

This is a repository copy of *Does 'online confidence' predict application success and later academic performance in medical school? A UK-based national cohort study.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/154193/>

Version: Accepted Version

---

**Article:**

Tiffin, Paul Alexander orcid.org/0000-0003-1770-5034 and Paton, Lewis William orcid.org/0000-0002-3328-5634 (2019) Does 'online confidence' predict application success and later academic performance in medical school? A UK-based national cohort study. *BMJ Open*. e034437. ISSN: 2044-6055

<https://doi.org/10.1136/bmjopen-2019-034437>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# **Does ‘online confidence’ predict application success and later academic performance in medical school? A UK-based national cohort study**

Paul A. Tiffin and Lewis W. Paton

## **Author details**

PAT: Department of Health Sciences, University of York & Hull York Medical School, Heslington, YO10 5DD, [paul.tiffin@york.ac.uk](mailto:paul.tiffin@york.ac.uk)

LWP: Department of Health Sciences, University of York, Heslington, YO10 5DD, [lewis.paton@york.ac.uk](mailto:lewis.paton@york.ac.uk)

## **Abstract**

**Objectives:** The UK Clinical Aptitude Test (UKCAT) previously piloted an assessment of 'online confidence', where candidates were asked to indicate how confident they were with their answers. This study examines the relationship between these ratings, the odds of receiving an offer to study medicine, and subsequent undergraduate academic performance.

**Design:** National cohort study.

**Setting:** UK undergraduate medical selection.

**Participants:** 56,785 UKCAT candidates who sat the test between 2013 and 2016.

**Primary outcome measures:** Two measures of 'online confidence' were derived: the well-established 'confidence bias', and; a novel 'confidence judgement' measure, developed using Item Response Theory in order to derive a more sophisticated metric of the ability to evaluate one's own performance on a task. Regression models investigated the relationships between these confidence measures, application success and academic performance.

**Results:** Online confidence was inversely related to cognitive performance. Relative underconfidence was associated with increased odds of receiving an offer to study medicine. For 'confidence bias' this effect was independent of potential confounders (OR 1.42, 1.09 to 1.86,  $p=0.01$ ). Whilst 'confidence judgement' was also a univariable predictor of application success (OR 1.22, 1.01 to 1.47,  $p=0.04$ ), it was not an independent predictor. 'Confidence bias', but not 'confidence judgment', predicted the odds of passing the first year of university at the first attempt, independently of cognitive performance, with relative underconfidence positively related to academic success (OR 3.24, 1.08 to 9.72,  $p=0.04$ ). No non-linear effects were observed, suggesting no 'sweet spot' exists in relation to online confidence and the outcomes studied.

**Conclusions:** Applicants who either appear underconfident, or are better at judging their own performance on a task, are more likely to receive an offer to study medicine. However, online confidence estimates had limited ability to predict subsequent academic achievement. Moreover, there are practical challenges to evaluating online confidence in high stakes selection.

### **Strengths and limitations of this study**

- This study used a large, national dataset.
- This is, to our knowledge, the first study to link ‘online confidence’ of medical school applicants to subsequent outcomes at medical school.
- This study introduces a novel, alternative approach to measuring ‘online confidence’, using Item Response Theory (IRT).
- The ‘online confidence’ tests were piloted in low-stakes conditions, and therefore it is unclear whether our results would generalise to high-stakes settings.
- It was not possible to link ‘online confidence’ with previously piloted ‘self-report’ confidence measures, and thus we could not compare confidence as a trait and as a meta-cognitive ability.

## Introduction

The safe and effective practice of medicine can be assumed to require an accurate appraisal of one's own abilities. Indeed, overconfidence can lead to diagnostic errors (1). Thus, the prospect of being able to accurately estimate ability is attractive to medical selectors.

Competition to study medicine is high; in the UK there are around 11 applications for every available place (2). As such, aptitude tests are commonly used as part of the selection procedure, partly in an attempt to effectively discriminate between large numbers of similarly high-performing candidates. Such cognitive tests scores are also intended to predict future academic achievement in medical education, and there is accumulating evidence that they do this to some extent (3-5).

There is an increased recognition of the essential role that 'non-academic' attributes play in a successful medical career. Indeed, most issues associated to professional malpractice are related to the personality and interpersonal functioning of the doctor involved, rather than a lack of clinical knowledge or skills (6). Consequently there has been a growing drive to evaluate 'personal qualities', aside from intellectual ability and academic achievement, as part of medical selection at different career stages (7-9). However, measuring such traits in a high-stakes selection situation poses a number of practical challenges (10).

In a wider educational context, the relationship between 'self-beliefs' and academic performance have been well researched, from the late 1970s (11, 12). These 'self-beliefs' predict academic achievement to varying degrees. A literature review reported that 'self-concept' had the lowest correlation, followed by 'academic anxiety' and then 'self-efficacy' (13). A fourth self-belief, self-confidence, is reported to be the best 'non-cognitive' predictor of future academic performance (13, 14).

The measurement of 'confidence' can be separated into two conceptually distinct approaches. Firstly, 'self-confidence' as a trait, is usually captured via responses to self-report questionnaires. Such self-reported 'confidence' may be considered a relatively stable

personality trait (15). Interestingly, higher scores on a questionnaire-based measure of self-confidence amongst medical school applicants have been shown to be associated with the reduced risk of an individual subsequently reporting experiencing health-related issues at medical school (16). Secondly, 'self-confidence' can be conceptualised as an ability to accurately (or otherwise) appraise one's own ability at a certain task (17). This is also often referred to in the literature as 'online confidence', frequently characterised by how well a test-taker is able to judge their own abilities on a written assessment. In this respect, online confidence could be conceptualised as a 'meta-cognitive' attribute (that is the ability to 'think about thinking'). That is, one which requires a cognitive judgment about one's own cognitive performance (18). In this sense the term 'non-cognitive', when used in conjunction with this trait, must be used cautiously, as there are suggestions that such self-appraisal has a cognitive component (14).

