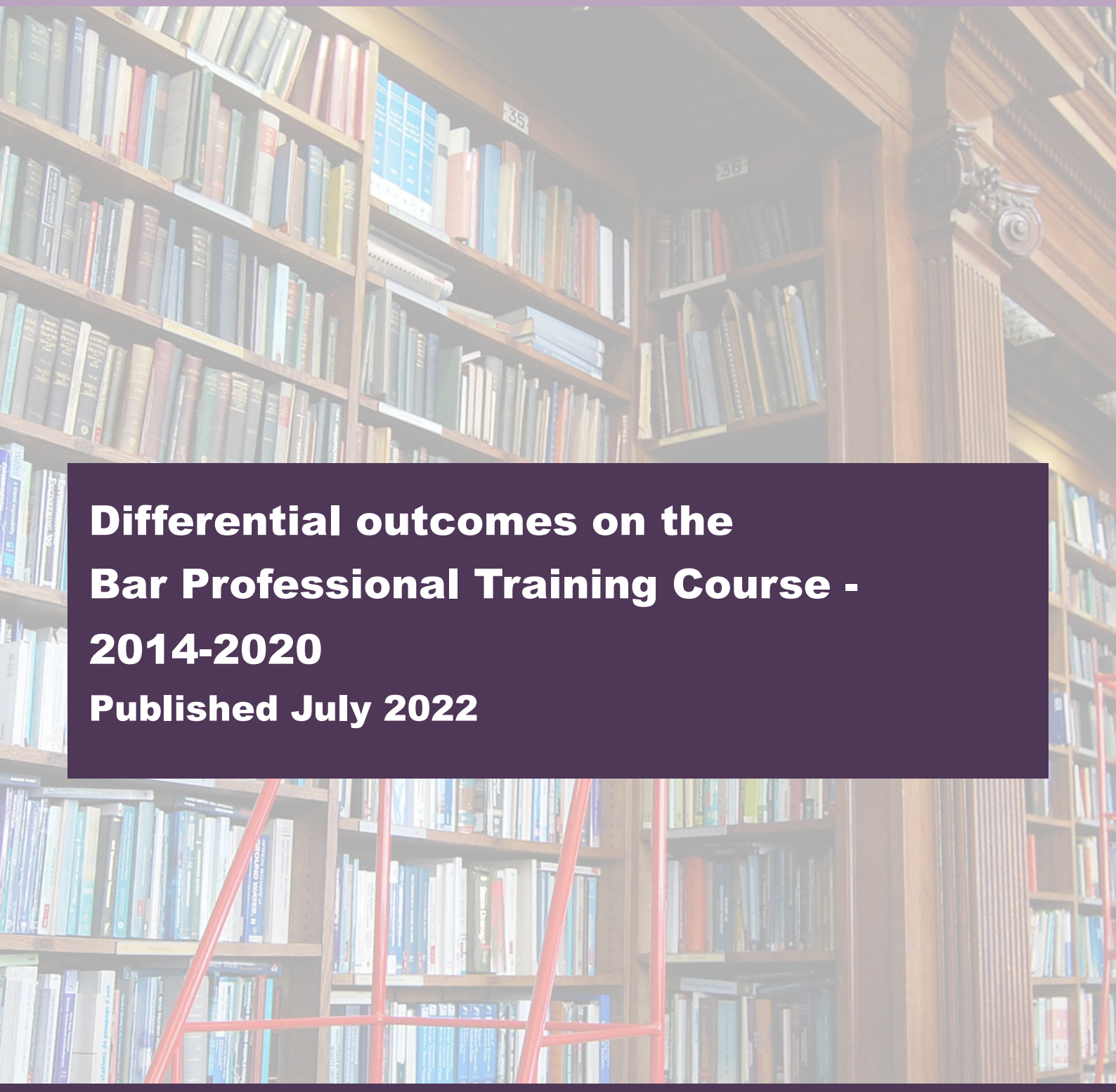


Differential outcomes on the Bar Professional Training Course - 2014-2020

Published July 2022

If you would like this report in an alternative format, please contact the BSB Research Team on 020 7611 1467 or email Research@BarStandardsBoard.org.uk

BAR STANDARDS BOARD

REGULATING BARRISTERS

Contents

Executive Summary

- This report presents research into differential outcomes for different demographic groups regarding the results from modules on the Bar Professional Training Course (BPTC) from 2014-2020. It has a particular focus on results for the centralised assessments, which are the three examinations that were set on behalf of the BSB by the Centralised Examination Board (CEB): Civil Litigation, Criminal Litigation and Professional Ethics.
- Some exploration of results across BPTC modules is presented, as is further analysis of how each centrally assessed module compares to other BPTC modules in terms of differential outcomes. This research follows on from research published in 2017 and helps us to fulfil our duties as a regulator.

Background

- To become a practising barrister in England and Wales prospective barristers are required to pass three stages of training - the academic, vocational, and professional components. The vocational stage of training from 2011/12-2019/20 was the BPTC.
- The three assessments ("Civil Litigation", "Criminal Litigation" and "Professional Ethics") were centrally set to ensure consistency across providers of vocational training for the Bar, and closer regulatory oversight of standards in knowledge subjects for prospective barristers.
- The BSB is committed to ensuring that access to training for the Bar is open to all equally. To fulfil this aim, it is important that we monitor the results from the centralised assessments, and other modules in vocational training and highlight any trends present. We have a particular interest in exploring whether the centralised assessments have a disproportionate impact on different groups sitting them, as these are set on behalf of the BSB, and are a key component of maintaining standards in vocational training for the Bar.
- Previous research on differential outcomes undertaken by the BSB suggested that, even once other variables are controlled for, ethnicity had a significant predictive value for module scores. The predictive value of ethnicity was larger for the modules centrally assessed by the BSB than for other modules, with minority ethnic background students scoring lower than White students.
- The format of the examinations for the centralised assessments changed in 2016/17. In addition to following up on the trends found previously, the change in format to the assessments gave this research an additional focus; to explore whether the change in format of the examinations on the centralised assessments had any impact on differential outcomes on these modules. The format of the exams introduced in 2016/17 was present until 2019/20, which was the last year of the BPTC.
- Centralised assessments in these subjects continue under the new Bar training courses that began in September 2020, but the examinations have been revised. The centralised assessment in Professional Ethics is now sat during pupillage and not as part of the vocational training stage. As such, the results of this research relate to examination formats that are no longer current. We will revisit this analysis in relation to the new-format exams in the future, and the findings of this report will provide a comparison so we will be able to determine the extent to which the reform of the exams has impacted on differential outcomes on the Bar training course.

This research

- This research involved the analysis of BPTC module results covering the 2014-2020 period. Only the first sittings for each year were included, as these had a greater number of students sitting them than second sittings. The analysis involved exploration of the data via some basic descriptive statistics (pass rate, mean scores and interquartile range), the development of more complicated multiple regression models that had module score as the variable being modelled, and a series of independent variables that were tested for the strength of the relationship they showed with module score. There were seven regression models developed;

- two for each centrally assessed module covering the 2014-2020 period (for each centrally assessed module there was one model controlling for individual sits, and one model investigating a broader before/after reform to the assessments trend); and
- a model comparing differential outcomes across modules only on the Spring 2019 and Summer 2020 sits.

