Policy and Equalising Opportunities (CEPEO) Working Paper No. 20-07, UCL Institute of Education** <https://EconPapers.repec.org/RePEc:ucl:cepeow:20-07>; 2020.
9. UCAS: *End of cycle report 2017: Qualifications and competition*. Cheltenham: UCAS [<https://www.ucas.com/data-and-analysis/ucas-undergraduate-releases/ucas-undergraduate-analysis-reports/2017-end-cycle-report>]; 2017.
10. Gill T, Rushton N: *The accuracy of forecast grades for OCR Alevels: Statistics Report Series No 26*. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/our-research/all-published-resources/statistical-reports/150215-the-accuracy-of-forecast%20-grades-for-ocr-a-levels-in-june-2012.pdf/>]; 2011.
11. Gill T, Chang Y: *The accuracy of forecast grades for OCR A levels in June 2012*. Cambridge: Cambridge Assessment Statistics Report Series No.64; 2013.
12. Gill T, Benton T: *The accuracy of forecast grades for OCR Alevels in June 2014: Statistics Report Series No 90*. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/Images/241261-the-accuracy-of-forecast-grades-for-ocr-a-levels-in-june-2014.pdf/>]; 2015.
13. UCAS: *Factors associated with predicted and achieved A level attainment, August 2016*. Cheltenham: UCAS: <https://www.ucas.com/file/71796/download?token=D4uuSzur>; 2016.

14. McManus IC, Woolf K, Dacre J: **The educational background and qualifications of UK medical students from ethnic minorities.** *BMC Medical Education* 2008, **8**: 21 (<http://www.biomedcentral.com/1472-6920/8/21>).
15. Gill T: **Methods used by teachers to predict final Alevel grades for their students.** *Research Matters (UCLES)* 2019, 33-42.
16. Child S, Wilson F: *An investigation of A level teachers' methods when estimating student grades.* Cambridge: Cambridge Assessment (Unpublished document, October 2015); 2015.
17. Lumb AB, Vail A: **Applicants to medical school: the value of predicted school leaving grades.** *Med Educ* 1997, **31**: 307-311.
18. Richardson PH, Winder B, Briggs K, Tydeman C: **Grade predictions for school-leaving examinations: do they predict anything?** *Med Educ* 1998, **32**: 294-297.
19. McManus IC, Richards P, Winder BC, Sproston KA, Styles V: **Medical school applicants from ethnic minorities: identifying if and when they are disadvantaged.** *Brit Med J* 1995, **310**: 496-500.
20. Boliver V: **How fair is access to more prestigious universities?** *British Journal of Sociology* 2013, **64**: 344-364.
21. Gill TRCV: *Predictive validity of level 3 qualifications: Extended Project, Cambridge Pre-U, International Baccalaureate, BTEC Diploma.* Cambridge: Cambridge Assessment: Cambridge Assessment Research Report; 2014.
22. Thomson D: *Moderating teaching judgments in 2020 [Blog post, 25th March 2020].* London: FFT Educational Lab: <https://ffteducationdatalab.org.uk/2020/03/moderating-teacher-judgments-in-2020/> (accessed 16th April 2020); 2020.
23. McManus IC, Dewberry C, Nicholson S, Dowell JS, Woolf K, Potts HWW: **Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: Meta-regression of six UK longitudinal studies.** *BMC Medicine* 2013, **11**:243: doi:10.1186/1741-7015-11-243.
24. Meng X-L, Rosenthal R, Rubin DB: **Comparing correlated correlation coefficients.** *Psychol Bull* 1992, **111**: 172-175.
25. McManus IC, Dewberry C, Nicholson S, Dowell J: **The UKCAT-12 study: Educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a collaborative study of twelve UK medical schools.** *BMC Medicine* 2013, **11** :244: doi:10.1186/1741-7015-11-244.
26. Wakeford R, Denney ML, Ludka-Stempien K, Dacre J, McManus IC: **Cross-comparison of MRCGP & MRCP(UK) in a database linkage study of 2,284 candidates taking both examinations: Assessment of validity and differential performance by ethnicity.** *BMC Medical Education* 2015, **15**.
27. McManus IC, Woolf K, Dacre J, Paice E, Dewberry C: **The academic backbone: Longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the Specialist Register in UK medical students and doctors.** *BMC Medicine* 2013, **11**:242: doi:10.1186/1741-7015-11-242.

- 1  
2  
3 28. Patterson F, Zibarras L, Ashworth V: **Situational judgement tests in medical education and**  
4 **training: Research, theory and practice: AMEE Guide No. 100.** *Med Teach* 2016, **38**: 3-17.  
5  
6  
7 29. McManus IC, Harborne A, Smith D, Horsfall HL, Devine O: **Exploring UK medical school**  
8 **differences: The *MedDifs* study of selection, teaching, student and F1 perceptions,**  
9 **postgraduate outcomes, and fitness to practise .** *BMC Medicine* 2019, In press.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6 *Supplementary Table 1: Comparison of predicted and forecasted grades in*  
7 *2009 and 2012.*  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

			Max	Over-		Under-			
Estimated grades			grade	predicted	Accurate	predicted	Population	Source	
<b>Predicted</b>	<b>October</b>	<b>2009</b>	<b>A</b>	<b>42%</b>	<b>52%</b>	<b>7%</b>	<b>UCAS</b>	Everett and Papageorgiou (2011)	
<b>Forecasted</b>	<b>May</b>	<b>2009</b>	<b>A</b>	<b>33%</b>	<b>55%</b>	<b>12%</b>	<b>OCR</b>	Gill and Rushton (2011)	
<i>Forecasted-Predicted</i>				<i>-9%</i>	<i>3%</i>	<i>5%</i>			
<b>Predicted</b>	<b>October</b>	<b>2012</b>	<b>A*</b>	<b>68%</b>	<b>20%</b>	<b>12%</b>	<b>UCAS</b>	UCAS (2017)	
<b>Forecasted</b>	<b>May</b>	<b>2012</b>	<b>A*</b>	<b>39%</b>	<b>48%</b>	<b>13%</b>	<b>OCR</b>	Gill and Chang (2013)	
<i>Forecasted-Predicted</i>				<i>-30%</i>	<i>29%</i>	<i>1%</i>			



*Supplementary Table 2: Comparison of predicted and attained A-level grades in medical school applicants, 2010-2018*

a) Counts of number of cases

		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	<b>200</b>	35	10	5	0	0	255 (0%)
	D	235	<b>610</b>	155	35	10	0	1045 (0%)
	C	635	1220	<b>2110</b>	505	95	5	4570 (2%)
	B	635	2095	4755	<b>7355</b>	1695	175	16715 (7%)
	A	430	1925	8785	35640	<b>61950</b>	12655	121390 (51%)
	A*	50	135	635	6025	42815	<b>43395</b>	93060 (39%)
	Total	2185	6020	16450	49570	106570	56235	237030
		(1%)	(3%)	(7%)	(21%)	(45%)	(24%)	

b) Percentages within predicted grades

		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	<b>79%</b>	14%	..	..	..	..	100%
	D	23%	<b>58%</b>	15%	3%	..	..	100%
	C	14%	27%	<b>46%</b>	11%	2%	..	100%
	B	4%	13%	28%	<b>44%</b>	10%	1%	100%
	A	0%	2%	7%	29%	<b>51%</b>	10%	100%
	A*	0%	0%	1%	7%	46%	<b>47%</b>	100%
	Total	1%	3%	7%	21%	45%	24%	100%

b) Percentages within predicted grades

		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	<b>9%</b>	1%	..	..	..	..	0%
	D	11%	<b>10%</b>	1%	0%	..	..	0%
	C	29%	20%	<b>13%</b>	1%	0%	..	2%
	B	29%	35%	29%	<b>15%</b>	2%	0%	7%
	A	20%	32%	53%	72%	<b>58%</b>	23%	51%
	A*	2%	2%	4%	12%	40%	<b>77%</b>	39%
	Total	100%	100%	100%	100%	100%	100%	100%

**Supplementary Table 3: Comparison of predicted and forecasted A-level grades in medical school applicants, 2010-2018**

Subject	N	Mean		Actual	r
		Predicted	Actual	minus Predicted (Pearson)	
Chemistry	62815	10.35	9.37	-0.98	0.623
Biology	61190	10.59	9.78	-0.82	0.632
Maths & Stats	54635	10.79	9.77	-1.02	0.600
Physics & Engineering	13870	10.67	9.52	-1.15	0.635
General Studies & Critical Thinking	6785	9.66	7.70	-1.96	0.534
Modern Languages	6720	10.59	9.74	-0.85	0.571
Psychology	6190	10.19	9.12	-1.07	0.631
Geography	4015	10.84	9.95	-0.89	0.538
History	3850	10.48	9.49	-0.99	0.546
English Literature & Language	3815	10.32	9.52	-0.80	0.681
Further Maths	2950	11.07	9.62	-0.80	0.681
Economics & Business Studies	2765	10.36	9.47	-0.89	0.577
Religious Studies	1890	10.45	9.40	-1.05	0.626
Art & Design	1035	10.60	10.03	-0.57	0.681
Latin & Classical Studies	675	10.74	9.65	-1.09	0.576
Music	640	10.49	9.51	-0.97	0.567
Sociology	525	9.51	8.49	-1.02	0.679
Computer Studies & ICT	475	9.89	8.82	-1.06	0.704
Physical Education	470	10.61	9.81	-0.80	0.610
Government & Politics	380	10.07	9.16	-0.91	0.656
Theatre Studies & Drama	260	10.14	9.02	-1.11	0.624
Science -- Misc & General	260	8.30	7.24	-1.06	0.821
Law	190	9.42	8.55	-0.87	0.766
Philosophy	155	10.37	9.06	-1.32	0.639
Classical Greek	115	10.90	9.98	-0.92	0.463
Media Studies	75	8.03	7.25	-0.78	0.798

*Supplementary Table 4: Comparison of predicted and attained EPQ grades in medical school applicants, 2010-2018*

a) EPQ: Counts of number of cases								
		Attained EPQ grade						
		E	D	C	B	A	A*	Total
Predicted EPQ grade	E	5	0	0	0	0	0	5 (0%)
	D	0	15	0	0	0	0	20 (0%)
	C	10	10	120	15	5	0	160 (2%)
	B	15	40	90	355	100	30	625 (7%)
	A	40	135	405	920	1970	1150	4620 (49%)
	A*	15	35	125	375	940	2420	3915 (42%)
	Total	85	240	740	1670	3010	3605	9345
		(1%)	(3%)	(8%)	(18%)	(32%)	(39%)	
b) EPQ: Percentages within predicted grades								
		Attained EPQ grade						
		E	D	C	B	A	A*	Total
Predicted EPQ grade	E	..	..	..	..	..	..	..
	D	..	..	..	..	..	..	..
	C	..	..	46%	..	..	..	100%
	B	..	13%	28%	44%	10%	1%	100%
	A	0%	2%	7%	29%	51%	10%	100%
	A*	..	0%	1%	7%	46%	47%	100%
	Total	1%	3%	7%	21%	45%	24%	100%

1 *Supplementary Table 5: Comparison of predicted and forecasted SQA*  
 2 *Highers and SQA Advanced Highers in medical school applicants, 2010-2018*  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16

17 **a) SQA Advanced Highers: Counts of number of cases**

		Attained SQA Highers							
		D7	C6	C5	B4	B3	A2	A1	Total
	D	<b>90</b>	0	0	0	0	0	0	95 (0.6%)
Predicted	C	20	<b>210</b>	<b>220</b>	15	10	10	0	485 (3.3%)
SQA	B	95	140	190	<b>455</b>	<b>490</b>	305	30	1700 (11.6%)
Highers	A	255	495	905	1405	2010	<b>5335</b>	<b>1955</b>	12360 (84.4%)
	Total	<b>465</b>	<b>845</b>	<b>1320</b>	<b>1875</b>	<b>2510</b>	<b>5645</b>	<b>1985</b>	14640 (100%)
	Total	3.2%	5.8%	9.0%	12.8%	17.2%	38.6%	13.6%	

31 **b) SQA Advanced Highers: Percentages within predicted grades**

		Attained SQA Highers							
		D7	C6	C5	B4	B3	A2	A1	Total
	D	<b>97%</b>	..	..	..	..	..	..	100%
Predicted	C	..	<b>43%</b>	<b>45%</b>	..	..	..	..	100%
SQA	B	6%	8%	11%	<b>27%</b>	<b>29%</b>	18%	2%	100%
Highers	A	2%	4%	7%	11%	16%	<b>43%</b>	<b>16%</b>	100%
	Total	3.2%	5.8%	9.0%	12.8%	17.2%	38.6%	13.6%	

*Supplementary Table 6: Correlation matrix of selection measures, undergraduate outcome measures, and postgraduate outcome measures (separated by grey lines for clarity). Cells indicate Pearson correlation and N. (NB presented as figure 3 in the main paper).*

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

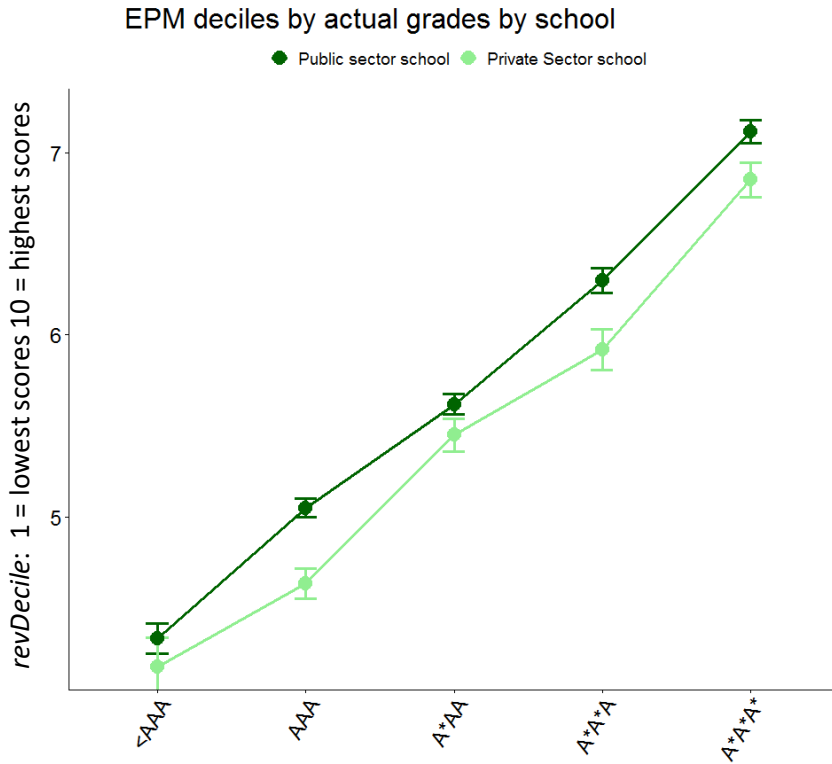
59

60

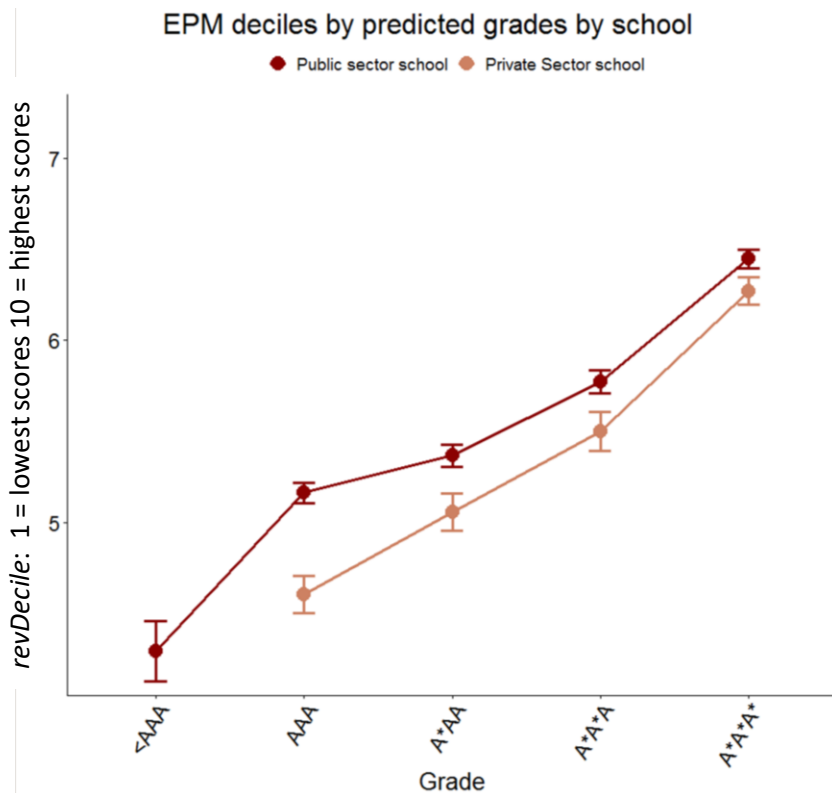
		Selection measures applicants					Undergraduate outcome measures			Postgraduate outcome measures	
		GCSE grades	Predicted Alevels	Alevel grades	UKCAT	BMAT	EPM	SJT	PSA	MRCP(UK) Part 1	MRCS Part A
	GCSE grades	1	0.452	0.421	0.265	0.223	0.180	0.190	0.201	0.212	0.173
	Predicted A-level grades	0.452	1	0.585	0.272	0.326	0.198	0.160	0.226	0.283	0.181
<b>Selection measures</b>	Attained A-level grades	0.421	0.585	1	0.326	0.416	0.297	0.195	0.306	0.421	0.358
<b>all applicants</b>	UKCAT total	0.265	0.272	0.326	1	0.483	0.115	0.243	0.238	0.200	0.181
	BMAT sections 1 and 2	0.223	0.326	0.416	0.483	1	0.089	0.239	0.321	0.378	0.319
	UKFPO EPM decile	0.180	0.198	0.297	0.115	0.089	1	0.319	0.470	0.509	0.535
<b>Undergraduate</b>	UKFPO SJT score	0.190	0.160	0.195	0.243	0.239	0.319	1	0.346	0.351	0.274
<b>Outcome measures</b>	PSA score	0.201	0.226	0.306	0.238	0.321	0.470	0.346	1	0.500	0.483
	MRCP(UK) Part 1	0.212	0.283	0.421	0.200	0.378	0.509	0.351	0.500	1	...
<b>Postgraduate</b>	MRCS Part A	0.173	0.181	0.358	0.181	0.319	0.535	0.274	0.483	...	1

Supplementary figure 1: Mean EPM revDeciles (95% CI) in relation to actual A-level grades (green) and predicted A-level grades (red), state sector schooling shown in darker colours and private sector schooling in paler colours.

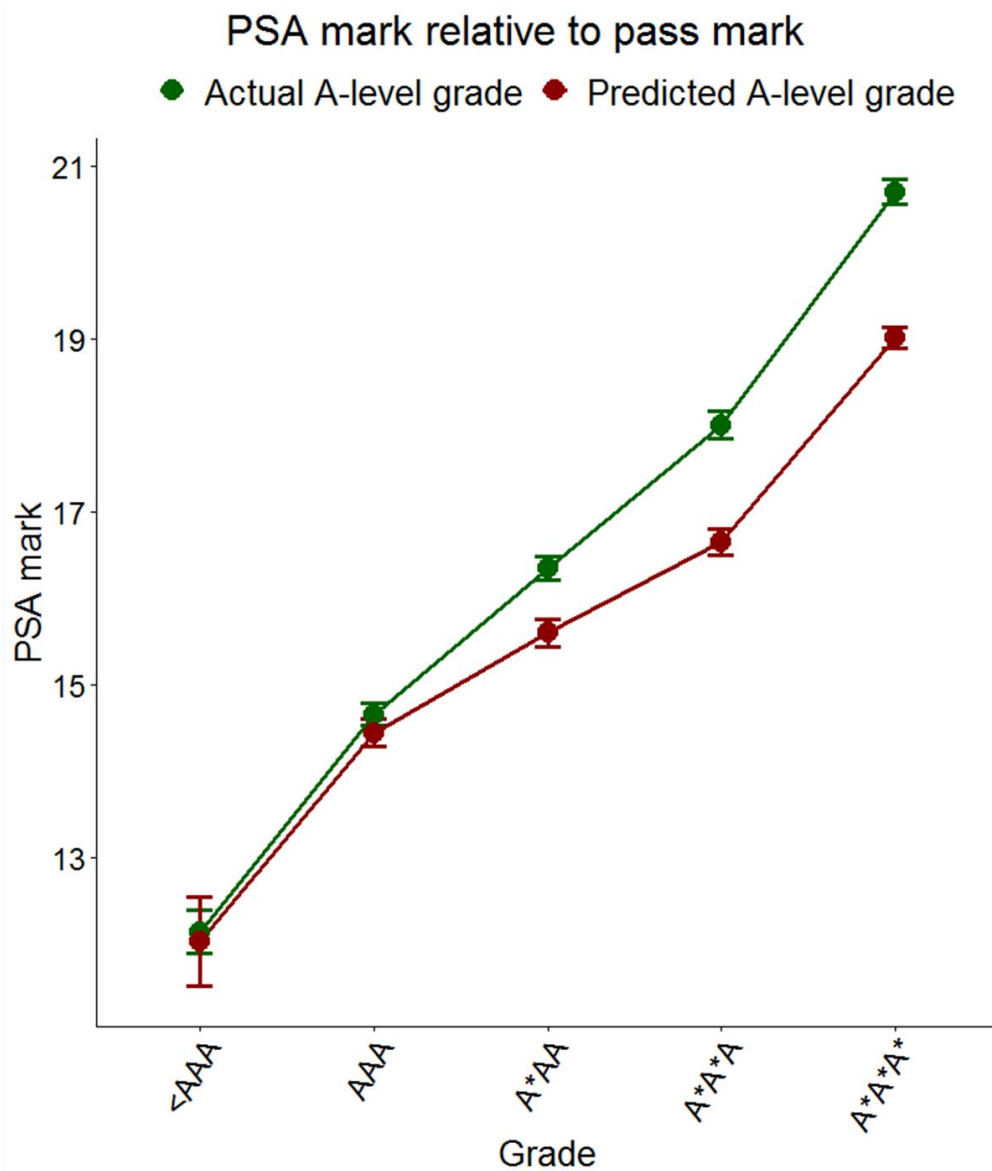
a.)



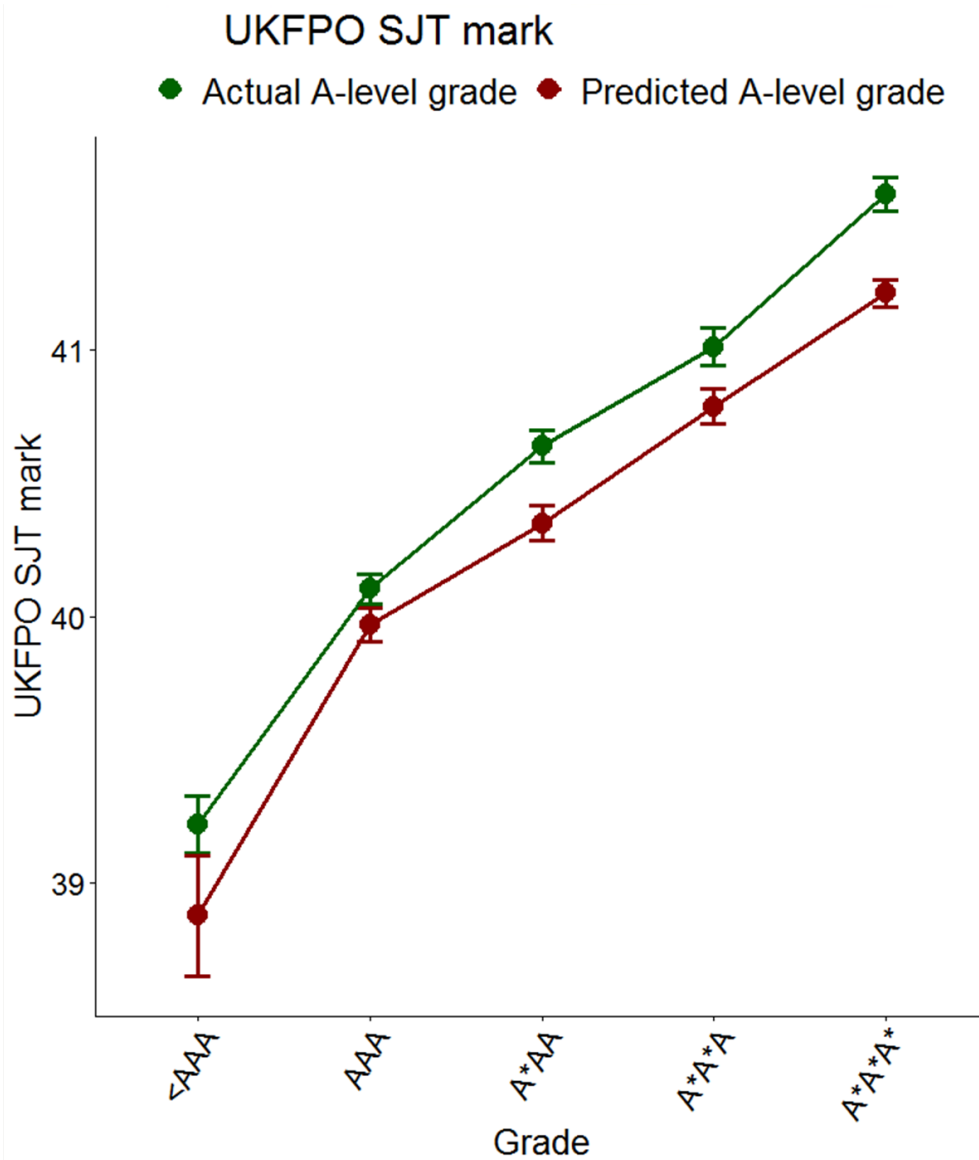
b.)



Supplementary figure 2: Mean PSA mark in relation to actual A-level grades (green) and predicted A-level grades (red)

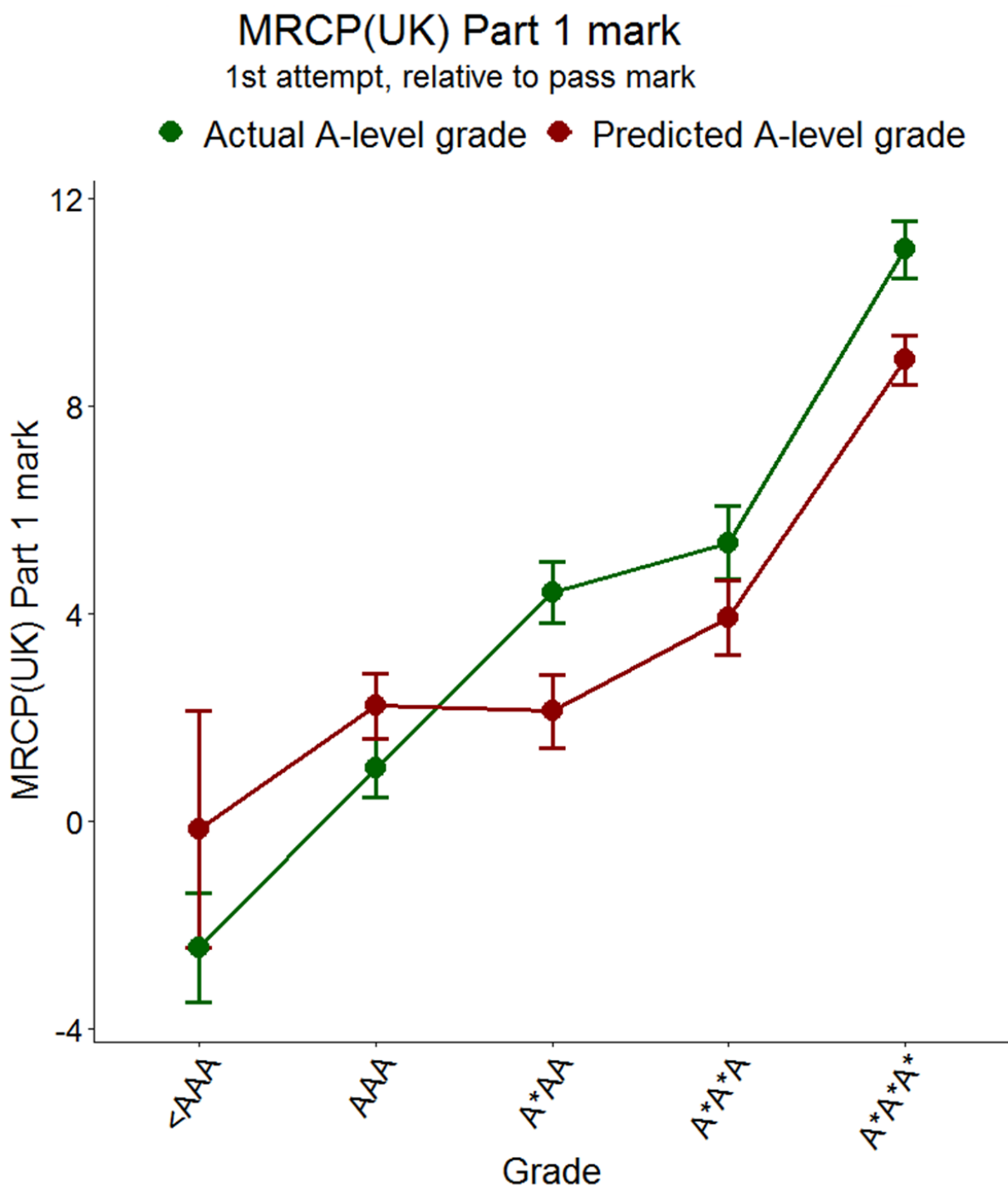


Supplementary figure 3: Mean SJT mark in relation to actual A-level grades (green) and predicted A-level grades (red)

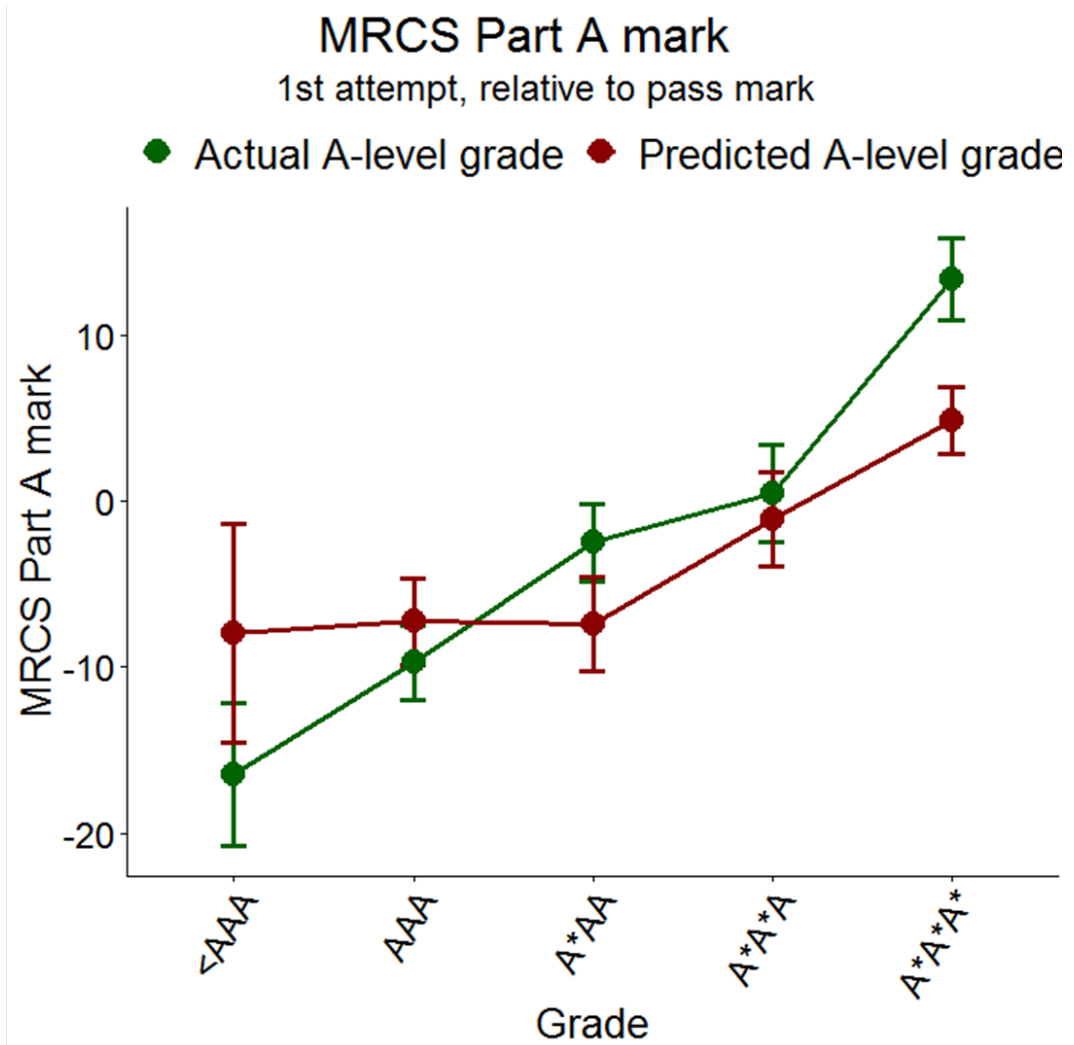




Supplementary figure 4: Mean MRCP(UK) Part 1 mark in relation to actual A-level grades (green) and predicted A-level grades (red)



Supplementary figure 5: Mean MRCS Part A mark in relation to actual A-level grades (green) and predicted A-level grades (red)



## Appendix:

### Are independent (private sector) schools more accurate in their A-level predictions?

A recurrent suggestion in the literature is that schools in the private sector (Independent Schools) are more accurate in their A-level predictions than those in the state sector. That suggestion raises many issues, not least concerned with social equity, but a key one to be resolved is whether the differences mainly are secondary to differences in overall attainment level, and as a result may be artefacts due to the ordinal nature of A-level grades and to A\* being the ceiling for A-level grades, and hence is a level beyond which candidates cannot reach. This appendix looks in detail at that question. The description is lengthy, technical, and partly didactic, and therefore has not been included in the main text or the main supplementary text. The conclusion is actually relatively simple: *Independent schools are not actually more accurate in their predictions, but they look that way because of having higher attained grades.*

*The data.* Appendix table 1 shows, in a similar format to those in the main paper, the A-level grades in P89 for applicants from state schools (defined as Comprehensives, Academies, Sixth Form Colleges, Tertiary Colleges and Technical Colleges) and private schools (defined as Independent Schools and Grammar Schools), with results restricted to first A-level attempts, duplicates removed, and where both A-level grades and predicted grades were available. Results are at the subject level, and therefore contain multiple subjects from individual applicants.

A simple glance at Appendix table 1 suggests that indeed Private Sector schools are more accurate, 53.3% of predictions being accurate compared to 45.5% of predictions from State Sector schools. Private schools also show a lower rate of over-prediction (40.5% compared with 48.0%), but not of under-prediction (6.2% vs 6.4%). These differences need however to be put into the context of higher overall attainment in private schools, where 29% of grades were A\* compared with 20% in state schools, a finding that reflects most private schools being selective and therefore inevitably taking higher ability entrants. Since attained A\* grades are more frequent in private schools, it is not surprising that predicted A\* grades are also more frequent in private schools, 48% vs 35%. The question therefore is whether the differences in accuracy are secondary to differences in overall performance. That question is best answered using *polychoric correlations*, which need description.

*Polychoric and tetrachoric correlations.* Polychoric and tetrachoric correlations are used frequently in psychometrics when dealing with binary and ordinal data. The need for them is shown by a simple 2x2 association table of the sort often tested using a chi-square test. Consider Appendix figure 1, which is a simple association table for characteristics P and Q in 100 individuals. 80% of cases have P present but only 50% of cases have Q present, meaning that the *marginal proportions* are not the same (80% vs 50%). A chi-square test is highly significant (chi-squared = 25, 1 df,  $p=0.0000006$ ) meaning that there is an association between P and Q. But what is the size of that association? Often in this situation a Pearson or Spearman correlation is calculated, and these give  $r_p=0.5$  and  $r_s=0.5$ , which suggests a moderately strong association.

However there is a problem in using the Pearson correlation, as a careful look at the table shows because the number of cases in which P is absent but Q is present, in the top right-hand corner, is zero. In other words the association could not be any stronger, but the correlation is still only 0.5, whereas a perfect correlation is usually taken as being 1. The problem arises because the marginal

1  
2  
3 proportions of P and Q are not the same, one being 0.5 and the other 0.8. If these two marginal  
4 proportions had been identical than all of the cases could have been on the diagonal and then the  
5 Pearson correlation would indeed have been 1. So what does one do in the case where the marginal  
6 proportions are not the same? The answer is another correlation developed by Pearson, called the  
7 *tetrachoric* or *polychoric* correlation for 2x2 or for larger tables respectively.  
8  
9

10 The tetrachoric correlation assumes that the data actually come from a bivariate normal distribution  
11 with some underlying correlation, and asks if that distribution were divided horizontally and  
12 vertically, what the correlation would have to be to create the contingency table that has been  
13 found. The lower part of Appendix figure 1 shows that diagrammatically<sup>1</sup>, the four quadrants  
14 containing the proportions of data in the contingency table. The calculation is easily carried out in  
15 the R function *polychor()* in the *polycor* library, and for the table in Appendix figure 1 it gives the  
16 answer that  $r_t=0.994$ , which effectively is  $r_t=1$ . The tetrachoric correlation therefore corresponds to  
17 our intuitive sense of what the correlation should be. The underlying bivariate normal distribution is  
18 assumed to have means of zero and standard deviations of one. *polychor()* then tells us that the  
19 thresholds for cutting the distribution need to be at 0.842 for P and 0 for Q. The threshold for Q at  
20 zero tells us that the cutting point is 0 standard deviations from the mean, and therefore 50% of  
21 cases are above the threshold and 50% below. The threshold for P is 0.842 standard deviations  
22 below the mean, and hence 20% of cases are below the threshold and 80% of cases above it. The  
23 marginal proportions of P and Q are then replicated.  
24  
25  
26

27 For a 2x2 table it is always possible to fit the tetrachoric correlation and the marginal proportions  
28 exactly. If the table is larger, giving a polychoric correlation, the marginal proportions and the cell  
29 frequencies cannot always be fitted exactly as the normal distribution may not be entirely  
30 appropriate, and in that case maximum likelihood estimates of the correlation and thresholds are  
31 found. The polychoric calculation for an  $m \times n$  table also provides a set of  $(m-1)$  and  $(n-1)$  thresholds  
32 for each of the variables, and it is possible to see if step sizes between the levels are equal.  
33 Polychoric correlations therefore are used for data where both measures are *ordinal* and for which it  
34 seems reasonable to assume an underlying latent distribution which is normal.  
35  
36  
37

38 *Polychoric correlations for A-level grades.* A-level grades are certainly at least ordinal in nature, but it  
39 is not clear that they are *equal interval*, the step from, say, D to C not necessarily being the same size  
40 as the step from B to A. Polychoric calculations allow the direct estimation of the step sizes between  
41 grades. If step sizes are not equal then many conventional statistics are not optimal. Equal interval  
42 scales are measures such as length, where the increments are identical in size (so the difference  
43 between, say, 2 cms and 3 cms is the same length as the difference between 10 cms and 11 cms).  
44 A-levels are often scored on a simple basis of allocating points, such as A\*=12, A=10, B=8, C=6, D=4  
45 and E=2 (and indeed we have done this elsewhere here), but that can sometimes be misleading in  
46 situations such as calculating correlations between actual and predicted grades, partly because  
47 marginal proportions are not the same, and partly because the data are *censored*, grades above A\*  
48 not being possible, however capable is a candidate, and hence over-prediction is not possible for  
49 estimated grades of A\*. In the case of a high ability group such as applicants to medical school the  
50 latter is problematic as state and private schools predict an A\* grade for 35% and 48% of exams. To  
51 put it another way, were a grade of A\*\* available then many examinees might have merited it [1],  
52 albeit probably more at private than state schools. There is also potentially a problem of computing  
53 total A-level scores (so that, say, AAA with 30 points is regarded as equivalent to A\*AB or A\*A\*C,  
54  
55  
56  
57  
58

59  
60  

---

<sup>1</sup> The correlation is actually drawn at 0.9 to make things pedagogically clearer, as a correlation of 1 is  
effectively a straight line.

which may not be exactly the case, although the approximation is probably good enough for most purposes).