The UK Clinical Aptitude Test (UKCAT) (19), subsequently renamed the University Clinical Aptitude Test (UCAT), introduced in 2006, is currently used as a component of selection by the majority of UK medical schools. It presently consists of four cognitive scales ('abstract reasoning', 'decision making', 'quantitative reasoning' and 'verbal reasoning') and a situational judgment test (SJT). The UKCAT Consortium replaced the 'decision analysis' subtest with the 'decision making subtest' in 2017, having piloted items from the latter in 2016. This change was made for several reasons. Given the overall ability level of the candidates it was desirable to have a subtest that discriminated more precisely at the upper end of performance (i.e. a test that was experienced as fairly challenging). Moreover, the design of the decision analysis subtest constrained trialling of new items. This led UKCAT (now UCAT) to have had concerns regarding overexposure of test content to potential candidates. It was also hoped that the decision making subtest would assess a relatively broader range of traits relating to decision making, as defined in the Selecting for Excellence Report (20).

The 'decision analysis' component of the UKCAT was based on a decoding task. That is, candidates were presented with a scenario and a set of codes (e.g. A = 'Never', B = 'Bad', C = 'Lawyer', D='Faithful', E='Employer', F='Good', G='Friends', and so on). The test items consisted of a series of statements (e.g. 'Bad lawyers are never loyal to their firm'). The test taker was then presented with a multiple choice of codes (e.g. 1. 'ABCDE', 2. 'ABDEF', 3. 'ACDEF', 4. 'ABCEG'). The code that best reflected the meaning of the statement had to be selected from this list by the test takers (in this example 'ABCDE'). For other items the decoding task was reversed and candidates had to select the most appropriate statement from a presented selection, having been given the code. Previous independent analysis demonstrated that the decision analysis had an acceptable level of reliability with Cronbach's alpha (binary version) and McDonald's omega values of 0.87 (21).

Between 2013 and 2016 the UKCAT was used to pilot an online 'confidence rating' within the 'decision analysis' section of the test. After each of the items presented in this section, candidates had to indicate their confidence that the answer they had provided was correct, using a scale from one to five. It was made clear to the candidates at the time that their responses were not going to be used in the selection process. An example item and corresponding confidence rating is shown in Figure 1. As can be seen, no guide or anchor points were provided to the candidates for the confidence rating.

[INSERT FIGURE 1 ABOUT HERE]

Previous work focussed on measuring online confidence has tended to take a relatively simplistic approach to calculating an estimate of this ability. This usually involves creating a 'confidence bias' score, which is defined as the average difference between the confidence self-ratings and the performance on the items (22). For example, on a five point confidence rating scale a test taker may choose the fourth point (e.g. 'fairly confident I got the question

right') which equates to a score of 0.75 on a scale of 0 to 1 (e.g. 'not at all confident' to 'almost sure'). Candidates, as is usual, would score 1 point for a correct answer and 0 for an incorrect response. In order to calculate confidence bias the confidence self-rating would be subtracted from this score. Thus, in this example a candidate would be allocated a score of 0.25: that is a score of 1 for a correct answer, minus a rating of 0.75 for a self-rating of 4/5 on the five point confidence scale. Had the candidate actually answered the item incorrectly the score would have been -0.75 (i.e. 0 for a wrong answer minus 0.75 for the confidence rating). The overall 'confidence bias' for each candidate is then calculated as the mean confidence bias score for all the items the test-taker responded to. This means that the theoretical range of confidence bias scores is from -1 to 1, where negative values suggest a tendency for over-confidence and positive values suggest under-confidence in one's own performance on the test.

'Confidence bias' calculated in this manner has previously been reported to be correlated with cognitive ability. Specifically, individuals who tend to report higher levels of confidence on a task have, on average, poorer performance (22). However, such an association may be vulnerable to potential confounding, given how this metric is derived. That is, an inverse relationship between 'confidence bias' and cognitive performance scores could, at least partly, be an artefact of the score achieved in the latter. That is, the confidence bias scores will be heavily dependent on performance on the cognitive task itself. Specifically, in this case, high bias scores are much more likely to be observed in those with relatively poor performance on the decision analysis subtest; thus, a higher frequency of incorrect responses will almost inevitably lead to more opportunities for 'over confidence' to be observed. Secondly, it is well recognised that individuals vary in their underlying tendency to respond to questionnaires by choosing central or extreme points on such scales. This is referred to as 'response style' and can be crudely categorised into a 'central tendency' or 'extreme response style', depending on the preference for mid-points or outer regions of scales respectively. A number of approaches have been suggested to adjust for response



style (23). Thirdly, online confidence, as an ability, could be confounded with the self-confidence as a trait if the mean (baseline) levels of confidence are not taken into account. That is, confidence, in this context could be considered an aspect of meta-cognition, specifically an ability to evaluate and reflect on one's performance on a task. This contrasts with a trait model of confidence, which would postulate that individuals have a generalised and consistent tendency to take either a positive or negative view of their own abilities. Thus any method to evaluate online confidence must attempt to differentiate between the 'trait' and 'ability' aspects of the construct.

Classical test theory relies on the raw summed scores to estimate the ability of a test-taker on an assessment measuring that trait, i.e. the sum of correct responses is assumed to be a 'sufficient statistic' by which to differentiate individuals in that respect. This assumes that all the test items are equally good at discriminating between candidates of differing abilities. In contrast, Item Response Theory (IRT) provides a more sophisticated approach to modelling the ability of a test-taker on an assessment and is able to accommodate items of differing discrimination (24). This permits a more nuanced metric of online confidence to be derived by evaluating the correlation between the prediction, derived from an IRT model, that a candidate would answer a particular item correctly, and their own judgment regarding whether they felt they had answered the question correctly. Thus, it would be anticipated that candidates with a well-developed ability to appraise their own test performance would have relatively high, positive, correlations between these two values. Conversely individuals with only a limited ability to judge their own ability would show little or no correlation. At extremes, theoretically, very overconfident candidates may show negative correlations between these two elements. We refer to this novel approach of measuring online confidence as estimating 'confidence judgment' in order to differentiate it from the conventional measure of 'confidence bias'. Thus, we define confidence judgment as the accuracy of one's own judgement in relation to one's ability at an online or written test. It can thus be conceptualised as an aspect of meta-cognition.

The presence of the UKCAT data relating to online confidence allowed us to conduct a study with the primary aim of evaluating the potential for the measurement of this construct to enhance medical selection. This included appraising the potential impact on the demographics of the UK medical student population if implemented. A secondary aim was to compare the properties of our novel metric of online confidence compared to the conventional approach. The study objectives were thus:

1. To evaluate the relationship between two measures of online confidence and the sociodemographic and educational characteristics of medical school applicants.
2. To model the relationship between online confidence and the likelihood of success at medical school application.
3. To model the relationship between online confidence and, for those successful applicants entering UKCAT consortium medical schools, academic performance in the first two years of undergraduate study.