Key findings

- Upon analysis of the data, it was found that the centralised assessments were the modules that consistently displayed the lowest mean scores for those sitting them. This was particularly the case for those with an upper second class and lower second class degree.
- Compared to examinations sat during 2014-2016 and following the introduction of the newer format examinations in 2017, there was a general trend of a drop in mean scores for sittings of each centrally assessed examination. The drop in mean scores is particularly seen in those with a first class or upper second class degree. The mean score for the centralised assessments also varied more widely between years than for other modules, as did the failure rate for those sitting the centrally assessed modules.
- The results of the regression models found that, as with in previous research on differential outcomes, ethnicity was found to be a statistically significant variable with a relatively large main effect size in relation to predicted score on Civil Litigation, Criminal Litigation, and Professional Ethics. Those from Asian/Asian British, Black/Black British, Mixed/Multiple ethnic backgrounds, and from other ethnic backgrounds were all predicted to do worse on the assessments than White students on each centralised assessment, even when controlling for other variables.
- First degree classification and first degree institution attended were also statistically significant variables with relatively large effect sizes across the Civil, Criminal and Professional Ethics models.
- The effect sizes of other variables in the centralised assessments models were generally smaller (except for those in older age groups to an extent). This suggests that differences in score have a stronger relationship with ethnicity and academic history than the other variables analysed.
- Overall results on a sit by sit basis suggest that the introduction of the newer format assessments did not appear to lead to a consistent change in differential outcomes on the centralised assessments for the demographic variables of age, disability, domicile, ethnicity, gender, parental university status, and type of school attended.
- When looking at the before/after assessment reform model for Professional Ethics, there were a few variables for which differential outcomes may have widened. These were age, English as a first language, gender, and whether a parent attended university.
- The level of differential outcomes on the centralised assessments was, overall, found to be in line with other BPTC modules for the demographic variables analysed. However, compared to other modules, the centralised assessments were linked with a higher level of differential outcomes related to academic history, as defined by the variables of first degree classification and university attended.
- The differences by ethnicity in differential outcomes between the centralised assessments and other BPTC modules were broadly similar. However, as the centrally assessed modules were more difficult to pass on average, the differences in outcomes by ethnicity had a larger impact on pass rates for the centralised assessments than for other modules - with higher proportions of students from minority ethnic backgrounds failing to pass the centralised examinations than other modules on the course.

Introduction

About the Bar Standards Board

1. The Bar Standards Board (BSB) is the regulator of barristers in England and Wales. The Legal Services Board oversees our activities. The BSB is a risk-based, transparent and proportionate regulator, targeting our work at the areas of most need in relation to our regulatory objectives. The BSB Handbook sets out the standards that the BSB requires the persons it regulates to comply with in order for the BSB to be able to meet our regulatory objectives. Among other responsibilities, the BSB is responsible for setting the education and training requirements for becoming a barrister.[1]

Background to the research

2. To become a practising barrister in England and Wales prospective barristers are required to pass three stages of training: the academic, vocational, and professional components.

3. The vocational stage of training from 2011/12 to 2019/20 was the Bar Professional Training Course (BPTC). It was designed to ensure that students acquired the skills, knowledge of procedure and evidence, attitudes, and competence to prepare them for the more specialised training of pupillage. From September 2020, vocational training for the Bar now consists of courses approved by the Bar Standards Board (BSB) following the Future Bar Training (FBT) programme of reforms.

4. This research focuses on modular results attained on the BPTC, with a particular focus on the centralised assessments (the three modules with examinations set centrally by the BSB). It was undertaken following previous research on differential outcomes by demographic groups on the BPTC, and is driven by the BSB's commitment and duty to ensure fairness.

5. The previous research on differential outcomes[2] suggested that, even once other variables are controlled for, ethnicity had a significant predictive value for BPTC average module scores, and that ethnicity and socio-economic status were both found to have a significant predictive value for success at obtaining pupillage. The predictive value of ethnicity was larger for the modules centrally assessed by the BSB than for other modules, with minority ethnic background students scoring lower than White students.

The structure of the BPTC

6. The modules on the BPTC could broadly be divided into those with 'knowledge based' assessments, and those with 'skills based' assessments.[3]

7. Knowledge areas:

   - Civil Litigation and evidence (hereafter referred to as 'Civil Litigation ' or more simply 'Civil')
   - Criminal Litigation, evidence and sentencing (hereafter referred to as 'Criminal Litigation' or more simply 'Criminal')
   - Professional ethics

8. Skills areas:

   - Advocacy (Advocacy 1, Advocacy 2 and Advocacy 3 modules)
   - Opinion writing

1 See https://www.barstandardsboard.org.uk/about-us/what-we-do.html for more on the BSB's responsibilities
2 The Bar Standards Board (2017). Exploring differential attainment at BPTC and Pupillage: A quantitative study. Accessed online here
3 In addition, there were two mandatory options modules which differed by provider.

- Drafting
- Conference skills
- Resolution of disputes out of court (ReDoC)

9. There were two sittings each year in each of the modules, a spring, and a summer sit, with those sitting in summer generally being those who were not successful in the spring sit, or those unable to take the spring sitting. The pass mark for all modules was set at 60 per cent - for the centralised assessments, standard setting was undertaken and the passing standards determined by this were translated to 60 per cent in order to fit with providers' student records systems.

10. From the 2016/17 academic year, candidates were allowed a maximum of three attempts at an assessment on the BPTC (not including where there were extenuating circumstances). Previously, candidates were allowed a maximum of two attempts at each assessment.

11. The Bar training courses provided from 2020/21 include new modules but the skills needed by prospective barristers are largely based around those assessed on the BPTC[4] and many newer assessments have much in common with those on the BPTC. This includes the centralised assessments, the passing of which remains a requirement for those studying to become a barrister in England & Wales.

The Centralised Assessments

12. Civil Litigation, Criminal Litigation and Professional Ethics have served, and still serve an important purpose in ensuring consistency across providers of vocational training for the Bar, and closer regulatory oversight of standards in knowledge subjects for prospective barristers.

13. The examinations for these modules are set on behalf of the BSB by the Central Examinations Board (CEB), which consists of a group of senior examiners, including experienced legal practitioners and academics. To determine the pass mark of the centralised examination papers, the BSB conducts standard setting. Those scoring below the pass mark after the standard setting will need to re-take the centralised examination they failed.

14. There have been three different formats employed for the centralised assessments since they were first implemented in 2011/12. These are as follows:

2011/12-2015/16 (older format)

- From 2011/12-2015/16 all three examinations consisted of both Multiple Choice Questions (MCQs) and Short Answer Questions (SAQs). Students taking the old format examinations had to pass both separate parts of the exam to pass the module overall. The SAQ part of each assessment was marked and moderated by the course providers.