*Fitting polychoric correlations to A-level grades from state and private schools.* The key question at present is whether private schools are more accurate in their predictions (53.3%) than state schools (45.5%) – see Appendix table 1. Accuracy can be considered in two ways, as the presence of systematic error (technically, ‘bias’), equivalent to rates of A\* etc being different in two groups, and random error, in terms of the correlation or lack of correlation between two sets of scores. Although the overall accuracy of private schools is *higher* than state schools, the correlation of predicted and actual grades is *lower* in private schools, with Pearson correlations of 0.635 in state schools and 0.552 in private schools (Appendix table 1), with a similar pattern for Spearman correlations. That suggests a potential problem in interpreting the data. Calculating the polychoric correlations suggests a very different picture, since the polychoric correlations in state schools ( $r_t = 0.717$ ) and private schools ( $r_t = 0.678$ ) are far more similar, particularly in comparison with the differences between the Pearson (or Spearman) correlations.

Interpreting the polychoric correlations is helped by a diagram. Appendix figure 2.a may look complex, but it summarises a lot of information about state sector applicants. The axes are on a normal distribution for the underlying latent scale, and so the units are standard deviations, from -4 to +4 SDs. Note these are not SDs for the raw data, but for the latent distribution. The polychoric correlation for the state sector is 0.717, and that is shown by the blue ellipse which is plotted to cover 99.9% of the data, which is reasonable given the large sample sizes. The dashed blue and yellow line on the diagonal is the line of equality for attained grades on the horizontal axis and predicted grades on the vertical axis. The vertical and horizontal lines show the thresholds separating the various A-level grades for attained and predicted grades. Appendix table 2 summarises the various thresholds and their intervals for state and private schools. As an example, for attained grades, the threshold separating A from A\* (Appendix table 2, row 4, column A:A\*) is 0.83, and so the vertical line in Appendix figure 2.a separating A from A\* is at 0.83. Similarly the horizontal line for predicted grades separating A from A\* is at 0.39 (row 2 in Appendix table 2). The intersection of these two lines is shown by a large red circle, which is *below* the blue-yellow dashed line, which indicates that the threshold for attained grades is higher than the threshold for predicted grades, so that it is easier to be predicted an A\* than to attain an A\*. The other vertical and horizontal lines show the thresholds between B and A (B:A), C and B (C:B), D and C (D:C) and E and D (E:D). As for A\*:A, all of the intersections, shown as red dots, are below the dashed blue-yellow line of equality, showing that predicted grades are always more generous than attained grades. Row 6 of Appendix table 2 shows that on average the threshold for attained grades is 0.73 SDs lower than for predicted grades. The coloured boxes in Appendix figure 2.a are equivalent to the coloured boxes in appendix table 1, with grey indicating accuracy, green and blue indicating under-estimation, and red and yellow over-estimation. More of the figure is red or yellow than is blue or green, indicating the overall over-estimation by predicted grades. It is also clear from the figure that the differences between the thresholds are not equal. The width of D, from E:D to D:C, is smaller than the width of A (from B:A to A:A\*), these values being shown in row 10 of Appendix table 2 for predicted grades and row 12 for attained grades. The widths of E and A\* cannot be calculated as they stop either at minus infinity or plus infinity. It is clear that the scale is not equal interval, with less change being required to move from D to C than from B to A. Statistical analyses should take care therefore in assuming that the usual A\* to E scale of grades is equal interval, and can be averaged.

The key question for this appendix is the extent to which state and private sector predictions are different. Appendix figure 2.b shows an equivalent plot to Appendix figure 2.a but for private sector

1  
2  
3 A-levels. At a glance it is not easy to see any obvious difference, but it is important to remember that  
4 the latent scales for both graphs each have a mean of zero and SD of one. However looking carefully  
5 shows that the threshold for attained grades at A\* is at 0.55 for private sector students compared  
6 with 0.83 for state sector students (see rows 4 and 5 of table Appendix table 2). The threshold is  
7 lower for private sector students and hence more of these students will attain an A\*, as is the case in  
8 Appendix table 2. All of the thresholds for the private sector students are actually moved to the left  
9 compared with state sector students (and compare the sizes of the A\*A\* boxes and the EE boxes in  
10 the two figures. Appendix figure 3 summarises the thresholds more clearly for attained and  
11 predicted grades in state and private sector schools. All thresholds are shown on the same  
12 horizontal scale. Attained grades for private schools are to the right of predicted grades, shown by  
13 the thin blue diagonal lines (meaning an attained A\* is harder to get than a predicted A\*), and the  
14 same pattern is seen for state schools, and shown by the thin diagonal red lines. Private school  
15 attained grades are also to the left of state school attained grades, shown by a thin purple line (with  
16 thresholds lower for private school students meaning that they get more A\* grades). Similarly,  
17 private school predicted grades are also to the left of state school predicted grades, also shown by a  
18 thin purple line. A key feature of Appendix figure 3 is that the blue diagonal lines are parallel, the  
19 red diagonal lines are parallel and the purple diagonal lines are nearly parallel, meaning that the  
20 relationships of grade boundaries are the same in private and state schools, and for attained and  
21 predicted grades, but are merely slid along relative to one another. The state and private schools  
22 are therefore handling predicted grades in a way that is similar, and they are similar related in each  
23 case to attained grades.

24  
25  
26  
27  
28  
29 The widths of the boxes in Appendix figure 2 are therefore very similar in state and private sector  
30 students, and are shown in rows 9 to 16 of Appendix table 2, particularly in rows 10 and 11, which  
31 compare predicted grades in state and private schools, and rows 12 and 13 which compare attained  
32 grades in private schools. The main difference between the two types of school is shown in the  
33 mean columns of rows 1 and 2 and rows 4 and 5, their mean differences being shown in the final  
34 column. Overall the state schools have thresholds which for predicted grades are on average are  
35 0.47 SDs higher and for attained grades are 0.42 grades higher than for private sector schools  
36 (meaning that higher grades are harder to attain). These values are very similar and suggest that  
37 predictions in the two types of school are being carried out in a similar way, but the overall ability of  
38 private school students is higher, and that is reflected in the attained and predicted ways to a similar  
39 extent.

40  
41  
42  
43 The private schools students are therefore about 0.44 SDs higher on the latent scale than the state  
44 school students. As a result it is possible to plot state and private schools on the same graph  
45 (Appendix figure 4), with the only difference being that the private schools are further along the  
46 diagonal towards the top right corner. That difference accounts for all of the differences in the  
47 private and state school students, with all other differences in Appendix table 1 being artefacts of  
48 the artificial ceiling of the range at A\*. To put it another way, were attained grades to be the same  
49 in state and private schools then the accuracy and the degree of over-estimation would be the same  
50 in the two types of schools.

51  
52  
53 Conventional statistics comparing attained and predicted grades at A-level are inherently misleading,  
54 and suggest differences between groups which are probably not present, meaning that great care  
55 must be taken in interpretation.

- 56  
57  
58 1. McManus IC, Woolf K, Dacre JE: **Even one star at A level could be "too little, too late" for**  
59 **medical student selection.** *BMC Medical Education* 2008, **8:16**  
60 (<http://www.biomedcentral.com/1472-6920/8/16>).

Appendix table 1. Predicted vs Attained A-level grades in applicants from a) State Sector schools (non-Private schools) and b) Independent (Private sector) schools.

a) State Sector: Counts of number of cases								
		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	140	30	5	5	0	0	180 (0%)
	D	210	420	125	20	5	0	780 (1%)
	C	535	1050	1545	400	55	5	3600 (2%)
	B	500	1735	3910	5190	1090	115	12540 (9%)
	A	270	1330	6250	24310	36915	7425	76495 (53%)
	A*	25	75	395	3950	24725	21410	50580 (35%)
	Total	1680	4645	12230	33870	62785	28960	144175
		(1%)	(3%)	(8%)	(23%)	(44%)	(20%)	
		Under	Prediction:	Over	Correlations:			
		6.4%	Accurate	48.0%	Pearson	Spearman	Polychoric	
			45.5%		0.635	0.590	0.717	
a) Private Sector: Counts of number of cases								
		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	15	0	0	0	0	0	15 (0%)
	D	15	55	15	5	0	0	85 (0%)
	C	50	85	200	40	10	0	385 (1%)
	B	60	185	430	1085	335	30	2130 (3%)
	A	65	300	1650	8785	19500	3935	34235 (49%)
	A*	5	20	115	1420	15270	16635	33455 (48%)
	Total	205	640	2405	11340	35115	20600	70305
		(0%)	(1%)	(3%)	(16%)	(50%)	(29%)	
		Under-	Prediction:	Over	Correlations:			
		estimate	Accurate	40.5%	Pearson	Spearman	Polychoric	
		6.2%	53.3%		0.552	0.523	0.678	

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

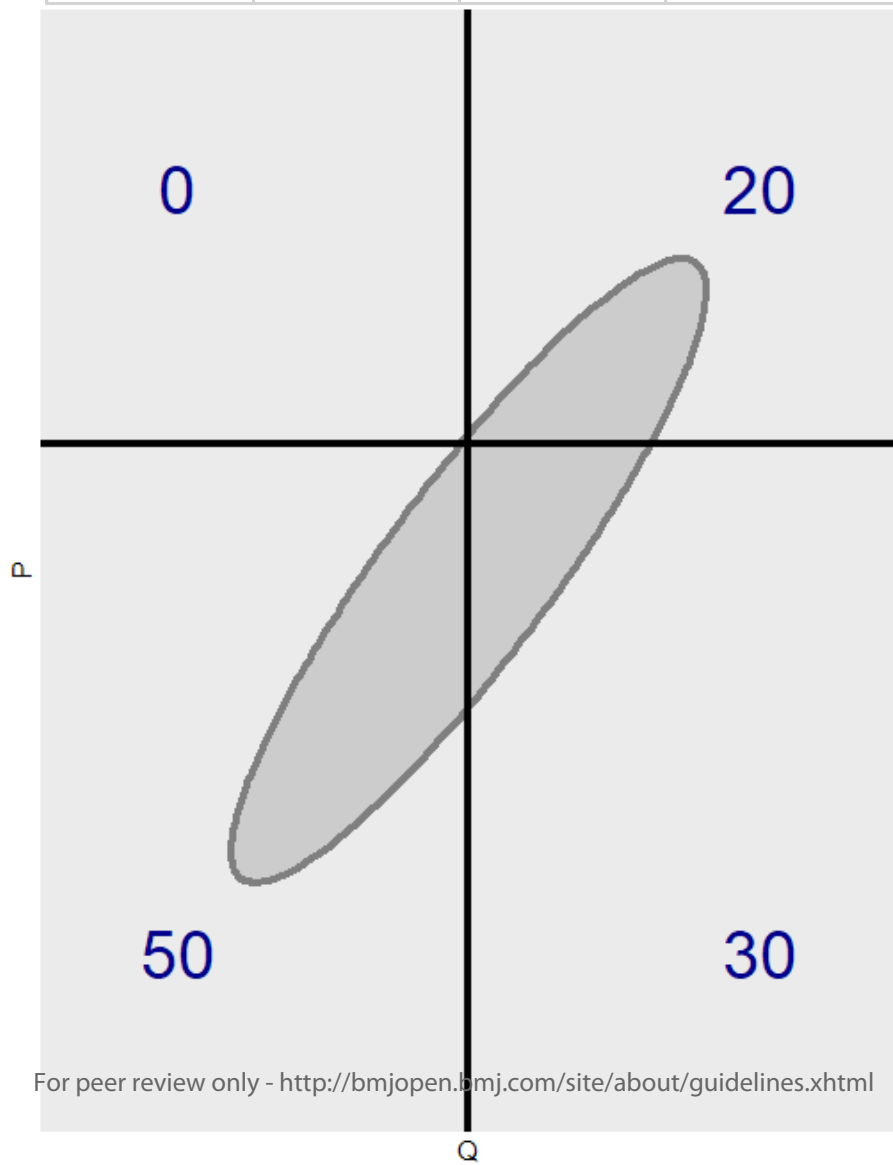
*Appendix table 2.* Thresholds, and intervals between thresholds, for the grades for applicants at State and Private schools. Values in bold show mean values across rows and down columns.

	<b>Ordinal</b>		E:D	D:C	C:B	B:A	A:A*	Mean	State minus Private
1	<b>Thresholds</b>								
2	Predicted	State	-3.11	-2.55	-1.89	-1.18	0.39	<b>-1.67</b>	0.47
3		Private	-3.51	-3.00	-2.47	-1.78	0.06	<b>-2.14</b>	
4	Attained	State	-2.31	-1.74	-1.13	-0.33	0.83	<b>-0.94</b>	0.42
5		Private	-2.57	-2.25	-1.68	-0.81	0.55	<b>-1.35</b>	
6	Predicted-Attained	State	-0.80	-0.81	-0.76	-0.85	-0.44	<b>-0.73</b>	0.02
7		Private	-0.76	-0.75	-0.79	-0.97	-0.48	<b>-0.75</b>	
8			<b>-2.18</b>	<b>-1.85</b>	<b>-1.45</b>	<b>-0.99</b>	<b>0.15</b>	<b>-1.26</b>	
9	<b>Threshold intervals</b>		D:C - E:D	C:B - D:C	B:A - C:B	A:A* - B:A			
10	Predicted	State	-0.57	-0.66	-0.71	-1.57		<b>-0.87</b>	0.02
11		Private	-0.51	-0.54	-0.69	-1.84		<b>-0.89</b>	
12	Attained	State	-0.57	-0.61	-0.80	-1.17		<b>-0.79</b>	-0.01
13		Private	-0.32	-0.58	-0.87	-1.36		<b>-0.78</b>	
14	Predicted-Attained	State	0.01	-0.05	0.09	-0.41		<b>-0.09</b>	-0.02
15		Private	-0.01	0.04	0.18	-0.48		<b>-0.07</b>	
16			<b>-0.33</b>	<b>-0.40</b>	<b>-0.47</b>	<b>-1.14</b>		<b>-0.58</b>	



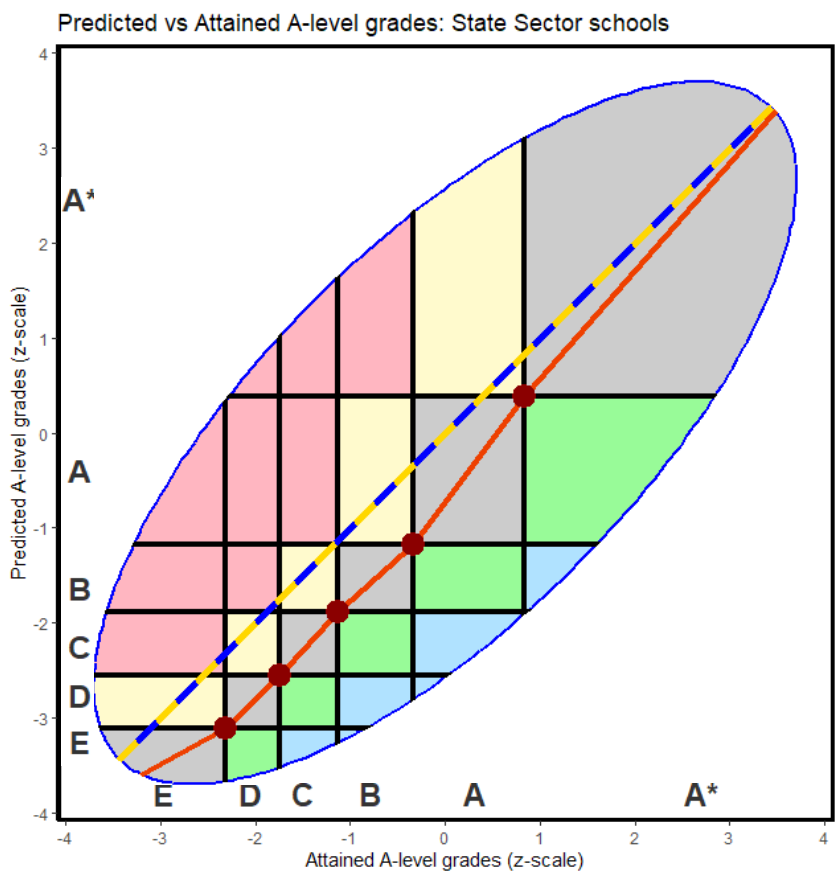
*Appendix figure 1.* Demonstration of how a conventional Pearson or Spearman correlation between binary variables P and Q cannot achieve a correlation of one when marginal proportions of P and Q differ. However the tetrachoric correlation is one, within calculation and rounding errors, being estimated from underlying latent correlation shown in the diagram, with thresholds at -0.842 and 0 for P and Q.

	Q absent	Q present	P totals
P absent	<b>0</b>	<b>20</b>	20
P present	<b>50</b>	<b>30</b>	80
Qtotals	50	50	100
Correlation	Pearson	Spearman	Tetrachoric
	0.5	0.5	0.994
Threshold	P	Q	
	-0.842	0	

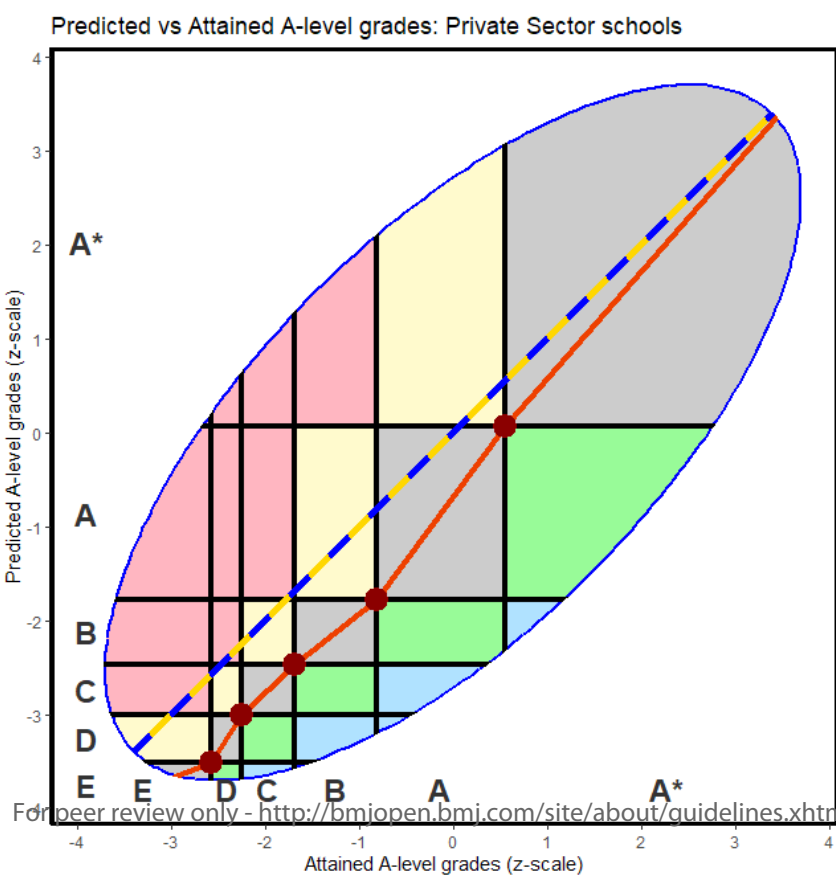


Appendix figure 2. Latent bivariate normal distribution of the relationship between attained A-level grades (horizontal) and predicted A-level grades (vertical). The correlation is represented by the blue ellipse. The dashed blue and yellow line is the line of equality of actual and attained grades. The vertical and horizontal black lines show the thresholds for the grades, shown as E, D, C, B, A and A\*. The solid red dots and red line show where the thresholds for a grade intersect, with all below the main diagonal. Colours indicate over-prediction (yellow and pink) and under-prediction (green and blue).

2.a



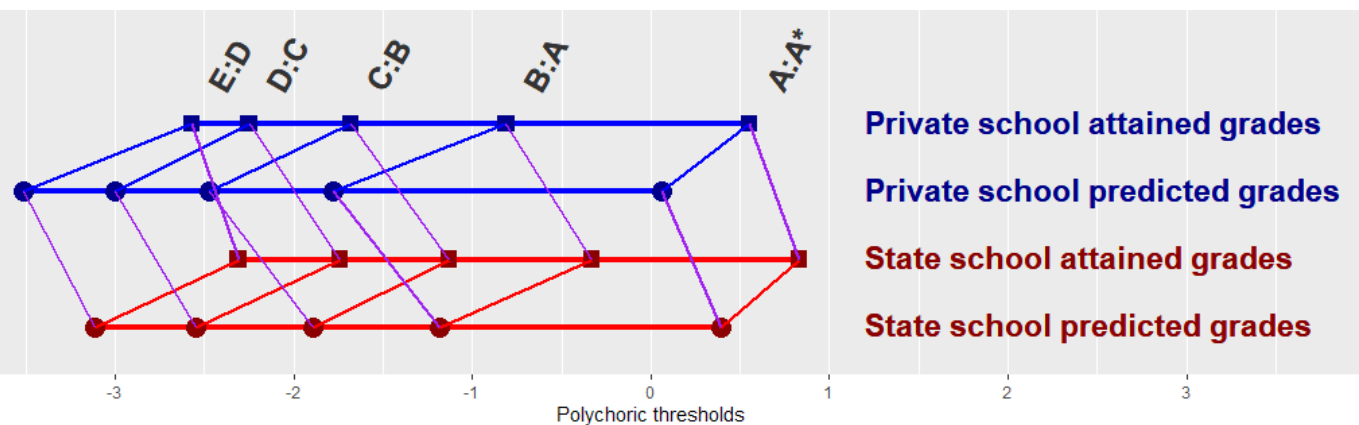
2.b



5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

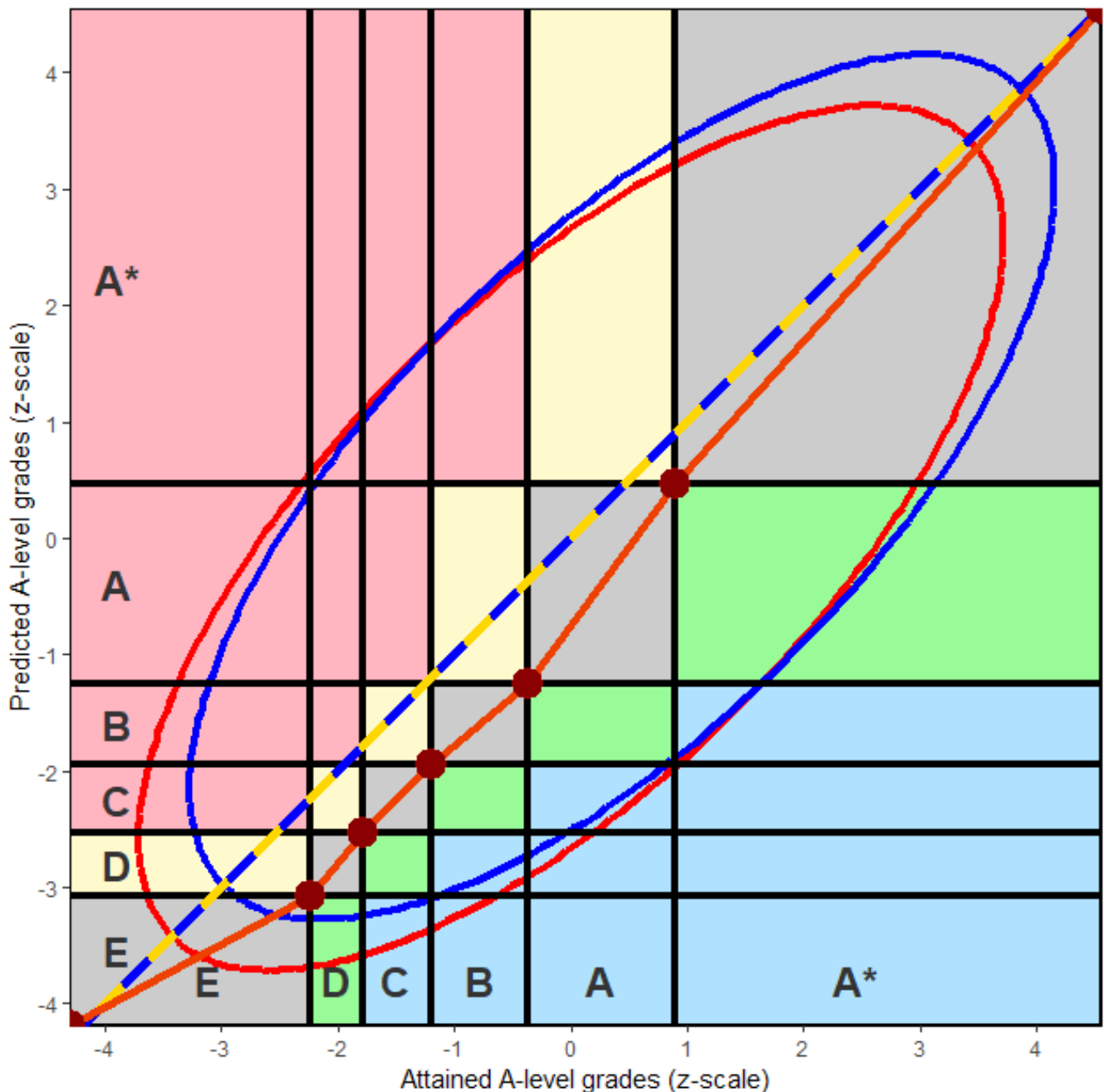
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Appendix figure 3. Summary of polychoric thresholds shown on the horizontal axis, for private (blue) and state (red) schools, for attained grades (squares) and predicted grades (circles). Narrower diagonal lines show the links between attained and predicted grades for private (blue) and state (red) schools. Purple diagonal lines link equivalent points for private and state schools (e.g. attained with attained grades and predicted with predicted grades).



Appendix figure 4. See Appendix figure 2 for the majority of conventions. The fitted ellipses for state sector schools (red) and private sector schools (blue) are shown separately, with the same grade thresholds for both schools. The latent bivariate normal distributions for the two types of school differ entirely in their mean scores, that for private sector schools being shifted up and to the right (by the same amount). The school types therefore differ only in their mean ability levels.

Predicted vs Attained grades: State Sector (red) & Private Sector (blue)



# BMJ Open

## The predictive validity of A-level grades and teacher-predicted grades in UK medical school applicants: A retrospective analysis of administrative data in a time of COVID-19

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-047354.R1
Article Type:	Original research
Date Submitted by the Author:	06-Jul-2021
Complete List of Authors:	McManus, Ian; UCL Medical School, Research Department of Medical Education Woolf, Katherine; UCL Medical School, Research Department for Medical Education Harrison, David; UCL Medical School, Research Department for Medical Education Tiffin, Paul; University of York, Department of Health Sciences; Hull York Medical School, Health Professions Education Unit Paton, Lewis; University of York, Cheung, Kevin; Cambridge Assessment Smith, Daniel; General Medical Council
<b>Primary Subject Heading</b>:	Medical education and training
Secondary Subject Heading:	Health services research
Keywords:	COVID-19, MEDICAL EDUCATION & TRAINING, EDUCATION & TRAINING (see Medical Education & Training)

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 The predictive validity of A-level grades and  
5 teacher-predicted grades  
6 in UK medical school applicants:  
7 A retrospective analysis of  
8 administrative data in a time of COVID-19  
9  
10  
11  
12  
13  
14  
15  
16

17 I C McManus <sup>a,b,e,\*</sup>

18 [i.mcmanus@ucl.ac.uk](mailto:i.mcmanus@ucl.ac.uk) ORCID 0000-0003-3510-4814

19 Katherine Woolf <sup>a,c,e</sup>

20 [k.woolf@ucl.ac.uk](mailto:k.woolf@ucl.ac.uk) ORCID 0000-0003-4915-0715

21 David Harrison <sup>a,e</sup>

22 [david.harrison@ucl.ac.uk](mailto:david.harrison@ucl.ac.uk) ORCID 0000-0002-6639-6752

23 Paul A Tiffin <sup>b,d,f</sup>

24 [paul.tiffin@york.ac.uk](mailto:paul.tiffin@york.ac.uk) ORCID 0000-0003-1770-5034

25 Lewis W Paton <sup>b,f</sup>

26 [lewis.paton@york.ac.uk](mailto:lewis.paton@york.ac.uk) ORCID 0000-0002-3328-5634

27 Kevin Yet Fong Cheung <sup>b,g</sup>

28 [Cheung.K@cambridgeenglish.org](mailto:Cheung.K@cambridgeenglish.org) ORCID 0000-0002-9548-2932

29 Daniel T. Smith <sup>a,b,h</sup>

30 [daniel.smith@gmc-uk.org](mailto:daniel.smith@gmc-uk.org) ORCID 0000-0003-1215-5811

31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

<sup>a</sup> UKMEDP089 project

<sup>b</sup> UKMEDP051 project

<sup>c</sup> UKMEDP089 Principal Investigator

<sup>d</sup> UKMEDP051 Principal Investigator

<sup>e</sup> Research Dept of Medical Education, University College London, London WC1E 6BT, UK

<sup>f</sup> Department of Health Sciences, University of York, Heslington, York, YO10 5DD, UK

<sup>g</sup> Cambridge Assessment, Shaftesbury Road, Cambridge CB2 8EA, UK

<sup>h</sup> General Medical Council, Regent's Place, 350 Euston Road, London NW1 3JN, UK

\* Author for correspondence: [i.mcmanus@ucl.ac.uk](mailto:i.mcmanus@ucl.ac.uk)

## Abstract

### Objectives

To compare in UK medical students the predictive validity of attained A-level grades and teacher-predicted A-levels for undergraduate and postgraduate outcomes. Teacher-predicted A-level grades are a plausible proxy for the teacher-estimated grades that replaced UK examinations in 2020 as a result of the COVID-19 pandemic. The study also models the likely future consequences for UK medical schools of replacing public A-level examination grades with teacher-predicted grades.

### Design

Longitudinal observational study using UK Medical Education Database (UKMED) data.

### Setting

UK medical education and training.

### Participants

Dataset 1: 81,202 medical school applicants in 2010 to 2018 with predicted and attained A-level grades. Dataset 2: 22,150 18-year old medical school applicants in 2010 to 2014 with predicted and attained A-level grades, of whom 12,600 had medical school assessment outcomes and 1,340 had postgraduate outcomes available.

### Outcome measures

Undergraduate and postgraduate medical examination results in relation to attained and teacher-predicted A-level results.

## Results

Dataset 1: Teacher-predicted grades were accurate for 48.8% of A-levels, over-predicted in 44.7% of cases and under-predicted in 6.5% of cases. Dataset 2: Undergraduate and postgraduate outcomes correlated significantly better with attained than with teacher-predicted A-level grades. Modelling suggests that using teacher-estimated grades instead of attained grades will mean that 2020 entrants are more likely to under-attain compared with previous years, 13% more gaining the equivalent of the lowest performance decile and 16% fewer reaching the equivalent of the current top decile, with knock-on effects for postgraduate training.

### Conclusions

The replacement of attained A-level examination grades with teacher-estimated grades as a result of the COVID-19 pandemic may result in 2020 medical school entrants having somewhat lower academic performance compared to previous years. Medical schools may need to consider additional teaching for entrants who are struggling, or who might need extra support for missed aspects of A-level teaching.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## Strengths and limitations of this study

- This is the first comparison of the predictive validity of teacher-predicted and attained A-level grades for performance in undergraduate and postgraduate assessments five to eight years later.
- The large sample size of all UK medical applicants from 2010 to 2018 provides adequate statistical power, and the complete population data means the results are unlikely to be biased.
- The teacher-predicted grades are those provided by schools as a part of university application, and probably form a good proxy for the “centre-assessment grades”, introduced by Ofqual during the Covid crisis of 2020.
- This study is with medical school applicants only, so that generalisability to students on other university courses is uncertain; however the over-prediction of grades we find in medical school applicants is similar to that found elsewhere for university applicants in general.

## Background

“... the ... exam hall [is] a level playing field for all abilities, races and genders to get the grades they truly worked hard for and in true anonymity (as the examiners marking don't know you). [... Now we] are being given grades based on mere predictions.” Yasmin Hussein, letter to *The Guardian*, March 29<sup>th</sup> 2020 <sup>1</sup>.

“[Let's] be honest, this year group will always be different...” Dave Thomson, blogpost on *FFT Educational Lab* <sup>2</sup>

“One headmistress commented that ‘entrance to university on teachers’ estimates may be fraught with unimagined difficulties’. ... If there is in the future considerable emphasis on school assessment, some work of calibration is imperatively called for.” James Petch, December 1964<sup>3</sup>.

UK schools closed on 20<sup>th</sup> March 2020 in response to the COVID-19 pandemic, and Key Stage 5 [Level 3] public examinations such as A-levels and SQA assessments were cancelled for summer 2020, and replaced by a complex system involving teacher assessments of the grades students would have achieved had they taken the examinations. A-levels and SQA assessments, like other national examinations in the UK, are normally set and marked anonymously by examination boards which are entirely separate from schools, and teachers usually play no part in this external assessment process. A-levels are good predictors of performance at university in general<sup>4</sup>, and at medical schools specifically<sup>5,6</sup>. Within this context the present paper compares achieved A-level grades with teacher-predicted grades, and in particular considers their relative predictive validities for educational outcomes at UK medical schools. The analyses were originally described in May 2020 and published as a preprint <sup>7</sup> while events were still ongoing and outcomes were not known. The present paper maintains much of that structure, and while mostly looking forward from 2020, also in part looks back from the perspective of 2021, meaning that past, present and future tenses are intermingled.

On April 3<sup>rd</sup> 2020 *Ofqual* (Office of Qualifications and Examinations Regulation) in England announced that A-level, GCSE and other exams under its purview, would be replaced by *Calculated Grades*, at the core of which are teachers' estimates of the grades that their students would attain (called *Centre Assessment Grades, CAGs*), which would then be moderated by *Ofqual* using a computer algorithm which included the prior performance of the school attended by candidates – see the Calculated Grades subsection below for details. The Scottish Qualification Authority (SQA) and other national bodies also announced similar processes for their examinations. Inevitably the announcement of Calculated Grades resulted in confusion and uncertainty in examination candidates, particularly those needing A-levels or SQA Advanced Highers<sup>8</sup> to meet conditional offers for admission to university in autumn 2020. Universities also faced a major problem for student selection, having had A-levels taken away, which are, “the single most important bit of information [used in selection]” <sup>8</sup>.

Some of the tensions implicit in Calculated Grades are well seen in the quotation above by Yasmin Hussein, a GCSE student in Birmingham, with its clear emphasis that a key strength of current

---

<sup>a</sup> SQA Highers are taken the year before (rather like AS-levels used to be) and therefore they will be available for 2020 applicants. Advanced Highers will not be available and will be estimated.

1 examination systems such as GCSEs, A-levels and similar qualifications, is their *anonymity* and  
2 *externality* with assessors who know nothing of the students whose work they are marking. In  
3 contrast the replacement of actual grades attained in the exam hall with what Hussein describes as  
4 'mere predictions' raises a host of questions, not the least being the possibility of bias when  
5 judgements are made by teachers.  
6  
7

## 8 Context of the current paper and the situation at the time of writing 9

10 Since the appearance of COVID-19 in Europe in early 2020, the situation has been and still is rapidly  
11 changing. As mentioned earlier, this paper was originally written in May 2020, but was revised and  
12 submitted to the journal, essentially as the preprint but with some additions, in November 2020  
13 when Europe was in the midst of a 'second wave' and England, Wales, Scotland and Northern  
14 Ireland, in a second national lockdown. The paper took almost six months to be reviewed, with  
15 revisions only being requested in May 2021 with the third UK national lockdown still not ended. To  
16 help the reader situate the current paper we explain briefly here what the exams situation was in  
17 the UK from April to August 2020, with more details provided in a postscript in Section 1 of the  
18 Supplementary Information.  
19

20  
21 University selection in the UK for admission in October 2020 began in the autumn, with medical  
22 school applicants submitting by October 15<sup>th</sup> to UCAS applications for four medical schools. Selection  
23 which may include interviews and other assessments is usually completed by the end of March with  
24 students being told of offers or rejections. Offers are usually conditional on A-levels and other  
25 qualifications to be taken in May, with results announced in August. In Spring 2020 as UK universities  
26 entered the final phases of the annual academic cycle of student selection, the present paper  
27 considered the potential problems of using teacher-estimated grades such as the Calculated Grades  
28 proposed by Ofqual, rather than attained grades obtained in the usual way via examinations. The  
29 pre-print of May 2020 was circulated primarily for information to medical school admissions tutors.  
30 By August 2020 some immediate effects on selection were shown when the algorithms used by  
31 regulators resulted in many students, particularly those from historically poorly performing schools,  
32 having their expected results adjusted downwards. This forced the Scottish Government, followed  
33 then by the English and Welsh Governments, to accept either teacher-estimated Centre Assessment  
34 Grades (CAGs) without moderation by an algorithm, or the Calculated Grade, whichever was the  
35 higher.  
36  
37  
38  
39

40 As expected in the pre-print, given that teacher-estimated grades were found to be higher than  
41 attained A-level grades, the scrapping of the algorithm resulted in a significant increase in grades  
42 compared to 2019<sup>b</sup>, with an immediate impact on the numbers of students meeting university  
43 conditional offers. Longer-term impacts are still to be seen, with some likely to result from the lower  
44 predictive validity of teacher-estimated grades, and a likely increase in under-performing students in  
45 medical schools and postgraduate training.  
46  
47

## 48 Medical school admissions 49

50 This paper mainly concentrates on medical school applications. UK medical education has a range of  
51 useful educational measures, including admissions tests during selection, and outcomes at the end  
52 of undergraduate training, which are linked together through UKMED (United Kingdom Medical  
53 Education Database; <https://www.ukmed.ac.uk/>). UKMED provides a sophisticated platform for  
54 assessing predictive validity in multiple entry cohorts in undergraduate and postgraduate training<sup>9</sup>.  
55 The current paper should also be read in parallel with a second study from some members of the  
56  
57

58  
59 <sup>b</sup> <https://ffteducationdatalab.org.uk/2020/08/gcse-and-a-level-results-2020-how-grades-have-changed-in-every-subject/>  
60

1 present team which assesses attitudes and perceptions to calculated grades and other changes in  
2 selection of current medical school applicants in the UKMACS (UK Medical Applicants Cohort Study)  
3 <sup>10 11</sup>.

4  
5 Fundamental questions about selection in 2020 concerned the likely nature of Calculated Grades,  
6 and the extent to which they would predict outcomes to the same extent as currently did *actual or*  
7 *attained grades*. The discussion will involve actual grades, and then four types of teacher-estimated  
8 grades: predicted grades (sent to UCAS at application to university), centre assessment grades (CAGs  
9 – submitted by schools to *Ofqual* in 2020) calculated grades (CAGs adjusted using an algorithm) and  
10 forecasted A-level grades (submitted by teachers to exam boards pre-2015 as a quality check for real  
11 exam grades). These related but different assessments are summarised in Box 1 below, together  
12 with final grades, which were the grades eventually accepted by UCAS and were the higher of the  
13 calculated grade or centre assessed grade. It should be noted that we have tried to use ‘teacher-  
14 predicted’ grades only to refer to the grades included as a part of the normal UCAS process, whereas  
15 the term teacher-estimated grades is used in a more generic sense.  
16  
17  
18  
19

## 20 Calculated grades

21 The status of calculated grades was made clear by *Ofqual* in April 2020:

22  
23 “The grades awarded to students will have equal status to the grades awarded in other years and  
24 should be treated in this way by universities, colleges and employers. On the results slips and  
25 certificates, grades will be reported in the same way as in previous years”. <sup>12</sup>, p.6.  
26

27 The decisions of *Ofqual* are supported by Ministerial statement, and universities and other bodies  
28 have little choice therefore but to abide by them, although that does not mean that other factors  
29 may not need to be taken into account in some cases, as often occurs when applicants do not attain  
30 the grades in conditional offers.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

None of the above means that calculated grades actually *will be* equivalent to conventional attained grades. Calculated grades will not actually *be* attained grades, they may well behave differently to attained grades, and in measurement terms they actually *are not* attained grades, even though in administrative and even in legal terms, by fiat, they have to be treated as equivalent. From the perspective of educational research, the key issue is the extent to which calculated grades actually will or can behave in an identical way to attained grades.

**Box 1: A-level grades: Actual, predicted, centre assessment, calculated, final, forecasted, and teacher-estimated grades**

**Actual or attained grades.** The grades awarded by examination boards/awarding organisations based on written and other assessments which are set and marked externally. Typically sat in *May and June of year 13*, with results announced in *mid-August*.

**Predicted grades.** Teacher estimates of the likely attained grades of candidates, provided to UCAS in the *first term of year 13*, and by *October 15<sup>th</sup>* for medical and some other applicants.