The exploration of these data also allowed us to consider the practicalities of implementing such a testing approach to medical selection.

## **Methods**

### *Data availability and preparation*

Applicants to UKCAT consortium medical schools must sit the UKCAT test during the calendar year of their application. The test may only be taken once per application cycle. Data were provided by UKCAT for 65,691 candidates who sat the test from 2013 to 2016. Raw data were placed by UKCAT in the Health Informatics Centre- a 'Safe Haven' hosted by the University of Dundee (25). The Safe Haven is a set of secure servers that allows management and analysis but where, for security, no individual data may be extracted, but only reports on aggregated data or the results of analyses may be outputted. Thus the data

were subsequently cleaned, managed, linked and analysed by the research team inside the Safe Haven. Table 1 and Figure 2 show the data that were available for this study.

[INSERT FIGURE 2 ABOUT HERE]

### *Predictor variables*

Approximately 10% of the candidates who completed the confidence ratings did not show any variation in their responses (i.e. they responded by choosing all ones, or all threes, and so on). These responses were deemed 'invalid' responses and were not further analysed. This left 56,785 candidates with 'valid' confidence ratings.

'Confidence bias' for each candidate was calculated as follows. For each item, the confidence rating (on a one to five scale) was converted to a confidence score on the unit scale (i.e. zero to one). This confidence score was then subtracted from the item score (zero for an incorrect answer, one for a correct answer). The results were then averaged within candidate to produce the 'confidence bias' score for each applicant, with higher scores representing relative underconfidence.

The novel 'confidence judgment' estimate for each candidate was calculated as follows. For each candidate, a 'two parameter logistic regression IRT' (2PL-IRT) model was used to predict the probability (or to be precise, the log of the odds- 'log odds') of the candidate getting each item correct (24). This probability is determined by three factors. Firstly, the candidate's ability, which is estimated using their performance across all items of the test, compared to the other individuals in the sample. Secondly the item's relative difficulty, as estimated using the performance of all candidates on that specific item. Finally that particular question's 'discrimination' (i.e. precision in differentiating between candidates at the ability level suggested by the item's difficulty). Once the probability (i.e. log odds) of each candidate getting each item correct had been estimated, the 'confidence judgment' for each candidate was calculated by computing the correlation between that candidate's log odds of getting each item correct and their own confidence rating. Note, in this context the use of the log

odds was preferable to the exponentiated odds ratio or probability, as the former behave in a more linear fashion than the latter two values. In order to differentiate the 'ability' aspects of online confidence from the 'trait' element (see earlier) the confidence self-ratings for each item were first adjusted by rescaling them in terms of their mean values ('elevation') and variance ('scatter') for each individual. This method has been previously shown to be useful for such purposes (26). In practice, such adjustment was done via the use of 'within person z-scores', which standardised the confidence self-ratings within each test-taker by subtracting the mean rating and dividing by the standard deviation. Thus each score represented a correlation coefficient for that individual. In this sense '1' would represent a perfect positive correlation between the respondent's estimation of their own ability at the test (ideal judgment), '0' as no correlation (i.e. no relationship) and '-1' as a perfect inverse correlation (i.e. the respondent was most confident about the items they were least likely to answer correctly).

### *Background variables*

We also had access to information relating to candidate sociodemographic characteristics (see Table 2). In line with previous research (3) we dichotomised gender, socioeconomic status (NS-SEC rating (27) one to three versus four or five), secondary school type attended (non-selective school, selective school (including state grammar schools)) and language status (native English speaker, English as a second language). As in previous studies relating to the UKCAT, a continuous metric of academic achievement in secondary (high) school was derived (5). This metric consisted of A-level performance and Irish and Scottish qualifications. In the UK, all secondary school grades are associated with a corresponding Universities and Colleges Admissions Service (UCAS) tariff. The metric of prior academic achievement was derived as the percentage of the maximum available UCAS tariff each candidate achieved. Only the three highest grades achieved were included. Any resits were excluded, as were qualifications in 'General Studies'. Standardised z-scores (i.e. mean of zero, standard deviation of one) within nationality were then calculated.

UKCAT scores for each candidate were available. This included overall performance ('UKCAT total score') and performance on each of the four subscales of the test. These five scores were standardised as z-scores according to the UKCAT year of sitting in order to allow cross-comparison across application year. Previously it has been shown that the UKCAT scores can be conceptualised as being divided into those related to verbal and non-verbal reasoning (21). The UKCAT total scores were made up of three non-verbal subtest scores (quantitative reasoning, abstract reasoning and decision analysis) and one verbal subtest (verbal reasoning). Therefore in order to obtain an overall estimate of cognitive ability a 'balanced' summary form of the test performance was derived by averaging the three standardised scores from the non-verbal subtests, adding the standardised verbal reasoning score and dividing this total by two.

#### *Outcome variables*

Data on whether an application made by a candidate resulted in an offer being made (or not) to study medicine at the university applied to were available for 2013 test takers only (18,985 applications). In this respect information was only available for medical schools in the UKCAT consortium, though this accounted for almost all the UK universities offering medicine courses at that time- that is roughly 30 of 35 medical schools (the numbers varied slightly during the study period).

In order to enter medical school, candidates must both receive an offer from a university and subsequently meet the conditions of that offer (e.g. achieving a certain level of secondary (high) school achievement). Of those who entered medical school, we also had access to performance in knowledge-based and skill-based end of year exams in the first year of university. These data were available for 1,252 and 854 candidates respectively. In order to allow cross-comparison across years and medical schools, these outcomes were standardised as z-scores according to the year of examination and institution of the candidate using the approach employed by previous studies using these outcomes (3, 9, 28).

## *Data analysis*

The data were placed in a safe haven, cleaned, managed and linked. All analyses were performed in the safe haven using Stata version 14 (29).