2016/17-2019/20 (newer format)

- Civil Litigation and Criminal Litigation: An examination of 75 MCQs in each subject. Machine marked by the BSB.
- Professional Ethics: An examination consisting of six SAQs (each consisting of two sub-parts). Marked by markers contracted by the BSB.

15. From 2020/21 a new format for assessment of the centralised assessments was introduced for Civil Litigation. Criminal Litigation retained the 75 question MCQ format. The centralised assessment in Professional Ethics was moved to the pupillage stage of training but Authorised Education Training Organisations (AETOs) are also expected to conduct their own assessment of Ethics during the

4 The Professional Statement describes the knowledge, skills and attributes that all barristers should have on "day one" of practice - https://www.barstandardsboard.org.uk/training-qualification/the-professional-statement.html

vocational stage.[5]

Previous research

16. In 2017 the BSB published research on differential outcomes on the BPTC.[6] The research included analysis of differential outcomes on the centralised assessments and other modules. The analysis found that the most significant factors in predicting student scores on the three centrally assessed modules (aggregated together) were first-degree classification and first-degree institution attended. It was also found that ethnicity was a significant predictor of centralised assessment module score, as was domicile and parental degree status.

17. In summary, the previous research found that on the centralised assessments, when controlling for the variables in the model:

- **Domicile –** Overseas domiciled students scored 1.8 points higher than UK or EU domiciled students.
- **Ethnicity –** Students from a minority ethnic background scored 4.7 points lower than White students.
- **First degree class -** Those with a first class or upper second class degree scored 11.3, and 6.5 points higher than those with a lower second class degree respectively.
- **First degree institution attended**
  - **Oxbridge –** Those who attended to Oxford or Cambridge University scored 10.9 points higher than those who attended non-Oxbridge or Russell Group universities.
  - **Russell Group -** students who went to Russell Group universities scored 5.6 points higher compared to those who attended universities outside Oxbridge or non-Oxbridge Russell Group universities.
- **Parental Degree -** students with at least one degree-educated parent scored 1.6 points higher than students with no degree-educated parents.

18. Two other models were developed to better understand the relationships between student characteristics and results on advocacy and on other modules. It was found that ethnicity, domicile, parental degree, study mode, degree class and degree institution were all significant predictors of advocacy results; and that ethnicity, degree class and degree institution were significant predictors of results in other modules.

19. The data used for the previous research on differential outcomes covered BPTC students enrolled between 2013/14 and 2015/16, meaning that results on the centralised examinations under the newer format exams from 2016/17 to 2019/20 were not included in the analysis.

This research

20. The BSB is committed to ensuring that access to training for the Bar is open to all equally. One of the stated aims is “Improving accessibility – so that the best candidates are able to train as barristers and that the Bar reflects the communities it serves’. In order to fulfil this aim, it is important that we monitor the results from the centralised assessments, and other assessments that are set and marked by the vocational AETOs, in order to highlight any trends present. We have a particular interest in exploring whether the centralised assessments have a disproportionate impact on different groups sitting them, as these examinations are set on behalf of the BSB, and are a key component of maintaining standards across providers of vocational training for the Bar.

21. This research involves analysis of differential outcomes on the centralised assessments from 2013/14

5 Details for the newer format can be found in Part 3D of the Bar Qualification Manual, found here.

6 The Bar Standards Board (2017). Exploring differential attainment at BPTC and Pupillage: A quantitative study. Accessed online here

to 2019/20. This is undertaken to better understand how the relationship between different variables and scores on the centralised assessments may have changed over time, and as a result of reforms to the assessments in 2016/17. It also involves analysis of differential outcomes on other modules on the BPTC in order to better understand how the centralised assessments compare to other modules in this regard. We aim to provide insight on the subject to the BSB, providers of Bar training, and others with an interest in the area.

22. This research involves an exploratory analysis, and the development of statistical models that can give us an estimate of the relationship between given characteristics and results on the centralised assessments; and between given characteristics and results on other modules.

Aims and objectives

23. The aims of this research are to:

- Compare differential outcomes on the centralised assessments under the new system brought in for the 2017 sits with the previous system in place until 2015/16.
- Look into differential outcomes on other modules on the BPTC and highlight any trends that may have occurred and how other modules compared to the centralised assessments.
- Report on other trends that may be of note following analysis of the data.

24. As a result of the above:

- To produce a report which can help with comparisons against future results in Bar training regarding differential outcomes on the centralised assessments and other modules.
- To provide evidence for ongoing monitoring and research on Bar Training for the BSB.

Methodology

Data

25. The dataset used consisted of data held by the BSB on student characteristics and performance on the BPTC. This included data submitted by those taking the BCAT along with student data provided by BPTC providers. The module results looked at were those of the ten BPTC modules that were the same between providers (the options modules were not included).

26. As already noted, in most years there were two sits for each module, with the second sits including more of those who had failed at least one attempt at the relevant module. As there were far fewer students at such sits, particularly those sitting the respective module for the first time, the second sits in each year were not included as part of the analysis. This was also due to the decision only to include data in the regression models for those taking their first sit at a module, in order to avoid having 'repeated measures' in the dataset used for the regression models. Given the sample size for the first sittings each year was large enough, this was the most straightforward approach.[7]

27. For the spring sits from 2014 to 2019 and for Summer 2020 there were around 1200-1500 students sitting each module for the first time.

28. The sittings that were part of the analysis were:

- Spring 2014, 2015, 2016, 2017, 2018, 2019; and
- Summer 2020 (this was the first sit of 2020).

7 Including repeated measures in a regression model can lead to errors in p-values that can indicate statistical significance inaccurately.

29. The format for the Spring 2014, 2015 and 2016 sits was of the old assessment system previously described (half MCQs and half SAQs). The format for the Spring 2017, 2018, 2019, and Summer 2020 sits was of the newer system. The 2012 and 2013 sits were not analysed, as the data collected when candidates sat the BCAT for these modules was missing (the BCAT started in April 2013 and ended in July 2022), and these missing data related to demographic variables that were used in the regression models detailed later in this report.

30. It should be noted that the majority of those taking the Summer 2020 sit for the centralised assessments did so online, owing to the Coronavirus pandemic. This makes the environment for this sit an outlier for most of those taking it. However, pass rates and module scores for candidates completing the exams were broadly in line with previous years/trends.

Variables of interest

31. The models used the following independent variables:

- **Age range** - Under 25; 25 to 34; 35 to 44; Over 45
- **Disability status** - Disability declared; No disability declared
- **Domicile** – UK/EU or Overseas
- **English as a first language -** English as a first language; Another language as a first language
- **Ethnicity** - Asian/Asian British; Black/Black British; Mixed/Multiple Ethnic Groups; White
- **First degree classification** – First class; Upper second class; Lower second class
- **Gender –** Female; Male
- **Mode of BPTC study –** Full time; Part time
- **Parental Degree** – Whether at least one parent/guardian had a degree: Yes; No
- **Type of school mainly attended between ages 11-18** - Fee-paying school; State school
- **Type of university attended** – Oxbridge; Russell Group (not including Oxbridge universities);[8] Other UK based university; Overseas university

32. The inclusion of variables with a substantial amount of missing data can result in errors in regression modelling and unrepresentative statistics of the 'true population' if the data are not missing in a random manner : For example, if those who did not provide information on degree classification were disproportionately males with a lower second class degree classification. In general, the data for most of the variables we looked at were almost 'complete'. There were a few exceptions to this.