**Centre assessment grades.** Used in the production of Calculated grades (see below). Provided by examination centres (typically schools) between 1<sup>st</sup> and 12<sup>th</sup> of June 2020, consisting of teacher-estimated grades and candidate rankings within examination centres.

**Calculated grades.** The final grades to be provided for candidates by exam boards for Summer 2020 assessments, in the absence of attained grades. Based on centre assessment grades, with final calculated grades involving standardisation/adjustment by exam boards using an algorithm. Calculated grades, “will have equal status to the grades awarded in other years and should be treated in this way by universities, colleges and employers” (Ofqual). These grades were often referred to as the ‘algorithm grades’ and were abandoned by the UK governments in August 2020

**Final grades.** The grades used by UCAS in the 2020 admissions cycle – the higher of the teacher estimated grade or the centre assessment grade

**Forecasted grades.** Prior to 2015, teachers, in *May of Year 13*, provided to exam boards a forecast of the likely grades of candidates along with rankings. Forecasted grades therefore take place later in the academic cycle than predicted grades, close to the time examinations are actually sat.

**Teacher-estimated grades.** Generic term used in this paper to refer to grades estimated by teachers. Includes predicted grades, centre assessment grades, calculated grades, and forecasted grades.

In April 2020 *Ofqual* issued guidance on how *calculated grades* would be provided for candidates for whom examinations have been cancelled. Essentially, teachers would be required, for individual candidates taking individual subjects within a *candidate assessment centre* (usually a school), to estimate *grades for candidates*, and then to *rank order* candidates within grades, to produce *centre assessment grades*. A statistical standardisation process would then be carried out centrally using a computer algorithm. Ranking is needed because standardisation, “will need more granular information than the grade alone” (<sup>12</sup> p.7), presumably to break ties at grade boundaries which occur because of standardisation. Standardisation, to produce *calculated grades*, would use an algorithm that took into account the typical distribution of results from that centre for that subject in the three previous years, along with aggregated centre data on SATS and previous exam attainment as in GCSEs<sup>c</sup>. This approach is consistent with *Ofqual’s* approach to standard-setting. Following Cresswell<sup>13</sup>, *Ofqual* has argued that during times of change in assessments, and perhaps more

<sup>c</sup> It was this standardisation process that Governments reversed in August 2020 after the protests against calculated grades.

1 generally, there should be a shift away from “comparable performance” (i.e. criterion-referencing),  
2 and that there is an “ethical imperative” to use “comparable outcomes” (i.e. norm-referencing) to  
3 minimise advantages and disadvantages to the first cohort taking a new assessment, as perhaps also  
4 for later cohorts as teachers improve at teaching new assessments <sup>14</sup>.

5  
6  
7 *Ofqual* said that centre assessment grades, the core of calculated grades, “are not the same as ...  
8 predicted grades provided to UCAS in support of university applications” <sup>15</sup>, (p.7). Predicted grades  
9 in particular are provided by schools in October of year 13 and centre assessment grades in  
10 May/June of year 13, seven months later, when *Ofqual* says that teachers should also consider  
11 classwork, bookwork, assignments, mock exams and previous examinations such as AS-levels (taken  
12 only by a minority of candidates now), but should *not* include GCSE results, or any student work  
13 carried out after 20<sup>th</sup> March. Whether centre assessment grades, or calculated grades – centre  
14 assessment grades moderated by the algorithm - will be fundamentally different from predicted  
15 grades is ultimately an empirical question, which should be answerable when UCAS data for 2020  
16 are available for medical school applicants in UKMED. In the meantime, and *it is a core and a*  
17 *reasonable assumption*, that centre assessment grades and hence calculated grades will probably  
18 correlate highly with earlier predicted grades, except for a small proportion of candidates who have  
19 improved dramatically from October 2019 to March 2020. Predicted grades, which have been  
20 collected for decades, should therefore act as a reasonable proxy in research terms for centre  
21 assessment grades and therefore calculated grades, particularly in the absence of any other  
22 information.  
23  
24  
25  
26

### 27 The rationale for using A-level grades in selection

28  
29 Stepping back slightly it is worth revisiting the reasons that A-levels exist and why universities use  
30 them in selection. A-levels assess at least three things: subject knowledge, intellectual ability, and  
31 study habits such as conscientiousness <sup>16</sup>. Knowledge and understanding of, say, chemistry is  
32 probably necessary for the high level study of medical science and medicine, to which it provides an  
33 underpinning, and experience suggests that students without such knowledge may have problems.  
34 A-levels also provide evidence for a student’s intellectual ability and capability of extended study at  
35 a high level. A-levels are regarded as a ‘gold standard’ qualification because of the rigour and  
36 objectivity of their setting and marking (see for example *Ofqual*’s ‘Reliability Programme’<sup>17</sup>). Their  
37 measurement is therefore *reliable*, and the presumption is that they are also *valid*, in some of the  
38 many senses of that word <sup>18-20</sup>, and as a result are *unbiased*. A crucial assumption is of *predictive*  
39 *validity*, that future outcomes at or after university are higher or better in those who have higher or  
40 better A-levels, as found both in predicting degree classes in general <sup>4 21 22</sup> and medical school  
41 performance in particular <sup>5 23</sup>. There is also an assumption of *incremental validity*, A-levels being  
42 better predictors than other measures<sup>6</sup>. At the other extreme, A-levels could be compared  
43 conceptually with, say, a mere assertion by a friend or colleague that, “Oh yes, they know lots of  
44 chemistry”. That is likely neither to be reliable, valid nor unbiased, and hence is a base metal  
45 compared with the gold standard of A-levels. The empirical question therefore is where on the  
46 continuum from gold to base metals, lie calculated grades or teacher-predicted grades.  
47  
48  
49

50  
51 The issue of predictive validity has been little discussed in relation to calculated grades, but in a *TES*  
52 (*Times Educational Supplement*) survey of teachers, there were comments that, “predictions and  
53 staff assessments would never have the same validity as an exam”, so that, “Predictions, past  
54 assessment data and mock data is not sufficient, and will never beat the real thing in terms of  
55 accuracy” <sup>24</sup>. The changes in university selection inevitably meant that difficult policy decisions  
56 needed to be made by universities and medical schools. Even in the absence of direct, high-quality,  
57 evidence, policy-makers still have an obligation to make decisions, and, therefore it is argued, must  
58 take theory, related evidence, and so on, into account <sup>25</sup>. This paper provides both a review of other  
59  
60

evidence, and also results on the related issue of predicted grades, which it will be argued are likely to behave in a way that is similar to calculated grades.

## Review of literature on predicted and forecasted grades

### Predicted grades in university selection

A notable feature of UK universities is that selection mostly takes place before A-levels or equivalent qualifications have been sat, so offers are largely conditional on later attained grades. As a result, UCAS application forms, since their inception in 1964, have included *predicted grades*, estimates by teachers of the A-level grades a student is likely to achieve. Admissions tutors also use other information in making conditional offers. A majority of applicants in England, applying in year 13 for university entry at age 18 will have taken GCSEs at age 16 in year 11, a few still take AS-levels in year 12, some students submit an EPQ (Extended Project Qualification), and UCAS forms also contain candidate statements and school references. Medical school applicants mostly also take admissions tests such as U(K)CAT or BMAT at the beginning of year 13, and many will take part in interviews or MMIs (multiple mini-interviews)<sup>d</sup>.

Predicted grades have always been controversial. A House of Commons Briefing Paper in 2019 noted that the UK was unusual among developed countries in using predicted grades<sup>e</sup>, and said that,

“The use of predicted grades for university admissions has been questioned for a long time. Many critics argue that predicted grades should not be used for university entry because they are not sufficiently accurate and it has been suggested that disadvantaged students in particular lose out under this system.”<sup>26</sup> p.4

Others have suggested that as well as being “biased”, “predicting A-level grades is clearly an imprecise science”<sup>27</sup> (p.418). There have been repeated suggestions over the years, none as yet successful, that predicted grades should be replaced with a PQA (Post-Qualification Applications) system. As Nick Hillman puts it,

“The oddity of our system is not so much that people apply before receiving their results; the oddity is that huge weight is put on predicted grades, which are notoriously unreliable. ... PQA could tackle this...”<sup>f</sup>.

The system of predicted grades is indeed odd, but also odd is the sparsity of academic research into predicted grades. The most important question that seems almost never to have been asked, and certainly not answered, is the fundamental one of whether it is predicted grades or actual grades which are better at predicting outcomes. Petch<sup>3</sup>, in his 1964 monograph which was one of the first serious discussions of the issues, considers that predicted and actual grades may be fundamentally different, perhaps being “complementary and not contradictory” (p.29), one being about scholarly attitude and the other about examination prowess, primarily because “the school knows the candidate as a pupil, knowledge not available to the examiners”. For Petch, either a zero correlation or a perfect correlation between predicted and actual grades would be problematic, the latter perhaps implying that actual grades might be seen as redundant (p.6).

The advent of *Ofqual's* calculated grades, which are in effect predicted grades carried out by teachers in a slightly different way, means there was a serious need in 2020 to know how effective

<sup>d</sup> see <https://www.medschools.ac.uk/studying-medicine/making-an-application/entry-requirements/>.

<sup>e</sup> <https://www.bbc.co.uk/news/education-44525719>

<sup>f</sup> <https://www.hepi.ac.uk/2019/08/14/pqa-just-what-does-it-mean/>



1 predicted grades were likely to be as a substitute for attained A-level grades, and the same concern  
 2 will apply in 2021, with *Ofqual* implementing a different model for teacher-estimated grades<sup>g</sup>. Are  
 3 teacher-predicted grades in fact ‘notoriously unreliable’, being ‘mere predictions’, or do they have  
 4 equivalent predictive validity as attained grades?  
 5

### 7 The research literature on predicted grades

8 As part of Section 1 of the Supplementary Information to this paper we have included a more  
 9 detailed overview of research studies on predicted grades. Here we will merely provide a brief set of  
 10 comments.  
 11

12 Most studies look at predictions at the level of individual exam subjects, which at A-level are graded  
 13 from E to A or, from 2010 onwards, from E to A\*. The most informative data show all combinations  
 14 of predicted grades against attained grades, and Figure 1 gives an example for medical school  
 15 applicants. Many commentators, though, look only at over-predictions (‘optimistic’) and under-  
 16 predictions (‘pessimistic’). Figure 2 summarises data from five studies of university applicants.  
 17 Accurate predictions occur in 52% of cases when A is the maximum grade and 17% when A\* is the  
 18 maximum grade (and with more categories accuracy is likely to be lower). Grades are mostly over-  
 19 predicted, in 42% of cases pre-2010 and 73% post-2010, with under-prediction rarer at 7% of cases  
 20 pre-2010 and 10% post-2010. A number of studies have reported that under-prediction is more  
 21 common in lower socio-economic groups, non-White applicants, and applicants from state school or  
 22 further education<sup>28-30</sup>. A statistical issue means such differences are not easy to interpret, as a  
 23 student predicted A\* cannot be under-estimated, and therefore under-estimation will inevitably be  
 24 more frequent in groups with lower overall levels of attainment. This issue is discussed and analysed  
 25 at length in Section 5 of the Supplementary Information in relation to applicants from private-sector  
 26 schools.  
 27

28 Some studies also consider grade-point predictions, the sum of grade scores for the three best  
 29 attaining subjects, scored A\*=12, A=10, B=8, etc<sup>h</sup>. In particular a large study by UCAS<sup>31</sup> showed that  
 30 applicants ‘missing their predictions’ (i.e. they were over-predicted) tended to have lower predicted  
 31 grades, lower GCSE attainment, were more likely to have taken physics, chemistry, biology and  
 32 psychology, and were from disadvantaged areas. To some extent the same statistical problems of  
 33 interpretation apply as with analysis at the level of individual exam subjects. For a number of years  
 34 UCAS only provided grade-point predictions, and they are included in the P51 data analysed below.  
 35

### 36 What are predicted grades and how are they made?

37 UCAS says that “A predicted grade is the grade of qualification an applicant’s school or college  
 38 believes they’re likely to achieve in positive circumstances.”<sup>i</sup> Later though, the document says  
 39 predicted grades should be, “**in the best interests of applicants** – fulfilment and success at college or  
 40 university is the end goal”, and “**aspirational but achievable** – stretching predicted grades are  
 41 motivational for students, unattainable predicted grades are not” (all emphases in original).  
 42 Predicted grades should be professional judgements and be data-driven, including the use of, “past  
 43 Level 2 and Level 3 performance, and/or internal examinations to inform ...predictions”.  
 44

45 <sup>g</sup> <https://www.gov.uk/government/publications/awarding-qualifications-in-summer-2021/awarding-qualifications-in-summer-2021>

46 <sup>h</sup> In some studies a scoring of A\*=6, A=5, B=4 is used. The 12,10,8... scoring was introduced so that AS levels,  
 47 weighted at half an A-level, could be scored as A=5, B=4 etc (there being no A\* grade at AS-level). For most  
 48 purposes A\*=12, A=10 ... is equivalent in all respects to A\*=6, A=5, etc, apart from a scaling factor.

49 <sup>i</sup> <https://www.ucas.com/advisers/managing-applications/predicted-grades-what-you-need-know> [Accessed  
 50 13th April 2020].  
 51

1 Few empirical studies have asked how teachers estimate grades, with not much progress since 1964  
2 when Petch said, "Little seems to be known about measures taken by schools to standardize  
3 evaluations of pupils"<sup>3</sup> (p.7). Two important exceptions are the studies of Child and Wilson<sup>j</sup> in 2015  
4 and Gill<sup>32</sup> in May 2018, with only the latter published. Gill sent questionnaires to selected OCR  
5 (Oxford, Cambridge and Royal Society of Arts) Examination Board exam centres concerning  
6 Chemistry, English Literature and Psychology exams. Teachers said the most important information  
7 used in predicting grades was performance in mock exams, observations of quality of work and  
8 commitment, oral presentation, the opinion of other teachers in the same subject and in other  
9 subjects, and the head of department. Some teachers raised concerns about the lack of high stakes  
10 for mock exams which meant that some students did not treat them seriously. AS-level grades were  
11 an important aid in making predictions, and there were concerns about the loss of AS-levels to help  
12 in prediction, as also mentioned elsewhere<sup>33</sup>, and that is relevant to 2020 where most candidates  
13 will not have taken AS-levels.  
14

15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Studies considered so far almost entirely are concerned with teacher predictions of A-level grades, since they are important for university admissions. More generally, studies looking at a wider range of teacher estimates, often in younger children, find a tendency for over-estimation across a range of skills<sup>34</sup>, with judgements often being systematically lower for marginalised learners<sup>35</sup>. A different position is taken in a genetically-informed study of twins, which suggests, in a forcefully worded conclusion, that, "Teachers can reliably and validly monitor students' progress, abilities and inclinations. ... For these reasons, we suggest that teacher assessments could replace some, or all, high-stakes exams"<sup>36</sup>. The study however uses only correlations as measures of accuracy, and cannot assess over- or under-estimation. Also, teacher ratings were only available at ages 7,11 and 14, at the same time as standardised tests are carried out, but were not available for GCSEs at age 16, or for A-levels and University Entrance at age 18, and as such are not informative for the purposes of the present study.

### Predicted grades in other Key Stage 5 qualifications than A-levels

40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Almost all studies on predicted grades have considered A-levels, with a few occasional exceptions looking at GCSEs. We know of no studies on the Extended Project Qualification (EPQ) in England, of Scottish Highers and Advanced Highers, or any other qualifications. Section 3 of the Supplementary Information includes data on both EPQ and SQA examinations.

### Forecasted grades

40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Until 2015, teachers in the May of school year 13 provided awarding organisations with *forecasted grades*, and those forecasts in part contributed to quality control of grades by the boards. Since forecasted grades were produced five to seven months after predicted grades, and closer to the exam date, they might be expected to be more accurate than predicted grades, being based on better and more recent information. Forecasted grades are important as they are more similar than predicted grades to the proposed calculated grades in the way they are calculated, and it is noted that "they may differ somewhat from the predicted grades sent to UCAS as part of the university application process"<sup>37</sup>. Three formal analyses are available, for candidates in 2009<sup>38</sup>, 2012<sup>39</sup> and 2014<sup>37</sup>, and four other studies from 1940<sup>40</sup>, 1963<sup>3</sup>, 1977<sup>41</sup> and 2018<sup>32</sup> are also available, with one post-2000 study before A\* grades were introduced and three after (Figure 2). Petch<sup>40</sup> also provides a very early description of forecasted grades, looking at teachers' predictions of pass or fail in School Certificate examinations in 1940, which also show clear over-prediction.

---

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

1 Forecasted A-level grades are similar in accuracy to predicted grades pre-2010 (42% vs 52%) but are  
2 more accurate post-2010 (47% vs 17%), in part due to a drop in accuracy of predicted grades when  
3 A\* grades are available. Despite there being *no aspirational or motivational reasons for teachers to*  
4 *over-predict forecasted grades*, particularly in the 1977 and 2018 studies, over-prediction  
5 nevertheless remains as frequent as with predicted grades (pre-2010: 39%; post-2010: 37%) and  
6 remains more common than under-prediction (pre-2010: 20%; post-2010 16%). Overall it is perhaps  
7 possible that calculated grades may be somewhat more accurate than predicted grades, but  
8 forecasted grades appear broadly in their behaviour to predicted grades. Two sets of forecasted  
9 grades are available for GCSEs<sup>42 43</sup>, and they show similar proportions of over and under-prediction  
10 as do results for A-levels. Over-prediction seems to be a feature of all predictions by teachers.

11 The three non-official studies of forecasted grades also asked teachers to rank-order candidates, a  
12 procedure which was included in calculated grades. The 1963 data<sup>3</sup> found a median correlation of  
13 rankings and exam marks within schools of 0.78, the 1977 data<sup>41</sup> a correlation of 0.66<sup>41</sup>, and the  
14 recent 2018 data<sup>32</sup> a correlation of about .82. The three estimates, mean  $r = 0.75$ , are somewhat  
15 higher than a meta-analytic estimate of .63 (SE = .03) for teachers' ability to predict academic  
16 achievement<sup>44</sup>.

17 The Gill study<sup>32</sup> is also of interest as one teacher commented on the difficulty of providing rankings  
18 with 260 students sitting one exam, and the author noted that, "it was easier for smaller centres to  
19 make predictions because they know individual students better" (p.42), with it also being the case  
20 that responses to the questionnaire were more likely to come from smaller centres. The 1963 study  
21 of Petch<sup>3</sup>, as well as commenting on "considerable divergencies ... in the methods by which  
22 estimates were produced" (p.27), as in the variable emphasis put on mock exams, also adds that,  
23 "some of the comments from schools suggested that at times there may be a moral ingredient  
24 lurking about some of the estimates"(p.28).

25 Overall it seems possible but unlikely that calculated grades might be more accurate than predicted  
26 grades, but they also make clear the problems shown by teachers in ranking and grading candidates.  
27 It also remains possible that examining boards have far more extensive and unpublished data on  
28 forecasted grades that they intend to use in assessing the likely effectiveness of calculated grades.

## 29 Applicants to medical school

30 So far, this review section has been entirely about university applicants across all subjects and the  
31 entire range of A-level grades. Only a handful of studies have looked at predicted grades in medical  
32 school applicants.

33 Lumb and Vail emphasised the importance of teacher-predicted grades since they determine in large  
34 part how shortlisting takes place<sup>45</sup>. In a study of 1995 applicants they found 52% of predictions were  
35 accurate, 41% were over-estimated and 7% under-estimated<sup>45</sup>, values very similar to those reported  
36 in university selection in general (Figure 2).

37 A study by one of the present team used path modelling to assess the causal inter-relationships of  
38 GCSE grades, predicted grades, receipt of an offer, attained A-level grades, and acceptance at  
39 medical school<sup>46</sup>. Predicted grades were related to GCSE grades (beta=0.89), and attained A-level  
40 grades were predicted by both GCSE grades (beta=0.44) and predicted A-level grades (beta=0.74).  
41 The study supports claims that teachers may well be using GCSE grades in part to provide predicted  
42 grades, which is perhaps not unreasonable given the clear correlation.

43 Richardson et al<sup>47</sup> in an important and seemingly unique study looked at the relative predictive  
44 validity of predicted as compared with attained A-level grades. Using a composite outcome of pre-  
45 clinical performance there was a minimal correlation with predicted grades ( $r=.024$ ) compared with  
46

1 a correlation of 0.318 ( $p < .001$ ) with attained A-level grades. To our knowledge this is the only study  
2 of any sort assessing the predictive validity of predicted vs attained A-level grades.  
3

## 4 The present study

5 Although calculated grades are novel and untested in their details, predicted grades have been  
6 around for half a century, and there is also a small literature on forecasted grades. This paper will try  
7 to answer several empirical questions about predicted grades, for which data are now available in  
8 UKMED. Predicted grades will then be used, *faute de mieux*, to make inferences about the likely  
9 consequence of using calculated grades.  
10  
11  
12

## 13 Empirical questions to be addressed.

### 14 *The relationship between predicted and attained grades in medical school applicants*

15 Few previous studies have looked in detail at this high-performing group of students. We will also  
16 provide brief results on Scottish Highers and Advanced Highers, and the EPQ (Extended Project  
17 Qualification), neither of which has been discussed elsewhere to our knowledge.  
18  
19  
20

### 21 *The predictive validity of predicted grades in comparison with attained grades*

22 A fundamental question concerning calculated grades is whether teacher-predicted grades are  
23 better or worse at predicting outcomes than are actual A-level grades. The relationship between  
24 predicted grades and actual grades cannot itself answer that question. Instead what matters is the  
25 relative performance of predicted and actual grades in predicting subsequent outcomes at the end  
26 of undergraduate or postgraduate training. The only relatively small study on this of which we are  
27 aware in medical students<sup>47</sup> found that only actual grades had predictive validity.  
28  
29  
30

## 31 Method

32 The method provided here is brief. A fuller description including a detailed table of measures can be  
33 found in Section 2 of the Supplementary Information. Overall the project is **UKMEDP112**, approved  
34 by the UKMED Research Group in May 2020, with data coming from two separate but related  
35 UKMED projects, both of which included predicted grades.  
36  
37  
38

39 Project **UKMEDP089**, “The UK Medical Applicant Cohort Study: Applications and Outcomes Study”,  
40 approved Dec 7<sup>th</sup>, 2018, with Dr Katherine Woolf as principal investigator, is an ongoing analysis of  
41 medical student selection as a part of UKMACS (UK Medical Applicant Cohort Study<sup>k</sup>). The data  
42 upload of 21<sup>st</sup> Jan 2020 included detailed information from UCAS and HESA on applicants for  
43 medicine from 2007 to 2018.  
44  
45

46 Project **UKMEDP051**, “A comparison of the properties of BMAT, GAMSAT and UKCAT”, approved  
47 Sept 25<sup>th</sup>, 2017, with Dr Paul Tiffin as principal investigator, is an ongoing analysis of the predictive  
48 validity of admissions tests and other selection methods such as A-levels and GCSEs in relation to  
49 undergraduate and postgraduate attainment. The present analysis used the download files dated  
50 13<sup>th</sup> May 2019<sup>l</sup>. UCAS data are included, although when the present analysis began the file had not  
51 yet included the detailed subject level information available in UKMEDP089<sup>m</sup>. Outcome data for the  
52 P51 dataset are extensive, and in particular undergraduate progression data are included, such as  
53  
54  
55  
56

57 <sup>k</sup> <https://ukmacs.wordpress.com/>

58 <sup>l</sup> UKCAT51\_APP\_ALL\_DATA\_13052019\_FILE1.SAV and UKCAT51\_APP\_ALL\_DATA\_13052019\_FILE2.SAV .

59 <sup>m</sup> An upload for P51 was made available on 20<sup>th</sup> April 2020 but was not included in the present analyses.  
60

1 UKFPO EPM and SJT, and PSA (Prescribing Safety Assessment), as well as performance on some  
2 postgraduate examinations (MRCP Part1 and MRCS Part A).  
3

4 Data from HESA and hence UKMED are required to be reported using their rounding and suppression  
5 criteria ([https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-](https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics)  
6 [anonymise-statistics](https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics)) and those criteria have been used for all UKMED data. In particular the  
7 presence of a zero or the absence of a percentage may not always mean that there are no  
8 individuals in a cell of a table, and all integers are rounded to the nearest 5.  
9

## 11 Results

12 A fuller description of the results can be found in Section 3 of the Supplementary Information.  
13

### 14 The relationships between predicted and actual grades in medical school 15 applicants.

#### 16 Predicted and actual A-level grades for individual A-level examinations

17 Figure 1 shows the relationship between predicted and attained A-level grades for 237,030  
18 examinations from 2010 to 2018 (i.e. assessments including A\* outcomes). 39.3% of predicted  
19 grades are A\* compared with 23.7% of attained grades. Figure 1.a shows predicted grades in  
20 relation to attained grades, with bold font for accurate predictions, green and blue shading for  
21 under-prediction and orange and red shading for over-prediction. Overall 48.8% of predicted grades  
22 are accurate, which is higher than for university applications in general (see Figure 2), reflecting the  
23 high proportion of A and A\* grades (69%). Over-prediction occurred in 44.7% of cases, and under-  
24 prediction in 6.5% of cases. Figure 1.b show the data as percentages. About a half of A\* predictions  
25 result in an attained A grade, and over a third of predicted A grades result in grade B or lower.  
26 Predicted and attained grades have a Pearson correlation of  $r = 0.63$ .  
27  
28  
29  
30  
31  
32  
33

#### 34 Differences between A-level subjects

35 There is little in the literature on the extent to which different A-level subjects may differ in the  
36 accuracy of their predictions, perhaps with different degrees of bias or correlation. Detailed results  
37 are presented in Section 3 of the Supplementary Information. Overall, Biology, Chemistry, Maths and  
38 Physics are very similar in terms of over-prediction and correlation with actual grades. However  
39 General Studies is particularly over-estimated compared with other subjects.  
40  
41  
42

#### 43 Extended Project Qualification (EPQ) and SQA Advanced Highers

44 Section 3 of the Supplementary Information contains information on these qualifications. SQA  
45 Advanced Highers, as well as the EPQ, show similar proportions of over-estimation as other  
46 qualifications (see Figure 2).  
47  
48

#### 49 Reliability of predicted and attained A-level grades

50 Considering the best three A-level grades, the reliability of an overall score can be calculated from  
51 the correlations of the individual subjects. For 66,006 candidates with at least three paired  
52 predicted and actual grades, Cronbach's alpha was 0.827 for actual grades and 0.786 for predicted  
53 grades, with a highly significant difference. The difference may in part reflect the higher proportion  
54 of A\* grades in predicted than actual grades, and hence a greater ceiling effect, but may also reflect  
55 greater measurement precision in the marking of actual A-levels.  
56  
57  
58  
59  
60

## How reliable are attained A-level grades?

Attained A-level grades, like any behavioural measurement are not perfectly reliable, in the sense that if a candidate took a parallel test containing equivalent but different items it is highly unlikely that they would get exactly the same mark as on the first attempt. They may, for instance, have been lucky (or unlucky) at their first attempt, being asked questions on topics which they happened to have studied or revised more (or revised less), and so on. Reliability is a technical subject<sup>n</sup> with many different approaches<sup>48 49</sup>. For continuous measures of raw scores, the reliability can be expressed as a coefficient such as alpha (and in one A-level maths test in 2011, alpha for the full test was about 0.97<sup>50</sup>, although it is suggested that value is unusually high). Boards though do not report raw scores, but instead award grades on a scale such as A\* to E. The 'classification accuracy' of grades is harder to estimate, and is greater with fewer grade points, wider grade intervals, and a wide spread of candidate ability<sup>50</sup>. There seem to be few published estimates of classification accuracy for A-levels (although they do exist for GCSEs and AS-levels<sup>50</sup>).

Estimating classification accuracy for the present high-attaining group of medical school applicants is not easy. A fundamental limit for any applicants is that predicted grades cannot possibly predict actual grades better than attained grades predict themselves (the reliability or classification accuracy). However, from considering the correlation of the three best predicted and actual grades it is unlikely that such a limit has currently been reached. The correlation of actual with predicted grades is .585, and the alpha reliabilities of .827 for actual grades and .786 for predicted grades (see above). The disattenuated correlation between predicted and actual grades is therefore  $.585 / (\sqrt{.827 \times .786}) = 0.726$ , which is substantially less than one, with predicted grades accounting for only about a half of the true variance present in actual grades. If the disattenuated correlation were close to one then it could be argued that predicted grades were doing as well as they could possibly do given that attained grades are not perfectly reliable, but that is clearly far from the case.

## True scores and actual scores

From a theoretical, psychometric, point of view it could be argued that it is neither actual nor predicted grades which need to be estimated for applicants, but their 'true ability scores', or the 'latent scores', to use the technical expressions, of which predicted and actual grades are but imperfect estimates. In an ideal world that would be the case, and a well-constructed exam tries to get as close as possible to true scores. However, it is not possible to know true scores (and if it were the boards would provide selectors with those scores). Selection itself does not work on true scores but on the actual grades that are written down, by teachers for predicted grades, and as grades on exam result certificates by boards. They are the currency in which transactions are conducted during selection, so that a predicted grade of less than a certain level means a candidate will not get a conditional offer, and likewise too low an actual grade means a candidate holding a conditional offer will be rejected. For that reason it is not strictly the correlation of predicted and actual grades which matters, the two measures being treated as symmetric, but the forward prediction of actual grades from predicted grades, i.e. the actual grades conditional on the predicted grades (as shown in figure 1b).

---

<sup>n</sup> See <https://www.gov.uk/government/publications/reliability-of-assessment-compedium> for a range of important papers commissioned and published by Ofqual.

## Predictive validity of predicted and attained A-level grades in medical students.

### Predictive validity in UKMEDP051

The version of the P51 data used here consists entirely of applicants applying to medical schools, but there is also follow-up into undergraduate and postgraduate training. Predicted A-level grades were available only for the UCAS application cycles of 2010 to 2014 (i.e. applying for university entry in October 2009, for the academic year 2010/11, etc), and consisted of a single score in the range 4 to 36 points, based on the sum of the three highest predicted grades, scored as A\*=12, A=10, etc. The modal score for 38,965 applicants was 30 (equivalent to AAA; mean=31.17; SD= 3.58; Median = 32; 5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup> and 95<sup>th</sup> percentiles= 26, 30, 34 and 36). For simplicity the study was restricted to applicants aged 18 in the year of application, who had both predicted and attained A-levels, which also ensured the sample contained only first applications for non-graduate courses, from candidates who had not taken pre-2010 A-levels, when A\* grades were not available. Overall, 22,955 applicants were studied. Other selection measures included were GCSEs (mean grade for best eight grades), as well as U(K)CAT and BMAT scores, based on the most recent attempt which for cases was also the first attempt. For simplicity we used the total of the four sub-scores of U(K)CAT, and the total of Section 1 and 2 scores for BMAT.

Follow-up is complicated as application cohorts enter medical school in different years, and spread out in time through medical school and training. Figure 3 uses an Ibry chart<sup>51-54</sup> to show the educational progression of typical 18-year old medical school entrants, through to postgraduate qualifications. There are however many variants on this theme. The horizontal axis shows academic years (September to August) and training years (August to July), with career stages, key events and measures used on the vertical axis, with coloured boxes indicating typical students, although there are many variants on entry and progression. The blue boxes show typical students on a five-year course who entered medical school in October 2010 at the age of 18. They would have taken GCSEs in June 2008 in school year 11, in the 2007/8 academic year, and some would have taken AS-levels in June 2009. Applicants would have taken aptitude tests in school year 13, most taking either U(K)CAT or BMAT but some taking both tests. U(K)CAT would have been taken between July and September 2009, and BMAT in November 2009. UCAS applications are submitted in October, with teachers providing teacher-estimated grades. Note that U(K)CAT results are known before UCAS applications, but BMAT results are not known until after application. A-levels would have been taken in May-June 2010, with results known in August 2010, and successful applicants entering medical school in Oct 2010. Students on a five year course would start the 2<sup>nd</sup> medical school year in Oct 2011, the 3<sup>rd</sup> and 4<sup>th</sup> years in 2012 and 2013, and during their final year beginning in Oct 2014 they would take the SJT and PSA tests and be awarded an EPM score, with graduation in May 2015. The first of the two Foundation years starts in August 2015, and core or specialist training begins in August 2017. Medical students at some schools take an optional or a compulsory intercalated BSc (iBSc) between years 2 and 3. As a result they are then a year later in progressing to the later stages, and are shown by the green boxes in figure 3. Although years are broadly divided into Basic Medical Science and Clinical stages, some medical schools have courses which are far more integrated<sup>55</sup>.

The above description is for 18-year olds entering the 2010 entry cohort. The present study included the 2010 to 2014 entry cohorts (shown by the solid black box in the lower left of figure 3). For simplicity the last of those cohorts is the only other one, the 2014 entrants having red boxes to show progression for a five year course, and orange for a six-year course including an iBSc. It should be re-emphasised that all career trajectories are idealized, and in reality students and doctors have many and varied training trajectories.

1 Data were available up until the 2018 academic year, and years after that are therefore shown  
2 greyed out in figure 3. Although all cohorts had data for EPM, SJT and PSA, the later entry cohorts  
3 are less likely to have postgraduate qualifications.  
4

5 Undergraduate outcome measures were for simplicity restricted to the deciles of the UKFPO's  
6 Educational Performance Measure (EPM), the raw score of the UKFPO's Situational Judgement Test  
7 (SJT), and the score relative to the pass mark of the Prescribing Safety Assessment (PSA), all at first  
8 attempt. Relatively few doctors, mostly from the earlier cohorts, had progressed through to  
9 postgraduate assessments, but sufficient numbers for analysis were present for MRCP(UK) Part 1  
10 and MRCS Part A, scores being analysed at the first attempt. It should be noted that while U(K)CAT,  
11 BMAT, PSA, SJT, and postgraduate assessments are *nationally standardised*, EPM deciles are *locally*  
12 *standardised* within medical schools.  
13  
14  
15

16 EPM, is a complicated measure summarising academic progression through the first four years of  
17 medical school, with individual medical schools deciding what measures to include<sup>56</sup>, and expressed  
18 as deciles *within* each school and graduating cohort year. EPM is here used as the main  
19 undergraduate outcome measure. EPM deciles are confusing, as UKFPO scores them in the reverse  
20 of the conventional order, the first decile being highest performance and the tenth the lowest<sup>o</sup>. Here  
21 for ease of interpretation we reverse the scoring in what we call *revDecile*, so that higher scores  
22 indicate higher performance. It should also be remembered that deciles are not an equal interval  
23 scale (figure 4)  
24  
25

26 Correlations between the measures are summarised in Figure 5. Large differences in Ns reflect some  
27 measures being used in applicants during *selection*, and others being outcome measures that are  
28 only present in *entrants*, as well as the smaller numbers of doctors who had progressed to  
29 postgraduate assessments. The distinction is emphasised by dividing the correlation matrix into  
30 three separate parts. Correlations of selection and outcome measures necessarily show *range*  
31 *restriction* because candidates have been selected on the basis of the selection measures, and  
32 likewise doctors taking postgraduate examinations may be self-selected for earlier examination  
33 performance.  
34  
35

36 Figure 5 contains much of interest (see also Section 3 of the Supplementary Information), but the  
37 most important question for present purposes is the extent to which Predicted and Attained A-level  
38 grades (shown in pink and green in Figure 5) differ in their prediction of the five outcome measures,  
39 remembering that undergraduate outcomes are typically five or six years after selection, and  
40 postgraduate outcomes are seven or eight years after selection.  
41  
42

43 Attained A-levels predict EPM with a simple Pearson correlation of  $r=0.297$  compared with a  
44 correlation of only 0.198 for predicted grades (simple correlations,  $r$ , are shown in blue in figure 5). N  
45 is large for these correlations and hence the difference, using a test for correlated correlations<sup>57</sup> is  
46 highly significant ( $Z=12.6$ ,  $p<10^{-33}$ ). Multiple regression (see Section 3 of the Supplementary  
47 Information) suggests that predicted grades may have a small amount of predictive variance which is  
48 not shared with attained A-levels. Figure 4 shows mean EPM *revDecile* scores in relation to actual  
49 and predicted A-levels. The slope of the line is clearly less for predicted A-levels, showing a less good  
50 prediction. It is also clear that attained grades predict well, A\*A\*A\* entrants scoring an average of  
51 two deciles higher at the end of the course than those with AAA grades, each extra grade raising  
52 average performance by about two-thirds of a decile. In contrast the slope is less for predicted  
53 grades, being slightly less than half a decile per predicted A-level grade. The broad pattern of results  
54  
55  
56  
57

58  
59 <sup>o</sup> <https://foundationprogramme.nhs.uk/wp-content/uploads/sites/2/2019/11/UKFP-2020-EPM-Framework-Final-1.pdf>,  
60



is similar for the other undergraduate outcomes, SJT and PSA, and is shown in section 3 of the Supplementary Information.

The two postgraduate outcome measures, MRCP(UK) Part 1 (Membership of the Royal Colleges of Physicians (UK) examination) Part 1, and MRCS (Membership of the Royal College of Surgeons) Part A, although both based on smaller, but still substantial, numbers of doctors, are still significant, actual grades correlating more highly with MRCP(UK) Part 1 ( $r=.421$ ) than do predicted grades ( $r=.283$ ;  $Z= 4.54$ ,  $p=.000055$ ). Likewise, actual grades correlate more highly with MRCS Part A ( $r=.421$ ) than do predicted grades ( $r=.358$ ;  $Z= 3.67$ ,  $p=.000238$ ).

The simple correlations ( $r$ ) in figure 5 are inevitably range restricted as A-level grades and predicted A-level grades have themselves been used as a part of the selection process. Taking range restriction into account using the method of Hunter, Schmidt and Le<sup>658</sup> (see also <sup>59</sup>), uses  $u_x$ , the ratio of standard deviations in the predictors in the unrestricted and the restricted population, values below one indicating more range restriction. Figure 5 shows  $u_x$  [ $uX$ ] at the bottom of the columns, and it can be seen that it is much lower for actual A-level grades than predicted A-level grades, suggesting that actual grades are more important in the selection process than are predicted grades. Construct-level predictive validity (CLPV)<sup>6</sup> can be calculated, taking reliability of measures into account, using .827 for attained A-levels and .785 for predicted A-levels (see earlier), with all other reliabilities set at 0.9 in the absence of better estimates. Note that the calculation, unlike that carried out previously <sup>6</sup>, for simplicity does not take censorship/ceiling effects of A-levels into account, and a fuller analysis will be presented elsewhere. The CLPV,  $\rho_{TPa}$  [shown as  $rTPa$  in figure 5], given the greater range restriction, are relatively higher for actual A-level grades than for predicted A-level grades. CLPV for predicting EPM is 0.403 for actual A-level grades compared with 0.251 for predicted A-level grades. For predicting postgraduate qualifications, CLPV for MRCP(UK) Part 1 and MRCS Part A are .601 and .519 for attained A-level grades compared with .360 and .216 respectively for predicted A-level grades.

There are suggestions that predicted grades may not be equivalent in candidates from state schools and private schools, grades being predicted more accurately in independent schools <sup>28,29</sup>. That is looked at in Section 5 of the Supplementary Information, and while there is clear evidence, as found before in the UKCAT-12 study <sup>60</sup>, that private school entrants underperform relative to expectations based on their A-levels, there is no evidence that predicted grades behave differently in candidates from private schools.

A practical question relevant to calculated grades concerns the extent to which, in the absence of attained A-level grades, other selection measures such as GCSEs, U(K)CAT and BMAT can replace the predictive variance of attained A-level grades. That will be considered for EPM where the sample sizes are large. Attained grades alone give  $r = 0.297$ , and predicted grades alone give  $r=.198$ , accounting for less than half as much outcome variance. Adding GCSEs to a regression model including just predicted grades increases multiple R to .225, and also including U(K)CAT and BMAT increases it to .231, which though is still substantially less than the .297 for attained A-levels alone. In the absence of attained A-level grades, prediction is improved by including GCSEs and U(K)CAT or BMAT, but the prediction still falls short of that for actual A-levels alone.