Univariable analyses between the confidence ratings and sociodemographic data were analysed via correlations, regressions and Kruskal-Wallis tests. Informed by these analyses, multivariable forwards stepwise regressions were built in order to predict the odds of success at application and the odds of passing the first year of university at the first attempt. As each candidate could apply to multiple medical schools, when looking at the relationship between the confidence scores and the odds of an offer, multilevel logistic regression was used, with offers conceptualised as nested within candidates. As missing data, in terms of item responses and sociodemographic variables, were relatively uncommon, we used listwise deletion to account for missingness. It should also be highlighted that as the predictor variables (UKCAT subtest scores etc.) were standardised according to the applicants, then the resulting regression or correlation coefficients did not require the usual correction approaches for the restriction of range due to the selection procedure (i.e. only applicants that were successful had academic outcomes observed) (30). Obtaining academic outcomes relied on the medical schools in the UKCAT consortium returning them annually. Thus, the missing data related to the academic outcomes were assumed to be either 'missing completely at random' (i.e. purely due to chance) or 'missing at random' (related to observed variables) and thus unlikely to threaten the validity of the results. Previously sensitivity analyses provided evidence that these academic outcomes, particularly in the first two years of undergraduate medical education, were likely to be largely missing in this way (3).

In order to evaluate the extent extreme response style (ERS) may influence or confound confidence bias scores, we estimated this trait separately. This was done by a previously outlined method whereby 'shadow indicators' were created (0 or 1) depending on whether an 'extreme' response was selected (in this case one or five) (23). This trait was separately

modelled for each candidate using a two parameter logistic IRT model with the 'shadow indicators' as binary indicators.

### *Patient and Public Involvement*

There was no patient or public involvement in this study.

## **Results**

As might be expected, there was at least modest correlation ( $r=0.24$ ) between 'confidence bias' and 'confidence judgement'. The mean 'confidence bias' score was  $-0.40$  (SD  $0.21$ ) and the mean 'confidence judgement' value was  $0.30$  (SD  $0.24$ ). Reliability indices, in terms of Cronbach's alpha, for the 'confidence bias' values related to the constituent items for the five forms of the test ranged from  $0.83$  to  $0.85$ . An equivalent reliability metric could not be calculated for confidence judgement as it is actually a 'within person' (rather than 'within test') correlation coefficient.

### *Relationship between online confidence and sociodemographic data*

Kruskall-Wallis testing showed that lower 'confidence bias' scores (i.e. overconfidence) were significantly ( $p<0.001$ ) associated with male gender, English as a second language (EASL), self-declared non-White relatively high and attendance at a non-selective secondary school. When controlling for score achieved on the 'decision analysis' subtest, this relationship remained significant ( $p<0.01$ ) only for male gender, EASL, and non-selective schooling. Similarly, poorer 'confidence judgement' (i.e. confidence ratings were less highly correlated with the actual probability of answering a question correctly) was observed for males, EASL and those reporting non-White ethnicity ( $p<0.0001$  in all cases). However, in the case of 'confidence judgement' lower values were also observed for candidates reporting to be from non-professional socio-economic backgrounds ( $p=0.0002$ ), suggesting poorer ability to evaluate one own performance. These associations remained significant at the  $p<0.001$  level after adjusting for performance on the 'decision analysis' with the exception of non-

professional background ( $p=0.5$ ). 'Confidence judgment' was not associated significantly with secondary school-type attended ( $p=0.2$ ).

Table 2 depicts the relationships between the two online confidence metrics and self-reported ethnicity. As can be seen, those self-reporting themselves as 'White' display the lowest levels of overconfidence (as measured by 'confidence bias') and also display the highest levels of accuracy in their own ability, as indexed by 'confidence judgment'. The differences between self-reported ethnic groups appears somewhat more marked for the 'confidence bias' scores, compared to the 'confidence judgment' estimates. However, in the former case these differences become statistically non-significant once performance on the decision analysis subtest are controlled for. For example, the most marked difference in both online confidence metrics can be seen between those reporting themselves as of 'white' ethnicity and those self-identifying as 'black'. Self-reported ethnicity ('white' vs 'black') can be predicted from confidence bias score (OR 0.18, 0.15 to 0.22,  $p<0.001$ ). However, once the influence of performance on decision analysis is controlled for this relationship with 'confidence bias' score disappears (OR 1.04, 0.84 to 1.29,  $p=0.7$ ). Thus, the inter-ethnic group differences for 'confidence bias' appear to be an artefact of performance on the related cognitive test (i.e. decision analysis). A similar magnitude of univariable relationship is observed between these two self-reported ethnic groups and 'confidence judgment' (OR 0.49, 0.42 to 0.57,  $p<0.001$ ). However, the relationship remains statistically significant even after the influence of performance on decision analysis is accounted for (OR 0.74, 0.64 to 0.87,  $p<0.001$ ).

#### *Relationship between online confidence and prior academic performance*

As can be seen from Table 3, both 'confidence bias' and 'confidence judgment' correlate positively with standardised secondary school performance and standardised UKCAT performance. The correlation between confidence bias and the standardised decision analysis was moderate ( $r=0.51$ ). This highlights the dependency of this traditional metric of 'online confidence' on task performance (see also earlier, in methods section). In contrast a



more modest correlation between this index of overall performance on this cognitive subtest and 'confidence judgement' was observed ( $r=0.16$ ), suggesting it is less dependent on performance at the related cognitive task.

#### *Online confidence and success at application to medical school*

Mann-Whitney U testing showed small yet statistically significant differences between the 2013 cohort (the only cohort for which we had data on application success) and later cohorts on online confidence and the 'decision analysis' scale.

Univariable multilevel logistic regression showed that those with higher 'confidence bias' scores (i.e. those who are relatively underconfident) were much more like to receive an offer to study medicine (OR 11.97, 9.55 to 15.01,  $p<0.001$ ). This can be interpreted as follows: for every standard deviation above the mean an applicant's 'confidence bias' score was, their odds of receiving an offer to study medicine increased by a factor of 12. After controlling for background variables, including cognitive ability, in a multivariable stepwise regression, underconfidence remained a statistically significant independent predictor of receiving an offer from a medical school (OR 1.42, 1.09 to 1.86,  $p=0.01$ ). The full results of this model are shown in Table 4.

'Confidence judgement' was also a univariable predictor of application success (OR 1.22, 1.01 to 1.47,  $p=0.04$ ), albeit of only borderline statistical significance. However, in contrast to 'confidence bias', once the potential influence of cognitive ability and background variables were controlled for 'confidence judgement' was not an independent, statistically significant, predictor of receiving an offer (OR 0.98, 0.80 to 1.20,  $p=0.8$ ).