33. The variables with a substantial proportion of missing data (over 10% for some sittings) were those for the classification of university attended (Russell Group/Oxbridge etc), the type of school attended, and whether at least one parent/guardian attended university. There was also a substantial amount of missing data for English as a first language for 2014.

34. These variables were cross-tabulated with others in the dataset in order to see whether the proportion of missing data were evenly spread across categories. This was indeed the case, and suggested that the data were missing at random. This meant that running a regression analysis that removed 'non-complete cases' (instances where a student had missing data in at least one category) could be undertaken.

8 The Russell Group of universities includes the Oxbridge universities. As the Oxbridge universities are presented as a separate category, the Russell Group category in this research does not include Oxbridge.

Overall approach for analysis of the data

Descriptive statistics

35. For the part of the analysis focused on descriptive statistics, mean score along with the interquartile range on each sit for each module, were analysed and are presented. The overall percentage failing each sit at each module is also provided, in Table 1.

36. Charts 1-7 and Table 1 only include statistics on those sitting the module for the first time. Data are only presented in cases where there are 30 or more students in the respective group (eg, there are 30 or more females with an upper second class degree sitting Professional Ethics in Spring 2014 etc). This is to avoid small sample sizes in the groups that are presented, as the smaller the group the more variable the results for the group.[9]

Regression analysis

37. For the part of this analysis focused on statistical modelling, this research made use of multiple linear regression analysis in order to enable analysis of different factors that may impact on module score to be considered simultaneously. This ensures the analysis can identify which factors have the strongest relationship with the outcome being analysed.

38. Regression analysis models the size of predictive relationships between one or more explanatory variable(s) and a single outcome variable. It provides an estimate of the size of and statistical significance of the modelled relationships, while controlling for the effects of other explanatory variables in the model.[10]

39. It should be noted that the sizes of the predictive relationships identified are statistical estimates and thus may be over or underestimated.

40. Regression model outputs are relatively easy to understand and make inferences from, if the key assumptions of the framework for undertaking such analyses are met. Multiple linear regression was the methodology followed in the previous research, and it was decided that this would also be the most appropriate approach taken for this analysis.

41. The format of the regression models used needed to be based around how best to answer the research questions and remain valid from a statistical point of view.

42. The principal questions to answer for this part of the research were:

- How have differential outcomes on the centralised assessments changed, particularly between the assessment format from 2012 to 2016, and that in place from 2017 to 2020?
- How do the centralised assessments compare to other modules in terms of differential outcomes?

43. **For the first question**, the analysis required a model that showed how things had changed over time in the relationship between module score on the centralised assessments and other variables. It is not a valid approach to run different models for different sittings and compare coefficients for the regression models between them, particularly if we want to understand the significance of any change over time. It is possible instead to run a model for each module for the entire dataset, including a variable for sitting, and interaction terms between this and variables that appear to show the strongest relationship with sitting.

9 For example, each individual in a group of 10 would represent 10% of the results for that group, whereas each individual in a group of 30 would represent around 3.33%. Resources detailing the Central Limit Theorem can provide greater elaboration on this topic as well.

10 An outcome variable is the variable we are interested in better understanding what influences the different outcomes/values of it. Explanatory variables are those variables we propose may influence the value of the outcome variable, and we undertake tests to determine whether this is the case.

44. Interaction terms are used in regression models for categorical variables when there is strong evidence that there is some interaction between two or more of the independent variables used in the model: eg the predicted effect of gender and sitting on score can be better modelled by accounting for an interaction that the variables of gender and sitting may have with one another (for example , there may be significantly different outcomes occurring for males sitting Professional Ethics in 2016 in comparison to males sitting Professional Ethics in 2014).

45. Another model for each centrally assessed module was also run, which followed the same methodology, but with a grouped sitting (2014-2016, 2017-2020) variable that was interacted with other variables instead of individual sit. This was aimed at presenting a simplified picture of before/after reform to the assessments.

46. To answer the second question, the analysis considered the two more recent first sittings of each year in the dataset, Spring 2019 and Summer 2020. The data for these were grouped together, and a multiple regression model was developed using module as an interaction variable with the other demographic variables, and module score as an outcome variable. This enables us to better understand how modules compare to one another in any additional relationship they may show with differential outcomes.

47. For this question, we are mostly interested in the level of differential outcomes on each module compared to other modules, and not differences in overall module mean scores. To undertake this piece of the analysis, the final result score was standardised to the mean for each sit for each module (this involves subtracting the mean value for the sit for the module from the individual students' score on the sitting for the module).

48. There were some repeated measures in this dataset, as most students in the dataset sat more than one module during the period. To better account for this, students who sat fewer than eight modules for the first time across the Spring 2019 and Summer 2020 sits were removed from the dataset (leaving clusters with more observations), and the p-values for the model were estimated in a different way compared to the other regression models. The p-values resulted from calculating 'robust standard errors' using the 'sandwich' package in R.

49. The stats package R was used for the analysis.

Model selection

50. It was decided that the simplest approach would be to use the same key variables for all modules, and across all sits for all models. There were more variables to choose from than were available for in the previous research, owing to more complete data this time around.

51. Inclusion of interaction terms. To start with, all except for degree institution attended[11] were interacted with sitting for the first regression model, and all variables were interacted with module for the second regression model. Interaction terms were included in the final models[12] if the variable for the interaction between the demographic variable and the interaction variable had p-value below 0.10 on the first iteration[13] of the model, and below 0.05 on the second iteration or resulted in a significant reduction in the sum of squares of the model when compared to the first iteration of the model (as determined by an ANOVA for model results).

Interpretation of regression models

52. Where differences or variables are described as 'statistically significant', this indicates that they have been found to be statistically significant at the 5 per cent significance level or below (the standard sig-

11 The inclusion of degree institution as an interaction term made for difficult to interpret terms, and did not significantly improve the model or help better understand the key variables of interest in this analysis.

12 Final model' refers to the models from which the results presented in this report are taken. Their development was an iterative process, as laid out in this section.

13 On the first run of the model.

nificance level for social research), as indicated by a p-value. A significance below 5 per cent would suggest there is less than a 5 per cent likelihood that the relationship observed between two or more variables can be explained by chance alone, given the data.

53. Regression models are suited to modelling uncertainty in processes where there is variability in some outcome.[14] The outputs of regression models include estimates of the effect of each explanatory variable on an outcome variable: These are known as **main effects, or model coefficients**. In the case of categorical variables (variables which have two or more categories[15]) they indicate the predicted effect of a category on some outcome in comparison to a **reference group**. This predicted effect is independent of other variables in the model (eg males in comparison to females, independent of the effect of height etc).