### Modelling the effect of only predicted grades being available for selection

In the context of the 2020 pandemic, an important question is the extent to which future outcomes may change as a result of selection being in terms of calculated grades. Calculated grades themselves were not known at the time of the study, but predicted grades are probably a reasonable surrogate for them in the first instance. A modelling exercise was therefore carried out whereby the numbers of students in the various EPM *RevDeciles* were tabulated in relation to predicted grades at five grade levels, 36 pts  $\equiv$  A\*A\*A\*, 34 pts  $\equiv$  A\*A\*A, 32 pts  $\equiv$  A\*AA, 30 pts  $\equiv$  AAA

1 and  $\leq 28$  pts  $\equiv \leq$  AAB, with the probability of each decile found for each predicted A-level band.  
2 Assuming that selection results in the usual numbers of entrants with grades of A\*A\*A\*, A\*A\*A, etc,  
3 but based on calculated grades rather than actual grades, the expected numbers of students in the  
4 various EPM deciles can be found. Figure 6 shows deciles as standard UKFPO deciles (1 = highest),  
5 UKFPO scores (43 = highest), and *RevDeciles*, (10 = highest). The blue column shows the actual  
6 proportions in the deciles based on attained A-level grades. Note that for various reasons there are  
7 not exactly equal proportions in the ten deciles<sup>p</sup>. Based on selection on attained A-level grades there  
8 are 7.2% of students in the lowest performing decile, compared with an expected proportion of 8.1%  
9 for selection on predicted grades, an increase of 0.9% percentage points, which is a relative increase  
10 of 13.0% in the proportion of the lowest decile, with an odds ratio of 1.141 of attaining the lowest  
11 decile. For the highest scoring decile, the proportion decreases from 10.1% with actual A-level  
12 grades to 8.8% if predicted A-level grades are used, an absolute decrease of 1.4%, and a relative  
13 decrease of 13.4% of top deciles, with an odds ratio of 0.853.

14  
15  
16  
17 Of course, the above calculations are based on the assumption that the 'deciles' for calculated  
18 grades are expressed at the same standard as currently. Were the outcomes to be re-standardised  
19 so that all deciles were equally represented then of course at finals no noticeable difference in  
20 performance would be present<sup>q</sup>. However the academic backbone would still be present, and overall  
21 poorer performance on statistically equated postgraduate exams<sup>61</sup> would be expected.

## 22 Discussion

23  
24  
25  
26  
27 The present data make clear that under a half of predicted grades are accurate, with 45% being  
28 higher than attained grades, and 17% being lower. The data also show that attained grades are far  
29 better predictors of medical school performance than are predicted grades, which account for only  
30 about a third as much outcome variance as attained grades. Attained grades are also more reliable  
31 than predicted grades.

32  
33  
34 Validation is the bottom line for all measures used during selection, and in the present case it is  
35 validation against assessment five to eight years down the line from the original A-levels, in both  
36 undergraduate and postgraduate assessments. That is strong support for what we have called 'the  
37 academic backbone', prior attainment providing the underpinning for later attainment, and hence  
38 there are correlations in performance at all stages of training from GCSEs through to medical  
39 degrees and on into postgraduate assessments<sup>5</sup>.

40  
41  
42 Our findings contradict suggestions that holistic judgments by teachers of predicted grades are  
43 better predictors of outcomes since teachers may know their students better than examiners. The  
44 immense efforts by exam boards and large numbers of trained markers to refine educational  
45 measurements is therefore gratifying and reassuring. Careful measurement does matter.

46  
47 An important question is whether there is some variance in predicted and actual grades which is  
48 complementary. We found that adding predicted grades to the model predicting outcomes  
49 improved the multiple correlation coefficient by only 0.05, accounting for only an additional 0.25%  
50 of variance)). This suggests that predicted grades may provide a very small amount of additional  
51 information in predicting outcomes. What that information might be is unclear, and it is possible  
52  
53

---

54 <sup>p</sup> In part this reflects the fact that some students, particularly weak ones, are given an EPM score, but then fail  
55 finals.

56 <sup>q</sup> This is based on deciles being calculated in a way that are equated to levels used in the present calculation.  
57 Of course if calculated strictly as deciles then of necessity 10% will still remain in the top decile, etc.. That  
58 difficulty of deciles will in large part be removed when the 2020 entrants graduate as UKMLA should be on  
59 stream by then.  
60

1 that it is what Petch called 'scholarly attitude'. At present though it is worth remembering that  
2 *examination* grades at A-level are primarily predicting further *examination* grades at the end of  
3 medical school, although EPM scores do include formal assessments of course work, and practical  
4 and clinical skills. If other outcome measures, perhaps to do with communication, caring or other  
5 non-cognitive skills were available then predicted grades might show a greater predictive value.  
6  
7

8 The present data inevitably have some limitations. There is little likelihood of bias since complete  
9 population samples have been considered, and there is good statistical power with large sample  
10 sizes. Inevitably not all outcomes can be considered, mainly because the cohorts analysed have not  
11 yet progressed sufficiently through postgraduate training. However, those postgraduate outcomes  
12 which are included do show substantial effects which are highly significant statistically.  
13  
14

15 Our questions about predicted grades have been asked in the practical context of the cancellation of  
16 A-level assessments and their replacement by calculated grades, as a result of the COVID-19  
17 pandemic. It seems reasonable to assume, given the literature on predicted grades, and particularly  
18 on forecasted grades, that calculated grades will probably have similar predictive ability to predicted  
19 grades, but perhaps be a little more effective due to occurring later in the academic cycle. Such a  
20 conclusion would be on firmer ground if exam boards had analysed the predictive validity of the data  
21 they had collected on forecasted grades, particularly in comparison with predicted and actual  
22 grades. Such data may exist, and if so then they need to be seen. In their absence, the present data  
23 may be the best available guesstimates of the likely predictive validity of calculated rather than  
24 actual grades.  
25  
26

27 A potential limitation of our study is that we do not include the calculated and final grades for  
28 students who applied for admission in 2020; however calculated and final grades for 2020 will be  
29 available in UKMED in 2021, and since that year group will also have the teacher-predicted grades  
30 submitted to UCAS, an immediate question of interest will be the extent of the correlation of the  
31 measures, and hence whether teacher-predicted grades are indeed a proxy for calculated grades.  
32 Having said that, it will not be possible to calculate the predictive validity of teacher-predicted and  
33 calculated grades for a number of years until the cohort progresses through undergraduate training.  
34 Medium- and long-term predictive validity inevitably take time to acquire, and practical decision-  
35 making sometimes has to be based on proxy, surrogate measures, with teacher-predicted grades at  
36 application to UCAS being a reasonable substitute. If it were the case that teacher-predicted grades  
37 for UCAS and teacher-estimated grades as a part of calculated grades were fundamentally  
38 discrepant then serious questions would be raised about one or other set of estimates. The same  
39 applies to the teacher-estimated grades being used as a substitute for A-levels in the summer of  
40 2021, which will apply to the cohort applying for entry to medical school in 2021.  
41  
42  
43  
44

## 45 Under-prediction

46 Under-prediction is a particular risk in cases where teachers do not know their students well, or in  
47 some cases perhaps underestimate their ability because of attitude, personal characteristics, or  
48 other factors. There is some evidence that teacher-assessed grades relate more to student  
49 personality than do grades in national examinations<sup>62 63</sup>, although effects were relatively weak. Any  
50 such biases are traditionally solved by the externality and objectivity of national examinations.  
51 Petch, once again, put it well, describing,  
52  
53

54 "instances, where, in the examination room, candidates have convinced the examiners that  
55 they are capable of more than their schools said that they were ... Paradoxical as it will  
56 seem, examiners are not always on the side of authority; an able rebel can find his wider  
57 scope within the so-called cramping confines of an examination." <sup>3</sup>(p.29).  
58  
59  
60

1 There is a clear echo here of the quote by Yasmin Hussein with which this paper began. Hussein's  
2 concerns are not alone, and the UKMACS study in April 2020 found concerns about fairness were  
3 particularly present in medical school applicants from non-selective schools, from Black Asian and  
4 Minority Ethnic (BAME) applicants, from female applicants, and from those living in more deprived  
5 areas<sup>10</sup>.  
6  
7

### 8 Effects of loss of schooling. 9

10 A further consideration is more general and asks what the broader effects of the COVID-19  
11 pandemic may be on medical education. Students at all levels of education have had teaching and  
12 learning disrupted, often extensively, and that is also true of all stages of medical education. The  
13 2020 cohort of applicants/entrants will not have been assessed formally at A-level. As well as  
14 meaning that they may only have calculated grades, which are likely to be less accurate, they also  
15 will have missed out on significant amounts of teaching. UK students who should have taken A-level  
16 exams in 2020 missed around 30 to 40 school days; those in the year below from whom 2021  
17 medical school entrants will be drawn, will have missed around 80 days. Burgess and Sievertsen<sup>64</sup>,  
18 using data from two studies<sup>65 66</sup>, estimate that 60 lost school days results in a reduction in  
19 performance of about 6% of a standard deviation, which they say is, "non-trivial" (and for  
20 comparison a rule of thumb is that students in school improve by about one third of a standard  
21 deviation in each school year<sup>67</sup>). These effects are likely to differ also by socio-economic  
22 background, particularly given variability in the effectiveness of home schooling. Applicants not  
23 taking A-levels will also suffer from the loss of the enhanced learning that occurs when learners are  
24 tested – the 'testing effect' – for which meta-analyses have found effect sizes of about 0.50<sup>68 69</sup>,  
25 which is also non-trivial. Taken overall, 2020 entrants to medical school, and perhaps those in 2021  
26 as well, may – without additional support - perform less well in the future as a result of missing out  
27 both on education and on its proper assessment.  
28  
29  
30  
31  
32

### 33 Conclusions 34

35 The events of 2020 as a result of the COVID-19 pandemic were extra-ordinary, and unprecedented  
36 situations occurred of which the cancellations of GCSE and A-level exam cancellations were but one  
37 example. The current study should not be seen as criticism of the response of *Ofqual* to that  
38 situation; given the circumstances in which it found itself, with examinations cancelled (when the  
39 Chair of *Ofqual*, Roger Taylor, had recommended socially-distanced or delayed exams), *Ofqual*'s  
40 solution to the problems had many obvious virtues. We began this paper by quoting a letter to a  
41 newspaper in March 2020 at the beginning of lockdown by a student taking GCSEs, and so it is  
42 probably appropriate to finish with a letter to a different newspaper by an A-level student. Written  
43 at the height of the A-levels crisis, in August 2020, it raises many subtle, important and mostly  
44 neglected questions, ones which researchers will need to grapple with in the future:  
45  
46  
47

48 *"Ofqual's grading system appears to be lacking in advocates. Blinded by rhetoric about what*  
49 *protesters call a 'classist' algorithm, key facts have been overlooked. It is very clear that*  
50 *teachers are shockingly bad at predicting grades; using teacher predictions there will be a*  
51 *12% inflation in higher grades compared with last year. While some centres predicted*  
52 *accurately, some centres predicted only the highest grades for their students. This U-turn*  
53 *from the government entails a huge injustice for the pupils who had fair and accurate*  
54 *predictions, as well as for those taking exams next year. In the zero-sum game of university*  
55 *applications, the results of these pupils make them appear weaker than they are.*  
56 *Irresponsible teachers who over-predicted their pupils' results ought to be ashamed that*  
57 *they too have thereby 'dashed the dreams' of many young people across the country. That it*  
58  
59  
60

1 is less obvious does not make it any less true.” (Letter to *The Times*, 19<sup>th</sup> August 2020, by Seb  
2 Bird, A-level student, Bristol)<sup>70</sup>.  
3

4 For most university applicants there already existed predicted grades from the previous autumn  
5 when UCAS applications were submitted, but they would have been on average half a grade or so  
6 too high, being aspirational as much as realistic, and also for medical students would have been  
7 made by October 2019, whereas calculated grades would be based on teacher predictions in May  
8 2020, albeit with several months of courses missing since March 2020.  
9

10  
11 In May 2020 we wrote that raw teacher-predicted grades would have wrecked much university  
12 planning, particularly coming so late in the year, after offers had been made, as numbers of  
13 acceptances would inevitably have been far too high<sup>7</sup>. That in fact happened, and quotas for  
14 university entries had to be abandoned in August 2020, including for medicine, and that had knock-  
15 on effects into first year university courses, and probably beyond. There was also a risk that  
16 predicted grades could have been systematically higher from some schools than others – the ones  
17 with a tendency to call all of their “geese” “swans” -- and that probably applies also to the centre  
18 assessment grades sent to examination boards and mostly eventually accepted without central  
19 standardisation in August 2020. The consequences of that will not become apparent for a few years.  
20  
21

22  
23 This paper has provided evidence that the grades awarded to medical applicants in summer 2020  
24 will probably not predict future outcomes with the same effectiveness as actual, attained grades,  
25 and that is a problem that universities and medical schools and postgraduate deaneries will have to  
26 work with, probably for many years as the 2020 cohort works through the system. It seems likely  
27 therefore, as Thomson has said, “... this year group will always be different...”<sup>2</sup>.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Figure captions

*Figure 1:* Predicted vs attained A-level grades for individual subjects in applicants to UK medical schools. Accurate predictions are in bold; yellow – over-estimates by 1 grade; orange – over-estimates by 2+ grades; green – under-estimates by 1 grade; blue – under-estimates by 2+ grades. a) Counts; b) attained grades as percentages within predicted grades.

*Figure 2:* Over-estimated, under-estimated and accurate predicted grades in various studies. Black font: predicted grades; red font: forecasted grades; yellow background: pre-2000; blue background: pre-2010; bold, underlined: averaged results post-2000.

*Figure 3:* An lbyr chart illustrating the progression of the 2010 to 2014 medical school entry cohorts through secondary schooling, application to medical school, undergraduate and post-graduate training, with the timing of key events shown. See text for further details.

*Figure 4:* Mean EPM *revDeciles* (95% CI) in relation to actual A-level grades (green) and predicted A-level grades (red)

*Figure 5:* Correlation matrix of selection measures, undergraduate outcome measures, and postgraduate outcome measures (separated by grey lines for clarity). Cells indicate simple Pearson correlations (*r*; in blue), construct-level predictive validity (*rTP $\alpha$* ; in red) and sample size (*N*; in black).

*Figure 6:* Predicted decile outcomes if selection were on Predicted A-level grades (blue) rather than actual A-level grades (orange).

## Acknowledgements

We are grateful to Paul Garrud, Gill Wyness, Paul Newton, Colin Melville and Christian Woodward for their comments on earlier versions of this manuscript, and to Jon Dowell, Peter Tang, Rachel Greatrix and other members of the UKMED Research Group and Advisory Board for their assistance in fast-tracking the pre-print of this paper, and for their comments on it. We also thank Tim Gill for providing us with an unpublished manuscript.

## Contributors

DTS prepared the data extracts, provided details on data sources and variable definitions where required and commented on manuscript drafts. ICM originated the idea for the study, and discussed it with other authors throughout the project. ICM wrote the first draft of the manuscript, and KW, DH, PAT, LP, KYFC and DTS have read, reviewed and commented on earlier drafts and contributed ideas, as well as approving the final draft, both of the preprint, and of the present paper.

## Funding

KW is a National Institute for Health Research (NIHR) Career Development Fellow (NIHR CDF-2017-10-008) and is principal investigator for the UKMACS and UKMEDP089 projects supported by the NIHR funding.

DH is funded by NIHR grant CDF-2017-10-008 to KW.

PAT's research time is supported by an NIHR Career Development Fellowship (CDF 2015-08-11), and PAT is also principal investigator for the UKMEDP051 project.

LWP is partly supported by NIHR grant CDF 2015-08-11 to PAT, and a portion of his research time is funded by the UCAT board.

ICM, KYFC and DTS have received no specific funding for this project.

## Disclaimers

KW and DH state that this publication presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

PAT states this research is supported by an NIHR Career Development Fellowship (CDF 2015-08-11). This paper presents independent research partly funded by the National Institute for Health Research. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

KYFC is employed as the Head of Marking and Results at Cambridge Assessment English. The views expressed are those of the authors and do not represent the views of Cambridge Assessment.

DTS is employed by the GMC as a data analyst working on the UKMED project. The views expressed here are his views and not the views of the GMC.

Data sources. UK Medical Education Database (UKMED) UKMEDP051 data extract generated on 13<sup>th</sup> May 2019. UKMEDP089 extract generated 21<sup>st</sup> January 2020. UKMEDP112 project, using UKMEDP051 and UKMEDP089 data, approved for publication on 29<sup>th</sup> May 2020. We are grateful to UKMED for the use of these data. However, UKMED bears no responsibility for their analysis or interpretation.

1 UKMEDP051 data includes information derived from that collected by the Higher Education Statistics  
2 Agency Limited (HESA) and provided to the GMC (HESA Data). Source: HESA Student Record  
3 2002/2003 to 2014/2015. Copyright Higher Education Statistics Agency Limited. The Higher  
4 Education Statistics Agency Limited makes no warranty as to the accuracy of the HESA Data, cannot  
5 accept responsibility for any inferences or conclusions derived by third parties from data or other  
6 information supplied by it.  
7  
8

9 UKMEDP051 and UKMEDP089 include Universities and Colleges Admissions Service (UCAS) data  
10 provided to the GMC (UCAS data). Source: UCAS (application cycles 2007 to 2018). Copyright  
11 Universities and Colleges Admissions Service (UCAS). UCAS makes no warranty as to the accuracy of  
12 the UCAS Data and cannot accept responsibility for any inferences or conclusions derived by third  
13 parties from data or other information supplied by it.  
14  
15

16 All data from HESA are required to be reported using their rounding and suppression criteria  
17 ([https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-](https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics)  
18 [statistics](https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics)) and we have applied those criteria to all UKMED-based tables and values reported here.  
19

## 20 21 Competing interests

22 ICM is a member of the UKMED Research Group and the UKMED Advisory Board, and is also on the  
23 UKMACS advisory group.  
24

25 PAT is a member of the UKMED Research Group. PAT has previously received research  
26 funding from the ESRC, the EPSRC, the Department of Health for England, the UCAT Board,  
27 and the GMC. In addition, PAT has previously performed consultancy work on behalf of his  
28 employing University for the UCAT Board and Work Psychology Group and has received travel  
29 and subsistence expenses for attendance at the UCAT Research Group.  
30  
31

32 KYFC is a member of the UKMED Research Group, and is an employee of Cambridge Assessment - a  
33 group of exam boards that owns and administers the BioMedical Admissions Test (BMAT); UK GCSEs  
34 and A-levels; and International GCSEs and A-levels.  
35  
36

37 DTS is a member of the UKMED Research Group and the UKMED Advisory Board and is employed by  
38 the GMC as a data analyst working on the UKMED project.  
39

40 KW, DH and LP declare no competing interests.  
41

## 42 43 Authorship

44 ICM conceived the idea for the study, conducted the statistical analysis and wrote the first draft of  
45 the paper. All authors contributed to subsequent drafts and approved the final version.  
46  
47

## 48 49 Ethical approval

50 Queen Mary Research Ethics Committee, University of London, agreed on 11 November 2015 that  
51 there was no need for ethical review of UK Medical Education Database research studies.  
52

53 [https://www.ukmed.ac.uk/documents/UKMED\\_research\\_projects\\_ethics\\_exemption.pdf](https://www.ukmed.ac.uk/documents/UKMED_research_projects_ethics_exemption.pdf)  
54

55 All data from HESA are required to be reported using their rounding and suppression criteria  
56 ([https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-](https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics)  
57 [statistics](https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics)) and we have applied those criteria to all UKMED-based tables and values reported here.  
58  
59  
60



## Provenance and peer review

Not commissioned; reviewed by the UKMED Research Subgroup and Advisory board, and to submitted to an externally peer reviewed journal. This paper was fast-tracked through the UKMED governance processes as with other COVID-19 related research projects elsewhere.

## Patient and public involvement.

No patient involvement.

## Data sharing statement

Researchers wishing to re-analyse the data used for this study can apply for access to the same datasets via UKMED ([www.ukmed.ac.uk](http://www.ukmed.ac.uk)).

## Figure captions

Figure 1: Predicted vs attained A-level grades for individual subjects in applicants to UK medical schools. Accurate predictions are in bold; yellow – over-estimates by 1 grade; orange – over-estimates by 2+ grades; green – under-estimates by 1 grade; blue – under-estimates by 2+ grades. a) Counts; b) attained grades as percentages within predicted grades.

Figure 2: Over-estimates, under-estimate and accurate predicted grades in various studies. Black font: predicted grades; red font: forecasted grades; yellow background: pre-2000; blue background: pre-2010; bold, underlined: averaged results post-2000.

Figure 3: An lbrary chart illustrating the progression of eighteen-year old 2010 to 2014 medical school entry cohorts through secondary schooling, application to medical school, undergraduate and post-graduate training, with the timing of key events shown. See text for further details.

Figure 4: Correlation matrix of selection measures, undergraduate outcome measures, and postgraduate outcome measures (separated by grey lines for clarity). Cells indicate Pearson correlations ( $r$ ) and  $N$ , as well as the construct-level predictive validity ( $rTPa$ ). Range restriction is indicated by  $uX$ . See text for details.

Figure 5: Predicted decile outcomes if selection were on Predicted A-level grades (blue) rather than actual A-level grades (orange).

## References

1. Hussein Y. Cancellation of GCSE is unfair to some students. *The Guardian* 2020(March 29th): <https://www.theguardian.com/world/2020/mar/29/cancellation-of-gcse-exams-unfair-to-some-students>.
2. Thomson D. Moderating teaching judgments in 2020 [Blog post, 25th March 2020]. London: FFT Educational Lab: <https://ffteducationdatalab.org.uk/2020/03/moderating-teacher-judgments-in-2020/> (accessed 16th April 2020) 2020.
3. Petch JA. School estimates and examination results compared. Manchester: Joint Matriculation Board 1964.
4. Higher Education Funding Council for England [HEFCE]. Differences in student outcomes: The effect of student characteristics. Data Analysis March 2018/05,. Bristol: HEFCE 2018.
5. McManus IC, Woolf K, Dacre J, et al. The academic backbone: Longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the Specialist Register in UK medical students and doctors. *BMC Medicine* 2013;11:242:doi:10.1186/741-7015-11-242.
6. McManus IC, Dewberry C, Nicholson S, et al. Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: Meta-regression of six UK longitudinal studies. *BMC Medicine* 2013;11:243:doi:10.1186/741-7015-11-243.
7. McManus IC, Woolf K, Harrison D, et al. Calculated grades, predicted grades, forecasted grades and actual A-level grades: Reliability, correlations and predictive validity in medical school applicants, undergraduates, and postgraduates in a time of COVID-19. *medRxiv* 2020;doi: <https://doi.org/10.1101/2020.06.02.20116830>
8. McKie A. Scrapped exams may spark UK admissions 'scramble'. *Times Higher Education* 2020;26th March 2020:9-9.
9. Dowell J, Cleland J, Fitzpatrick S, et al. The UK medical education database (UKMED): What is it? Why and how might you use it? *BMC Medical Education* 2018;18 (DOI 10.1186/s12909-017-1115-9)(6):1-8.
10. Woolf K, Harrison D, McManus IC. The attitudes, perceptions and experiences of medical school applicants following the closure of schools and cancellation of public examinations due to the COVID-19 pandemic in 2020. *medRxiv* 2020;submitted
11. Woolf K, Harrison D, McManus C. The attitudes, perceptions and experiences of medical school applicants following the closure of schools and cancellation of public examinations in 2020 due to the COVID-19 pandemic: a cross-sectional questionnaire study of UK medical applicants. *BMJ open* 2021;11(3):e044753.
12. Ofqual. Summer 2020 grades for GCSE, AS and A level, Extended Project Qualification and Advanced Extension Award in maths: Guidance for teachers, students, parents and carers. Coventry: Ofqual: Ofqual/20/6607/2 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/877842/Summer\\_2020\\_grades\\_for\\_GCSE\\_AS\\_A\\_level\\_EPQ\\_AEA\\_in\\_maths\\_-\\_guidance\\_for\\_teachers\\_students\\_parents.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/877842/Summer_2020_grades_for_GCSE_AS_A_level_EPQ_AEA_in_maths_-_guidance_for_teachers_students_parents.pdf) (accessed 3rd April 2020) 2020.
13. Cresswell M. Heaps, prototypes and ethics: The consequence of using judgments of student performance to set examination standards in a time of change. London: Institute of Education 2003.
14. Ofqual. Setting GCSE, AS and A Level Grade Standards in Summer 2014 and 2015. London: <https://www.gov.uk/government/publications/setting-gcse-and-a-level-grade-standards-in-summer-2014-and-2015> [accessed 18th April 2020] 2020.
15. Ofqual. Summer 2020 grades for GCSE, AS and A level, Extended Project Qualification and Advanced Extension Award in maths: Information for Heads of Centre, Heads of Department and teachers on the submission of Centre assessment grades. Coventry: Ofqual: Ofqual/20/6607/1 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/877930/Summer\\_2020\\_grades\\_for\\_GCSE\\_AS\\_A\\_level\\_EPQ\\_AEA\\_in\\_maths\\_-\\_guidance\\_for\\_heads\\_of\\_centres.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/877930/Summer_2020_grades_for_GCSE_AS_A_level_EPQ_AEA_in_maths_-_guidance_for_heads_of_centres.pdf) (accessed 3rd April 2020) 2020.

16. McManus IC, Powis DA, Wakeford R, et al. Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *BMJ* 2005;331:555-59.
17. Opposs D, He Q. The reliability programme: Final Report. Coventry: Office of Qualifications and Examinations Regulation (<http://www.ofqual.gov.uk/files/reliability/11-03-16-Ofqual-The-Final-Report.pdf>) 2011.
18. Downing SM. Validity: On the meaningful interpretation of assessment data. *ME* 2003;37:830-37.
19. Association AER, Association AP, Education NCoMi. Standards for educational and psychological testing. Washington, DC: American Educational Research Association 2014.
20. Kane MT. Validating the interpretations and uses of test scores. *Journal of Educational Measurement* 2014;50(1):1-73.
21. Bekhradnia B, Thompson J. Who does best at University? London: Higher Education Funding Council England (<http://webarchive.nationalarchives.gov.uk/20081202000732/http://hefce.ac.uk/Learning/whodoes/>) 2002.
22. Higher Education Funding Council for England [HEFCE]. Differences in degree outcomes: The effect of subject and student characteristics. Issues Paper 2015/21. Bristol: HEFCE 2015.
23. McManus IC, Smithers E, Partridge P, et al. A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ* 2003;327:139-42.
24. Lough C. GCSEs: Only 39% teachers think 2020 grades fair for all: Plan for teacher-assessed GCSE and A-level grades prompts concerns about potential teacher bias, TES survey of 19,000 finds. *TES (Times Educational Supplement)* 2020;13th May 2020:<https://www.tes.com/news/coronavirus-gcses-only-39-teachers-think-2020-grades-fair-all>.
25. Lilford R. Policy makers should use evidence, but what should they do in an evidence vacuum? *ARC West Midlands News Blog [NIHR Applied Research Collaboration, West Midlands]* 2020;2(4 (April 24th, 2020)):1-2 (<https://arcwm.files.wordpress.com/2020/04/arc-wm-newsblog-20-04-24.pdf>).
26. Hubbles S, Bolton P. The review of university admissions [Briefing Paper Number 8538, 10 April 2019]. London: House of Commons (<https://researchbriefings.files.parliament.uk/documents/CBP-8538/CBP-8538.pdf>) 2019.
27. Snell M, Thorpe A, Hoskins S, et al. Teachers' Perceptions and A-Level Performance: Is There Any Evidence of Systematic Bias? *Oxford Review of Education* 2008;34(4):403-23.
28. Everett N, Papageorgiou J. Investigating the Accuracy of Predicted A Level Grades as part of 2009 UCAS Admission Process. London: Department for Business, Innovation and Skills 2011.
29. Wyness G. Predicted grades: Accuracy and impact. A report of University and College Union. London: University and College Union ([https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted\\_grades\\_report\\_Dec2016.pdf](https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted_grades_report_Dec2016.pdf)) 2016.
30. UCAS. End of cycle report 2017: Qualifications and competition. Cheltenham: UCAS (<https://www.ucas.com/data-and-analysis/ucas-undergraduate-releases/ucas-undergraduate-analysis-reports/2017-end-cycle-report>) 2017.
31. UCAS. Factors associated with predicted and achieved A level attainment, August 2016. Cheltenham: UCAS: <https://www.ucas.com/file/71796/download?token=D4uuSzur> 2016.
32. Gill T. Methods used by teachers to predict final Alevel grades for their students. *Research Matters (UCLES)* 2019(28):33-42.
33. Walland E, Darlington E. Insights on trends in AS Levels, the EPQ and Core Maths: summary report. Cambridge: 35859 /id}. Cambridge Assessment: <https://www.cambridgeassessment.org.uk/.../527125-insights-on-trends-in-as-levels-the-epq-and-core-maths-summary-report.pdf> 2019.
34. Urhahne D, Wijnia L. A Review on the Accuracy of Teacher Judgments. *Educational Research Review* 2020:100374.
35. Meissel K, Meyer F, Yao ES, et al. Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education* 2017;65:48-60.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
36. Rimfeld K, Malanchini M, Hannigan LJ, et al. Teacher assessments during compulsory education are as reliable, stable and heritable as standardized test scores. *Journal of Child Psychology and Psychiatry* 2019;60(12):1278-88.
  37. Gill T, Benton T. The accuracy of forecast grades for OCR Alevels in June 2014: Statistics Report Series No 90. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/Images/241261-the-accuracy-of-forecast-grades-for-ocr-a-levels-in-june-2014.pdf>] 2015.
  38. Gill T, Rushton N. The accuracy of forecast grades for OCR Alevels: Statistics Report Series No 26. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/our-research/all-published-resources/statistical-reports/150215-the-accuracy-of-forecast%20-grades-for-ocr-a-levels-in-june-2012.pdf/>] 2011.
  39. Gill T, Chang Y. The accuracy of forecast grades for OCR A levels in June 2012: Statistics Report Series No.64. Cambridge: Cambridge Assessment Statistics Report Series No.64 2013.
  40. Petch JA. Fifty years of examining: The Joint Matriculation Board, 1903-1953.
  41. Murphy RJL. Teachers' assessments and GCE results compared. *Educational Research* 1979;22(1):54-59.
  42. Gill T, Chang Y. The accuracy of forecast grades for OCR GCSEs in June 2013: Statistics Report Series No 89. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/Images/241260-the-accuracy-of-forecast-grades-for-ocr-gcses-in-june-2013.pdf>] 2015.
  43. Gill T, Benton T. The accuracy of forecast grades for OCR GCSEs in June 2014: Statistics Report Series No 91. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/Images/241265-the-accuracy-of-forecast-grades-for-ocr-gcses-in-june-2014.pdf>] 2015.
  44. Südkamp A, Kaiser J, Möller J. Accuracy of Teachers' Judgments of Students' Academic Achievement: A Meta-Analysis. *Journal of Educational Research* 2012;104(3):743-62.
  45. Lumb AB, Vail A. Applicants to medical school: the value of predicted school leaving grades. *Med Educ* 1997;31:307-11.
  46. McManus IC, Richards P, Winder BC, et al. Medical school applicants from ethnic minorities: identifying if and when they are disadvantaged. *Brit Med J* 1995;310:496-500.
  47. Richardson PH, Winder B, Briggs K, et al. Grade predictions for school-leaving examinations: do they predict anything? *Med Educ* 1998;32:294-97.
  48. Wilmut J, Wood R, Murphy R. A review of research into the reliability of examinations: A discussion paper prepared for the School Curriculum and Assessment Authority. Nottingham: School of Education [available at [www.gov.uk/systems/uploads](http://www.gov.uk/systems/uploads)] 1996.
  49. Bramley T, Dhawan V. Estimates of Reliability of Qualifications. Cambridge: Cambridge Assessment:[[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/578868/2011-03-16-estimates-of-reliability-of-qualifications.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/578868/2011-03-16-estimates-of-reliability-of-qualifications.pdf)] 2010.
  50. Wheadon C, Stockford I. Classification accuracy and consistency in GCSE and A level examinations offered by the Assessment and Qualifications Alliance (AQA) November 2008 to June 2009. Office of Qualifications and Examinations Regulation: Coventry (<http://www.ofqual.gov.uk/files/reliability/11-03-16-AQA-Classification-Accuracy-and-Consistency-in-GCSE-and-A-levels.pdf>) 2011.
  51. Marey EJ. La Méthode graphique dans les sciences expérimentales et particulièrement en physiologie et en médecine. Paris 1878.
  52. Wainer H, Harik P, Neter J. Visual Revelations: Stigler's Law of Eponymy and Marey's Train Schedule: Did Serjev Do It Before Ibry, and What About Jules Petiet? *Chance* 2013;26(1):53-56.
  53. Tufte ER. The visual display of quantitative information. Cheshire, CT: Graphics Press 2018.
  54. Garrud P, McManus IC. Impact of accelerated, graduate-entry medicine courses: a comparison of profile, success, and specialty destination between graduate entrants to accelerated or standard medicine courses in UK. *BMC Medical Education* 2018;18:250:<https://doi.org/10.1186/s12909-018-1355-3>.

- 1 55. Devine OP, Harborne AC, Horsfall HL, et al. The Analysis of Teaching of Medical Schools (AToMS)  
2 survey: an analysis of 47,258 timetabled teaching events in 25 UK medical schools relating to  
3 timing, duration, teaching formats, teaching content, and problem-based learning. *BMC*  
4 *Medicine* 2020;18(126):<https://doi.org/10.1186/s12916-020-01571-4>.
- 5 56. Curtis S, Smith D. A comparison of undergraduate outcomes for students from gateway courses  
6 and standard entry medicine courses. *BMC Medical Education*  
7 2020;20(4):<https://doi.org/10.1186/s12909-019-1918-y>.
- 8 57. Meng X-L, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. *Psychological*  
9 *Bulletin* 1992;111(1):172-75.
- 10 58. Hunter JE, Schmidt FL, Le H. Implications of direct and indirect range restriction for meta-analysis  
11 methods and findings. *Journal of Applied Psychology* 2006;91(3):594-612.
- 12 59. Fife DA, Mendoza JL, Terry R. Revisiting Case IV: A reassessment of bias and standard errors of  
13 Case IV under range restriction. *British Journal of Mathematical and Statistical Psychology*  
14 2013;66(3):521-42.
- 15 60. McManus IC, Dewberry C, Nicholson S, et al. The UKCAT-12 study: Educational attainment,  
16 aptitude test performance, demographic and socio-economic contextual factors as  
17 predictors of first year outcome in a collaborative study of twelve UK medical schools. *BMC*  
18 *Medicine* 2013;11 :244:doi:10.1186/741-7015-11-244.
- 19 61. McManus IC, Chis L, Fox R, et al. Implementing statistical equating for MRCP(UK) Parts 1 and 2.  
20 *BMC Medical Education* 2014;14(204):<http://www.biomedcentral.com/1472-6920/14/204>;  
21 doi:10.1186/472-6920-14-204.
- 22 62. Papageorgiou KA, Likhonov M, Costantini G, et al. Personality, Behavioral strengths and  
23 difficulties and performance of adolescents with high achievements in science, literature, art  
24 and sports. *Personality and Individual Differences* 2020;160:109917.
- 25 63. Zimmermann F, Schütte K, Taskinen P, et al. Reciprocal effects between adolescent externalizing  
26 problems and measures of achievement. *Journal of educational psychology* 2013;105(3):747.
- 27 64. Burgess S, Sievertsen HH. Schools, skills, and learning: The impact of COVID-19 on education.  
28 <https://voxeu.org/article/impact-COVID-19-education> (1st April 2020; accessed 31st May  
29 2020)2020.
- 30 65. Carlsson M, Dahl GB, Öckert B, et al. The effect of schooling on cognitive skills. *Review of*  
31 *Economics and Statistics* 2015;97(3):533-47.
- 32 66. Lavy V. Do differences in schools' instruction time explain international achievement gaps?  
33 Evidence from developed and developing countries. *The Economic Journal*  
34 2015;125(588):397-424.
- 35 67. Hanushek EA, Woessman L. The economic impacts of learning losses (OECD Education Working  
36 Paper). *Education Working Papers (OECD Publishing, Paris)* 2020;No. 225 doi:  
37 <https://doi.org/10.1787/21908d74-e>
- 38 68. Rowland CA. The effect of testing versus restudy on retention: A meta-analytic review of the  
39 testing effect. *Psychological Bulletin* 2014;140(6):1432-63.
- 40 69. Yang C, Luo L, Vadillo MA, et al. Testing (quizzing) boosts classroom learning: A systematic and  
41 meta-analytic review. *Psychological Bulletin* 2021
- 42 70. Bird S. A-levels fiasco. *The Times* 2020 August 19th.