#### *Online confidence and academic performance during medical school*

Mann-Whitney U testing showed that those individuals with data on academic performance in medical school had statistically significantly higher confidence bias, confidence judgement and decision analysis scores than the dataset as a whole.

The only statistically significant association between 'confidence bias' and undergraduate performance was that observed for the odds of passing the end of year one at the first attempt (OR 4.37, 1.54 to 12.42,  $p=0.006$ ). That is, those who reported relative underconfidence (as indexed by averaging one point above the mean on the 'confidence bias' score) had over four times the odds of passing the year at first attempt, compared to those with average scores. This effect remained statistically significant when controlling for performance on the 'decision analysis' subtest of the UKCAT (OR 3.24, 1.08 to 9.72,  $p=0.04$ ). As the outcome was categorical, no easily interpretable R statistic (representing the proportion of variance explained by the predictor variables) was available, in contrast to continuous variables modelled using linear regression. However, an analogous statistic for logistic regression exists in 'McFadden's pseudo- $R^2$ ' (31). This reflects the amount of variance explained in the hypothetical latent variable, postulated to be underlying the observed responses. In the case of predicting the odds of passing year 1 at first attempt both performance on the decision analysis subtest and the confidence bias scores had relatively low pseudo- $R^2$  values, though the addition of the latter into the logistic regression equation doubled its magnitude. Specifically, the pseudo  $R^2$  value for predicting the odds of passing year 1 at first attempt from the standardised decision analysis score was 0.0057. This increased to 0.011 when the confidence bias scores were entered. Thus, at least for this specific academic outcome the tradition measure of online confidence appeared to show some incremental predictive validity over the linked cognitive task.

Once the potential influence of other background variables were controlled for, 'confidence bias' scores were no longer significantly predictive of the odds of passing year one of medical school at the first attempt (OR 3.08, 0.72 to 13.21,  $p=0.1$ ). We observed no significant relationships between 'confidence bias' score and the odds of passing year two at first attempt, nor with performance on knowledge or skills-based assessments in either of the first two years of medical school.

The relationship between ‘confidence judgment’ and subsequent academic performance was even weaker. We observed no statistically significant associations, or even modest trends, between confidence judgement scores and any of the academic outcomes we had access to.

## **Discussion**

In these analyses we examined the relationship between ‘online confidence’, sociodemographic data, the probability of an offer of a place to study medicine and subsequent academic performance in medical school. We were able to use an existing metric of online confidence, as utilised in the original pilot study and provisional descriptive analysis by Pearson Vue (32). In addition we were able to explore the use of an experimental method to evaluate ‘confidence judgement’, using a relatively sophisticated approach to modelling self-appraisal of ability at a cognitive test.

In line with previous research we found that overconfidence was inversely related to cognitive performance (14). In this study we observed that underconfidence, as measured by the ‘confidence bias’ score, was associated with an increased odds of receiving an offer to study medicine. Moreover this effect remained significant once the impact of their background demographic and academic variables were taken into consideration.

Additionally, ‘confidence bias’ was modestly linked to subsequent academic performance, in the sense that those that were relatively underconfident had higher odds of passing the first year of undergraduate medicine at first sitting. This association was relatively independent of cognitive ability, though not of the influence of sociodemographic background variables.

The novel measure of ‘confidence judgement’ showed somewhat different properties to that of ‘confidence bias’, despite the scores derived from the two approaches correlating moderately. ‘Confidence judgement’ showed a much weaker relationship with cognitive ability, which is undoubtedly advantageous when trying to delineate between cognitive ability and this potential aspect of meta-cognition. There was also somewhat different relationship

with self-reported ethnicity. Specifically, inter-group ethnic differences were not quite as marked for 'confidence judgement', compared to 'confidence bias'. However, in the case of the latter, the observed relationship appeared to be totally accounted for by a candidate's performance on the decision analysis subtest. This was not so with the 'confidence judgement' measure, which may be picking up genuine differences between self-reported ethnic groups in terms of meta-cognitive style.

Importantly, in contrast to 'confidence bias', there were no observed relationships between 'confidence judgement' and undergraduate academic performance. However 'confidence judgement' was significantly related to the odds of receiving an offer from medical school, though this effect was not independent of other cognitive, educational and background factors.

#### *Possible interpretations*

The association between underconfidence and an increased chance of receiving an offer for medical school was independent of other educational, cognitive and sociodemographic factors. It could be that candidates who appear overconfident are less likely to receive offers since the vast majority of UK medical schools still use face-to-face interviews and/or group exercises as part of the selection process.

'Confidence judgement' had lower observed correlation with cognitive tasks than 'confidence bias'. This suggests that 'confidence judgement' (i.e. a candidate's judgement of their own ability) is less of a proxy for overall ability at a cognitive task than 'confidence bias' (whether a candidate is on average over or under confident). Furthermore, 'confidence bias' may be measuring other irrelevant constructs, such as 'extreme response style' (see Table 3). It may also be that the 'confidence judgement' estimates are less prone to confounding with general intellectual ability. Indeed, it could be conceptualised as a more nuanced meta-cognitive skill focussed on being able to appraise the relative difficulty of test items in relation to one's own ability in the domain being tested.

It could be hypothesised that the lower observed correlations for 'confidence judgement' are that it is less 'reliable', in some sense. As 'confidence judgement' is itself an individual correlation coefficient, traditional measures of reliability are not applicable. However, it is possible that 'confidence judgement' captures relatively less information on each candidate than 'confidence bias' does. However, we note that we did observe significant associations between 'confidence judgement' and the odds of receiving an offer to study medicine, as well as differences in the measure across ethnicities. This suggests the measure is able to discriminate between individuals. Nevertheless, we cannot rule out that the lower correlations observed in some cases are due to 'lower reliability'.

It could be hypothesised that future doctors should ideally show neither overconfidence nor underconfidence. However, we did not observe any non-linear effects for 'confidence bias' in our analyses. This would suggest that no such 'sweet spot' exists, at least in relation to the outcome measures examined in this study. Indeed, in this context, the facet of over or underconfidence may be better conceptualised as a component of interpersonal competency. That is, individuals who are more narcissistic, and hence may be 'overconfident' are likely to encounter more interpersonal difficulties in the workplace (33). However, recent findings have (depressingly) suggested that narcissism, as a trait, may be associated with greater academic success than an individual's cognitive ability might otherwise suggest (34). On the other hand, one study reported that higher self-rated confidence in medical undergraduates was associated with poorer academic performance (9).