54. The model coefficients presented all indicate the predicted average effect of the presence of a variable in relation to a reference group on module score. For example, a main effect/coefficient of 5.0 for Group 2 in comparison to a reference level of Group 1 would suggest that those in Group 2 score around 5.0 points higher than those in Group 1 on average when holding all other variables constant. A p-value of less than 0.05, would additionally indicate that, based on the data, there is a less than 5 per cent chance that the association indicated by the coefficient is due to chance alone.

55. The **intercept variable** is the predicted outcome for a case where all of the variables in the model are at their reference level. However, the inclusion of **interaction terms** makes regression models more difficult to interpret. Including interaction terms results in three different types of coefficients in the model; the **intercept**, the **main effects**, and the **interaction effects**. We have aimed to present these in a way that is aimed at being relatively easy to interpret in the main body of the report. Further information on how to interpret interactions in the model is given in the appendices.

56. It should be noted that the resulting coefficients from the models are estimates only. Full model summaries are provided in a spreadsheet made available alongside this report.[16]

Limitations

57. For the regression modelling results, coefficients are only reported in the main body of this report where the variables were statistically significant, and based on the data, would appear to be those most strongly associated with module score. In the interests of transparency, this research did not identify sparser models that excluded non-significant predictors.

58. Regression models offer a statistical estimate of the relationships between variables based on the data available. They are a simplification of reality and the numerous factors that in some way relate to module score. There is much variability in module score unaccounted for by the models, and so uncertainty in the models needs to be considered when assessing the results.

14 The measures of uncertainty in the regression models (such as p-values and confidence intervals) largely relate to the variation in the data caused by such factors.

15 For example, a variable of age range with categories of 18-24, 25-34 etc, would be a categorical variable.

16 Which can be found next to the report at https://www.barstandardsboard.org.uk/news-publications/research-and-statistics/bsb-research-reports.html

Results

Descriptive statistics for all modules

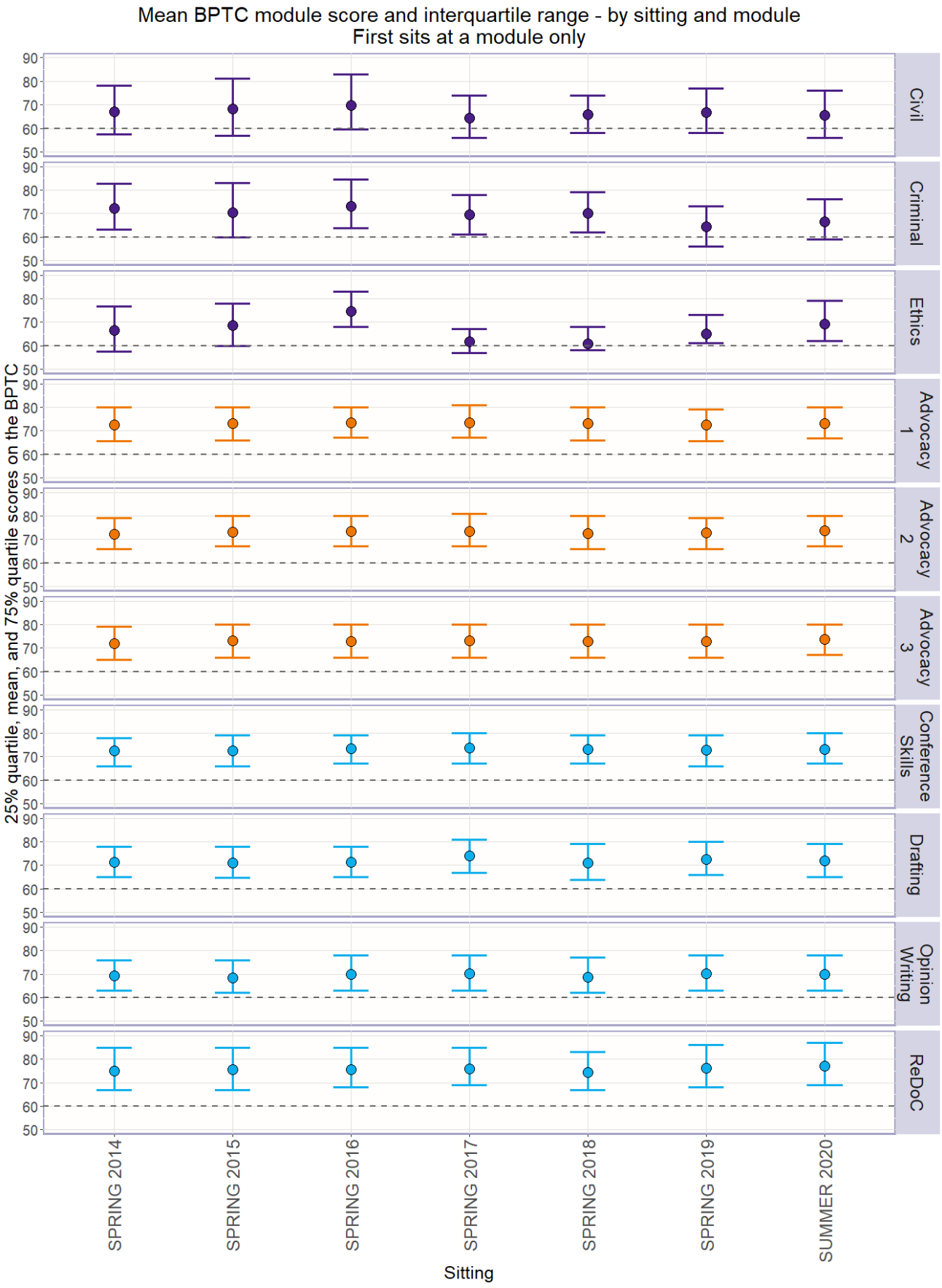
59. The below series of charts show how the centralised assessments compared to other modules on the BPTC (not including the options modules). Represented on each one is the mean score and interquartile range across sittings for the modules for the relevant group depending on the chart. The figures related to those sitting the module for the first time only, as otherwise the scores may be biased by some modules having more second and third sitters than others, and this skewing the score downwards. Results for a group are only presented where there are 30 or more students in the group.

60. One of the major trends seen across the charts is that it is generally the case that the mean scores seen on the centrally assessed modules are lower than that seen in other modules. There also appears to have been a drop in mean scores seen for the centrally assessed modules following the introduction of the newer format examinations in 2017. The mean score for the centralised assessments does also appear to vary more widely between years than that seen for other BPTC modules.

Overall

61. Chart 1 below shows the overall picture from Spring 2014 to Summer 2020 sittings in terms of the mean score and interquartile range across BPTC modules for those sitting each module for the first time. The interquartile range consists of the range between the 25th and 75th percentiles, meaning 25 per cent of candidates score lower than the lowest score in this range, and 25 per cent of candidates score above the highest score in this range.