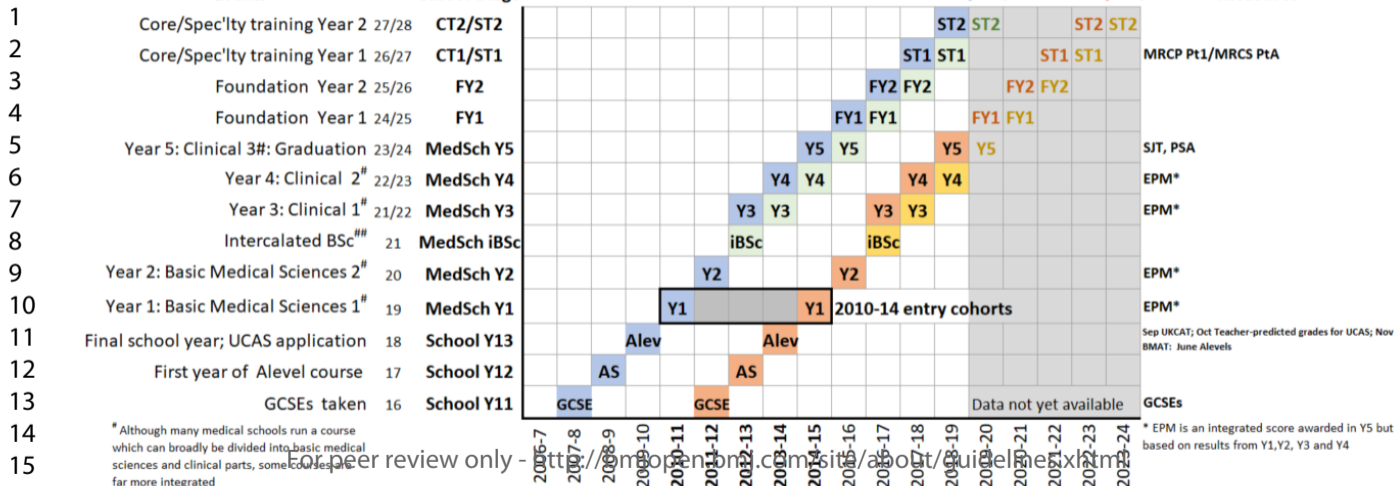
		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades (points)	E (2 pts)	<b>200</b>	35	10	5	0	0	255 (0%)
	D (4 pts)	235	<b>610</b>	155	35	10	0	1045 (0%)
	C (6 pts)	635	1220	<b>2110</b>	505	95	5	4570 (2%)
	B (8 pts)	635	2095	4755	<b>7355</b>	1695	175	16715 (7%)
	A (10 pts)	430	1925	8785	35640	<b>61950</b>	12655	121390 (51%)
	A* (12 pts)	50	135	635	6025	42815	<b>43395</b>	93060 (39%)
	Total	2185	6020	16450	49570	106570	56235	237030
		(1%)	(3%)	(7%)	(21%)	(45%)	(24%)	

Percentages within predicted grades

		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades (points)	E (2 pts)	<b>79%</b>	14%	..	..	..	..	100%
	D (4 pts)	23%	<b>58%</b>	15%	3%	..	..	100%
	C (6 pts)	14%	27%	<b>46%</b>	11%	2%	..	100%
	B (8 pts)	4%	13%	28%	<b>44%</b>	10%	1%	100%
	A (10 pts)	0%	2%	7%	29%	<b>51%</b>	10%	100%
	A* (12 pts)	0%	0%	1%	7%	46%	<b>47%</b>	100%
	Total	1%	3%	7%	21%	45%	24%	100%

Key:	Blue background: Pre-2010 results	Yellow background: PBMbOpen results				Page 34 of 87	
	Red font: Forecasted grades	<b><u>Bold, underlined: Averaged results, post 2000</u></b>					
1	Study	Context	Year	A-level range	Under-estimated	Over-estimated	
2	<b>University applicants overall: A-levels, etc</b>				"Pessimistic"	"Optimistic"	
3	Everett & Papageorgiou (2011) [24]	Predicted Grades	2009	A-E	7%	52%	42%
4	UCAS [27]	Predicted Grades	2012	A*-E	12%	20%	68%
5	Wyness (2016) [25]	Predicted Grades	2013-15	A*-E	9%	16%	75%
6	UCAS [27]	Predicted Grades	2016		9%	16%	74%
7	UCAS [27]	Predicted Grades	2017		10%	16%	73%
8	Petch (1953) [33]	Forecasted Grades	1940	School Cert Pass/Fail	2%	89%	9%
9	Petch (1964) [4]	Non-official forecasted Grades	1963	A+B/C+D/E/O/F	18%	43%	39%
10	Murphy (1979) [34]	Non-official forecasted Grades	1977	A-E	29%	27%	44%
11	Gill and Rushton (2011) [31]	Forecasted Grades	2009	A-E	12%	55%	33%
12	Gill and Chang (2013) [32]	Forecasted Grades	2012	A*-E	13%	48%	39%
13	Gill and Benton (2015) [30]	Forecasted Grades	2014	A*-E	14%	43%	43%
14	Gill (2019) [28]	Non-official forecasted Grades	2018	A*-E	20%	45%	35%
15		<b><u>Mean Predicted Grades</u></b>	Pre-2010	A-E	<b>7%</b>	<b>52%</b>	<b>42%</b>
16		<b><u>Mean Forecasted Grades</u></b>	Pre-2010	A-E	<b>20%</b>	<b>42%</b>	<b>39%</b>
17		<b><u>Mean Predicted Grades</u></b>	Post-2010	A*-E	<b>10%</b>	<b>17%</b>	<b>73%</b>
18		<b><u>Mean Forecasted Grades</u></b>	Post-2010	A*-E	<b>15%</b>	<b>46%</b>	<b>39%</b>
19							
20	<b>Medical school applicants and students: Alevels and other qualifications</b>						
21	Students: Lumb & Vail [38]	Predicted grades	1995	A-E	7%	52%	41%
22	Applicants: Alevels (this study)	Predicted grades	2010-18	A*-E	7%	49%	45%
23	Applicants: EPQ (this study)	Predicted grades	2010-19	A*-E	14%	52%	34%
24	Applicants: SQA Adv. Highers (this study)	Predicted grades	2010-18	A-D	3%	60%	38%
25		<b><u>Mean Predicted Grades (Medics)</u></b>			<b>8%</b>	<b>53%</b>	<b>39%</b>
26							
27	Gill and Chang (2015) [36]	Forecasted GCSE Grades	2013	A*-G	12%	47%	41%
28	Gill and Benton (2015) [35]	Forecasted GCSE Grades	2014	A*-G,U	14%	44%	42%
29		<b><u>Mean Forecasted GCSE grades</u></b>			<b>13%</b>	<b>45%</b>	<b>42%</b>

5 year course  
6 year course  
5 year course  
6 year course



\* Although many medical schools run a course which can broadly be divided into basic medical sciences and clinical parts, some courses are far more integrated

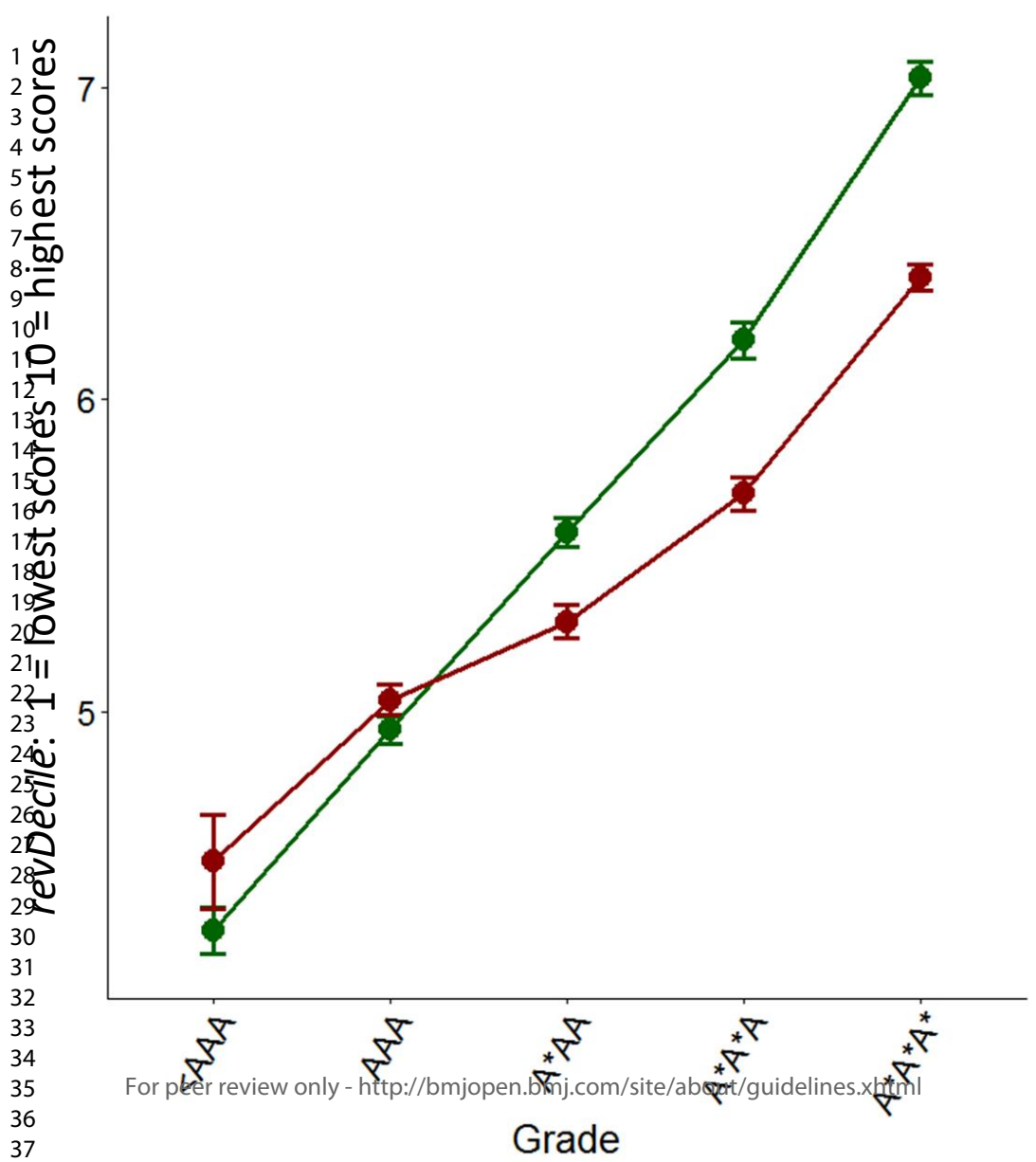
### Intercalated/Integrated BScs are compulsory in some medical schools and optional in others

\* EPM is an integrated score awarded in Y5 but based on results from Y1, Y2, Y3 and Y4

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

Year (Academic: Sept to August; Training: August to July)





		Selection measures in all applicants					Undergraduate outcome measures			Postgraduate outcome measures		
		GCSE grades	Predicted Alevels	Alevel grades	UKCAT	BMAT	EPM	SJT	PSA	MRCP(UK) Part 1	MRCSC Part A	
1			<b>r 1</b>	<b>0.452</b>	<b>0.421</b>	<b>0.265</b>	<b>0.223</b>	<b>0.180</b>	<b>0.190</b>	<b>0.201</b>	<b>0.212</b>	<b>0.173</b>
2	GCSE grades	N	22150	22150	22145	4935	12230	12185	12265	890	430	
3												
4	Predicted A-level grades	<b>r 0.452</b>	<b>1</b>	<b>0.585</b>	<b>0.272</b>	<b>0.326</b>	<b>0.198</b>	<b>0.160</b>	<b>0.226</b>	<b>0.283</b>	<b>0.181</b>	
5	Selection measures in all applicants	N	22150	22955	22520	5225	12560	12515	12600	910	440	
6												
7	Attained A-level grades	<b>r 0.421</b>	<b>0.585</b>	<b>1</b>	<b>0.326</b>	<b>0.416</b>	<b>0.297</b>	<b>0.195</b>	<b>0.306</b>	<b>0.421</b>	<b>0.358</b>	
8		N	22150	22955	22520	5225	12560	12515	12600	910	440	
9												
10	UKCAT total	<b>r 0.265</b>	<b>0.272</b>	<b>0.326</b>	<b>1</b>	<b>0.483</b>	<b>0.115</b>	<b>0.243</b>	<b>0.238</b>	<b>0.200</b>	<b>0.181</b>	
11		N	22145	22520	5225	5080	12385	12340	12420	900	435	
12												
13	BMAT sections 1 and 2	<b>r 0.223</b>	<b>0.326</b>	<b>0.416</b>	<b>0.483</b>	<b>1</b>	<b>0.089</b>	<b>0.239</b>	<b>0.321</b>	<b>0.378</b>	<b>0.319</b>	
14		N	4935	5225	5225	5080	4850	4840	4875	450	240	
15												
16	UKFPO EPM decile	<b>r 0.180</b>	<b>0.198</b>	<b>0.297</b>	<b>0.115</b>	<b>0.089</b>	<b>1</b>	<b>0.319</b>	<b>0.470</b>	<b>0.509</b>	<b>0.535</b>	
17		<b>rTPa</b>	<b>0.213</b>	<b>0.251</b>	<b>0.403</b>	<b>0.149</b>	<b>0.101</b>	-	-	-	-	-
18	Undergraduate outcome measures	N	12230	12560	12385	4850	12515	12505	905	440		
19												
20	UKFPO SJT score	<b>r 0.190</b>	<b>0.160</b>	<b>0.195</b>	<b>0.243</b>	<b>0.239</b>	<b>0.319</b>	<b>1</b>	<b>0.346</b>	<b>0.351</b>	<b>0.274</b>	
21		<b>rTPa</b>	<b>0.223</b>	<b>0.203</b>	<b>0.267</b>	<b>0.310</b>	<b>0.267</b>	-	-	-	-	-
22		N	12185	12515	12340	4840	12515	12475	905	435		
23												
24	PSA score	<b>r 0.201</b>	<b>0.226</b>	<b>0.306</b>	<b>0.238</b>	<b>0.321</b>	<b>0.470</b>	<b>0.346</b>	<b>1</b>	<b>0.500</b>	<b>0.483</b>	
25		<b>rTPa</b>	<b>0.236</b>	<b>0.287</b>	<b>0.415</b>	<b>0.305</b>	<b>0.360</b>	-	-	-	-	-
26		N	12265	12600	12600	12420	12505	12475	910	440		
27												
28	MRCP(UK) Part 1	<b>r 0.212</b>	<b>0.283</b>	<b>0.421</b>	<b>0.200</b>	<b>0.378</b>	<b>0.509</b>	<b>0.351</b>	<b>0.500</b>	<b>1</b>	...	
29		<b>rTPa</b>	<b>0.272</b>	<b>0.360</b>	<b>0.601</b>	<b>0.273</b>	<b>0.398</b>	<b>0.586</b>	<b>0.391</b>	<b>0.576</b>	-	-
30	Postgraduate outcome measures	N	890	910	900	450	905	905	910	-	10	
31												
32	MRCSC Part A	<b>r 0.173</b>	<b>0.181</b>	<b>0.358</b>	<b>0.181</b>	<b>0.319</b>	<b>0.535</b>	<b>0.274</b>	<b>0.483</b>	...	<b>1</b>	
33		<b>rTPa</b>	<b>0.196</b>	<b>0.216</b>	<b>0.519</b>	<b>0.282</b>	<b>0.313</b>	<b>0.618</b>	<b>0.306</b>	<b>0.575</b>	-	-
34		N	430	440	435	240	440	435	440	10	-	
35												
36	Undergraduate (n=3)	<b>r 0.190</b>	<b>0.195</b>	<b>0.266</b>	<b>0.199</b>	<b>0.216</b>	-	-	-	-	-	
37		<b>rTPa</b>	<b>0.224</b>	<b>0.247</b>	<b>0.362</b>	<b>0.255</b>	<b>0.243</b>	-	-	-	-	
38	Postgraduate (n=2)	<b>r 0.193</b>	<b>0.232</b>	<b>0.390</b>	<b>0.191</b>	<b>0.349</b>	<b>0.522</b>	<b>0.313</b>	<b>0.492</b>	-	-	
39		<b>rTPa</b>	<b>0.234</b>	<b>0.288</b>	<b>0.560</b>	<b>0.278</b>	<b>0.356</b>	<b>0.602</b>	<b>0.349</b>	<b>0.576</b>	-	-
40	Undergraduate and Postgraduate (n=5)	<b>r 0.191</b>	<b>0.210</b>	<b>0.315</b>	<b>0.195</b>	<b>0.269</b>	-	-	-	-	-	
41		<b>rTPa</b>	<b>0.228</b>	<b>0.263</b>	<b>0.441</b>	<b>0.264</b>	<b>0.288</b>	-	-	-	-	
42												
43	Range restriction (uX) = SD(entrants)/SD(applicants)	<b>uX</b>	<b>0.955</b>	<b>0.958</b>	<b>0.888</b>	<b>0.890</b>	<b>0.994</b>	-	-	-	-	-
44												
45		<b>uX</b>	<b>0.833</b>	<b>0.954</b>	<b>0.883</b>	<b>0.842</b>	<b>1.055</b>	<b>0.962</b>	<b>0.997</b>	<b>0.957</b>	-	-
46												
47		<b>uX</b>	<b>0.985</b>	<b>0.998</b>	<b>0.835</b>	<b>0.761</b>	<b>1.123</b>	<b>0.946</b>	<b>0.958</b>	<b>0.927</b>	-	-

		UKFPO		BMJ Open Selection grades:			Absolute	Relative	
	Decile	score	<i>RevDecile</i>	Attained	Predicted	Odds Ratio	difference	increase	
1	Worst	10	34	1	7.2%	8.1%	1.141	0.9%	13.0%
2		9	35	2	9.4%	10.6%	1.135	1.1%	12.0%
3		8	36	3	10.1%	11.1%	1.107	1.0%	9.5%
4		7	37	4	10.7%	11.2%	1.052	0.5%	4.6%
5		6	38	5	10.7%	10.8%	1.003	0.0%	0.3%
6		5	39	6	10.6%	10.4%	0.978	-0.2%	-2.0%
7		4	40	7	10.7%	10.4%	0.970	-0.3%	-2.7%
8		3	41	8	10.3%	9.7%	0.935	-0.6%	-5.8%
9		2	42	9	10.2%	9.1%	0.882	1.1%	-10.7%
10		1	43	10	10.1%	8.8%	0.853	-1.4%	-13.4%
11	Best								

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

*The predictive validity of A-level grades and teacher-predicted grades in UK medical school applicants: A retrospective analysis of administrative data in a time of COVID*

13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

## Supplementary information

17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

### Contents

20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

1. Supplementary literature review, including a summary of events from March to November 2020.	2
2. Supplementary Methods including a table of measures .....	11
3. Supplementary Results including the Extended Project Qualification (EPQ) and SQA Advanced Highers. ....	13
4. Supplementary Tables 1, 2, 3, 4, 5 & 6 and Supplementary Figures 1, 2, 3, 4 & 5 .....	20
5. <i>Appendix</i> : Are independent (private sector) schools more accurate in their A-level predictions?..	32
6. <i>Appendix</i> : Appendix Tables 1 & 2 and Appendix Figures 1, 2, 3 & 4. ....	36
References .....	43

43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

*Note*: The Supplementary Material contains extended versions of the Literature Review from the main paper, the Method section of the main paper, and the Results section of the main paper. In order to maintain the flow and continuity of the supplementary material, some of the material is duplicated from the main paper. In addition to the supplementary literature review, methods and results, there is also an appendix which looks in depth at the issue of whether independent (private sector) schools are more accurate in their A-level predictions.

## 1. Supplementary literature review, including a summary of events from March to November 2020.

*Overview of literature on predicted, forecasted and attained A-level grades.* The majority of studies reported here are also discussed in the main paper, in much more abbreviated form, but here are described more discursively.

### *University applications in general*

Petch in 1964<sup>1</sup> did what Wilmut has described as “one of the earliest and most celebrated studies of teacher estimates of examination result”<sup>2</sup> (p.60), describing how Petch found, “grade agreement in about 43% of cases, but the examination grade was higher than the teacher estimate in 18% of cases, but lower in 39% of cases, sometimes heavily so”.

Two other early studies were by Murphy in the first of which in 1979 he compared actual and predicted grades both for A-levels and, unusually, for O-levels (the predecessor of GCSEs)<sup>3</sup>, including two-way tables of predicted vs actual grades. Of 291 results the predicted grades were accurate in 27% of cases, over-predictions in 44% and under-prediction in 29% of cases. Teachers were also asked to provide a rank order of students, and overall these correlated 0.6 with rank order in the examination, although individual teachers showed a range of correlations from just less than zero through to more than 0.9. Murphy’s 1981 study drew on application forms submitted to UCAS (now UCAS) by 15,109 candidates, of which “a large number included teachers’ pre-examination estimates of A-level grades” (with predicted grades being A, A/B, B, B/C, C etc). Results were broken down by exam board and also by subject. The overall correlation of predicted and actual grades was 0.66, with Physics, Chemistry and French showing the highest correlations. The study also looked at A-level – O-level correlations<sup>4</sup>. Although described as predicted grades, these data are actually best described as being *forecasted grades*.

More recent studies have mostly been concerned with the relationship of attained A-level grades and the predicted A-level grades entered on UCAS application forms by teachers. UCAS changed the way it collected such data in 2009, so that for UK-domiciled applicants subject-level predicted grades were available, rather than as earlier when predicted grades were only available as total point scores<sup>5</sup>. For various reasons, not all A-levels have predicted grades. Most analyses are for candidates across all ability levels. Note that A\* grades were only introduced in 2010.

In a study of 2009 applicants<sup>5</sup>, overall *accuracy* at the subject level for A-levels for 219,744 A-levels was 52%, with predicted and attained grade being the same. In 42% of cases predicted grades were over-estimates, and in only 7% were they under-estimates. A grades tended to be predicted more accurately but that in part reflects that A grades cannot be under-predicted (or E grades over-predicted).

Female candidates showed a slight tendency for grades to be more accurately predicted (52.3% vs 51.1% in males). Socio-economic group showed strong relationships to accuracy, with 58% accurate predictions in the Higher Managerial group and 43% in the Routine group, but that in part reflects different actual A-level achievement (58% of Managerial candidates receiving an A grade compared to 33% of Routine candidates). The Higher Managerial group had the greatest over-prediction and the Routine group the highest under-prediction. Considering ethnicity, 53% of White applicants had accurate predictions compared with 47% of Asian ethnicity, and 39% of those of Black ethnicity. Centre (school) was related to accuracy, with 64% accuracy in Independent schools, 47% in state schools, and 40% in those in Further or Higher education. The authors note that multivariate analyses are probably needed to tease apart the relationships between the various correlates of accuracy. Other analyses looked at disability, region, and nation within the UK. Number of choices also related to accuracy, applicants making four choices being more accurate than those making five choices, but it was suggested that was because of the majority of the former being higher attainers applying to Medicine, Dentistry or Veterinary Medicine. The paper concluded that it is difficult to

1  
2  
3 separate out the various factors involved in accuracy, not least because of the ceiling and floor  
4 effects for high and low attainers <sup>5</sup>.

5  
6 Wyness <sup>6</sup> analysed aggregated data provided by UCAS for the applicants from 2013-15, and hence A\*  
7 grades were included in the analysis. Overall only 16.1% of grades were accurately predicted, a much  
8 lower figure than the earlier study using 2009 data <sup>5</sup>, perhaps because of the inclusion of the new A\*  
9 grades. 8.54% of grades were under-predicted, while 75.4% of grades were over-predicted. As with  
10 the 2009 data, there was a clear relationship between over-prediction and attained grade, although  
11 it is noted that there are strong ceiling effects at work. As with the 2009 study, independent schools  
12 provided the most accurate predictions. Applicants from disadvantaged backgrounds showed  
13 moderate to severe over-prediction. Asian and Black applicants were also more likely to be severely  
14 over-predicted. There were no differences between male and female applicants. The report is  
15 particularly interesting as it looks at prediction in high ability students, defined as AAB or above. The  
16 difference between the most and least disadvantaged in this group is much smaller, with 44.0%  
17 overpredicted in the most disadvantaged and 47.4% in the least disadvantaged. There was some  
18 evidence that under-predicted applicants tended to show under-matching (i.e. entering less  
19 competitive universities than their actual grades might predict). Further analyse and discussion of  
20 these data are provided elsewhere <sup>7,8</sup>.

21  
22  
23 UCAS in 2017 provided some limited data on over-prediction and under-prediction of A-levels since  
24 the introduction of A\* grades, with data for 2012, 2016 and 2017 <sup>9</sup>. Overall 19.5%, 16.3% and 16.0%  
25 of predictions were accurate, with over-prediction in 68.4%, 74.3% and 73.3% of cases, and under-  
26 prediction in 11.8%, 9.0% and 10.4% (figures from EoC17\_Figure7\_9\_database.csv<sup>a</sup>). UCAS  
27 commented that, the gap between achieved and predicted A-level grades, “continues to widen”  
28 (p.23), although a comparison of 2016 and 2017 results concluded that there was little effect  
29 due to the reforms in A-levels that took place in 2017.

30  
31 Not all studies have used the *predicted grades* provided to UCAS for use by universities in selection,  
32 which for medical school applicants would have been by mid-October). Until 2015 teachers were  
33 also asked, by the end of the following May, just before A-levels were sat, to provide *forecasted*  
34 *grades* to Awarding Organisations, and those grades then contributed in part to decisions on  
35 grading. Forecasted grades are clearly of particular interest given proposals for calculated grades to  
36 be based on estimates of performance by schools during May. Three analyses are available, for  
37 candidates taking A-levels in 2009 <sup>10</sup>, 2012 <sup>11</sup> and 2014 <sup>12</sup> which are before and after A\* grades were  
38 introduced. A primary interest must be the comparison of these forecasted grades with the more  
39 usually studied predicted grades, described earlier for 2009 <sup>5</sup> and 2012 <sup>9</sup>. Note that the studies of  
40 forecasted grades are only for OCR (Oxford, Cambridge and Royal Society of Arts Examination Board)  
41 and hence include all A-level candidates, whereas the studies of predicted grades are for university  
42 applicants. Supplementary table 1 compares the two sets of predictions. In 2009 there is little  
43 difference between predicted and forecasted grades in accuracy, with a small diminution of over-  
44 predictions. The picture three years later, in 2012 after A\* grades have been introduced, is rather  
45 different. Forecasted grades have an accuracy of 48% compared with only 20% for predicted grades.  
46 Taken overall it is difficult to reconcile the two studies which are only three years apart. Based on  
47 the 2009 data it would seem that predictions in May are no more accurate than those in October,  
48 whereas the 2012 data suggest that May predictions are much more accurate than October  
49 predictions. Having said that, even in May 2012, slightly less than a half of forecasted grades are  
50 accurate, with the same grade as in October.

51  
52 It should be noted, as pointed out earlier, that the early studies by Murphy should probably be  
53 regarded as being of forecasted and not predicted grades.

54  
55  
56  
57  
58  
59  
60  

---

<sup>a</sup> <https://www.ucas.com/file/140426/download?token=tUxAGXtt>

1  
2  
3 *Grade point predictions.* The analyses described so far have been at the level of A-level subjects.  
4 Students mostly take three or sometimes more A-levels, and universities usually look at the three  
5 best grades attained. Scoring grades as A\*=12, A=10, B=8, C=6, D=4 and E=2 then a candidate  
6 passing three A-levels will score between 6 and 36 points for their three best grades<sup>b</sup>. Two studies<sup>10</sup>  
7<sup>11</sup> have pointed out the difficulty of using totalled points. As an example, a candidate predicted AAA  
8 will be predicted 30 points but may attain grades AAA or grades A\*A\*D; both are equally accurate in  
9 point terms but not in grade terms. Total predicted points are important in that UCAS for a number  
10 of years only provided total predicted points for the best three A-levels, without subjects or  
11 individual grades being specified<sup>c</sup>.  
12

13  
14 UCAS in 2016 reviewed predicted and actual A-level grade points in applicants from 2010 to 2015<sup>13</sup>  
15 considering the best three grades attained. Achieved grades were one or two grades in total lower  
16 for attained than predicted grades. About a half of applicants in 2015 missed predicted total grades  
17 by two or more grades (e.g. ABB rather than AAA), a proportion that had increased by a third since  
18 2010. Simple analyses in particular showed that missing predicted grades was associated with having  
19 *lower* predicted grades overall (as in the earlier analyses at the subject level). Multivariate analyses  
20 i.e. taking other factors into account, found missing predicted grades was associated with having  
21 *higher* predicted grades, lower GCSE attainment, taking biology, chemistry and maths, having Asian,  
22 Black, Mixed and Other ethnicity, coming from disadvantaged areas, being female, and having '[pre-  
23 A-level]unconditional offers'. Of particular interest is the relationship to GCSE grades, which have a  
24 strong relationship to A-level attainment<sup>14</sup> which is clearly seen in the UCAS data (see their figures 5  
25 and 6).  
26

#### 27 *What are predicted grades and how are they made?*

28  
29 UCAS, in its document, "Predicted grades – what you need to know"<sup>d</sup> says that "A predicted grade is  
30 the grade of qualification an applicant's school or college believes they're likely to achieve in positive  
31 circumstances." Later the document says predicted grades should be, "**in the best interests of**  
32 **applicants** – fulfilment and success at college or university is the end goal ", and "**aspirational but**  
33 **achievable** – stretching predicted grades are motivational for students, unattainable predicted  
34 grades are not" (all emphases in original). It also says that grades should be "determined by  
35 professional judgement" and be data-driven, including "past Level 2 and Level 3 performance,  
36 and/or internal examinations to inform your predictions".  
37

38  
39 Gill<sup>15</sup> has described the relatively sparse literature on how teachers estimate grades. Gill's own  
40 study followed the methodology of Child and Wilson<sup>e</sup> although that study is not in the public  
41 domain. Gill sent questionnaires in May to selected OCR exam centres concerning Chemistry, English  
42 Literature and Psychology, and as well as estimating grades teachers were also asked to rank within  
43 grades, the method currently being adopted by Ofqual for calculated grades<sup>f</sup>. Teachers also  
44  
45

---

46  
47 <sup>b</sup> Some studies, including my own earlier ones, score A\*=6, B=5, etc.. Such schemes became less popular with  
48 the advent of AS-grades, which were scored as half of an A-level, and hence it made sense to double the points  
49 available for a full A-level so that totals remained integer. With the near disappearance now of AS-levels that  
50 rationale makes less sense.

51 <sup>c</sup> Earlier studies, such my 1991 cohort, had to extract predicted grades from UCAS references, and hence they  
52 are often embedded in free text, making it difficult to match them up with specific A-level subjects.

53 <sup>d</sup> <https://www.ucas.com/advisers/managing-applications/predicted-grades-what-you-need-know> [Accessed  
54 13th April 2020].

55 <sup>e</sup> Child S, Wilson F. An investigation of A level teachers' methods when estimating student grades. Cambridge:  
56 Cambridge Assessment (Unpublished document, October 2015).

57 <sup>f</sup> One teacher refused to take part because of the difficulty of ranking 260 students sitting one exam. Another  
58 teacher commented, "it was easier for smaller centres to make predictions because they know individual  
59 students better" (p.42). The paper in fact comments that, "Responses to the questionnaire were more likely to  
60 come from smaller centres. ... [T]he maximum centre size amongst the sample data was only 40 for Chemistry

1  
2  
3 indicated the evidence they had used for each decision. The response rate was extremely low (2.8%).  
4 About 45% of forecasted grades were accurate (which is similar to the 48% in supplementary table  
5 1). Detailed A-level raw marks were also available and could be correlated with rankings, giving  
6 correlations of .87, .76 and .83 for the three subjects. Those correlations are high, and certainly are  
7 higher than a meta-analytic estimate of the effect size for teachers predicting academic achievement  
8 in pupils of 0.63 (SE=.03), although there was substantial heterogeneity. They are also higher than  
9 Murphy's 1979 estimate of 0.66 for the correlation of rankings and exam marks<sup>3</sup>. The most  
10 important information said by teachers to be used when predicting grades was performance in mock  
11 exams, and observations of quality of work and commitment, with oral presentation also important.  
12 Amongst other topics written in, the most important was the opinion of other teachers both in the  
13 same subject and other subjects, including the head of department. Other teachers raised concerns  
14 about the lack of high stakes for mock exams which meant that students did not treat them  
15 seriously. There were also concerns about the loss of AS-levels to help in prediction.  
16

17  
18 *Other examinations.* We know of no studies that have looked at accuracy of prediction of Scottish  
19 Highers or Advanced Highers, of the EPQ (Extended Project Question) used in England, or of other  
20 examinations carried out in the UK.  
21

### 22 *Applications to medical school*

23 Relatively few studies have looked at predicted grades in medical school applicants, although those  
24 studies do show a tendency to ask rather more stretching questions, perhaps because of the  
25 different interests of the researchers, and the specificity of the course and its outcomes.  
26

27 Lumb and Vail pointed out that predicted grades are particularly important in the shortlisting phase  
28 of medical student selection<sup>16</sup>. They studied 1661 applications in 1995 to a single medical school  
29 who had estimated grades for 5053 A-levels, 52% of predictions being accurate, 41% were over-  
30 estimated and 7% under-estimated<sup>16</sup>. The authors presented an ROC curve (but not the area under  
31 the curve), and concluded that, "... selectors for medical schools can have some confidence in the  
32 accuracy of predictions and we should therefore continue to use them ... [for] selecting the doctors  
33 of the future." (p.311).  
34

35  
36 Richardson et al, studied 721 entrants from 1991 to 1994 to a single medical school<sup>17</sup>. Unusually  
37 they looked at predictive validity, assessing how well predicted and actual A-level grades related to a  
38 composite outcome on the pre-clinical course. Predicted and actual A-level grades showed a minimal  
39 correlation ( $r=0.024$ ), but selection would have imposed range restriction. Pre-clinical exam  
40 performance correlated 0.318 ( $p<.001$ ) with attained A-level grades, but only 0.041 (NS) with  
41 predicted A-level grades. This is a rare study in which predictive validity was assessed and it implied  
42 that selection should be on actual grades rather than predicted grades, concluding in contra-  
43 distinction to Lumb and Vail that, "medical school admissions panels would be well advised to take  
44 the predicted grade with a sizeable pinch of salt" (p.296).  
45  
46  
47

48 A third study, by one of the present team, took a different approach, using path modelling to assess  
49 the causal inter-relationships between GCSE grades, predicted A-level grades, receipt of an offer,  
50 actual A-level grades, and acceptance at medical school in an original sample size of 6901 applicants  
51 to five English medical schools<sup>18</sup>. A-level estimates were predicted by GCSE grades ( $\beta=0.89$ ), with  
52 attained A-level grades predicted by both GCSE grades ( $\beta=0.44$ ) and predicted A-level grades  
53 ( $\beta=0.74$ ). A substantive question of interest was whether the paths in the model differed  
54 between White and non-White candidates, with it being shown that none of the relationships  
55 described showed ethnic differences (although non-white candidates were significantly less likely  
56  
57  
58

---

59 (compared with 423 amongst all centres), 26 for English Literature (compared with 180) and 32 for psychology  
60 (compared with 378)."



1  
2  
3 than White candidates to receive an offer based on predicted A-level grades). Although the study  
4 reported no follow-up into the medical course, this dataset is analysed further below to assess  
5 predictive validity for postgraduate examination performance.  
6

7 *A comment on issues in studying predicted A-level grades.*

8 Although predicted A-level grades have been an integral part of university application and selection  
9 in the UK for four decades, obtaining data on them is less than easy. Early studies, including my own,  
10 as well as those of other medical researchers, simply resorted to having researchers transcribe  
11 grades from paper UCCA and UCAS application forms, although often that was not easy in earlier  
12 forms as the predictions were often embedded in the free text of the Referee's Statement. Until  
13 2009 UCAS only recorded the summed score of the best three A-levels, so that study of specific  
14 subjects was not possible. Even now obtaining UCAS data on predicted grades is less than easy, and  
15 Boliver in 2013 comments, "It would have been desirable to include predicted A-level grades...  
16 . Unfortunately UCAS are unable to provide this information in microdata form because of uncertainty  
17 about its validity in the case of applicants whose application is not linked to a school or college ...  
18 (personal communication from UCAS)." <sup>19</sup> Similarly Wyness in 2016 in her study of three years of  
19 UCAS data comments that, "The data are aggregate (for reasons of privacy)" <sup>6</sup>, which means of  
20 course that proper analyses at the level of individual participants are not possible. There is an irony  
21 here in that of course all universities have access to predicted grades provided by UCAS as a part of  
22 the admissions process, but subsequently obtaining those data for research is often very difficult.  
23 The data for the present study are the result of an important collaboration between UKMED and  
24 UCAS, with UCAS providing detailed information on applicants to UK medical schools for inclusion in  
25 the database, which is hosted in a safe haven to ensure strict controls on access; we are very  
26 grateful to UCAS for that collaboration without which the present study would not be possible.  
27  
28  
29  
30  
31

32 *A summary of events surrounding the cancellation of Alevels from March to November 2020.*

33 The research for the present paper was carried out in April and May 2020, in parallel with the study  
34 of attitudes and responses of medical school applicants to the cancellation of A-level examinations  
35 and their replacement by 'Calculated Grades'<sup>20,21</sup>. The main bulk of the present paper was written  
36 between March and June 2020, with a preprint being published in June 2020<sup>22</sup>. Key findings from  
37 this paper and the accompanying applicant attitudes paper were presented to UK medical  
38 admissions tutors at a meeting of the MSC-SA (Medical Schools Council Selection Alliance) on 6<sup>th</sup>  
39 May 2020, and drafts of the two papers distributed. The present paper is in large part a statement of  
40 how we understood the situation in May 2020, with a few amendments for clarity.  
41  
42  
43

44 Inevitably events, in large part political but also with many practical ramifications for medical schools  
45 and student selection, continued onwards from June 2020, particularly with the publication of Alevel  
46 results in August 2020, and the abandonment of the algorithm for calculating A-level grades, with its  
47 inevitable fallout. The following paragraphs provide a summary of events, both going forward and  
48 also, to some extent, looking back to March 2020 as a result of documents published in September  
49 and October 2020.  
50

51 In July and August 2020 things moved rapidly, with dramatic changes taking place. It would have  
52 been extremely confusing and probably misleading to have tried to incorporate those changes into  
53 the text of the main paper. Instead we hope this postscript will give readers a sense of what  
54 happened, to what extent events were correctly or incorrectly predicted by us, what impact the  
55 present paper may have had, and what may be the implications for the future.  
56  
57  
58  
59  
60

1  
2  
3 The story is best told chronologically and we mostly use reports from newspapers, and refer  
4 interested readers to a brief summary on Wikipedia<sup>g</sup> and a journalistic review on the BBC website<sup>h</sup>.  
5  
6  
7

### 8 **The awarding of A-level grades in 2020: the story from March to November 2020**

9

10 As described in our main paper, on 20<sup>th</sup> March 2020 public examinations in the UK including A-levels  
11 were cancelled. On 3<sup>rd</sup> April *Ofqual* announced that exam grades in England would be replaced with  
12 Calculated Grades. Calculated Grades were to consist of Centre Assessment Grades (also called  
13 Centre Assessed Grades or CAGs), estimated by teachers that centres (mostly schools and colleges)  
14 would submit to *Ofqual*. *Ofqual* would moderate these CAGs using an algorithm – the details of  
15 which had not yet been published but, it was stated, would be based on the prior performance of  
16 pupils within schools. The Scottish Qualification Authority (SQA), Qualifications Wales, and the  
17 Northern Irish Council for Curriculum, Examination and Assessment (CCEA) also announced that they  
18 would use a broadly similar approach to that of *Ofqual*.  
19

20  
21 Schools (centres) had to return their teacher-estimated CAGs to *Ofqual* in June 2020. On June 16<sup>th</sup> a  
22 report in *The Times* said that “Teachers have marked too generously in allocating GCSE and A-level  
23 grades this year, research suggests” (*The Times*, 16<sup>th</sup> June), the article being based on a report from  
24 FFT Education Datalab, which actually had only asked about GCSEs, and had no data on A-levels<sup>i</sup>. In  
25 July a *Guardian* article reported a statement from *Ofqual* that “a substantial number of students  
26 would receive at least one adjusted grade – usually downwards – as a result of a standardisation  
27 process” although they “sought to allay fears that certain groups of pupils, ... could be disadvantaged  
28 by calculated grades. *Ofqual* said their analysis had found no evidence of widening of gaps in  
29 attainment”. (*The Guardian (G)*, 21<sup>st</sup> July).  
30  
31  
32

33 SQA results in Scotland are announced a week before those in England, so the SQA results  
34 announced on August 5<sup>th</sup> 2020 gave a preview of what was to come the following week in the rest of  
35 the UK. The Scottish results were immediately controversial when it emerged that the moderation of  
36 teacher-estimated grades (CAGs) by an algorithm had resulted in a quarter of grades being adjusted  
37 downwards. The Scottish Education Secretary, John Swinney, said that without those adjustments  
38 the pass rates would be up on the previous year by 14% for Highers and 13.4% for Advanced  
39 Highers. He added that, “... these robust processes mean we have upheld standards... All exam  
40 systems rely on an essential process known as moderation to uphold standards. This ensures an A  
41 grade is the same in every part of the country, making the system fair for everyone, and across all  
42 years.” (*G*, 4<sup>th</sup> August).  
43  
44  
45

46 Teachers, students, parents and the media were unhappy with the moderation. A *Daily Telegraph*  
47 editorial entitled “Exam moderation is a gross injustice” attacked statistical modelling in general, and  
48 the SQA process in particular, which “gives poorer marks to children living in deprived areas ...  
49 [without] recognition of individuals who buck the general trend” (*Daily Telegraph (DT)*, 5<sup>th</sup> Aug), and  
50 asked “Is the same fiasco about to be inflicted on A-level students in England and Wales?”. By 11<sup>th</sup>  
51 August, students in Scotland were protesting on the streets, Nicola Sturgeon, the First Minister, was  
52 apologising for the exam results debacle, and the Scottish government was facing a vote of no  
53  
54  
55

56  
57 <sup>g</sup> [https://en.wikipedia.org/wiki/2020\\_UK\\_GCSE\\_and\\_A-Level\\_grading\\_controversy](https://en.wikipedia.org/wiki/2020_UK_GCSE_and_A-Level_grading_controversy)

58 <sup>h</sup> Coughlan, S. “Coronavirus: The story of the big U-turn of the summer”, <https://www.bbc.co.uk/news/education-54103612>

59 <sup>i</sup> <https://ffteducationdatalab.org.uk/2020/06/gcse-results-2020-a-look-at-the-grades-proposed-by-schools/>  
60

1  
2  
3 confidence (G, 11<sup>th</sup> Aug). Expectations in the media of problems with A-levels were also growing. On  
4 12<sup>th</sup> August it was announced in Scotland that the teacher-estimated CAGs downgraded by the SQA  
5 algorithm during moderation “would be reinstated” (G, 12<sup>th</sup> Aug).  
6

7  
8 In an attempt to prevent problems with A-levels, on August 12<sup>th</sup> 2020 the English education  
9 secretary, Gavin Williamson announced “a triple lock” for A-level students, whereby students could  
10 accept their Calculated Grade results, use the results of mock exams (practice exams which students  
11 take in schools), or use the results of real exams due to take place in Autumn 2020 after the start of  
12 the university academic year (*The Times*, (T), 12<sup>th</sup> August; T, 13<sup>th</sup> August). Protests were immediate  
13 as mock exams vary immensely, and many schools had been encouraged to cancel mock exams as a  
14 part of the Covid lockdown in March 2020.  
15

16  
17 A-level Calculated Grades (i.e. the teacher-estimated CAGs adjusted by the algorithm during  
18 moderation) were announced on 13<sup>th</sup> August 2020, and UCAS announced which students had  
19 obtained places at their chosen university based on these Calculated Grades. An immediate problem  
20 arose: following the Scottish government’s reversal, students in Scotland now had unadjusted SQA  
21 grades, which were higher on average and gave them an advantage over applicants with A-level  
22 Calculated Grades, which had been adjusted (G, 13<sup>th</sup> Aug). University admission processes were also  
23 becoming embroiled in confusion, and although universities had, “reassured ministers that they will  
24 ‘soften’ the grades they normally require” (T, 13<sup>th</sup> August), by the next day universities were accused  
25 of being inflexible (G, 14<sup>th</sup> Aug).  
26  
27

28  
29 It soon became apparent that schools in the private, fee-paying sector, had probably benefited from  
30 the algorithm, primarily because statistical predictions were less accurate for the small class sizes  
31 more likely to be found in private schools, and in those cases the teacher-estimated CAGs had been  
32 allowed to stand unadjusted. Although the Prime Minister Boris Johnson defended the system  
33 saying, “Let’s be in no doubt about it: the exam results that we’ve got today are robust, they’re  
34 good, they’re dependable for employers” (G, 14<sup>th</sup> August), many backbench MPs were in revolt,  
35 having been deluged with complaints from constituents (T, 14<sup>th</sup> August). The next day Gavin  
36 Williamson said, “No U-turn. No change” (T, 15<sup>th</sup> August), and although he did agree to waive fees  
37 for appeals against Calculated Grades, he insisted that the grades themselves would not change in  
38 order to avoid the grade inflation that had occurred in Scotland (T, 15<sup>th</sup> August).  
39  
40

41  
42 Meanwhile the *Ofqual* algorithm, published in a document over 300 pages in length, was being  
43 dissected carefully, and when one headteacher anonymously shared their school’s results, the  
44 problems became particularly apparent<sup>1</sup>. In the previous three years at their school, 12.5% of pupils  
45 had achieved A\* and none had got a U; however the algorithm meant only 3.7% of students  
46 (equivalent to just one student) received an A\* - much below the historic 12.5%. The algorithm also  
47 resulted in one student being awarded a U, despite no students at that school having received a U  
48 previously. The weekend newspapers attacked the government, “which deserved a U grade for this  
49 debacle” (*Sunday Times* (ST), 16<sup>th</sup> August). GCSE results, due on Aug 20<sup>th</sup>, were also on the horizon,  
50 with similar problems predicted (T, 15<sup>th</sup> Aug, p.14; *Observer* (O), 16<sup>th</sup> August). Students in England  
51 demonstrated outside the Department for Education in London. *Ofqual* also announced guidance on  
52 the role of mock exams in appeals only to withdraw it a few hours later (T, 17<sup>th</sup> August).  
53  
54  
55  
56

57  
58 <sup>1</sup> Hern,Alex. (2020) “Do the maths: analysis shows why England’s grading system is both imprecise and unfair”, *Guardian*,  
59 15<sup>th</sup> August, 2020, p.13; the analyses are based on Thomson, Dave, “A-Level results 2020: How have grades been  
60 calculated?”, 13<sup>th</sup> August 2020, <https://ffteducationdatalab.org.uk/2020/08/a-level-results-2020-how-have-grades-been-calculated/>.