We also noted that all standardised subtest scores on the UKCAT were independent and statistically significant predictors of the odds of an offer of a place to study at medical school, with very similar effects sizes (Table 4). Presumably this observation reflects the common use of the UKCAT summed total score within the admissions process, which would, in effect, give each component an equal weighting.

Formal testing showed statistically significant differences between the 2013 cohort and later cohorts. However, the actual effect sizes were trivially small and may have been caused by differences in the demographics.

### *Implications for policy*

In theory, the ability to evaluate under or overconfidence, as distinct from cognitive ability, in medical applicants should add value within the selection process. In particular the construct could predict more distal, non-academic, aspects of performance. There would, however, appear to be a number of practical challenges with implementing either 'confidence bias' or 'confidence judgement' assessments in high stakes selection. Firstly, it was noted in these data that a proportion of candidates did not vary in their responses in relation to their perceived confidence (i.e. there was zero variance). Thus such candidates provided no information about how they perceived their confidence related to their actual ability. It is difficult to see how such invalid response patterns could be discouraged or mitigated in practice. Indeed, if candidates believe that underconfidence was more desirable than overconfidence, they may deliberately rate their confidence rather lower than they might otherwise. It is difficult to safeguard against such social desirability bias. Moreover, overall, there seemed to be a stronger relationship between online confidence and success at application than with subsequent academic performance, though this could have been partly an artefact of 'restriction of range' as a result of the selection process (35). This would have been mainly, statistically, addressed via standardising the predictor variables according to the applicant, rather than entrant pool (30). Thus, it would seem unwise to implement a selection measure where success at application was not reasonably mirrored by performance on subsequent work related metrics. There would also be challenges, related to the wider context to how the resulting scores might be interpreted and used by selectors. For example, would the ratings be treated as a criterion or norm referenced measure? How might the demographic associations with online confidence, highlighted in the results of the

present study, present results, be handled, to prevent certain population groups being disadvantaged?

When considering the introduction of new metrics into the selection process, it is important that they add value above and beyond those metrics already in use. A particular problem with using the conventional estimate of 'confidence bias' is that it correlates at least moderately with actual cognitive ability in the domain tested (in this case  $r=0.51$ ). Therefore, there are some doubts about the incremental value that 'confidence bias' ratings would add to the selection process, at least in terms of academic performance. Nevertheless, the estimation of online confidence may still prove useful. It could still be used in selection if the practical challenges of measuring online confidence in a high stakes setting could be overcome. There may also be potential to use online confidence as a future research tool, exploring the link between self-appraisal and actual clinical practice in medical staff. With the increased focus on assessing the inter-personal competence of medical applicants it may also be worth revisiting some of the earlier work around 'emotional intelligence' as an applied ability (36). That is, being able to accurately identify emotional states in oneself and others, and also being able to respond to them effectively. Such traits can be evaluated, to some extent, via situational judgment tests, though more resource intensive approaches may be required to increase the precision by which such 'non-cognitive' attributes are measured. Thus, it may be more desirable, from a personnel selection perspective, to focus on the personal qualities associated with trait confidence and inter-personal effectiveness, in contrast to online confidence, as an aspect of meta-cognition.

#### *Potential strengths and limitations*

There were three key limitations with this study. Firstly, self-report confidence measures have previously been piloted as part of the UKCAT (9). However, we could not link the data relating to online confidence presented in this paper to the self-reported confidence, as there was a negligible overlap between candidates who had undertaken both the measures. Thus, we could not compare confidence as a trait and as a meta-cognitive ability.

Secondly, whilst confidence has previously been explored as a predictor of academic performance, one could argue that it would be of most interest to link such measures to interpersonal functioning, such as fitness to practise issues. However at present, such outcomes are not freely available. In the future it may be possible to link measures of confidence to these outcomes. However fitness to practice issues are a relatively infrequent occurrence and therefore study power would remain challenging. However, it may be feasible to link metrics of confidence with other outcomes related to interpersonal functioning, such as 'high fidelity' simulations featuring role played patients. It is possible that these first two limitations could be addressed if the data available here were included in the UK Medical Education Database (UKMED) (37), the UK's national data repository for medical trainees.

Thirdly, these tests were piloted in low stakes conditions. It is thus unclear the extent to which they may generalise to high stakes settings. Indeed, although it is indisputably desirable to select candidates on personal qualities beyond cognitive and academic ability, there are substantial challenges with evaluating such traits in a high-stakes setting (10). The most obvious threat to the validity of the scores from such assessments is proposed by faking-effects or social desirability bias. However, there are also some deeper philosophical issues, which surfaced in previous debates relating to the overall concept and measurement of 'emotional intelligence', of which the ability to appraise one's own ability could be considered a putative facet. Indeed, there were two main criticisms of the concept of 'emotional intelligence' as an ability. Firstly, tests of such personal qualities tried to capture 'typical performance', that is an individual's general behavioural tendency. This is in contrast to tests of 'maximum performance', such as those related to academic or cognitive ability, where a test taker would attempt to achieve the highest score possible (38, 39). Another related criticism of the idea that such personal qualities could be conceptualised as abilities was the problem of deriving a scoring rubric. That is, tests evaluating such traits generally cannot be considered to have 'right' or 'wrong' answers, in the same way that cognitive



assessments do. This leads to difficulties in deriving valid and fair scoring systems, with all options, including expert and normative-based keys having flaws (40).

Other limitations to this study are worth noting. These include the fact that only applications to UKCAT consortium universities could be observed, though these represented most medical schools at the time of the study. The number of UK medical schools operating in the UK varied slightly during the study period (for example, Aston and the University of Central Lancashire both launched in 2015). However, the vast majority of medical schools were part of the UKCAT consortium at that time, with roughly 30 of the approximately 35 British medical schools having membership. However, the few universities using the Biomedical Admissions Test (BMAT), the main alternative selection assessment, included Oxford and Cambridge. Therefore, it is therefore not possible to rule out that those who apply solely to non-UKCAT consortium medical schools are more or less confident than those applicants present in this study. Nevertheless, our experience of working with medical selection assessments is that almost all candidates who sit the BMAT also sit the UKCAT, though the converse is not true. This is presumably the case because medical school applicants would generally not want their options at application limited to the small number of universities that use the BMAT as their selection assessment. As with previous studies relating to the UKCAT, a reliance on local, medical school assessments, where the definition of 'knowledge' and 'skills' based assessments are not operationalised and defined, limits the inferences we can draw from such results (3, 28).