Chart 1. Mean BPTC module score and interquartile range (IQR) – by sitting and module – First sits only

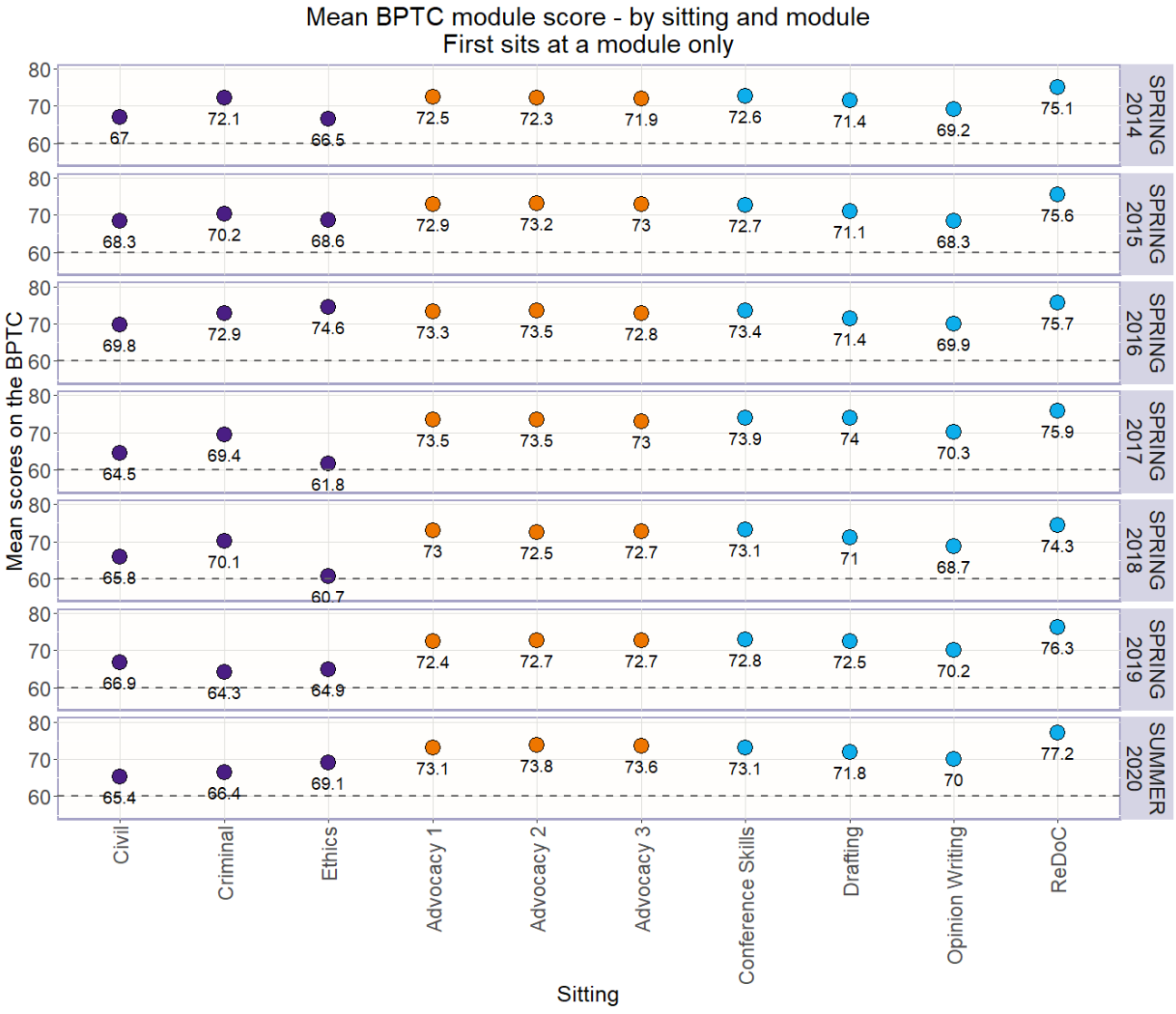


62. **Non-centrally assessed modules -** The mean score and interquartile ranges for non-centrally assessed modules do not appear noticeably to change between years. The mean score for Opinion Writing generally seems to be slightly lower than that seen for the other non-centrally assessed modules, and that seen for ReDoC is a bit higher.

63. **Centralised assessments –** The centralised assessments are the only modules for which the 25th percentile overall is below the pass mark of 60 across at any sit. This is seen for Civil Litigation for all sittings, Criminal Litigation for Spring 2019 and Summer 2020, and Professional Ethics for Spring 2014, 2017 and 2018. The mean score and interquartile ranges for the centrally assessed modules have changed more over time than that seen for other BPTC modules. This is highlighted in the chart below.

64. The overall mean scores on the centralised assessments under the older format assessments (Spring 2014- 2016) could be said to be more in line with that of the other BPTC modules, albeit with a wider interquartile range. Spring 2017 onwards sees more divergence in mean score between the centralised assessments and other BPTC modules, particularly for Civil and Professional Ethics from 2017, and Criminal Litigation from 2019.

Chart 2. Mean BPTC score by sitting and module – First sits only



65. The above does not tell the entire story in terms of the proportion not passing each assessment however, as under the assessment format in place prior to Spring 2017, students needed to score above the pass mark in both the MCQ and SAQ sections of the examinations. The below table shows the proportion of students sitting the module for the first time who did not pass by sit.

Table 1. Proportion of students sitting each module for the first time who did not pass by sit.

Module	SPRING 2014	SPRING 2015	SPRING 2016	SPRING 2017	SPRING 2018	SPRING 2019	SUMMER 2020
Civil	30%	39%	33%	34%	30%	32%	31%
Criminal	19%	34%	25%	19%	19%	34%	27%
Ethics	29%	41%	26%	37%	27%	21%	16%
Advocacy 1	8%	7%	5%	5%	6%	7%	4%
Advocacy 2	6%	5%	4%	6%	7%	5%	3%
Advocacy 3	6%	6%	6%	6%	7%	6%	5%
Conference Skills	4%	4%	3%	4%	4%	5%	4%
Drafting	7%	11%	10%	6%	11%	7%	8%
Opinion Writing	13%	16%	13%	11%	15%	13%	11%
ReDoC	13%	10%	10%	9%	11%	9%	7%

66. As can be seen, the introduction of the newer format assessments did not lead to a notable change in failure rates for Civil or Criminal Litigation in the year they were introduced. There was an increase seen for Ethics, but a subsequent reduction in the following years. This may be related to the way Ethics was taught taking time to catch up with the requirements of the new format exam – more teaching time became allocated for the teaching of Ethics from 2018 onwards. There was also an increase in the failure rate seen for Criminal Litigation in Spring 2019 compared to Spring 2018 – this may partly be related to a push to remove questions from our question bank that were less effective at discerning between low and high ability candidates.