1  
2  
3 On August 18<sup>th</sup> the government scrapped *Ofqual's* algorithm and reverted to unadjusted teacher-  
4 estimated CAGs (*G*, 18<sup>th</sup> August). *The Times*, normally a supporter of the Conservative Party, simply  
5 called its main editorial, "Another Fine Mess" (*T*, 18<sup>th</sup> August). The chairman and chief executive of  
6 *Ofqual* were criticised for having little experience of education (*T*, 19<sup>th</sup> August), and the Chief  
7 Executive eventually resigned on August 25<sup>th</sup>.  
8  
9

10 Several other problems now emerged, not the least being that universities would not know the (now  
11 unadjusted CAG) grades for several days. Once universities did receive these grades, they found that  
12 they did not have enough places to honour all the offers they had made students months earlier.  
13 This was because universities typically make more offers than they have places, knowing that a  
14 significant number of students will not meet those offers when they achieve lower exam grades than  
15 the Predicted Grades estimated by their teachers and submitted to UCAS when they apply to  
16 university. But now with the teacher-estimated CAGs replacing exam grades, and the resulting  
17 increase in the percentage of A and A\* grades from 28% to 38%, many more students than expected  
18 did in fact meet their university offers (*T*, 18<sup>th</sup> August). This caused problems for some students  
19 whose Calculated Grades (the grades adjusted by the algorithm) had been too low for their first  
20 choice university and so had accepted offers from their second choice, but now they found their  
21 unadjusted teacher-estimated CAGs enabled them to meet their first choice offer they inevitably  
22 wanted to go there. Other students seemed to have been left in limbo, needing to delay entry until  
23 the next year, with a potential knock-on effect for students taking A-levels in 2021 (*T*, 19<sup>th</sup> August).  
24 Some groups of A-level students had clearly fallen through the net and were in limbo, such as who  
25 were home-schooled or private A-level students, having no teachers to estimate their grades (*G*, 21<sup>st</sup>  
26 August).  
27  
28  
29  
30

31 Regarding medical schools specifically, *The Guardian* reported that the Norwich Medical School at  
32 the University of East Anglia had 185 places and a possible overshoot of 50, a 27% increase,  
33 emphasising that with medical school numbers more strictly limited than other university places and  
34 costing £50,000 a year this would have clear financial implications (*G*, 19<sup>th</sup> August). A news story  
35 based in part on the current research as published in pre-print on *medRxiv*, suggested that more  
36 medical students may be liable to drop out, a co-chairman of the Medical Schools Council suggesting  
37 that "we are going to have, on average, students with lower grades than in previous years" (*DT*, 20<sup>th</sup>  
38 August). A similar concern was also raised by headteachers in relation to GCSE grades, which it had  
39 now been announced would also be based on unadjusted teacher estimates, with fears that  
40 students, "could end up on unsuitable courses post-16 which could set them up for failure" (*G*, 21<sup>st</sup>  
41 August).  
42  
43  
44

45 The cap on university student numbers was released on 17<sup>th</sup> August, the medical school cap  
46 following on 20<sup>th</sup> August, resulting in a bulge of new undergraduates. There was also a bulge in  
47 admissions to sixth form colleges as a result of unadjusted teacher-estimated GCSE grades being  
48 higher than the expected exam grades (*G*, 21<sup>st</sup> August). There are financial implications for  
49 educational institutions but no-one would know the size of the problem until UCAS released its entry  
50 statistics for October 2020.  
51  
52

53 The controversies rumbled on into September 2020 as the House of Commons Select Committee on  
54 Education heard evidence from *Ofqual* and other bodies. *Ofqual* put out a lengthy statement on 2<sup>nd</sup>  
55 September in evidence to the Committee<sup>k</sup>. It set out the history according to the regulator, stating  
56 that in March its advice to the Secretary of State had been that, "the best option in terms of valid  
57  
58

59 <sup>k</sup> <https://www.gov.uk/government/news/written-statement-from-chair-of-ofqual-to-the-education-select-committee>,  
60 "Written statement from Chair of Ofqual to the Education Select Committee", 2<sup>nd</sup> Sept 2020.

1  
2  
3 qualifications would be to hold exams in a socially distanced manner”; however, “The decision to use  
4 a system of statistical standardised teacher assessments was taken by the Secretary of State and  
5 issued as a direction to *Ofqual*”. In reviewing the failure of the system, the conclusion was reached  
6 that, “a ‘better’ algorithm would not have made the outcomes significantly more acceptable. The  
7 inherent limitations of the data and the nature of the process were what made it unacceptable.... *To*  
8 *try to deliver comparable qualification results in the absence of students having taken any*  
9 *assessments (examinations) proved to be an impossible task*” (our emphasis). Cambridge  
10 Assessment’s submission to Select Committee provides a detailed timeline of collaborative efforts to  
11 inform decision making.<sup>l</sup> With the model running and results being calculated, from late July through  
12 August, Cambridge Assessment worked with Ofqual and DfE to understand possible unfairness in the  
13 outcomes, and to put in place adequate remedy. No doubt the post-mortem will continue for a long  
14 while.  
15  
16  
17

18 In autumn 2020, the needs of the next year began to be considered. One group, including several  
19 university Vice-Chancellors, argued that A-levels should be cancelled once more and replaced by  
20 teacher-estimated grades (*T*, 2<sup>nd</sup> October). Others argued more radically that teacher-estimated  
21 grades should permanently replace A-level examinations. On 12<sup>th</sup> October 2020 the Secretary of  
22 State announced that A-level exams would go ahead in England in summer 2021 with some minor  
23 changes, including being three weeks later than usual, and with results announced a few days later  
24 than typical<sup>m</sup>. Contingency measures would be in place for possible disruption, but were yet to be  
25 described, although a leaked newspaper report suggested that they might include more formal mock  
26 exams as a back-up when following earlier leaked reports (*G*, 10<sup>th</sup> October). Perhaps the most  
27 interesting comment by Williamson was that, “Exams are the fairest way of judging a student’s  
28 performance ...”, with its tacit acceptance that teacher-estimated grades perhaps actually were not  
29 fair to many students.  
30  
31  
32

33 To date other UK countries have taken different approaches. On 7<sup>th</sup> October the Scottish  
34 Government announced the cancellation of National 5 exams in 2021 although Higher and Advanced  
35 Higher examinations would be taken (BBC, 7<sup>th</sup> October). A month later on 10<sup>th</sup> November the Welsh  
36 Government announced that GCSE and A-level examinations in Wales would be cancelled, with  
37 grades being based on classroom assessments instead (BBC, 10<sup>th</sup> November). On the same day  
38 Education Minister Peter Weir announced that GCSE and A-levels examinations would be taken in  
39 Northern Ireland (BBC, 10<sup>th</sup> November). In November 2020 the Secretary of State announced his  
40 intention to consider post-qualification university admissions  
41 (<http://researchbriefings.files.parliament.uk/documents/CBP-8538/CBP-8538.pdf>).  
42  
43  
44

45 This postscript has been relatively brief, given the complexity of the events, and it has not attempted  
46 to summarise events occurring in 2021, with teachers now mostly but not entirely responsible for  
47 awarding Alevel grades, but under some control by Ofqual. That story is probably too long, and not  
48 yet complete, for it to be included here. The history of the events of 2020 may however help those  
49 new to the issues to navigate through the major changes that occurred. In research terms, in  
50 medical education, higher education more generally, and in secondary education, there seems little  
51 doubt that researchers will be following in detail the outcomes for the cohorts affected by the  
52 dramatic changes which resulted in a giant, unplanned, experiment, where notional grades awarded  
53  
54  
55  
56

57  
58 <sup>l</sup> <https://committees.parliament.uk/writtenevidence/11358/default/>, “Written evidence submitted by  
59 Cambridge Assessment”, 2<sup>nd</sup> Sept 2020.

60 <sup>m</sup> <https://www.gov.uk/government/news/students-to-be-given-more-time-to-prepare-for-2021-exams>, “Students to be  
given more time to prepare for 2021 exams”, 12<sup>th</sup> October 2020.

were probably different in many cases to what students would have been awarded in normal circumstances.

## 2. Supplementary Methods including a table of measures

Data for the present study comes from two separate primary sources:

“P89”. UKMED project UKMEDP089, “The UK Medical Applicant Cohort Study: Applications and Outcomes Study”, approved Dec 7<sup>th</sup>, 2018, with Dr Katherine Woolf as principal investigator, is an ongoing analysis as a part of UKMACS (UK Medical Applicant Cohort Study). Data are primarily concerned with the process of selection. In particular in the upload of 21<sup>st</sup> Jan 2020<sup>n</sup> there is detailed information from UCAS on all applicants to medical schools from 2007 to 2018, including all attained Key Stage 5 (Level 3) qualifications (e.g. A-levels and SQA) as well as teacher predicted grades for individual Key stage 5 qualifications.

“P51”. UKMED project UKMEDP051, “A comparison of the properties of BMAT, GAMSAT and UKCAT”, approved Sept 25<sup>th</sup>, 2017, with Dr Paul Tiffin as principal investigator, is an ongoing analysis of the predictive validity of admissions tests and other selection methods such as A-levels and GCSEs in relation to undergraduate and postgraduate attainment. A major feature of the study is the inclusion of data from UCAS, although in the 13<sup>th</sup> May 2019 data upload, which was used here<sup>o</sup>, UCAS predicted grades were only available as a composite, 18-point score, for application years 2010 to 2014. A new upload of the data in late April 2020 will provide more detailed information, but that will require quite extensive coding, etc., making it similar to the qualifications data for applicants in P89. For the present data upload, predicted A-level grades are in the old UCAS format consisting of a single number from 6 to 18 (i.e. 3 Es to 3 A\*s using A\*=6 coding). Outcome data for the P51 dataset are more extensive, and in particular include data for end of undergraduate training, including the UKFPO EPM measures, the UKFPO SJT as well as PSA (Prescribing Safety Assessment). Some data are available for later postgraduate examinations, but numbers inevitably are small for cohorts entering medical school in 2011 onwards.

*A-level grade scoring.* In both P89 and P51, A-level grades are expressed numerically on a standard scale of A\*=12 points, A=10, B=8, C=6, D=4 and E=2, or have been rescaled to that score.

The table below provides a detailed description of the source and coding of the measures used in the analyses:

Measure name	Description	Derivation	Source
GCSE grades	Average GCSE score from the best 9 GCSEs	The sum of the nine best grades (counting double science as two separate GCSEs)/the number of GCSEs included – i.e. 9 or fewer. This is the methodology used by the UKCAT-12 study (McManus, I.C., Dewberry, C., Nicholson, S. and Dowell, J.S. (2013) “The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of 12 UK medical schools”. <i>BMC Medicine</i> 11 (1), p. 244. ISSN 1741-7015. <a href="http://bmcmmedicine.biomedcentral.com/articles/10.1186/1741-7015-11-244">http://bmcmmedicine.biomedcentral.com/articles/10.1186/1741-7015-11-244</a>	GCSEs supplied to UKMED by UKCAT from data obtained from UCAS. We used the GCSEs associated with the first application.

<sup>n</sup> OUTPUT\_UCAS\_QUALS\_DEC\_20200121\_1.TXT and OUTPUT\_UCAS\_QUALS\_DEC\_20200121\_2.TXT (both dated 21/1/2020), and OUTPUT\_UCAS\_QUALS\_VER.TXT (dated 5/12/2019).

<sup>o</sup> UKCAT51\_APP\_ALL\_DATA\_13052019\_FILE1.SAV and UKCAT51\_APP\_ALL\_DATA\_13052019\_FILE2.SAV (both dated 13/5/2019).

		Further details are available in McManus IC, Dewberry C, Nicholson S, Dowell J: <i>The UKCAT-12 Study: Technical Report. 2012</i> , UKCAT Consortium: Nottingham, <a href="https://www.ucat.ac.uk/media/1185/ukat-technicalreport-march2012-withbackgroundandsummary-sep2013v3.pdf">https://www.ucat.ac.uk/media/1185/ukat-technicalreport-march2012-withbackgroundandsummary-sep2013v3.pdf</a>	
Predicted A-level grades	Predicted grades are provided by an applicant's teachers on the UCAS form the autumn prior to sitting A-levels.	UCAS supplied a score from the A-level grades declared by the applicant's teachers on the application. The highest three grades are considered only, adding up the following points per grade: A* = 6, A = 5, B = 4, C = 3, D = 2, E = 1. AS levels are not included. This variable was only available for 18-year-old applicants domiciled in England, Northern Ireland and Wales. To bring this into line with attained grades the number was multiplied by 2 to give a maximum of 36.  The value is taken from the 1st UCAS application present in the UCAS extract.	UCAS supplied data
Attained A-level grades	Total score from the best 3 A-levels.	Sum of the three highest A-level grades. Assign point scores to A-Level Grades in 2-point increments (A*=12, A=10, B=8, C=6, D=4, E=2, else=0). This is the methodology used by the UKCAT-12 study (see above). By using the best three A-levels we are able to accommodate the differing numbers of A-levels taken by applicants.  All grades were from 2010 onwards, when A* grades were available at Alevel, prior to that the maximum possible mark being A.	HESA qualifications data
UKCAT total	Total score on UKCAT test	A total scale score is generated by summing the individual scale scores of Verbal Reasoning, Decision Making, Quantitative Reasoning and Abstract Reasoning.  We used the score from the last available attempt which was therefore that associated with admission.	UCAT known as UKCAT at the time these data were created.  UCAT publish technical reports with the details. See: <a href="https://www.ucat.ac.uk/about-us/technical-reports/">https://www.ucat.ac.uk/about-us/technical-reports/</a>
BMAT sections 1 and 2	Combined BMAT score	BMAT_SECTION1 measures Aptitude and Skills. BMAT_SECTION2 measures Scientific Knowledge and Applications.	BMAT published details of the test. See: <a href="https://foundationprogramme.nhs.uk/faqs/educational-performance-measure-epm-faqs/">https://foundationprogramme.nhs.uk/faqs/educational-performance-measure-epm-faqs/</a>
UKFPO EPM decile	Medical school performance relative to peers ranked in deciles.	Students in a graduating cohort are ranked on their medical school performance (Educational Performance Measure, EPM). Individual schools decide which assessments to include in the EPM that meet the specified criteria and are required to consult with students and publish on their website the assessments included in that score.	UKFPO. Contained in the FP table in UKMED. See: Educational Performance Measure (EPM) 2019 Framework 2019. <a href="http://www.foundationprogramme.nhs.uk/sites/default/files/2018-07/UKFP%202019%20EPM%20Framework%20Final%200.pdf">http://www.foundationprogramme.nhs.uk/sites/default/files/2018-07/UKFP%202019%20EPM%20Framework%20Final%200.pdf</a>
UKFPO SJT score		The Situational Judgement Test (SJT) is a final year undergraduate test that assesses individuals' reactions to a number of hypothetical role-relevant scenarios, which reflect situations candidates are likely to encounter as a doctor. It seeks to provide a reliable measurement of the following non-academic domains: Coping with pressure, Working effectively as part of a team, Effective communication, Problem solving, and Commitment to professionalism.	UKFPO. Contained in the FP table in UKMED. See ISFP Project. Situational Judgement Test <a href="https://isfp.org.uk/sjt/">https://isfp.org.uk/sjt/</a> 14 <sup>th</sup> February 2019.
PSA score	Score relative to pass on	The British Pharmacological Society and MSC Assessment developed the Prescribing Safety Assessment (PSA) that allows all UK medical	MSC Assessment provide these data to UKMED. See:

	first attempt at PSA	students to demonstrate their competencies in relation to the safe and effective use of medicines. We used the score relative to the pass mark, as the pass mark varies by diet. We used the first attempt at the exam.	<a href="https://prescribingsafetyassessment.ac.uk/">https://prescribingsafetyassessment.ac.uk/</a>
MRCP (UK) Part 1	Score relative to pass on first attempt at MRCP (UK) Part 1	MRCP Part 1 is the entry-level examination accessible to doctors with a minimum of 12 months' postgraduate experience in medical employment. It covers a broad range of topics to ensure the level of knowledge is appropriate to physicians at the beginning of postgraduate training. We used the score relative to the pass mark, as the pass mark varies by diet. We used the first attempt at the exam.	The Royal Colleges provide data annually to the GMC for quality assurance purposes. The collection notices are published by year (see: <a href="https://www.gmc-uk.org/education/reports-and-reviews/progression-reports/downloads-resources-and-briefing-notes">https://www.gmc-uk.org/education/reports-and-reviews/progression-reports/downloads-resources-and-briefing-notes</a> ; e.g. Medical Royal College & Faculty Exam Data (2015) available at: <a href="https://www.gmc-uk.org/-/media/documents/exams-data-project---data-submission-briefing-note_pdf-56793364.pdf">https://www.gmc-uk.org/-/media/documents/exams-data-project---data-submission-briefing-note_pdf-56793364.pdf</a>  The data in UKMED is the variable EXAM_TOTAL_MARKS ; see <a href="https://www.ukmed.ac.uk/documents/UKMED_data_dictionary.pdf">https://www.ukmed.ac.uk/documents/UKMED_data_dictionary.pdf</a>
MRCS Part A	Score relative to pass on first attempt at MRCS Part A	The Intercollegiate MRCS Part A is designed to test knowledge of both applied basic science and principles of surgery in general to a level that a surgical trainee should have. It is a five-hour MCQ exam consisting of two papers taken on the same day. We used the score relative to the pass mark, as the pass mark varies by diet. We used the first attempt at the exam.	Details of data source etc are the same as for the MRCP examination (see above)

*Rounding and suppression criteria.* All data from HESA are required to be reported using their rounding and suppression criteria (<https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics>) and although not all data in the current study use HESA measures we have nevertheless applied the HESA criteria to all UKMED-based tables and values reported in this study. It should be noted in particular that the presence of a zero or a zero percentage may not always mean that there are no individuals in a cell of a table. All Ns are rounded to the nearest 5 which should easily flag up that rounding has been applied, all counts ending in 0 or 5. Percentages are only reported when the number of participants is greater than 22.5.

### 3. Supplementary Results including the Extended Project Qualification (EPQ) and SQA Advanced Highers.

*Predicted and actual grades for Key Stage 5 qualifications.*



1  
2  
3 *Predicted and actual grades for individual A-levels.* Supplementary table 2 shows the relationship  
4 between predicted and attained A-level grades for 237,030 individual examinations from 2010 to  
5 2018. Supplementary table 2.a shows frequencies in the various combinations, with bold values in  
6 grey boxes on the diagonal indicating accurate prediction of grades, green and blue indicating under-  
7 prediction by 1 or 2 grades, and orange and red indicating over-prediction by 1 or 2 grades. Overall  
8 48.8% of predicted grades are accurate. Under-prediction occurs by one grade in 35.7% of cases, and  
9 by two or more grades in 9.0% of cases. Over-prediction is by one grade for 6.3% of A-levels, and  
10 0.1% by two or more grades. It should be remembered that since the median grade for actual A-  
11 level grades is A, then over-prediction in such cases can only be by a maximum of one grade, since  
12 A\* is the highest grade.  
13

14  
15 Supplementary tables 2.b and 2.c show the data of supplementary table 2.a as percentages. As has  
16 been pointed out<sup>5</sup> percentages within predicted grades and percentages within actual grades have  
17 different interpretations and uses. Both are presented here, but from the perspective of admissions  
18 tutors perhaps the most useful are those in supplementary table 2.b of percentages within predicted  
19 grades in relation to actual grades, as they show the likelihood that a predicted grade will actually  
20 manifest as particular actual grades. About a half of A\* predictions actually gain an A grade, and over  
21 a third of predicted A grades result in a grade B or lower.  
22

23 Allocating points on the basis of A\*=12, A=10, B=8, C=6, D=4 and E=2, predicted grades show  
24 systematic *bias*, the mean prediction of 10.53 points being systematically higher than the mean  
25 actual grade of 9.55 points, the difference of 0.98 points being about half of an A-level grade, and  
26 can be seen in the greater numbers in red and orange cells in supplementary table 2.a (over-  
27 prediction, 45%) than in the blue and green cells (under-prediction, 6%).  
28

29 Despite the bias, predicted grades overall show a reasonable *correlation* with actual grades, with a  
30 Pearson  $r_p$  of 0.624 and a Spearman correlation  $r_s$  of 0.581. Both predicted and actual grades are  
31 skewed because of censorship, values above A\* not being possible. A tetrachoric or polychoric  
32 correlation fits an underlying latent normal distribution into account, accepting that row and column  
33 totals may not be equally spaced, being ordinary in nature<sup>23</sup>. Using the *polychor()* function in *R* the  
34 polychoric correlation,  $r_t$  is somewhat higher at 0.716 (SE 0.002), and is probably the best estimate  
35 of the true extent of correlation.  
36

37 *Differences between A-level subjects.* A-levels in different subjects may show differences in their  
38 degrees of bias or correlation. Subjects were divided into 26 broad groups (see supplementary table  
39 3), with the Modern Languages group including 21 languages.  
40

41 Supplementary table 3 shows the mean predicted points, the mean actual points, actual minus  
42 predicted points, and the Pearson correlation of predicted and actual points. Subjects are sorted by  
43 the number of examination entries, and values are colour coded on a green-yellow-red scale, green  
44 indicating higher predicted and actual grades, a smaller difference between predicted and actual  
45 grades (i.e. less bias), and higher correlation of predicted and actual grades.  
46

47 Considering the four major subjects of chemistry, biology, maths and physics, differences between  
48 actual and predicted grades are very similar (-1.15 to -0.98) indicating a bias of about 1 point (i.e.  
49 half of a grade) and very similar correlations of 0.600 to 0.635. Amongst other subjects there is  
50 inevitably greater variation in those subjects taken less frequently. Of particular interest, given that  
51 some medical schools use it for selection, is General Studies, which has the largest difference of  
52 predicted and actual grades of -1.96 points, equivalent to a whole grade. The smallest bias is for art  
53 and design subjects at -.57 points, perhaps indicating the role of an in-course portfolio in these  
54 subjects giving teachers a better sense of how students are performing. Correlations of predicted  
55 and actual grades are mostly very similar, although the lower correlations are for general studies,  
56 modern languages, geography, history, economics, music and classics, and, as mentioned, for  
57 general studies.  
58  
59  
60

### *Total predicted and actual points, correlations between grades and reliability of measures*

*Reliability of actual and predicted A-levels.* The reliability of total points from the three best actual and predicted A-levels was calculated by randomly sampling a pair of grades from the best three and finding the correlation. Cronbach's alpha for the three totalled grades could then be calculated from the standard formula,  $\text{Alpha} = 3.r/(1+2.r)$  where  $r$  is the mean correlation, and is equivalent to a single randomly sampled correlation between a pair of grades since any pair should give similar results. Analysis was restricted to the 66,006 candidates who had at least three paired predicted and actual grades. For actual grades  $r=0.615$  (SE .003) giving  $\text{alpha}=0.827$ , while for predicted grades  $r=0.550$  (SE = .004) and hence  $\text{alpha} = 0.786$ . Given the standard errors, the correlation between grades is clearly substantially lower for predicted than actual grades, and the same must be true of alpha. Interpreting the difference is not entirely straightforward, since on the one hand more predicted grades are at A\*, meaning that there should be fewer non-identical grades, but range restriction might also result in a lower correlation. In terms of mechanism, teachers may collaborate in producing predicted grades<sup>15</sup>, and such non-independence would increase correlations and increase alpha. However teachers may also spend less time making judgements than do A-level examiners, and hence there should be lower correlations. On balance it seems that the most likely conclusion is that estimated grades are somewhat less reliable than actual grades, but there is clearly a need for more complex modelling of the reliability of actual and estimated grades.

*Predicted and actual grades for Extended Project Qualification (EPQ).* The English EPQ has become popular qualification for medical school applicants, being taken by 18616 applicants over the years 2018 to 2018, about 2100 applicants a year (perhaps 10% of all applicants). There is evidence that it has predictive validity for degree outcomes<sup>23</sup>. At present it is not known if it predicts outcomes in application or at medical school. Supplementary table 4 shows the relationship between actual and predicted grades. Grades are over-estimated in 33.7% of cases, under-estimated in 14.0% and accurate in 52.3% of cases, the mean score difference, the bias, being 0.805, which is a little under half a grade. Pearson's correlation is  $r_p=.459$ , Spearman's correlation is  $r_s=.457$ , whereas the polychoric correlation is somewhat higher at  $r_t=.569$ .

*Predicted and actual grades for SQA Advanced Highers.* SQA Advanced Highers, as with SQA Highers, are scored both as simple literals (A, B, C D) and as a more extended scoring (A1, A2, B3, B4, C5, C6, D7), although predicted grades are only in terms of literals. Supplementary tables 5.a and 5.b show, that A grades are more frequent in predicted than in attained grades. Using literals, 59.8% of predictions are accurate, 37.7% are over-estimated, and 2.6% are under-estimated, and for literal grades the bias was 0.976 points, equivalent to half a grade. Correlations of predicted grades with literal attained grades were  $r_p=.407$  and  $r_s=.357$ , whereas with extended grades were  $r_p=.409$  and  $r_s=.355$ . Polychoric correlations were  $r_t=.575$  for literal grades and  $r_t=.587$  for extended grades, again showing the similarity across the two grading schemes.

*Summary.* Taking all the exam types together, A-Levels, EPQ and SQA Advanced Highers, it is generally clear that predicted grades are usually about a half-grade higher than actual grades. Where grades are not accurate there are about four times as many grades over-estimated as under-estimated.

### *Predictive validity of predicted and attained A-level grades.*

A key question throughout discussions of calculated grades is whether grades estimated by teachers are better or worse at predicting outcomes than are actual A-level grades. That question is answered not in terms of how well predicted grades relate to actual A-level grades, but by assessing how well predicted and actual grades predict subsequent outcomes during undergraduate and postgraduate training. It should also be said that it is not entirely self-evident that teachers' grades will be less good, and in the context of GCSEs rather than A-levels, Thomson said, "It is possible, in theory at least, that teacher judgements may be more reliable than exam grades, particularly in those subjects where exam reliability is lower"<sup>24</sup>, with "more reliable" being somewhat ambiguous and perhaps

also meaning more valid as well as more reliable in the narrow statistical sense. Questions about predictive validity can be answered by the P51 dataset.

*Predictive validity in P51.* The P51 UKMED data includes only applicants applying for medical schools. Predicted A-level grades were available only for the UCAS application cycles of 2010 to 2014, and consisted of a single score in the range 2 to 18 points, based on the three highest predictions scored as A\*=6, A=5, etc.. The modal score for 38964 applicants was 15 (equivalent to AAA; mean=15.88; SD= 1.79; Median = 16; 5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup> and 95<sup>th</sup> percentiles= 13, 15, 17 and 18). Some older applicants had only pre-A\* A-levels, and it was also desirable to restrict the analysis to standard applicants in their first year of application, and so only those aged 18 in the UCAS year were included. For multiple reasons not all applicants had both predicted grades and attained A-level grades, and analysis was restricted to the 22954 applicants with both predicted and attained grades. Other selection measures which were included in the analysis are GCSEs (mean grade for best eight grades), as well as U(K)CAT and BMAT scores, which are based on the most recent attempt which in most of the present cases is also the first attempt. For simplicity we used the total of the four sub-scores for U(K)CAT, and for BMAT the total of the Section 1 and 2 scores. No GAMSAT scores were available for this age-group.

Outcome measures are complicated as different application cohorts enter medical school and graduate at different times, and lags within the system mean that not all outcome measures are available. In this UKMED data extract, applicants to UCAS in 2010 entered the medical register from 2015-18, 2011 applicants in 2016-8, 2012 applicants in 2017-18 and 2013 applicants in 2018. Applicants for 2014 would only have qualified in 2019 but the UKMED dataset did not yet include that years, and some earlier entrants would also be expected to qualify after 2018. For simplicity, outcome measures were restricted to the deciles of the UKFPO's Educational Performance Measure (EPM), the raw score of the UKFPO's Situational Judgement Test (SJT), and the score relative to the pass mark of the Prescribing Safety Assessment (PSA), all at first attempt, as these are the main outcomes from undergraduate training. Insufficient numbers of doctors had progressed further in postgraduate training to make analysis meaningful in this data extract.

Supplementary table 6 (presented also in the main paper) summarises the correlation matrix of the various measures. It is important to note that the large differences in Ns are primarily because some measures are present in applicants and used during *selection*, and others are undergraduate outcome measures from medical school, which of necessity are only present in *entrants*, and some are postgraduate outcome measures, only present in *graduates*, not all cohorts yet having reached that stage. The three parts of the correlation matrix are separated to clarify the distinction. Correlations of selection and outcome measures necessarily show range restriction because candidates have been selected on the basis of these measures, and in the case of graduates, selected and self-selected, so that they are less variable than would be the case in an unrestricted population of applicants. The most important question for these data is the extent to which Predicted and Attained A-level grades (shown in pink and green in Supplementary table 6) differ in how much they predict the three outcome measures, which typically are taken five or six years later.

*Prediction of Educational Performance Measure (EPM).* EPM is probably the most important outcome measure since it integrates educational performance across assessments for all but the final year of the undergraduate course<sup>p</sup>. Note that deciles are confusing, as UKFPO scores them in the reverse of the usual order, the first decile being highest performance and the tenth the lowest. Here for ease of interpretation we reverse the scoring in what we call *revDecile*, so that higher *revDeciles* indicate higher performance. EPM is a summary of outcome across assessments within a medical school, expressed as deciles of achievement within each school. EPM is predicted  $r=0.297$  by attained A-level grades but only  $r=0.198$  by predicted grades. Although in absolute terms those

---

<sup>p</sup> <https://foundationprogramme.nhs.uk/wp-content/uploads/sites/2/2019/11/UKFP-2020-EPM-Framework-Final-1.pdf>

1  
2  
3 correlations may seem small it must be remembered that they are range restricted, and the  
4 construct level predictive validity, taking into account range restriction and measurement error is  
5 likely to be much higher<sup>25</sup>. N is large for these correlations and hence the differences are highly  
6 significant using Meng and Rosenthal's test for correlated correlations<sup>26</sup>,  $Z = 12.6$ , with  $p < 10^{-33}$ .  
7 Although predicted grades predict less well than attained grades, they may predict differently, and  
8 hence contribute something over and above attained grades in predicting outcome? Entering  
9 predicted grades after attained grades in a multiple regression shows a highly significant but small  
10 additional prediction of predicted grades (beta=.052, compared with beta=.269 for attained grades).  
11 Attained grades are therefore substantially better at predicting undergraduate outcome, but  
12 predicted grades may have a small amount of variance which is not shared with attained A-levels.  
13  
14

15 *Can other measures replace attained A-level grades for predicting EPM?* In the absence of attained  
16 grades, to what extent can other selection measures such as GCSE grades, U(K)CAT and BMAT  
17 replace the predictive variance in attained A-level grades? Regressing EPM on just predicted grades  
18 gives multiple R = .198, compared with an R of 0.297 when regressed on just actual grades. Adding  
19 GCSEs to Predicted grades increases R to .225, while also including U(K)CAT and BMAT increases R to  
20 .231, although that is still far short of the .297 from A-levels alone. Interestingly if Actual Grades are  
21 now added in to the equation as well, R increases to .308, which is higher than the R for just actual  
22 grades. Exploration suggests that the effect is due to the additional effect of GCSEs grades compared  
23 with just having attained A-level grades in the model (R=.306; Beta(attained grades)=.268,  
24 beta(GCSES)=.077). Overall therefore if only Predicted Grades are available, an improved prediction  
25 is obtain by including GCSEs and U(K)CAT/BMAT, although the model still falls short of that of actual  
26 A-levels in terms of prediction.  
27  
28

29 *Private and State Sector schooling and EPM.* The UKCAT-12 study<sup>27</sup> found that medical students  
30 educated in the private sector performed less well at medical school than those educated in the  
31 state sector with equivalent A-level grades. It is important to replicate that finding in the present  
32 data, and to explore the extent to which there are effects related to predicted as opposed to  
33 attained grades. Overall 6149 (26.8%) of students were educated in the private sector, compared  
34 with 16805 (73.2%) in the state sector. Supplementary figure 1 plots *revDecile* in relation to attained  
35 and predicted grades, separately by private and state education. Visually it is immediately clear that  
36 there is an overall main effect of schooling, the lines for private sector schools (pale green and pale  
37 red) being below those for state schools. Note that the point for private schools with predicted  
38 grades <AAA is missing, as N was very small, because of few private schools predicting grades below  
39 AAA. Considering just attained grades, regression showed effects of both A-level grade ( $b=.299$  (SE  
40 .008)<sup>9</sup>,  $\beta=.301$ ,  $t=35.24$ ,  $p < 10^{-100}$ ) and private schooling ( $b= -.292$  (SE=.053),  $\beta= -.047$ ,  $t=-5.478$ ,  
41  $p=4 \times 10^{-8}$ ), but the addition of an interaction was not significant ( $t=0.746$ ,  $p=.455$ ) meaning that the  
42 slopes in supplementary figure 1.a 1.b are the same. A similar analysis for predicted grades found  
43 effects of predicted grade ( $b=.213$  (SE .009),  $\beta=.201$ ,  $t=22.94$ ,  $p < 10^{-100}$ ) and private schooling ( $b= -$   
44  $.256$  (SE .055),  $\beta= -.041$ ,  $t=-4.679$ ,  $p=0.000003$ ), but the addition of an interaction was not  
45 significant ( $t=0.680746$ ,  $p=.455$ ), again meaning that the slopes are similar in the two types of school  
46 in supplementary figure 1.b. The standard errors for the effects of private schooling suggest that the  
47 difference between the slope is similar for actual and predicted grades.  
48  
49

50 Supplementary table 6 contains a number of other interesting features. [Note that the main paper  
51 has some extended descriptive statistics and additional comments in the text].  
52

53 *Other outcome measures in relation to actual and predicted A-levels.* There are four other outcome  
54 variables, two undergraduate and two postgraduate. For the undergraduate measures, PSA mark  
55 (supplementary figure 2) and SJT score (supplementary figure 3), both correlate more strongly with  
56  
57

58 <sup>9</sup> Actual and Predicted grades are scored on the basis of A\*=12, A=10 etc so are in the range 6 to 36 for three  
59 best grades.  $b=.299$  therefore means an increase of 0.3 deciles per step on the A-level grade score, and  
60 therefore a full A-level grade (e.g A\*AA compared with AAA is 0.6 EPM deciles higher).

attained A-level grades than predicted A-levels (PSA:  $Z = 10.31$ ,  $p < 10^{-23}$ ; SJT  $Z = 4.38$ ,  $p = 0.000012$ ). The two postgraduate outcome measures, are based on smaller, but still substantial, numbers of doctors, MRCP(UK) Part 1 being taken by 910 doctors, and MRCS Part A by 440 doctors. Both outcomes have higher correlations with attained A-level grades than predicted grades, MRCP(UK) Part 1 correlating 0.421 with actual A-level grades (supplementary figure 4), and 0.283 with predicted grades ( $Z = 4.54$ ,  $p = .000055$ ), and MRCS Part A correlating 0.421 with actual grades (supplementary figure 5) compared with 0.358 with predicted grades ( $Z = 3.67$ ,  $p = .000238$ ). The five outcome measures therefore show the same broad pattern of results.

*Correlations of outcome measures and the status of the SJT.* The five outcome measures correlate well with each other (mean  $r = .420$ )<sup>r</sup>, as might be expected given the academic backbone<sup>29</sup>. Noteworthy is the relatively low correlation of SJT with EPM (.319) and PSA (.346), compared with the correlation of EPM and PSA (.470). That pattern is repeated when postgraduate exams are included, the four non-SJT assessments showing a higher correlation (mean  $r = .499$ ) than the correlations of the four non-SJT assessments with SJT (mean  $r = .322$ ). Overall that suggests that SJT may be measuring a construct that is different in part from the other more academic assessments, and that will need investigating more closely in the future. It is also of interest when considering predicted grades that SJT correlates only slightly better with actual grades than predicted grades (.195 vs .160), compared with the other four outcomes (.297 vs .198; .306 vs .226; .421 vs .283; and .358 vs .181; mean  $r = .346$  vs .222) raising the possibility that predicted grades may include some non-academic variance which then is predictive for SJT. That can be tested by regressing SJT on actual and predicted grades, when including predicted grades increases R from .195 to .206. The model including both grade types, shows an effect of actual grades (beta=.153,  $t = 14.8$ ,  $p = 10^{-49}$ ) and an effect of predicted grades (beta=.077,  $t = 7.42$ ,  $p = 1.2 \times 10^{-13}$ ), so that the beta effect of predicted grades is 50% of that for actual grades, compared with the earlier regression for deciles, where the beta of .052 for predicted grades is only 19% of the beta of .269 for attained grades.

The present SJT test is administered at the time of graduation. There is also a separate SJT administered as a part of the U(K)CAT tests, which was only introduced in 2014, and none of that cohort have outcome variables in the present data set. However it is of interest that, for the 4286 applicants in 2014 with U(K)CAT SJT, there is a correlation of .145 with Actual A-levels and .127 with predicted A-levels ( $Z = 1.28$ ,  $p = 0.192$ ). Overall it is possible that SJT tests are behaving differently to academic outcomes, despite moderately strong correlations of SJT with other academic outcomes. SJT tests are, “designed to assess for key attributes ... including commitment to professionalism, coping with pressure, effective communication, patient focus, and working effectively as part of a team”<sup>30 31</sup>.

*Correlations of A-levels with GCSEs, U(K)CAT and BMAT.* Without going into details, attained A-levels correlate more strongly with U(K)CAT and BMAT ( $r = .326$  and  $.416$ ) than do predicted A-levels ( $r = .272$  and  $.326$ ), suggesting that admissions tests are particularly assessing academic attainment. However GCSE grades show the reversed pattern and correlated *more strongly with predicted A-levels* (0.452) than with attained A-level grades (0.421), perhaps implying that teachers in part use GCSE grades to make predictions (as has been found in a previous study<sup>18</sup>).

*Correlations of admissions tests with outcome measures.* Neither of the two admissions tests, U(K)CAT and BMAT, has a strong prediction of EPM ( $r = .115$  and  $.089$  respectively), and both clearly

---

<sup>r</sup> Note that there are too few doctors who took both MRCP(UK) Part 1 and MRCS Part A to be able to calculate a correlation. Elsewhere we have looked at the relatively rare groups of doctors taking both MRCP(UK) and MRCS, and shown high correlations between performance on the two assessments<sup>28</sup>. Wakeford R, Denney ML, Ludka-Stempien K, et al. Cross-comparison of MRCS & MRCP(UK) in a database linkage study of 2,284 candidates taking both examinations: Assessment of validity and differential performance by ethnicity. *BMC Medical Education* 2015;15(1 (doi:10.1186/s12909-014-0281-2)), making it likely that the same would also apply to MRCP(UK) Part 1 and MRCS Part A.