## **Conclusions**

We observed relatively few associations between online confidence in medical applicants and the outcomes of interest, though underconfidence seemed a relatively robust predictor of success at application. However, implementing such measures in high-stakes selection situations may not be feasible or desirable at present. There are issues in relation to both the practicality of such an approach as well as questions related to the added value that such a metric may add to the current battery of tests already employed by many medical schools in

Western countries. Thus, in practice it may be more fruitful to concentrate on the attempted measurement of other aspects of personal qualities that may be deemed important to future clinical performance, in particular those that are less strongly related to cognitive ability.

## **Declarations**

### *Ethics approval and consent to participate*

Ethical approval for this study was not required as it relied on the analysis of de-identified routinely collected data. This was confirmed in writing by Durham University's School for Health Ethics Committee.

### *Consent for publication*

Not applicable.

### *Data sharing statement*

The data used in this study may be made available from UCAT on an application being approved by the UCAT Board.

### *Funding*

PAT receives funding for his academic time via an NIHR Career Development Fellowship. This paper represents independent research part-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. The UKCAT Board provided funding for a portion of LWP's time on this project.

### *Competing interests*

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

at the UKCAT Research Group. LWP has received travel and subsistence expenses for attendance at the UKCAT consortium meeting. The UKCAT Board partly funded this research but did not take an active role in determining the study design or reporting the results.

#### *Author Contributions*

PAT lead on study conception, data analysis and drafting the manuscript. LWP contributed to data analysis, interpretation and drafting the manuscript. Both authors approve the manuscript and agree with its submission to BMJ open. Both authors agree to be accountable for all aspects of the work.

#### *Transparency statement*

The lead author (the manuscript's guarantor) affirms the manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

#### *Acknowledgements*

We thank Rachel Greatrix from UKCAT for her assistance with accessing the data used in the study. We are also grateful to Brad Wu from Pearson VUE for supervising the original pilot and his helpful comments on an earlier version of this manuscript.

#### **List of abbreviations**

CB: Confidence bias

CJ: Confidence judgement

IRT: Item response theory

OR: Odds ratio

UCAS: Universities and Colleges Admissions Service

UKCAT: UK Clinical Aptitude Test

## References

1. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121(5):S2-S23.
2. The Medical Schools Council. Selecting for Excellence: End of Year Report- 2013. London: The Medical Schools Council; 2013.
3. Tiffin PA, Mwandigha LM, Paton LW, et al. Predictive validity of the UKCAT for medical school undergraduate performance: a national prospective cohort study. *BMC Med*. 2016;14(1):140.
4. Donnon T, Paolucci E, Violato C. The predictive validity of the MCAT for medical school performance medical board licensing examinations: a meta-analysis of the published research. *Acad Med*. 2007;82:100 - 6.
5. McManus I, 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 Med*. 2013;11:244.
6. Tiffin PA, Paton LW, Mwandigha LM, et al. Predicting fitness to practise events in international medical graduates who registered as UK doctors via the Professional and Linguistic Assessments Board (PLAB) system: a national cohort study. *BMC Med*. 2017;15(1):66.
7. MacKenzie RK, Dowell J, Ayansina D, et al. Do personality traits assessed on medical school admission predict exit performance? A UK-wide longitudinal cohort study. *Adv Health Sci Educ Theory Pract*. 2017;22(2):365-85.

8. Smith DT, Tiffin PA. Evaluating the validity of the selection measures used for the UK's foundation medical training programme: a national cohort study. *BMJ open*. 2018;8(7).
9. Finn GM, Mwandigha L, Paton LW, et al. The ability of 'non-cognitive' traits to predict undergraduate performance in medical schools: a national linkage study. *BMC Med Educ*. 2018;18(1):93.
10. Niessen ASM, Meijer RR. On the Use of Broadened Admission Criteria in Higher Education. *Perspect Psychol Sci* 2017;12(3):436-48.
11. Shavelson RJ, Hubner JJ, Stanton GC. Self-concept: Validation of construct interpretations. *Rev Educ Res*. 1976;46(3):407-41.
12. Bandura A, Walters RH. Social learning theory: Prentice-hall Englewood Cliffs, NJ; 1977.
13. Stankov L, Lee J. Quest for the best non-cognitive predictor of academic achievement. *Educ Psychol*. 2014;34(1):1-8.
14. Stankov L, Morony S, Lee YP. Confidence: the best non-cognitive predictor of academic achievement? *Educ Psychol*. 2014;34(1):9-28.
15. Stankov L, Lee YP. Confidence and cognitive test performance. *Educ Psychol*. 2008;100:961-76.
16. Paton LW, Tiffin PA, Smith D, et al. Predictors of fitness to practise declarations in UK medical undergraduates. *BMC Med Educ*. 2018;18(1):68.
17. Stankov L, Kleitman S, Jackson SA. Measures of the trait of confidence. *Measures of Personality and Social Psychological Constructs*: Elsevier; 2015. p. 158-89.
18. Moore DA, Healy PJ. The trouble with overconfidence. *Psychol Rev* 2008;115(2):502.
19. UKCAT Consortium. About the UK Clinical Aptitude Test (UKCAT) [Available from: <https://www.ucat.ac.uk/ucat/>]. Accessed 19<sup>th</sup> November 2019.
20. Medical Schools Council. Selecting for Excellence: Final Report. London: The Medical Schools Council; 2014.