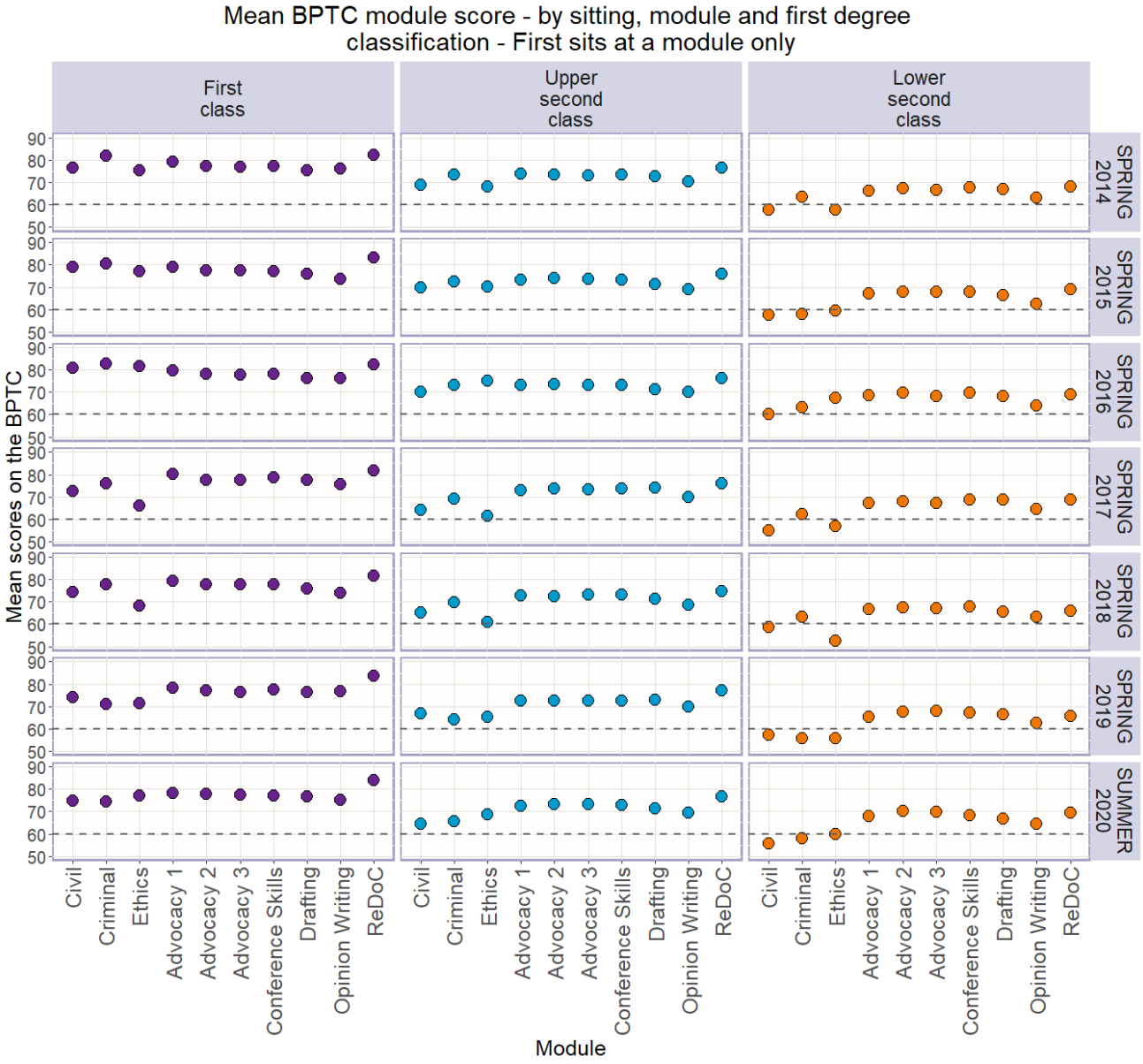
67. In relation to the percentage not passing the assessment at the first attempt for each of the centralised assessments, Civil shows the most stable trend over time, while the percentage figure has fluctuated more for Criminal Litigation and Professional Ethics. The proportion of those not passing the centrally assessed modules at the first attempt in the sittings given above is notably higher than that seen for other modules, suggesting that the centralised assessments were, overall, the most difficult to pass assessments on the BPTC.

Degree classification

68. When further disaggregating results from 2014 to 2020 for those sitting the relevant module for the first time by first degree classification, the effect of the introduction of the newer format centralised assessments can again be seen. The relationship between first degree classification and module scores is also shown.

69. Results for those with a first class degree or upper second class degree on the centralised assessments were broadly in line with those seen across other BPTC modules from 2014 to 2016. Upon the introduction of the newer format assessments, there was a drop in mean score across all three centralised assessments, particularly on Professional Ethics. For those with a first class degree, results for the centralised assessments for Summer 2020 looked more in line with that seen on the other BPTC modules once again.

Chart 3. Mean BPTC module score – by sitting, module and degree classification – First sits only