1  
2  
3 correlate less with EPM than does attained A-levels,  $r=.297$ , despite A-levels showing range  
4 restriction due to a ceiling effect at A\*. PSA and SJT though show a somewhat different picture. PSA  
5 correlates more highly with BMAT ( $r=.321$ ) than with U(K)CAT ( $r=.238$ ), and the correlation with  
6 BMAT is higher than that with attained A-levels ( $r=.306$ ). In contrast U(K)CAT and BMAT both  
7 correlate similarly with SJT ( $r=.243$  and  $.249$ ), and both correlations are higher than with attained A-  
8 levels ( $r=.195$ ). BMAT and U(K)CAT both show correlations with the two postgraduate outcomes  
9 (0.200 and 0.378 for MRCP(UK) Part 1 and 0.181 and 0.319 for MRCS Part A, but both are lower than  
10 the correlations with A-levels (0.421 and 0.358). Taken overall, BMAT has somewhat higher  
11 correlations with the five outcome measures (mean  $r = .269$ ) than does U(K)CAT (mean  $r = .195$ ) but  
12 both correlate less with outcomes than do attained A-levels (mean  $r=.315$ ). U(K)CAT correlates at a  
13 similar level to predicted A-levels (mean  $r=.209$ ) but BMAT at a somewhat higher level.  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27 4. Supplementary Tables 1, 2, 3, 4, 5 & 6 and Supplementary Figures  
28 1, 2, 3, 4 & 5  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 *Supplementary Table 1: Comparison of predicted and forecasted grades in*  
7 *2009 and 2012.*  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

			Max	Over-		Under-			
Estimated grades			grade	predicted	Accurate	predicted	Population	Source	
<b>Predicted</b>	<b>October</b>	<b>2009</b>	<b>A</b>	<b>42%</b>	<b>52%</b>	<b>7%</b>	<b>UCAS</b>	Everett and Papageorgiou (2011)	
<b>Forecasted</b>	<b>May</b>	<b>2009</b>	<b>A</b>	<b>33%</b>	<b>55%</b>	<b>12%</b>	<b>OCR</b>	Gill and Rushton (2011)	
<i>Forecasted-Predicted</i>				<i>-9%</i>	<i>3%</i>	<i>5%</i>			
<b>Predicted</b>	<b>October</b>	<b>2012</b>	<b>A*</b>	<b>68%</b>	<b>20%</b>	<b>12%</b>	<b>UCAS</b>	UCAS (2017)	
<b>Forecasted</b>	<b>May</b>	<b>2012</b>	<b>A*</b>	<b>39%</b>	<b>48%</b>	<b>13%</b>	<b>OCR</b>	Gill and Chang (2013)	
<i>Forecasted-Predicted</i>				<i>-30%</i>	<i>29%</i>	<i>1%</i>			



*Supplementary Table 2: Comparison of predicted and attained A-level grades in medical school applicants, 2010-2018*

a) Counts of number of cases

		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	<b>200</b>	35	10	5	0	0	255 (0%)
	D	235	<b>610</b>	155	35	10	0	1045 (0%)
	C	635	1220	<b>2110</b>	505	95	5	4570 (2%)
	B	635	2095	4755	<b>7355</b>	1695	175	16715 (7%)
	A	430	1925	8785	35640	<b>61950</b>	12655	121390 (51%)
	A*	50	135	635	6025	42815	<b>43395</b>	93060 (39%)
	Total	2185	6020	16450	49570	106570	56235	237030
		(1%)	(3%)	(7%)	(21%)	(45%)	(24%)	

b) Percentages within predicted grades

		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	<b>79%</b>	14%	..	..	..	..	100%
	D	23%	<b>58%</b>	15%	3%	..	..	100%
	C	14%	27%	<b>46%</b>	11%	2%	..	100%
	B	4%	13%	28%	<b>44%</b>	10%	1%	100%
	A	0%	2%	7%	29%	<b>51%</b>	10%	100%
	A*	0%	0%	1%	7%	46%	<b>47%</b>	100%
	Total	1%	3%	7%	21%	45%	24%	100%

b) Percentages within predicted grades

		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	<b>9%</b>	1%	..	..	..	..	0%
	D	11%	<b>10%</b>	1%	0%	..	..	0%
	C	29%	20%	<b>13%</b>	1%	0%	..	2%
	B	29%	35%	29%	<b>15%</b>	2%	0%	7%
	A	20%	32%	53%	72%	<b>58%</b>	23%	51%
	A*	2%	2%	4%	12%	40%	<b>77%</b>	39%
	Total	100%	100%	100%	100%	100%	100%	100%

**Supplementary Table 3: Comparison of predicted and forecasted A-level grades in medical school applicants, 2010-2018**

Subject	N	Mean	Mean	Actual	r
		Predicted	Actual	minus Predicted (Pearson)	
Chemistry	62815	10.35	9.37	-0.98	0.623
Biology	61190	10.59	9.78	-0.82	0.632
Maths & Stats	54635	10.79	9.77	-1.02	0.600
Physics & Engineering	13870	10.67	9.52	-1.15	0.635
General Studies & Critical Thinking	6785	9.66	7.70	-1.96	0.534
Modern Languages	6720	10.59	9.74	-0.85	0.571
Psychology	6190	10.19	9.12	-1.07	0.631
Geography	4015	10.84	9.95	-0.89	0.538
History	3850	10.48	9.49	-0.99	0.546
English Literature & Language	3815	10.32	9.52	-0.80	0.681
Further Maths	2950	11.07	9.62	-0.80	0.681
Economics & Business Studies	2765	10.36	9.47	-0.89	0.577
Religious Studies	1890	10.45	9.40	-1.05	0.626
Art & Design	1035	10.60	10.03	-0.57	0.681
Latin & Classical Studies	675	10.74	9.65	-1.09	0.576
Music	640	10.49	9.51	-0.97	0.567
Sociology	525	9.51	8.49	-1.02	0.679
Computer Studies & ICT	475	9.89	8.82	-1.06	0.704
Physical Education	470	10.61	9.81	-0.80	0.610
Government & Politics	380	10.07	9.16	-0.91	0.656
Theatre Studies & Drama	260	10.14	9.02	-1.11	0.624
Science -- Misc & General	260	8.30	7.24	-1.06	0.821
Law	190	9.42	8.55	-0.87	0.766
Philosophy	155	10.37	9.06	-1.32	0.639
Classical Greek	115	10.90	9.98	-0.92	0.463
Media Studies	75	8.03	7.25	-0.78	0.798

*Supplementary Table 4: Comparison of predicted and attained EPQ grades in medical school applicants, 2010-2018*

a) EPQ: Counts of number of cases								
		Attained EPQ grade						
		E	D	C	B	A	A*	Total
Predicted EPQ grade	E	5	0	0	0	0	0	5 (0%)
	D	0	15	0	0	0	0	20 (0%)
	C	10	10	120	15	5	0	160 (2%)
	B	15	40	90	355	100	30	625 (7%)
	A	40	135	405	920	1970	1150	4620 (49%)
	A*	15	35	125	375	940	2420	3915 (42%)
<b>Total</b>		85	240	740	1670	3010	3605	9345
		(1%)	(3%)	(8%)	(18%)	(32%)	(39%)	
b) EPQ: Percentages within predicted grades								
		Attained EPQ grade						
		E	D	C	B	A	A*	Total
Predicted EPQ grade	E	..	..	..	..	..	..	..
	D	..	..	..	..	..	..	..
	C	..	..	46%	..	..	..	100%
	B	..	13%	28%	44%	10%	1%	100%
	A	0%	2%	7%	29%	51%	10%	100%
	A*	..	0%	1%	7%	46%	47%	100%
<b>Total</b>		1%	3%	7%	21%	45%	24%	100%

1 *Supplementary Table 5: Comparison of predicted and forecasted SQA*  
 2 *Highers and SQA Advanced Highers in medical school applicants, 2010-2018*  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16

17 **a) SQA Advanced Highers: Counts of number of cases**

		Attained SQA Highers							
		D7	C6	C5	B4	B3	A2	A1	Total
	D	<b>90</b>	0	0	0	0	0	0	95 (0.6%)
Predicted	C	20	<b>210</b>	<b>220</b>	15	10	10	0	485 (3.3%)
SQA	B	95	140	190	<b>455</b>	<b>490</b>	305	30	1700 (11.6%)
Highers	A	255	495	905	1405	2010	<b>5335</b>	<b>1955</b>	12360 (84.4%)
	Total	<b>465</b>	<b>845</b>	<b>1320</b>	<b>1875</b>	<b>2510</b>	<b>5645</b>	<b>1985</b>	14640 (100%)
	Total	3.2%	5.8%	9.0%	12.8%	17.2%	38.6%	13.6%	

31 **b) SQA Advanced Highers: Percentages within predicted grades**

		Attained SQA Highers							
		D7	C6	C5	B4	B3	A2	A1	Total
	D	<b>97%</b>	..	..	..	..	..	..	100%
Predicted	C	..	<b>43%</b>	<b>45%</b>	..	..	..	..	100%
SQA	B	6%	8%	11%	<b>27%</b>	<b>29%</b>	18%	2%	100%
Highers	A	2%	4%	7%	11%	16%	<b>43%</b>	<b>16%</b>	100%
	Total	3.2%	5.8%	9.0%	12.8%	17.2%	38.6%	13.6%	

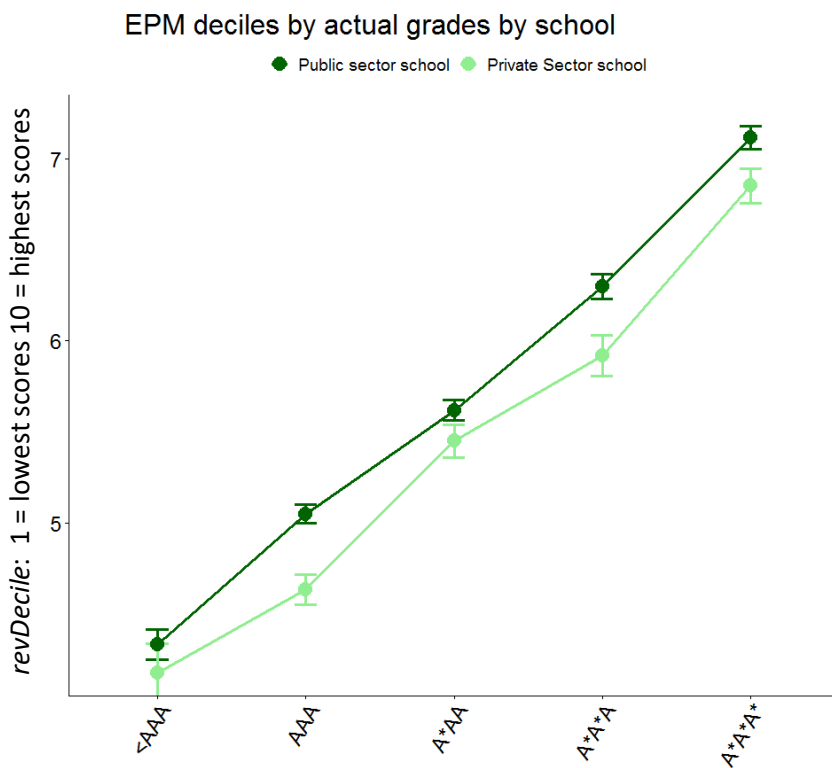
**Supplementary Table 6: Correlation matrix of selection measures, undergraduate outcome measures, and postgraduate outcome measures (separated by grey lines for clarity). Cells indicate Pearson correlation and N. (NB presented as figure 3 in the main paper).**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

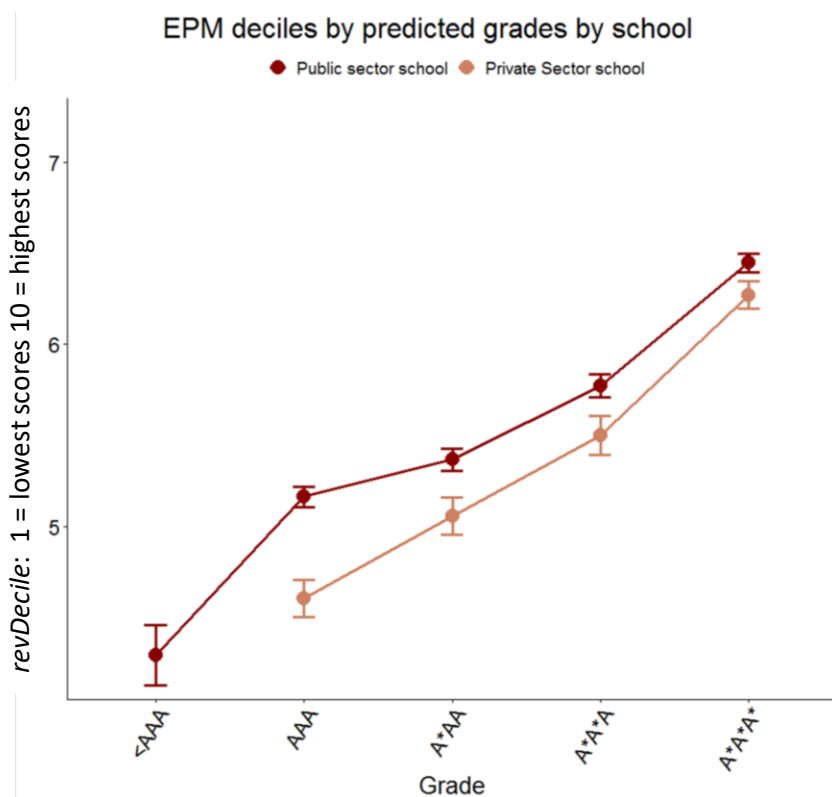
		Selection measures applicants					Undergraduate outcome measures			Postgraduate outcome measures	
		GCSE grades	Predicted Alevels	Alevel grades	UKCAT	BMAT	EPM	SJT	PSA	MRCP(UK) Part 1	MRCS Part A
	GCSE grades	<b>1</b>	<b>0.452</b>	<b>0.421</b>	<b>0.265</b>	<b>0.223</b>	<b>0.180</b>	<b>0.190</b>	<b>0.201</b>	<b>0.212</b>	<b>0.173</b>
	Predicted A-level grades		<b>1</b>	<b>0.585</b>	<b>0.272</b>	<b>0.326</b>	<b>0.198</b>	<b>0.160</b>	<b>0.226</b>	<b>0.283</b>	<b>0.181</b>
<b>Selection measures</b>	Attained A-level grades			<b>1</b>	<b>0.326</b>	<b>0.416</b>	<b>0.297</b>	<b>0.195</b>	<b>0.306</b>	<b>0.421</b>	<b>0.358</b>
<b>all applicants</b>	UKCAT total				<b>1</b>	<b>0.483</b>	<b>0.115</b>	<b>0.243</b>	<b>0.238</b>	<b>0.200</b>	<b>0.181</b>
	BMAT sections 1 and 2					<b>1</b>	<b>0.089</b>	<b>0.239</b>	<b>0.321</b>	<b>0.378</b>	<b>0.319</b>
	UKFPO EPM decile						<b>1</b>	<b>0.319</b>	<b>0.470</b>	<b>0.509</b>	<b>0.535</b>
<b>Undergraduate</b>	UKFPO SJT score							<b>1</b>	<b>0.346</b>	<b>0.351</b>	<b>0.274</b>
<b>Outcome measures</b>	PSA score								<b>1</b>	<b>0.500</b>	<b>0.483</b>
	MRCP(UK) Part 1									<b>1</b>	...
<b>Postgraduate</b>	MRCS Part A									...	<b>1</b>

Supplementary figure 1: Mean EPM revDeciles (95% CI) in relation to actual A-level grades (green) and predicted A-level grades (red), state sector schooling shown in darker colours and private sector schooling in paler colours.

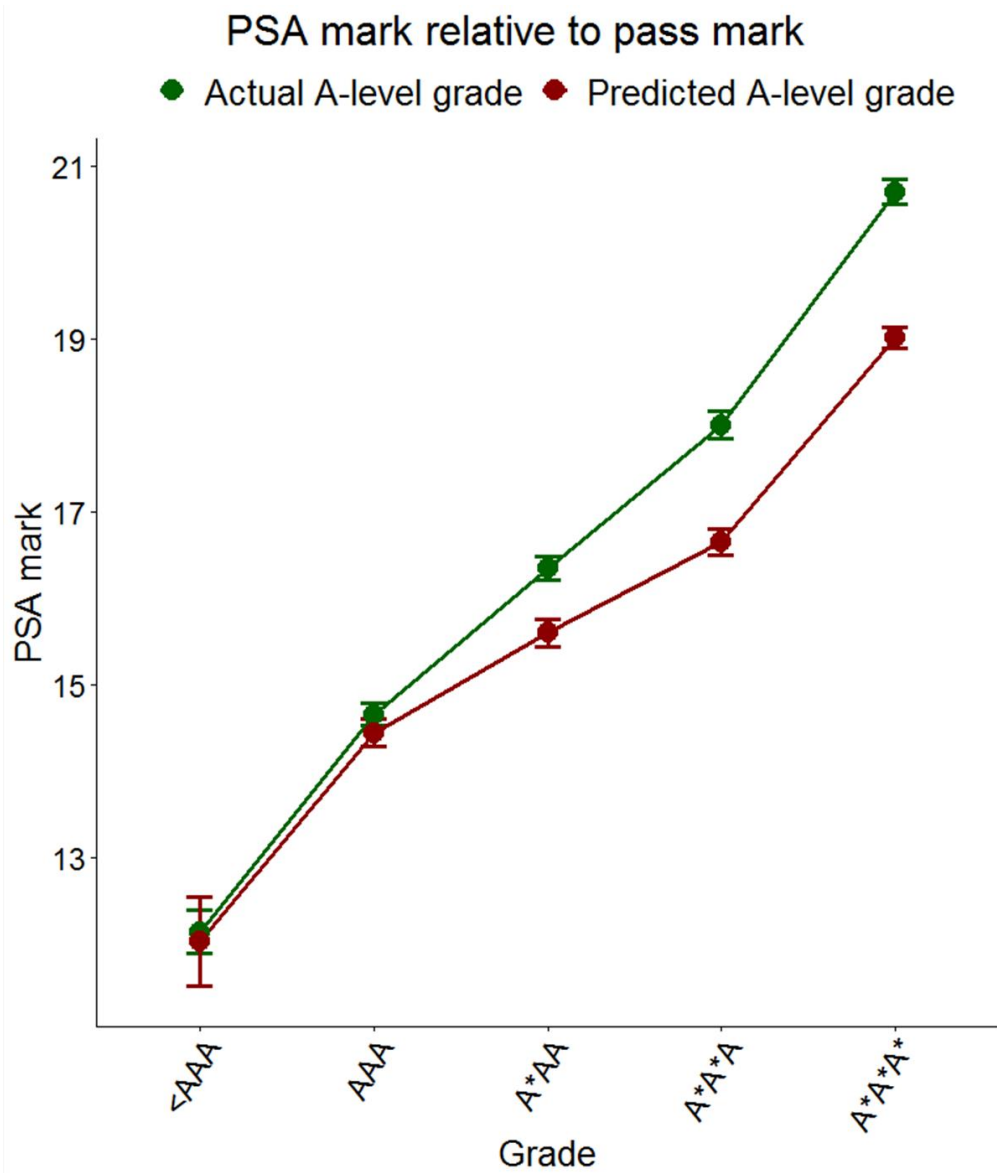
a.)



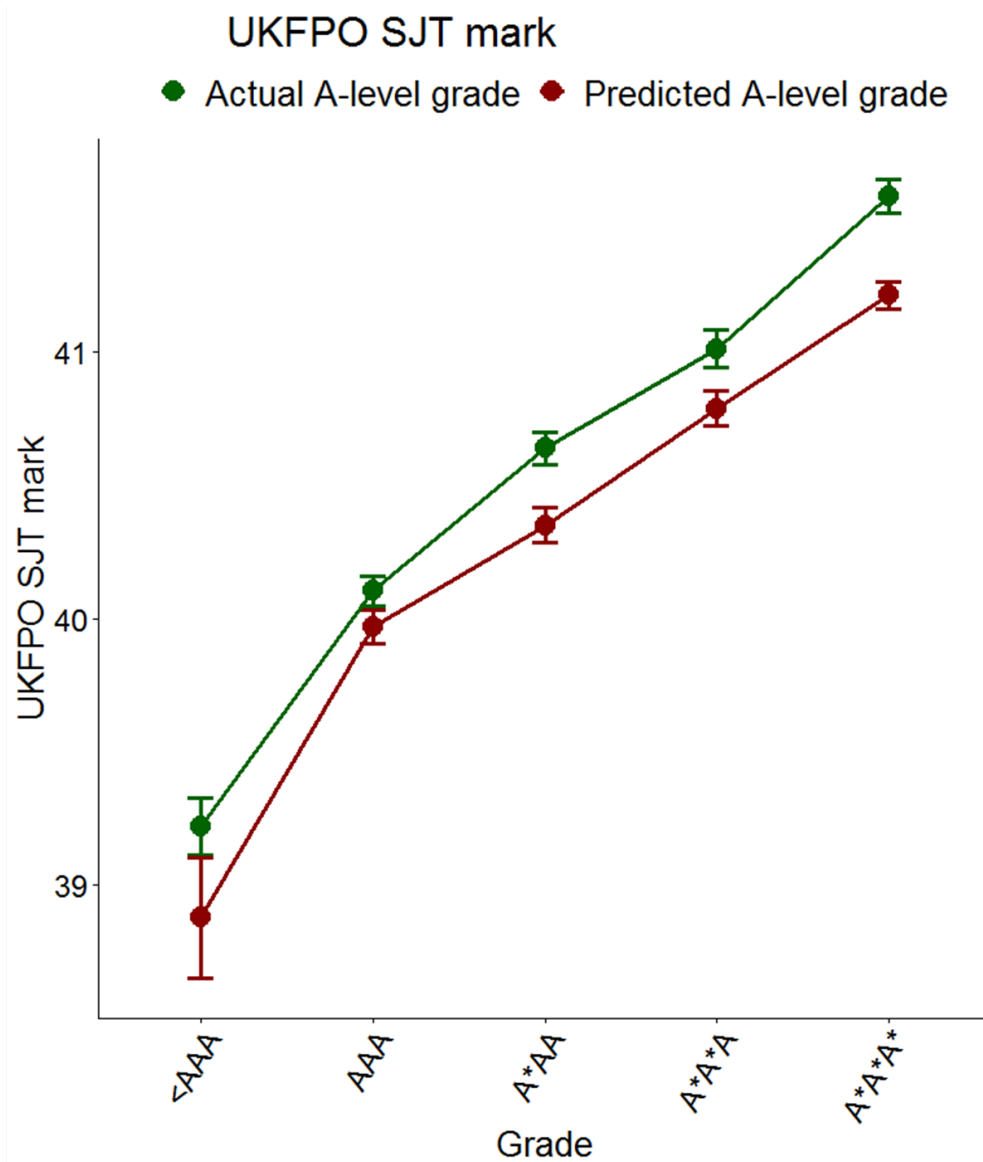
b.)



Supplementary figure 2: Mean PSA mark in relation to actual A-level grades (green) and predicted A-level grades (red)

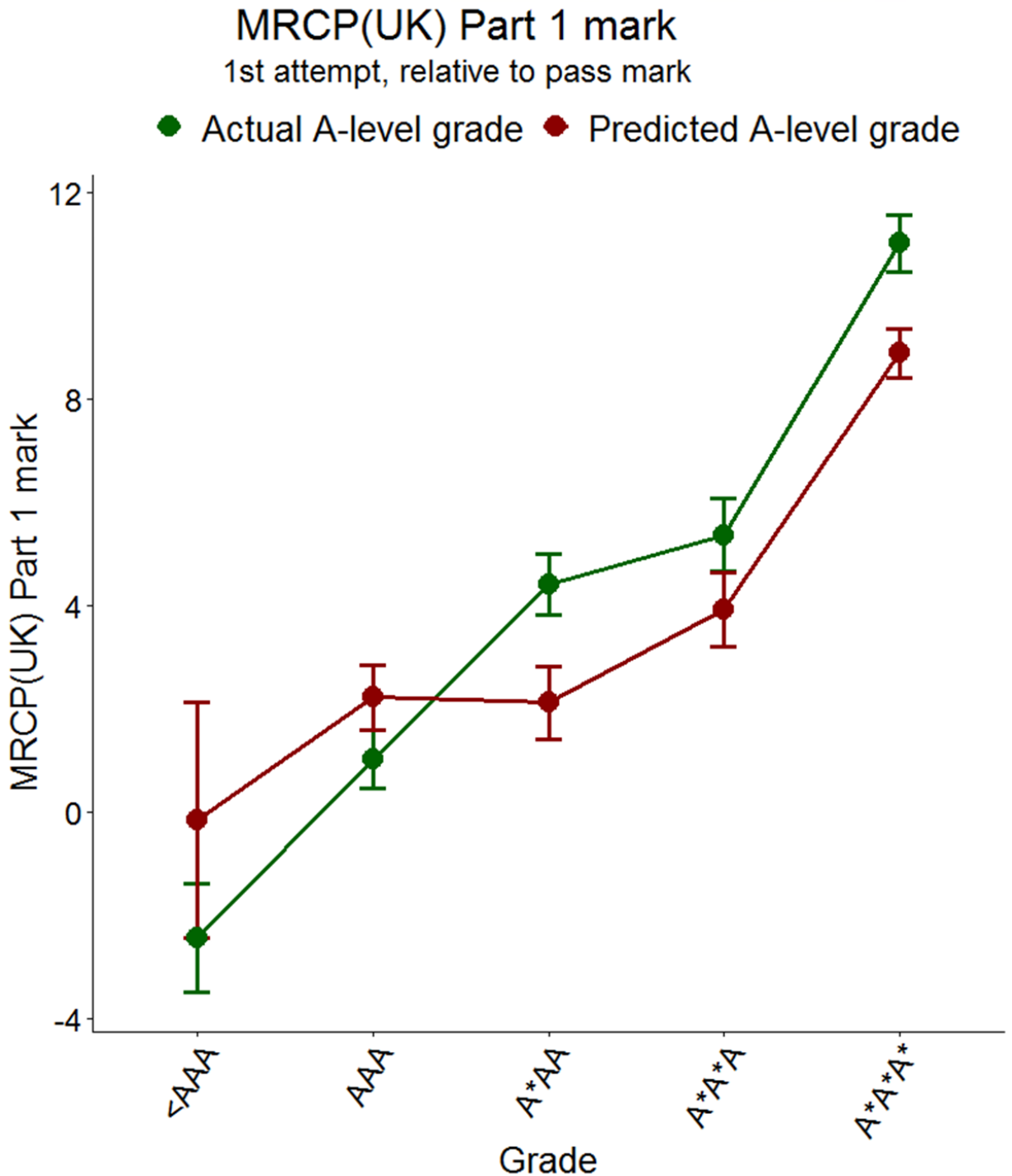


Supplementary figure 3: Mean SJT mark in relation to actual A-level grades (green) and predicted A-level grades (red)

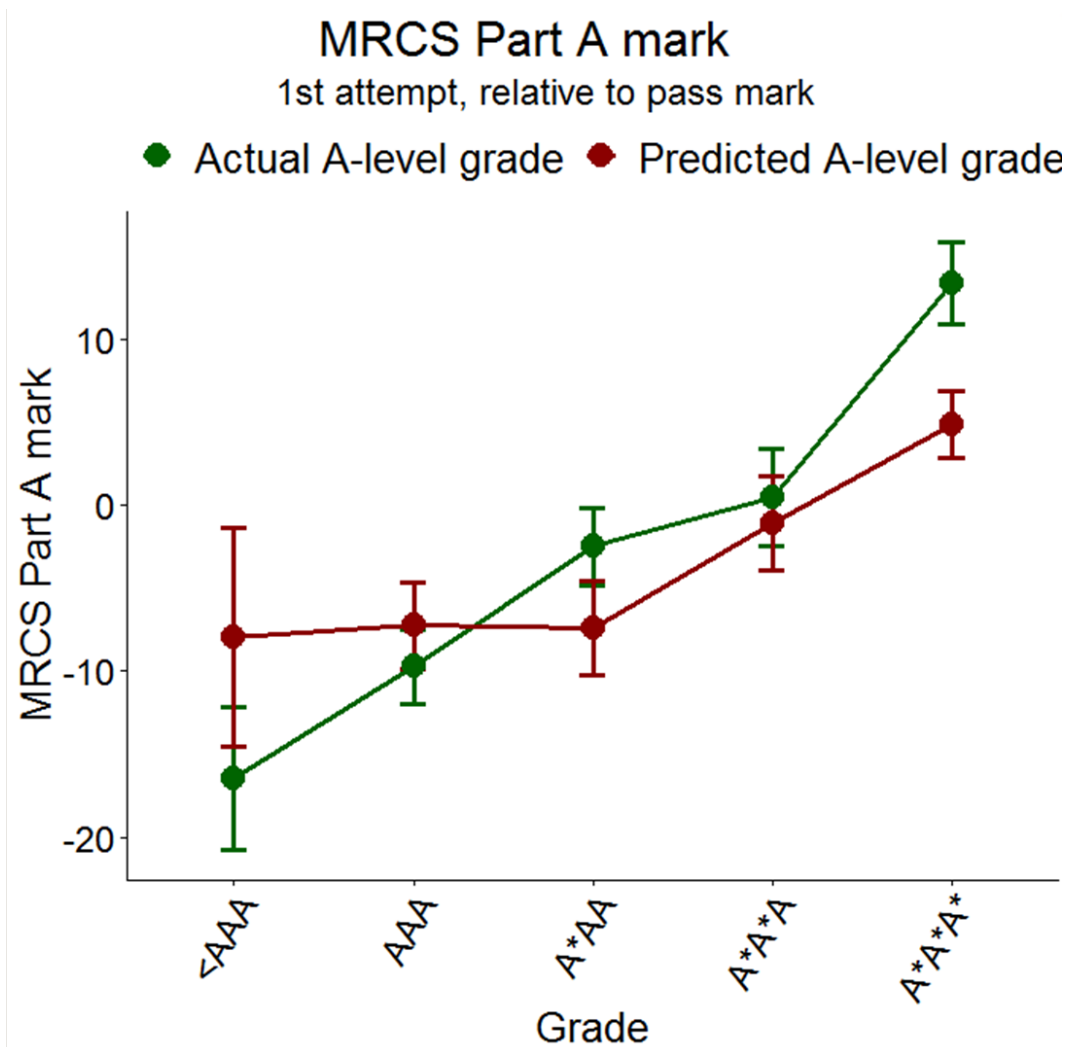




Supplementary figure 4: Mean MRCP(UK) Part 1 mark in relation to actual A-level grades (green) and predicted A-level grades (red)



Supplementary figure 5: Mean MRCS Part A mark in relation to actual A-level grades (green) and predicted A-level grades (red)



## 5. Appendix: Are independent (private sector) schools more accurate in their A-level predictions?

A recurrent suggestion in the literature is that schools in the private sector (Independent Schools) are more accurate in their A-level predictions than those in the state sector. That suggestion raises many issues, not least concerned with social equity, but a key one to be resolved is whether the differences mainly are secondary to differences in overall attainment level, and as a result may be artefacts due to the ordinal nature of A-level grades and to A\* being the ceiling for A-level grades, and hence is a level beyond which candidates cannot reach. This appendix looks in detail at that question. The description is lengthy, technical, and partly didactic, and therefore has not been included in the main text or the main supplementary text. The conclusion is actually relatively simple: *Independent schools are not actually more accurate in their predictions, but they look that way because of having higher attained grades.*

*The data.* Appendix table 1 shows, in a similar format to those in the main paper, the A-level grades in P89 for applicants from state schools (defined as Comprehensives, Academies, Sixth Form Colleges, Tertiary Colleges and Technical Colleges) and private schools (defined as Independent Schools and Grammar Schools), with results restricted to first A-level attempts, duplicates removed, and where both A-level grades and predicted grades were available. Results are at the subject level, and therefore contain multiple subjects from individual applicants.

A simple glance at Appendix table 1 suggests that indeed Private Sector schools are more accurate, 53.3% of predictions being accurate compared to 45.5% of predictions from State Sector schools. Private schools also show a lower rate of over-prediction (40.5% compared with 48.0%), but not of under-prediction (6.2% vs 6.4%). These differences need however to be put into the context of higher overall attainment in private schools, where 29% of grades were A\* compared with 20% in state schools, a finding that reflects most private schools being selective and therefore inevitably taking higher ability entrants. Since attained A\* grades are more frequent in private schools, it is not surprising that predicted A\* grades are also more frequent in private schools, 48% vs 35%. The question therefore is whether the differences in accuracy are secondary to differences in overall performance. That question is best answered using *polychoric correlations*, which need description.

*Polychoric and tetrachoric correlations.* Polychoric and tetrachoric correlations are used frequently in psychometrics when dealing with binary and ordinal data. The need for them is shown by a simple 2x2 association table of the sort often tested using a chi-square test. Consider Appendix figure 1, which is a simple association table for characteristics P and Q in 100 individuals. 80% of cases have P present but only 50% of cases have Q present, meaning that the *marginal proportions* are not the same (80% vs 50%). A chi-square test is highly significant (chi-squared = 25, 1 df,  $p=0.000006$ ) meaning that there is an association between P and Q. But what is the size of that association? Often in this situation a Pearson or Spearman correlation is calculated, and these give  $r_p=0.5$  and  $r_s=0.5$ , which suggests a moderately strong association.

However there is a problem in using the Pearson correlation, as a careful look at the table shows because the number of cases in which P is absent but Q is present, in the top right-hand corner, is zero. In other words the association could not be any stronger, but the correlation is still only 0.5, whereas a perfect correlation is usually taken as being 1. The problem arises because the marginal proportions of P and Q are not the same, one being 0.5 and the other 0.8. If these two marginal proportions had been identical than all of the cases could have been on the diagonal and then the

Pearson correlation would indeed have been 1. So what does one do in the case where the marginal proportions are not the same? The answer is another correlation developed by Pearson, called the *tetrachoric* or *polychoric* correlation for 2x2 or for larger tables respectively.

The tetrachoric correlation assumes that the data actually come from a bivariate normal distribution with some underlying correlation, and asks if that distribution were divided horizontally and vertically, what the correlation would have to be to create the contingency table that has been found. The lower part of Appendix figure 1 shows that diagrammatically<sup>5</sup>, the four quadrants containing the proportions of data in the contingency table. The calculation is easily carried out in the R function *polychor()* in the *polycor* library, and for the table in Appendix figure 1 it gives the answer that  $r_t=0.994$ , which effectively is  $r_t=1$ . The tetrachoric correlation therefore corresponds to our intuitive sense of what the correlation should be. The underlying bivariate normal distribution is assumed to have means of zero and standard deviations of one. *polychor()* then tells us that the thresholds for cutting the distribution need to be at 0.842 for P and 0 for Q. The threshold for Q at zero tells us that the cutting point is 0 standard deviations from the mean, and therefore 50% of cases are above the threshold and 50% below. The threshold for P is 0.842 standard deviations below the mean, and hence 20% of cases are below the threshold and 80% of cases above it. The marginal proportions of P and Q are then replicated.

For a 2x2 table it is always possible to fit the tetrachoric correlation and the marginal proportions exactly. If the table is larger, giving a polychoric correlation, the marginal proportions and the cell frequencies cannot always be fitted exactly as the normal distribution may not be entirely appropriate, and in that case maximum likelihood estimates of the correlation and thresholds are found. The polychoric calculation for an  $m \times n$  table also provides a set of  $(m-1)$  and  $(n-1)$  thresholds for each of the variables, and it is possible to see if step sizes between the levels are equal. Polychoric correlations therefore are used for data where both measures are *ordinal* and for which it seems reasonable to assume an underlying latent distribution which is normal.

*Polychoric correlations for A-level grades.* A-level grades are certainly at least ordinal in nature, but it is not clear that they are *equal interval*, the step from, say, D to C not necessarily being the same size as the step from B to A. Polychoric calculations allow the direct estimation of the step sizes between grades. If step sizes are not equal then many conventional statistics are not optimal. Equal interval scales are measures such as length, where the increments are identical in size (so the difference between, say, 2 cms and 3 cms is the same length as the difference between 10 cms and 11 cms). A-levels are often scored on a simple basis of allocating points, such as A\*=12, A=10, B=8, C=6, D=4 and E=2 (and indeed we have done this elsewhere here), but that can sometimes be misleading in situations such as calculating correlations between actual and predicted grades, partly because marginal proportions are not the same, and partly because the data are *censored*, grades above A\* not being possible, however capable is a candidate, and hence over-prediction is not possible for estimated grades of A\*. In the case of a high ability group such as applicants to medical school the latter is problematic as state and private schools predict an A\* grade for 35% and 48% of exams. To put it another way, were a grade of A\*\* available then many examinees might have merited it<sup>32</sup>, albeit probably more at private than state schools. There is also potentially a problem of computing total A-level scores (so that, say, AAA with 30 points is regarded as equivalent to A\*AB or A\*A\*C, which may not be exactly the case, although the approximation is probably good enough for most purposes).

---

<sup>5</sup> The correlation is actually drawn at 0.9 to make things pedagogically clearer, as a correlation of 1 is effectively a straight line.