21. Tiffin PA. Understanding the Dimensionality and Reliability of the Cognitive Scales of the UK Clinical Aptitude test (UKCAT): Summary Version of the Report 2013 [Available from: <https://www.ucat.ac.uk/media/1182/understanding-the-dimensionality-and-reliability-of-the-cognitive-scales-of-the-ukcat.pdf>]. Accessed 19<sup>th</sup> November 2019.
22. Burns KM, Burns NR, Ward L. Confidence—More a Personality or Ability Trait? It Depends on How It Is Measured: A Comparison of Young and Older Adults. *Front Psychol.* 2016;7(518).
23. Wetzel E, Böhnke JR, Rose N. A Simulation Study on Methods of Correcting for the Effects of Extreme Response Style. *Educ Psychol Meas.* 2016;76(2):304-24.
24. Franzen MD. Item Response Theory. In: Kreutzer JS, DeLuca J, Caplan B, editors. *Encyclopedia of Clinical Neuropsychology*. New York, NY: Springer New York; 2011. p. 1362-.
25. University of Dundee. HIC Safe Haven [Available from: [www.dundee.ac.uk/hic/hicsafehaven/](http://www.dundee.ac.uk/hic/hicsafehaven/)] Accessed 11<sup>th</sup> November 2019.
26. McDaniel MA, Psotka J, Legree PJ, et al. Toward an understanding of situational judgment item validity and group differences. *J Appl Psychol.* 2011;96(2):327-36.
27. Statistics. OfN. The National Statistics Socio-economic classification (NS-SEC) [Available from: <https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatistics socioeconomicclassificationnssecbasedonsoc2010>]. Accessed 19<sup>th</sup> November 2019.
28. 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(5).
29. StataCorp LP. *Stata MP for Windows 64-bit*. College Station, TX: StataCorp; 2016.
30. Muthén LK, Muthén BO, Asparouhov T. Chapter 10: Missing data. *Regression and Mediation analysis using MPLUS*. Los Angeles, CA.: Muthén & Muthén; 2016.

31. McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka P, editor. *Frontiers in Econometrics*. Cambridge, MA: Academic Press; 1973. p. 105-42.
32. Pearson VUE. Confidence Rating Data and Score. London: Pearson VUE; 2017 March 2017.
33. Penney LM, Spector PE. Narcissism and Counterproductive Work Behavior: Do Bigger Egos Mean Bigger Problems? *Int J Select Assess*. 2002;10(1-2):126-34.
34. Lin Y, Clough P, Welch J, Papageorgiou K. Individual Differences in Mental Toughness Associate with Academic Performance and Income. *Pers Individ Diff*. 2017;113:178-83.
35. Zimmermann S, Klusmann D, Hampe W. Correcting the predictive validity of a selection test for the effect of indirect range restriction. *BMC Med Educ*. 2017;17:246.
36. Cherry MG, Fletcher I, O'Sullivan H, Dorman T. Emotional intelligence in medical education: a critical review. *Med Educ*. 2014;48(5):468-78.
37. UK Medical Education Database. [Available from: <https://www.ukmed.ac.uk>]. Accessed 19<sup>th</sup> November 2019.
38. Brody N. What cognitive intelligence is and what emotional intelligence is not. *Psychol Inq*. 2004;15(3):234-8.
39. Sackett PR, Zedeck S, Fogli L. Relations between measures of typical and maximum job performance. *J Appl Psychol*. 1988;73(3):482-6.
40. MacCann C, Roberts RD, Matthews G, Zeidner M. Consensus scoring and empirical option weighting of performance-based Emotional Intelligence (EI) tests. *Pers Individ Diff*. 2004;36(3):645-62.

## **Figure legends**

Figure 1: An example of an item from the Decision Analysis section of the UK Clinical Aptitude Test within the confidence rating pilot.

Figure 2: Data flow chart for the study.



## Tables

Variable	Number of candidates with valid data
UKCAT scores	65,691
'Valid' confidence ratings	56,785
Gender	65,691
Socioeconomic background	49,943
School type attended	10,053
English language status	55,212
Secondary (high) school qualifications	11,384
University application information	18,985
Academic outcome year 1 medical school	1,252
Academic outcome year 2 medical school	854

Table 1. Numbers of candidates with data available for the analyses.

Self-reported ethnicity	Mean CB Score (SD)	Mean CJ Score (SD)
White (n= 26,859)	-0.01 (0.19)	0.32 ( 0.24)
Asian (n=15,928)	-0.04 (0.22)	0.28 (0.25)
Black (n=3,866)	-0.08 (0.22)	0.28 (0.24)
Mixed (n=2,001)	-0.02 (0.21)	0.30 (0.24)
Other (n=1,305)	-0.05 (0.22)	0.29 (0.24)

Table 2: Confidence bias (CB) and confidence judgment (CJ) scores according to self-reported ethnicity. CB scores less than zero represent overconfidence, with larger negative values indicating more overconfidence. CJ scores represent the ability for candidates to appraise their own performance, with a CJ of one indicating perfect judgement.

<b>Variable</b>	<b>Correlation with CB (n=56,785)</b>	<b>Correlation with CJ (n=56,479)</b>
Standardised secondary school qualification tariff	0.16	0.06
Standardised 'decision analysis' score	0.51	0.16
Standardised 'abstract reasoning' score	0.23	0.07
Standardised 'quantitative reasoning' score	0.23	0.09
Standardised 'verbal reasoning' score	0.21	0.09
Standardised total UKCAT score	0.39	0.14
'Balanced' UKCAT score	0.34	0.12
Extreme response style	-0.46	-0.01

Table 3: Correlation between educational performance (standardised A-level tariff and UKCAT performance) and confidence ratings (confidence bias (CB) and confidence judgement (CJ)).

<b>Predictor variable</b>	<b>Odds Ratio (OR)</b>	<b>95% Confidence Interval for OR</b>	<b>p</b>
<b>Confidence bias</b>	1.48	1.15 to 1.91	0.002
<b>Male gender</b>	0.69	0.64 to 0.76	<0.001
<b>Standardised ‘decision analysis’ score</b>	1.49	1.39 to 1.60	<0.001
<b>Standardised ‘verbal reasoning’ score</b>	1.51	1.43 to 1.59	<0.001
<b>Standardised ‘quantitative reasoning’ score</b>	1.40	1.32 to 1.49	<0.001
<b>Standardised ‘abstract reasoning’ score</b>	1.44	1.37 to 1.51	<0.001
<b>Standardised secondary school qualification tariff</b>	1.48	1.39 to 1.58	<0.001
<b>‘Non-white’ ethnicity</b>	0.89	0.81 to 0.97	0.012
<b>Non-selective secondary (high) school attended</b>	0.92	0.84 to 1.00	0.046

Table 4. Results from the multivariable multilevel models predicting an offer from ‘confidence bias’ scores in candidates (n=7,870), controlling for the influence of other statistically significant predictor variables.