1  
2  
3 *Fitting polychoric correlations to A-level grades from state and private schools.* The key question at  
4 present is whether private schools are more accurate in their predictions (53.3%) than state schools  
5 (45.5%) – see Appendix table 1. Accuracy can be considered in two ways, as the presence of  
6 systematic error (technically, ‘bias’), equivalent to rates of A\* etc being different in two groups, and  
7 random error, in terms of the correlation or lack of correlation between two sets of scores. Although  
8 the overall accuracy of private schools is *higher* than state schools, the correlation of predicted and  
9 actual grades is *lower* in private schools, with Pearson correlations of 0.635 in state schools and  
10 0.552 in private schools (Appendix table 1), with a similar pattern for Spearman correlations. That  
11 suggests a potential problem in interpreting the data. Calculating the polychoric correlations  
12 suggests a very different picture, since the polychoric correlations in state schools ( $r_t = 0.717$ ) and  
13 private schools ( $r_t = 0.678$ ) are far more similar, particularly in comparison with the differences  
14 between the Pearson (or Spearman) correlations.  
15  
16  
17

18 Interpreting the polychoric correlations is helped by a diagram. Appendix figure 2.a may look  
19 complex, but it summarises a lot of information about state sector applicants. The axes are on a  
20 normal distribution for the underlying latent scale, and so the units are standard deviations, from -4  
21 to +4 SDs. Note these are not SDs for the raw data, but for the latent distribution. The polychoric  
22 correlation for the state sector is 0.717, and that is shown by the blue ellipse which is plotted to  
23 cover 99.9% of the data, which is reasonable given the large sample sizes. The dashed blue and  
24 yellow line on the diagonal is the line of equality for attained grades on the horizontal axis and  
25 predicted grades on the vertical axis. The vertical and horizontal lines show the thresholds  
26 separating the various A-level grades for attained and predicted grades. Appendix table 2  
27 summarises the various thresholds and their intervals for state and private schools. As an example,  
28 for attained grades, the threshold separating A from A\* (Appendix table 2, row 4, column A:A\*) is  
29 0.83, and so the vertical line in Appendix figure 2.a separating A from A\* is at 0.83. Similarly the  
30 horizontal line for predicted grades separating A from A\* is at 0.39 (row 2 in Appendix table 2). The  
31 intersection of these two lines is shown by a large red circle, which is *below* the blue-yellow dashed  
32 line, which indicates that the threshold for attained grades is higher than the threshold for predicted  
33 grades, so that it is easier to be predicted an A\* than to attain an A\*. The other vertical and  
34 horizontal lines show the thresholds between B and A (B:A), C and B (C:B), D and C (D:C) and E and D  
35 (E:D). As for A\*:A, all of the intersections, shown as red dots, are below the dashed blue-yellow line  
36 of equality, showing that predicted grades are always more generous than attained grades. Row 6 of  
37 Appendix table 2 shows that on average the threshold for attained grades is 0.73 SDs lower than for  
38 predicted grades. The coloured boxes in Appendix figure 2.a are equivalent to the coloured boxes in  
39 appendix table 1, with grey indicating accuracy, green and blue indicating under-estimation, and red  
40 and yellow over-estimation. More of the figure is red or yellow than is blue or green, indicating the  
41 overall over-estimation by predicted grades. It is also clear from the figure that the differences  
42 between the thresholds are not equal. The width of D, from E:D to D:C, is smaller than the width of A  
43 (from B:A to A:A\*), these values being shown in row 10 of Appendix table 2 for predicted grades and  
44 row 12 for attained grades. The widths of E and A\* cannot be calculated as they stop either at minus  
45 infinity or plus infinity. It is clear that the scale is not equal interval, with less change being required  
46 to move from D to C than from B to A. Statistical analyses should take care therefore in assuming  
47 that the usual A\* to E scale of grades is equal interval, and can be averaged.  
48  
49  
50  
51  
52  
53  
54

55 The key question for this appendix is the extent to which state and private sector predictions are  
56 different. Appendix figure 2.b shows an equivalent plot to Appendix figure 2.a but for private sector  
57 A-levels. At a glance it is not easy to see any obvious difference, but it is important to remember that  
58 the latent scales for both graphs each have a mean of zero and SD of one. However looking carefully  
59 shows that the threshold for attained grades at A\* is at 0.55 for private sector students compared  
60

1  
2  
3 with 0.83 for state sector students (see rows 4 and 5 of table Appendix table 2). The threshold is  
4 lower for private sector students and hence more of these students will attain an A\*, as is the case in  
5 Appendix table 2. All of the thresholds for the private sector students are actually moved to the left  
6 compared with state sector students (and compare the sizes of the A\*A\* boxes and the EE boxes in  
7 the two figures. Appendix figure 3 summarises the thresholds more clearly for attained and  
8 predicted grades in state and private sector schools. All thresholds are shown on the same  
9 horizontal scale. Attained grades for private schools are to the right of predicted grades, shown by  
10 the thin blue diagonal lines (meaning an attained A\* is harder to get than a predicted A\*), and the  
11 same pattern is seen for state schools, and shown by the thin diagonal red lines. Private school  
12 attained grades are also to the left of state school attained grades, shown by a thin purple line (with  
13 thresholds lower for private school students meaning that they get more A\* grades). Similarly,  
14 private school predicted grades are also to the left of state school predicted grades, also shown by a  
15 thin purple line. A key feature of Appendix figure 3 is that the blue diagonal lines are parallel, the  
16 red diagonal lines are parallel and the purple diagonal lines are nearly parallel, meaning that the  
17 relationships of grade boundaries are the same in private and state schools, and for attained and  
18 predicted grades, but are merely slid along relative to one another. The state and private schools  
19 are therefore handling predicted grades in a way that is similar, and they are similar related in each  
20 case to attained grades.  
21  
22  
23  
24

25 The widths of the boxes in Appendix figure 2 are therefore very similar in state and private sector  
26 students, and are shown in rows 9 to 16 of Appendix table 2, particularly in rows 10 and 11, which  
27 compare predicted grades in state and private schools, and rows 12 and 13 which compare attained  
28 grades in private schools. The main difference between the two types of school is shown in the  
29 mean columns of rows 1 and 2 and rows 4 and 5, their mean differences being shown in the final  
30 column. Overall the state schools have thresholds which for predicted grades are on average are  
31 0.47 SDs higher and for attained grades are 0.42 grades higher than for private sector schools  
32 (meaning that higher grades are harder to attain). These values are very similar and suggest that  
33 predictions in the two types of school are being carried out in a similar way, but the overall ability of  
34 private school students is higher, and that is reflected in the attained and predicted ways to a similar  
35 extent.  
36  
37  
38

39 The private school students are therefore about 0.44 SDs higher on the latent scale than the state  
40 school students. As a result it is possible to plot state and private schools on the same graph  
41 (Appendix figure 4), with the only difference being that the private schools are further along the  
42 diagonal towards the top right corner. That difference accounts for all of the differences in the  
43 private and state school students, with all other differences in Appendix table 1 being artefacts of  
44 the artificial ceiling of the range at A\*. To put it another way, were attained grades to be the same  
45 in state and private schools then the accuracy and the degree of over-estimation would be the same  
46 in the two types of schools.  
47  
48  
49

50 In conclusion, conventional statistics comparing attained and predicted grades at A-level are  
51 inherently misleading, and suggest differences between groups which are probably not present,  
52 meaning that great care must be taken in interpretation.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

6. *Appendix: Appendix Tables 1 & 2 and Appendix Figures 1, 2, 3 & 4.*

Appendix table 1. Predicted vs Attained A-level grades in applicants from a) State Sector schools (non-Private schools) and b) Independent (Private sector) schools.

a) State Sector: Counts of number of cases								
		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	140	30	5	5	0	0	180 (0%)
	D	210	420	125	20	5	0	780 (1%)
	C	535	1050	1545	400	55	5	3600 (2%)
	B	500	1735	3910	5190	1090	115	12540 (9%)
	A	270	1330	6250	24310	36915	7425	76495 (53%)
	A*	25	75	395	3950	24725	21410	50580 (35%)
	Total	1680	4645	12230	33870	62785	28960	144175
		(1%)	(3%)	(8%)	(23%)	(44%)	(20%)	
		Under	Prediction: Accurate	Over	Correlations:			
		6.4%	45.5%	48.0%	Pearson	Spearman	Polychoric	
					0.635	0.590	0.717	
a) Private Sector: Counts of number of cases								
		Attained Alevel grades						
		E	D	C	B	A	A*	Total
Predicted Alevel grades	E	15	0	0	0	0	0	15 (0%)
	D	15	55	15	5	0	0	85 (0%)
	C	50	85	200	40	10	0	385 (1%)
	B	60	185	430	1085	335	30	2130 (3%)
	A	65	300	1650	8785	19500	3935	34235 (49%)
	A*	5	20	115	1420	15270	16635	33455 (48%)
	Total	205	640	2405	11340	35115	20600	70305
		(0%)	(1%)	(3%)	(16%)	(50%)	(29%)	
		Under- estimate	Prediction: Accurate	Over	Correlations:			
		6.2%	53.3%	40.5%	Pearson	Spearman	Polychoric	
					0.552	0.523	0.678	



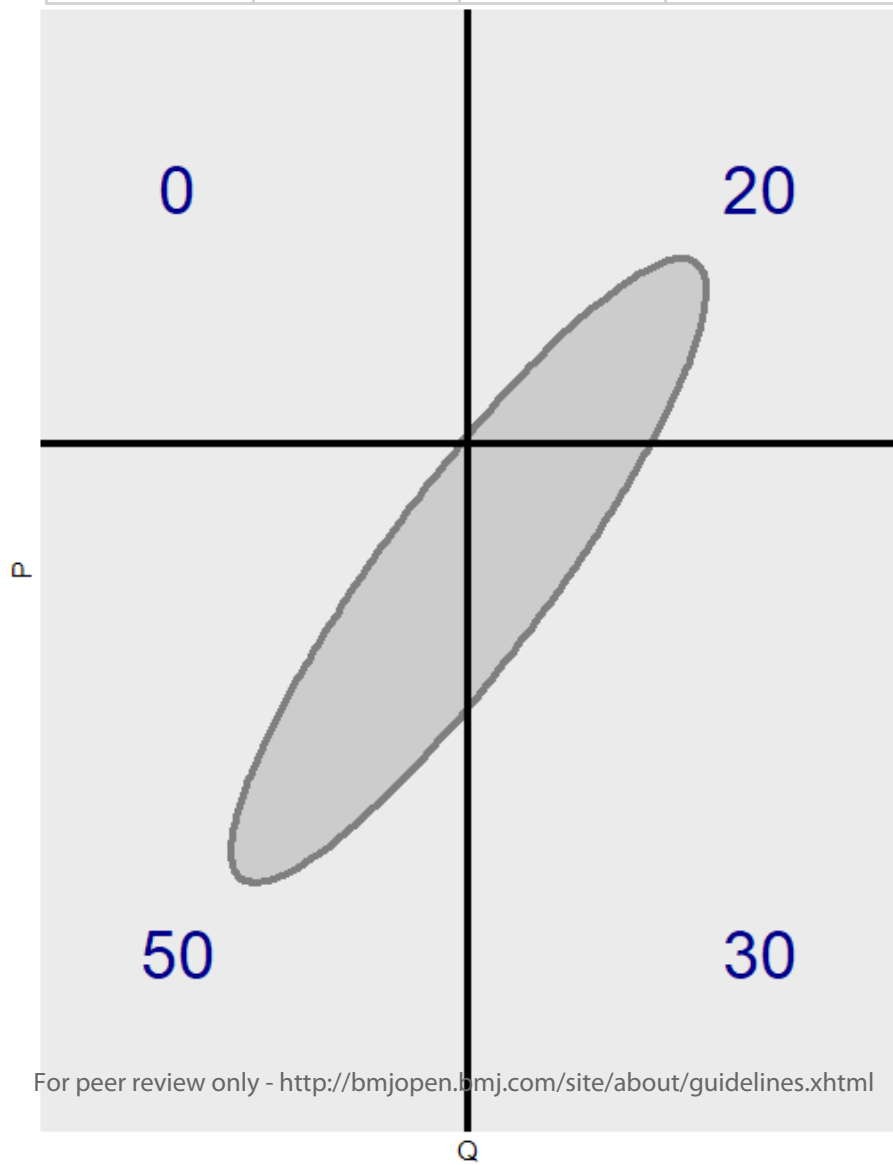
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

*Appendix table 2.* Thresholds, and intervals between thresholds, for the grades for applicants at State and Private schools. Values in bold show mean values across rows and down columns.

	<b>Ordinal</b>		E:D	D:C	C:B	B:A	A:A*	Mean	State minus Private
1	<b>Thresholds</b>								
2	Predicted	State	-3.11	-2.55	-1.89	-1.18	0.39	<b>-1.67</b>	0.47
3		Private	-3.51	-3.00	-2.47	-1.78	0.06	<b>-2.14</b>	
4	Attained	State	-2.31	-1.74	-1.13	-0.33	0.83	<b>-0.94</b>	0.42
5		Private	-2.57	-2.25	-1.68	-0.81	0.55	<b>-1.35</b>	
6	Predicted-Attained	State	-0.80	-0.81	-0.76	-0.85	-0.44	<b>-0.73</b>	0.02
7		Private	-0.76	-0.75	-0.79	-0.97	-0.48	<b>-0.75</b>	
8			<b>-2.18</b>	<b>-1.85</b>	<b>-1.45</b>	<b>-0.99</b>	<b>0.15</b>	<b>-1.26</b>	
9	<b>Threshold intervals</b>		D:C - E:D	C:B - D:C	B:A - C:B	A:A* - B:A			
10	Predicted	State	-0.57	-0.66	-0.71	-1.57	<b>-0.87</b>	0.02	
11		Private	-0.51	-0.54	-0.69	-1.84	<b>-0.89</b>		
12	Attained	State	-0.57	-0.61	-0.80	-1.17	<b>-0.79</b>	-0.01	
13		Private	-0.32	-0.58	-0.87	-1.36	<b>-0.78</b>		
14	Predicted-Attained	State	0.01	-0.05	0.09	-0.41	<b>-0.09</b>	-0.02	
15		Private	-0.01	0.04	0.18	-0.48	<b>-0.07</b>		
16			<b>-0.33</b>	<b>-0.40</b>	<b>-0.47</b>	<b>-1.14</b>	<b>-0.58</b>		

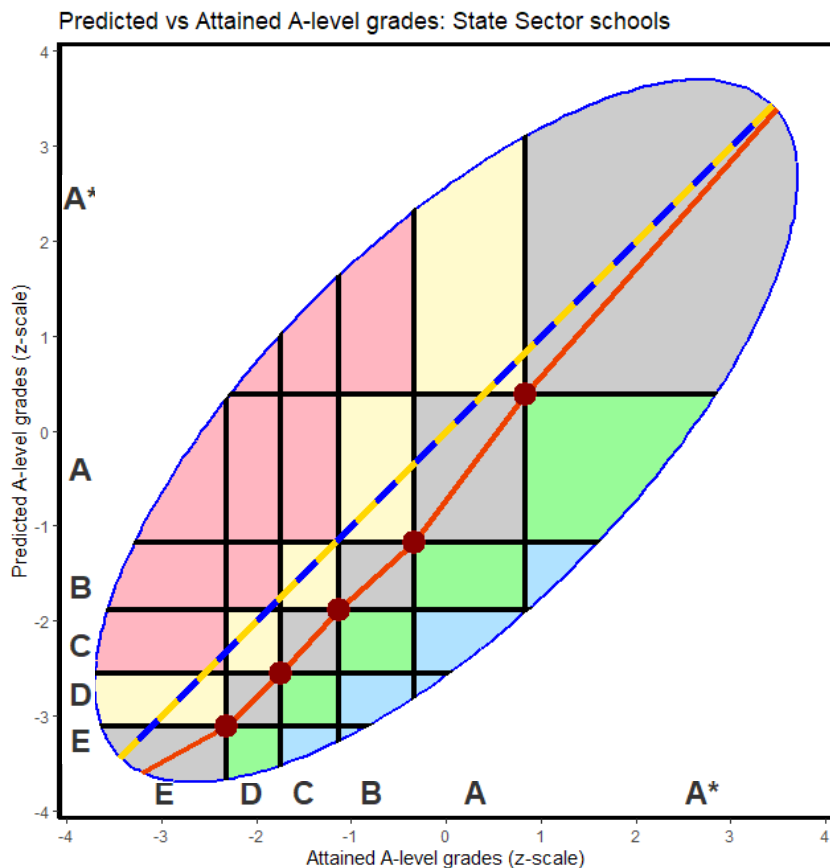
*Appendix figure 1.* Demonstration of how a conventional Pearson or Spearman correlation between binary variables P and Q cannot achieve a correlation of one when marginal proportions of P and Q differ. However the tetrachoric correlation is one, within calculation and rounding errors, being estimated from underlying latent correlation shown in the diagram, with thresholds at -0.842 and 0 for P and Q.

	Q absent	Q present	P totals
P absent	<b>0</b>	<b>20</b>	20
P present	<b>50</b>	<b>30</b>	80
Qtotals	50	50	100
Correlation	Pearson	Spearman	Tetrachoric
	0.5	0.5	0.994
Threshold	P	Q	
	-0.842	0	

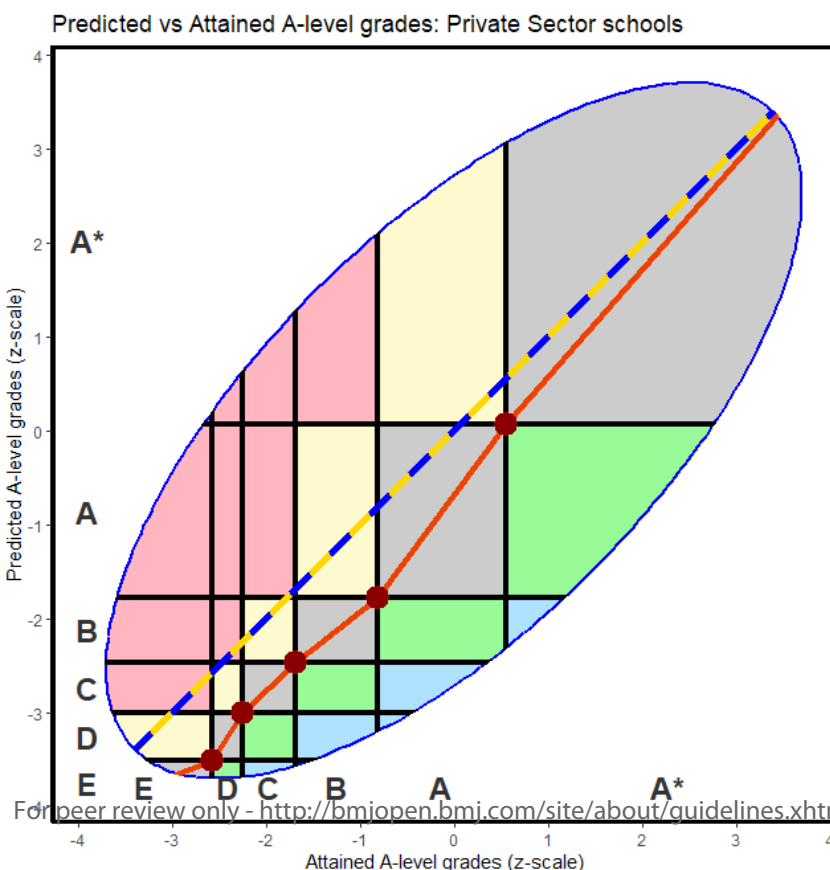


Appendix figure 2. Latent bivariate normal distribution of the relationship between attained A-level grades (horizontal) and predicted A-level grades (vertical). The correlation is represented by the blue ellipse. The dashed blue and yellow line is the line of equality of actual and attained grades. The vertical and horizontal black lines show the thresholds for the grades, shown as E, D, C, B, A and A\*. The solid red dots and red line show where the thresholds for a grade intersect, with all below the main diagonal. Colours indicate over-prediction (yellow and pink) and under-prediction (green and blue).

2.a



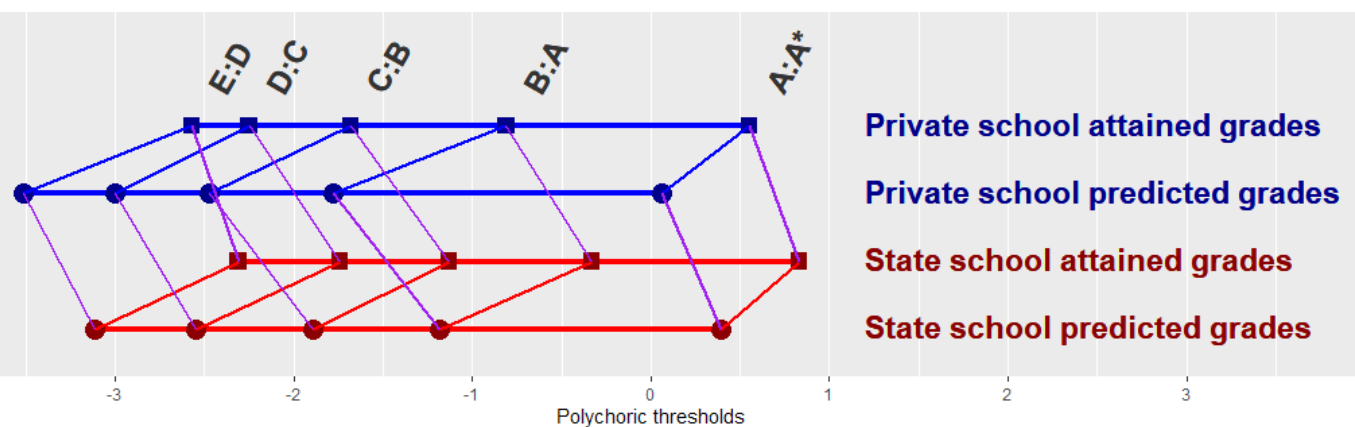
2.b



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

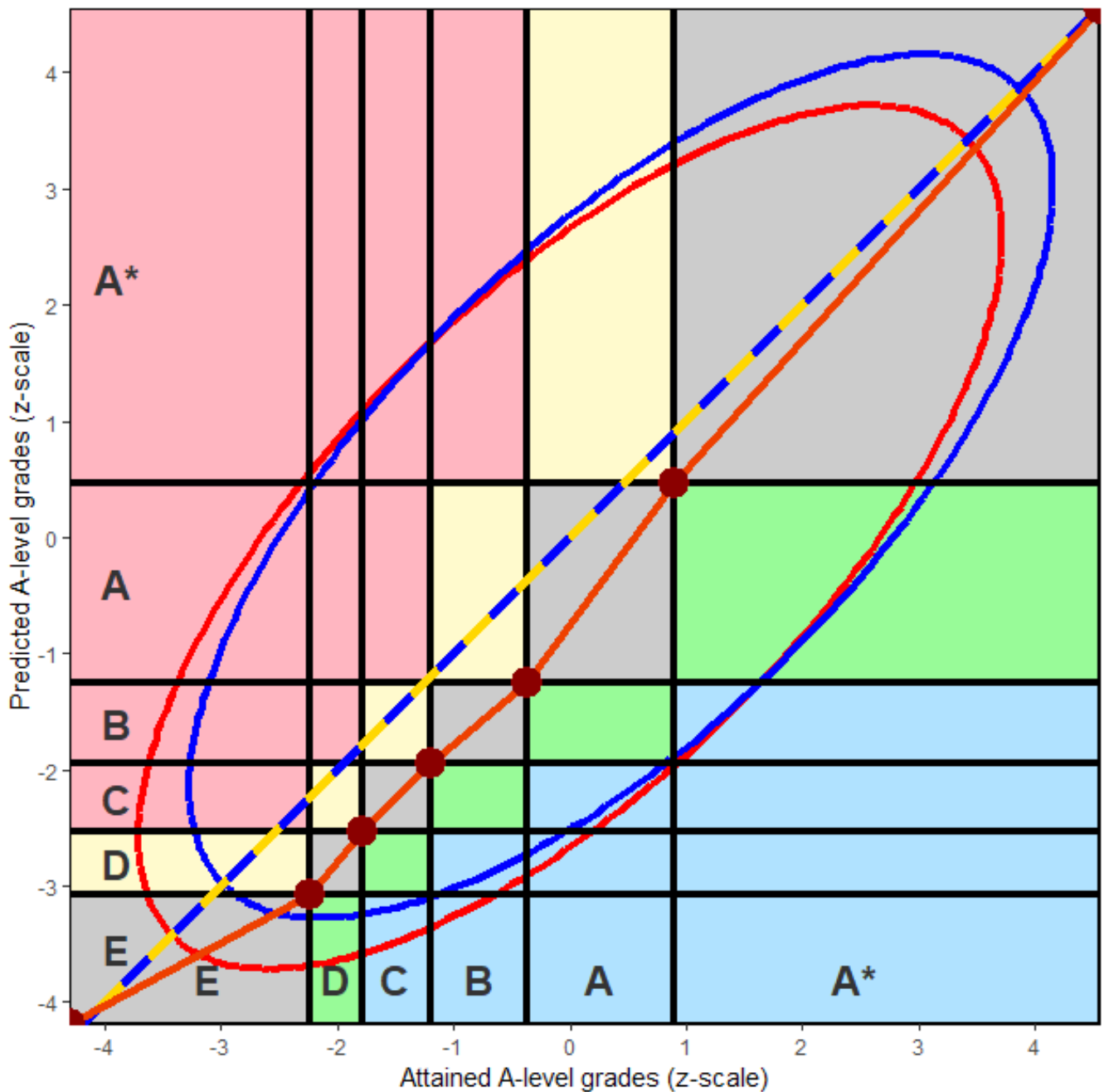
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Appendix figure 3. Summary of polychoric thresholds shown on the horizontal axis, for private (blue) and state (red) schools, for attained grades (squares) and predicted grades (circles). Narrower diagonal lines show the links between attained and predicted grades for private (blue) and state (red) schools. Purple diagonal lines link equivalent points for private and state schools (e.g. attained with attained grades and predicted with predicted grades).



Appendix figure 4. See Appendix figure 2 for the majority of conventions. The fitted ellipses for state sector schools (red) and private sector schools (blue) are shown separately, with the same grade thresholds for both schools. The latent bivariate normal distributions for the two types of school differ entirely in their mean scores, that for private sector schools being shifted up and to the right (by the same amount). The school types therefore differ only in their mean ability levels.

Predicted vs Attained grades: State Sector (red) & Private Sector (blue)



## References

1. Petch JA. School estimates and examination results compared. Manchester: Joint Matriculation Board 1964.
2. Wilmot J. Experiences of summative teacher assessments in the UK. London: Qualifications and Curriculum Authority 2011.
3. Murphy RJL. Teachers' assessments and GCE results compared. *Educational Research* 1979;22(1):54-59.
4. Murphy RJL. O-level grades and teachers' estimates as predictors of the A-level results of UCCEA applicants. *British Journal of Educational Psychology* 1981;51(1):1-9.
5. Everett N, Papageorgiou J. Investigating the Accuracy of Predicted A Level Grades as part of 2009 UCAS Admission Process. London: Department for Business, Innovation and Skills 2011.
6. Wyness G. Predicted grades: Accuracy and impact. A report of University and College Union. London: University and College Union ([https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted\\_grades\\_report\\_Dec2016.pdf](https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted_grades_report_Dec2016.pdf)) 2016.
7. Wyness G. Rules of the game: Disadvantaged students and the university admissions process. London: The Sutton Trust 2017.
8. Murphy R, Wyness G. Minority Report: the impact of predicted grades on university admissions of disadvantaged groups. London: **Centre for Education Policy and Equalising Opportunities (CEPEO) Working Paper No. 20-07, UCL Institute of Education** <https://EconPapers.repec.org/RePEc:ucl:cepeow:20-07> 2020.
9. UCAS. End of cycle report 2017: Qualifications and competition. Cheltenham: UCAS [<https://www.ucas.com/data-and-analysis/ucas-undergraduate-releases/ucas-undergraduate-analysis-reports/2017-end-cycle-report>] 2017.
10. Gill T, Rushton N. The accuracy of forecast grades for OCR Alevels: Statistics Report Series No 26. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/our-research/all-published-resources/statistical-reports/150215-the-accuracy-of-forecast%20-grades-for-ocr-a-levels-in-june-2012.pdf/>] 2011.
11. Gill T, Chang Y. The accuracy of forecast grades for OCR A levels in June 2012: Statistics Report Series No.64. Cambridge: Cambridge Assessment Statistics Report Series No.64 2013.
12. Gill T, Benton T. The accuracy of forecast grades for OCR Alevels in June 2014: Statistics Report Series No 90. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/Images/241261-the-accuracy-of-forecast-grades-for-ocr-a-levels-in-june-2014.pdf>] 2015.
13. UCAS. Factors associated with predicted and achieved A level attainment, August 2016. Cheltenham: UCAS: <https://www.ucas.com/file/71796/download?token=D4uuSzur> 2016.
14. McManus IC, Woolf K, Dacre J. The educational background and qualifications of UK medical students from ethnic minorities. *BMC Medical Education* 2008;8: 21 (<http://www.biomedcentral.com/1472-6920/8/21>)
15. Gill T. Methods used by teachers to predict final Alevel grades for their students. *Research Matters (UCLES)* 2019(28):33-42.
16. Lumb AB, Vail A. Applicants to medical school: the value of predicted school leaving grades. *Med Educ* 1997;31:307-11.
17. Richardson PH, Winder B, Briggs K, et al. Grade predictions for school-leaving examinations: do they predict anything? *Med Educ* 1998;32:294-97.
18. McManus IC, Richards P, Winder BC, et al. Medical school applicants from ethnic minorities: identifying if and when they are disadvantaged. *Brit Med J* 1995;310:496-500.
19. Boliver V. How fair is access to more prestigious universities? *British Journal of Sociology* 2013;64(2):344-64.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
20. Woolf K, Harrison D, McManus IC. The attitudes, perceptions and experiences of medical school applicants following the closure of schools and cancellation of public examinations due to the COVID-19 pandemic in 2020. *medRxiv* 2020;submitted
  21. Woolf K, Harrison D, McManus C. The attitudes, perceptions and experiences of medical school applicants following the closure of schools and cancellation of public examinations in 2020 due to the COVID-19 pandemic: a cross-sectional questionnaire study of UK medical applicants. *BMJ open* 2021;11(3):e044753.
  22. McManus IC, Woolf K, Harrison D, et al. Calculated grades, predicted grades, forecasted grades and actual A-level grades: Reliability, correlations and predictive validity in medical school applicants, undergraduates, and postgraduates in a time of COVID-19. *medRxiv* 2020;doi: <https://doi.org/10.1101/2020.06.02.20116830>
  23. Gill T, Rodeiro, C.V. Predictive validity of level 3 qualifications: Extended Project, Cambridge Pre-U, International Baccalaureate, BTEC Diploma. Cambridge: Cambridge Assessment: Cambridge Assessment Research Report 2014.
  24. Thomson D. Moderating teaching judgments in 2020 [Blog post, 25th March 2020]. London: FFT Educational Lab: <https://ffteducationdatalab.org.uk/2020/03/moderating-teacher-judgments-in-2020/> (accessed 16th April 2020) 2020.
  25. McManus IC, Dewberry C, Nicholson S, et al. Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: Meta-regression of six UK longitudinal studies. *BMC Medicine* 2013;11:243:doi:10.1186/741-7015-11-243.
  26. Meng X-L, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. *Psychological Bulletin* 1992;111(1):172-75.
  27. McManus IC, Dewberry C, Nicholson S, et al. The UKCAT-12 study: Educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a collaborative study of twelve UK medical schools. *BMC Medicine* 2013;11 :244:doi:10.1186/741-7015-11-244.
  28. Wakeford R, Denney ML, Ludka-Stempien K, et al. Cross-comparison of MRCGP & MRCP(UK) in a database linkage study of 2,284 candidates taking both examinations: Assessment of validity and differential performance by ethnicity. *BMC Medical Education* 2015;15(1 (doi:10.1186/s12909-014-0281-2))
  29. McManus IC, Woolf K, Dacre J, et al. The academic backbone: Longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the Specialist Register in UK medical students and doctors. *BMC Medicine* 2013;11:242:doi:10.1186/741-7015-11-242.
  30. Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Medical Teacher* 2016;38(1):3-17.
  31. McManus IC, Harborne A, Smith D, et al. Exploring UK medical school differences: The *MedDifs* study of selection, teaching, student and F1 perceptions, postgraduate outcomes, and fitness to practise. *BMC Medicine* 2019;In press
  32. McManus IC, Woolf K, Dacre JE. Even one star at A level could be "too little, too late" for medical student selection. *BMC Medical Education* 2008;8:16 (<http://www.biomedcentral.com/1472-6920/8/16>)
  1. Petch JA. School estimates and examination results compared. Manchester: Joint Matriculation Board 1964.
  2. Wilmot J. Experiences of summative teacher assessments in the UK. London: Qualifications and Curriculums Authority 2011.
  3. Murphy RJL. Teachers' assessments and GCE results compared. *Educational Research* 1979;22(1):54-59.

- 1
  - 2
  - 3
  4. Murphy R, J. L. O-level grades and teachers' estimates as predictors of the A-level results of UCCA applicants. *British Journal of Educational Psychology* 1981;51(1):1-9.
  - 5
  - 6
  - 7
  - 8
  - 9
  - 10
  - 11
  - 12
  - 13
  - 14
  - 15
  - 16
  - 17
  - 18
  - 19
  - 20
  - 21
  - 22
  - 23
  - 24
  - 25
  - 26
  - 27
  - 28
  - 29
  - 30
  - 31
  - 32
  - 33
  - 34
  - 35
  - 36
  - 37
  - 38
  - 39
  - 40
  - 41
  - 42
  - 43
  - 44
  - 45
  - 46
  - 47
  - 48
  - 49
  - 50
  - 51
  - 52
  - 53
  - 54
  - 55
  - 56
  - 57
  - 58
  - 59
  - 60
4. Murphy R, J. L. O-level grades and teachers' estimates as predictors of the A-level results of UCCA applicants. *British Journal of Educational Psychology* 1981;51(1):1-9.
  5. Everett N, Papageorgiou J. Investigating the Accuracy of Predicted A Level Grades as part of 2009 UCAS Admission Process. London: Department for Business, Innovation and Skills 2011.
  6. Wyness G. Predicted grades: Accuracy and impact. A report of University and College Union. London: University and College Union ([https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted\\_grades\\_report\\_Dec2016.pdf](https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted_grades_report_Dec2016.pdf)) 2016.
  7. Wyness G. Rules of the game: Disadvantaged students and the university admissions process. London: The Sutton Trust 2017.
  8. Murphy R, Wyness G. Minority Report: the impact of predicted grades on university admissions of disadvantaged groups. London: **Centre for Education Policy and Equalising Opportunities (CEPEO) Working Paper No. 20-07, UCL Institute of Education** <https://EconPapers.repec.org/RePEc:ucl:cepeow:20-07> 2020.
  9. UCAS. End of cycle report 2017: Qualifications and competition. Cheltenham: UCAS [<https://www.ucas.com/data-and-analysis/ucas-undergraduate-releases/ucas-undergraduate-analysis-reports/2017-end-cycle-report>] 2017.
  10. Gill T, Rushton N. The accuracy of forecast grades for OCR Alevels: Statistics Report Series No 26. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/our-research/all-published-resources/statistical-reports/150215-the-accuracy-of-forecast%20-grades-for-ocr-a-levels-in-june-2012.pdf/>] 2011.
  11. Gill T, Chang Y. The accuracy of forecast grades for OCR A levels in June 2012: Statistics Report Series No.64. Cambridge: Cambridge Assessment Statistics Report Series No.64 2013.
  12. Gill T, Benton T. The accuracy of forecast grades for OCR Alevels in June 2014: Statistics Report Series No 90. Cambridge: Cambridge Assessment [<https://www.cambridgeassessment.org.uk/Images/241261-the-accuracy-of-forecast-grades-for-ocr-a-levels-in-june-2014.pdf>] 2015.
  13. UCAS. Factors associated with predicted and achieved A level attainment, August 2016. Cheltenham: UCAS: <https://www.ucas.com/file/71796/download?token=D4uuSzur> 2016.
  14. McManus IC, Woolf K, Dacre J. The educational background and qualifications of UK medical students from ethnic minorities. *BMC Medical Education* 2008;8: 21 (<http://www.biomedcentral.com/1472-6920/8/21>)
  15. Gill T. Methods used by teachers to predict final Alevel grades for their students. *Research Matters (UCLES)* 2019(28):33-42.
  16. Lumb AB, Vail A. Applicants to medical school: the value of predicted school leaving grades. *Med Educ* 1997;31:307-11.
  17. Richardson PH, Winder B, Briggs K, et al. Grade predictions for school-leaving examinations: do they predict anything? *Med Educ* 1998;32:294-97.
  18. McManus IC, Richards P, Winder BC, et al. Medical school applicants from ethnic minorities: identifying if and when they are disadvantaged. *Brit Med J* 1995;310:496-500.
  19. Boliver V. How fair is access to more prestigious universities? *British Journal of Sociology* 2013;64(2):344-64.
  20. Woolf K, Harrison D, McManus IC. The attitudes, perceptions and experiences of medical school applicants following the closure of schools and cancellation of public examinations due to the COVID-19 pandemic in 2020. *medRxiv* 2020;submitted
  21. Woolf K, Harrison D, McManus C. The attitudes, perceptions and experiences of medical school applicants following the closure of schools and cancellation of public examinations in 2020 due to the COVID-19 pandemic: a cross-sectional questionnaire study of UK medical applicants. *BMJ open* 2021;11(3):e044753.
  22. McManus IC, Woolf K, Harrison D, et al. Calculated grades, predicted grades, forecasted grades and actual A-level grades: Reliability, correlations and predictive validity in medical school



- 1  
2  
3 applicants, undergraduates, and postgraduates in a time of COVID-19. *medRxiv* 2020;doi:  
4 <https://doi.org/10.1101/2020.06.02.20116830>
- 5  
6 23. Gill T, Rodeiro, C.V. Predictive validity of level 3 qualifications: Extended Project, Cambridge Pre-  
7 U, International Baccalaureate, BTEC Diploma. Cambridge: Cambridge Assessment:  
8 Cambridge Assessment Research Report 2014.
- 9  
10 24. Thomson D. Moderating teaching judgments in 2020 [Blog post, 25th March 2020]. London: FFT  
11 Educational Lab: [https://ffteducationdatalab.org.uk/2020/03/moderating-teacher-](https://ffteducationdatalab.org.uk/2020/03/moderating-teacher-judgments-in-2020/)  
12 [judgments-in-2020/](https://ffteducationdatalab.org.uk/2020/03/moderating-teacher-judgments-in-2020/) (accessed 16th April 2020) 2020.
- 13  
14 25. McManus IC, Dewberry C, Nicholson S, et al. Construct-level predictive validity of educational  
15 attainment and intellectual aptitude tests in medical student selection: Meta-regression of  
16 six UK longitudinal studies. *BMC Medicine* 2013;11:243;doi:10.1186/741-7015-11-243.
- 17  
18 26. Meng X-L, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. *Psychological*  
19 *Bulletin* 1992;111(1):172-75.
- 20  
21 27. McManus IC, Dewberry C, Nicholson S, et al. The UKCAT-12 study: Educational attainment,  
22 aptitude test performance, demographic and socio-economic contextual factors as  
23 predictors of first year outcome in a collaborative study of twelve UK medical schools. *BMC*  
24 *Medicine* 2013;11 :244;doi:10.1186/741-7015-11-244.
- 25  
26 28. Wakeford R, Denney ML, Ludka-Stempien K, et al. Cross-comparison of MRCGP & MRCP(UK) in a  
27 database linkage study of 2,284 candidates taking both examinations: Assessment of validity  
28 and differential performance by ethnicity. *BMC Medical Education* 2015;15(1  
29 (doi:10.1186/s12909-014-0281-2))
- 30  
31 29. McManus IC, Woolf K, Dacre J, et al. The academic backbone: Longitudinal continuities in  
32 educational achievement from secondary school and medical school to MRCP(UK) and the  
33 Specialist Register in UK medical students and doctors. *BMC Medicine*  
34 2013;11:242;doi:10.1186/741-7015-11-242.
- 35  
36 30. Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and  
37 training: Research, theory and practice: AMEE Guide No. 100. *Medical Teacher* 2016;38(1):3-  
38 17.
- 39  
40 31. McManus IC, Harborne A, Smith D, et al. Exploring UK medical school differences: The *MedDifs*  
41 study of selection, teaching, student and F1 perceptions, postgraduate outcomes, and  
42 fitness to practise. *BMC Medicine* 2019;In press
- 43  
44 32. McManus IC, Woolf K, Dacre JE. Even one star at A level could be 'too little, too late' for medical  
45 student selection. *BMC Medical Education* 2008;8:16  
46 (<http://www.biomedcentral.com/1472-6920/8/16>)
- 47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**STROBE Statement**

Checklist of items that should be included in reports of observational studies

Section/Topic	Item No	Recommendation	Reported on Page No
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	2 – “ Longitudinal observational study using UK Medical Education Database (UKMED) data.”
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	Yes page 2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Yes page 9 & 10 in “The rationale for using A-level grades in selection” and “Review of literature on predicted and forecasted grades”
Objectives	3	State specific objectives, including any prespecified hypotheses	Yes page 14 – in “Empirical questions to be addressed.”
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	Summary on page 14 – reader directed to Supplementary Information for more information.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up and data collection	Years covered by data sets and descriptions of them given on page 14. Reader directed to Supplementary Information for more information.
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 14 data sources for participants are described.
		<i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	
Variables	7	(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed	A table of measures and outcomes has been included in the Supplementary
		<i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	
		Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	1

Information

3	Data sources/measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	See table of measures and outcomes in the Supplementary Information
5	Bias	9	Describe any efforts to address potential sources of bias	
7	Study size	10	Explain how the study size was arrived at	Page 4 we state full population data so unlikely to be biased.
11	Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Page 17 in Predictive validity of predicted and attained A-level grades in medical students describes how the various outcome variable were handled.
16	Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	Pages 18 and 19
17			(b) Describe any methods used to examine subgroups and interactions	Not applicable
18			(c) Explain how missing data were addressed	No missing data
20			(d) Cohort study—If applicable, explain how loss to follow-up was addressed Case-control study—If applicable, explain how matching of cases and controls was addressed	Not applicable
22			(e) Describe any sensitivity analyses Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	Not applicable

Section/Topic	Item No	Recommendation	Reported on Page No
<b>Results</b>			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable
		(b) Give reasons for non-participation at each stage	Not applicable
		(c) Consider use of a flow diagram	An Ibray chart is provided in figure 3
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential	9

36/bmjopen-2020-047354 on 16 December 2021. Downloaded from <http://bmjopen.bmj.com/> on April 19, 2024 by guest. Protected by copyright.

confounders

(b) Indicate number of participants with missing data for each variable of interest

Not applicable

(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)

Not applicable

*Cohort study*—Report numbers of outcome events or summary measures over time

9,10

*Case-control study*—Report numbers in each exposure category, or summary measures of exposure

Not applicable

*Cross-sectional study*—Report numbers of outcome events or summary measures

Figure 4 contains Ns

(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included

Figure 4 gives raw Pearson correlations (r; in blue), construct-level predictive validity

(b) Report category boundaries when continuous variables were categorized

Not applicable

(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period

Page 20 gives details of potential impact on EPM of using grades not based on actual exam performance.

Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

Not applicable

## Discussion

Key results 18 Summarise key results with reference to study objectives

Yes page 20

36bmjopen-2020-047354 on 16 December 2021. Downloaded from <http://bmjopen.bmj.com/> on April 19, 2024 by guest. Protected by copyright.

1	Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Yes page 21
2				
3	Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Yes pages 20, 21
4				
5			Discuss the generalisability (external validity) of the study results	Yes
6				mentions
7				that same
8				issues apply
9	Generalisability	21		to 2021
10				applicant
11				cohort –
12				page 21
13				
14				
15	<b>Other Information</b>			
16				
17	Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Yes page 32
18				Funding

19 *\*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.*

20

21 **Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).

For peer review only