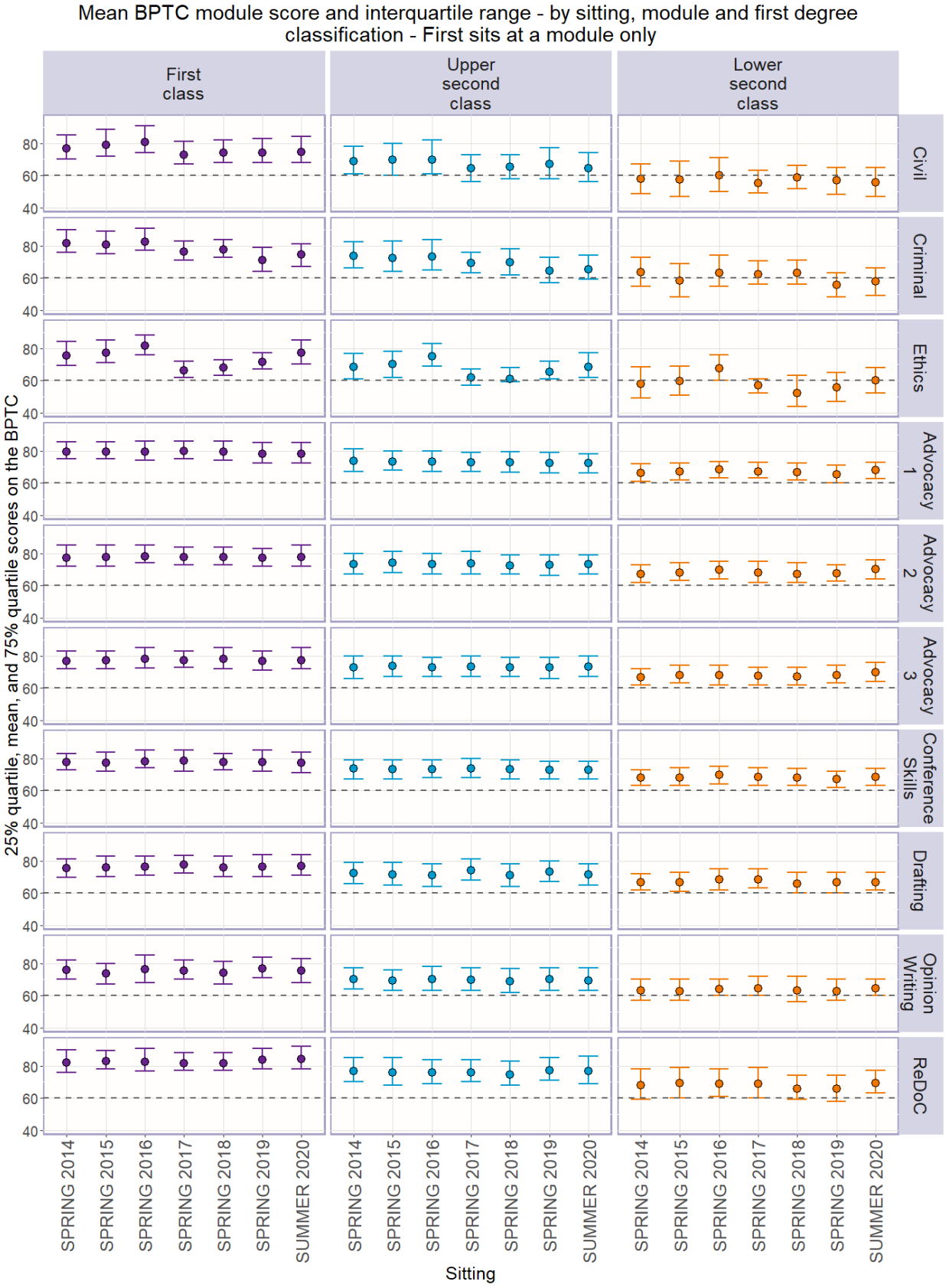


70. For those with a lower second class degree, the introduction of the newer format assessments in 2017 had slightly less of a clear impact for those sitting the centrally assessed modules for the first time, although perhaps did lead to a slight reduction in mean score in more sittings than not. The mean score for those with a lower second class degree has been below the pass mark of 60 for the following sittings and modules:

- **Civil** – Spring 2014, 2015, 2017, 2018, 2019 and Summer 2020
- **Criminal** – Spring 2015, 2019 and Summer 2020
- **Professional Ethics** - Spring 2014, 2015, 2017, 2018 and 2019

71. Trends in scores by sitting for each module can also be seen in the chart below, which more clearly highlights the drop in scores following the introduction of the newer format assessments for the centrally assessed modules, and a more stable trend for other modules.

Chart 4. Mean BPTC module score and IQR – by sitting, module and first degree classification – First sits only



72. Upon the introduction of the newer format assessments, the 25th percentile for those with an upper second-class degree was lower than the pass mark for sittings for Civil for every year from 2017 on-wards, for Criminal for 2019 and 2020, and for Professional Ethics for 2017 and 2018. This is not the case for any of the other modules.

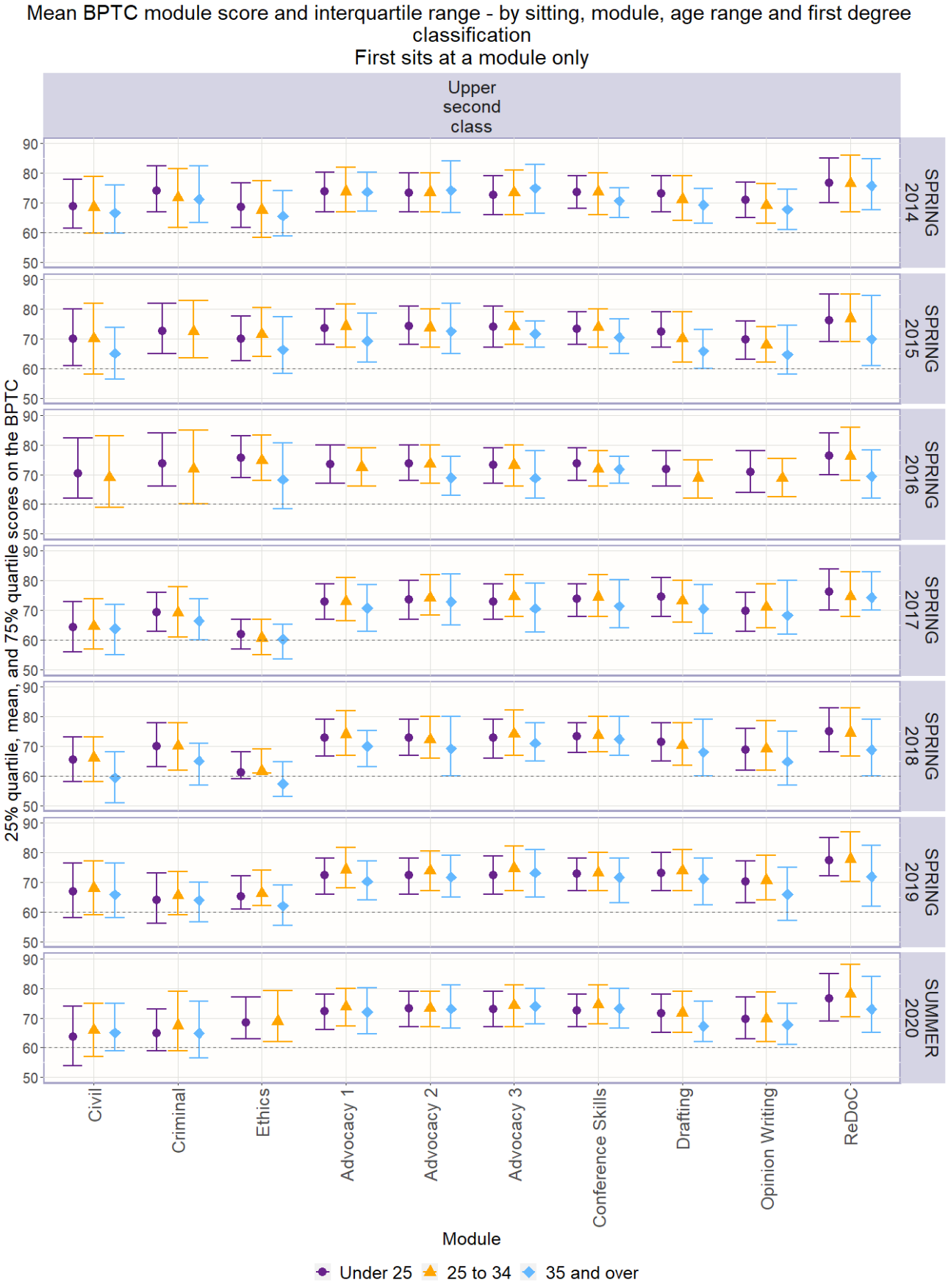
73. In addition, on the other BPTC modules, for no sit is the mean score for those with a lower second class degree below the pass mark of 60. For Opinion Writing the 25th percentile with this degree classification scored below the pass mark for Spring 2014, 2015, 2018, and 2019.

74. The above would suggest that the centralised assessments were more difficult to pass than the non-centrally assessed modules for those with an upper second class degree and lower second class degree across the sittings analysed. This particularly appears to be the case following the introduction of the newer format assessments. The same can be said for those with a first class degree in some sittings.

Age

75. For age, it is only for those with an upper second class degree where there are enough candidates aged over 35 sitting assessments each year that we are able to have them as a comparative group. It appears that as a group, those aged over 35, in comparison to those aged under 25 and 25-34 may be more likely to have a slightly lower mean score for more sittings than not across modules. There is no clear trend in terms of difference in results between those aged under 25 and those aged 25-34. This is highlighted in the below chart.

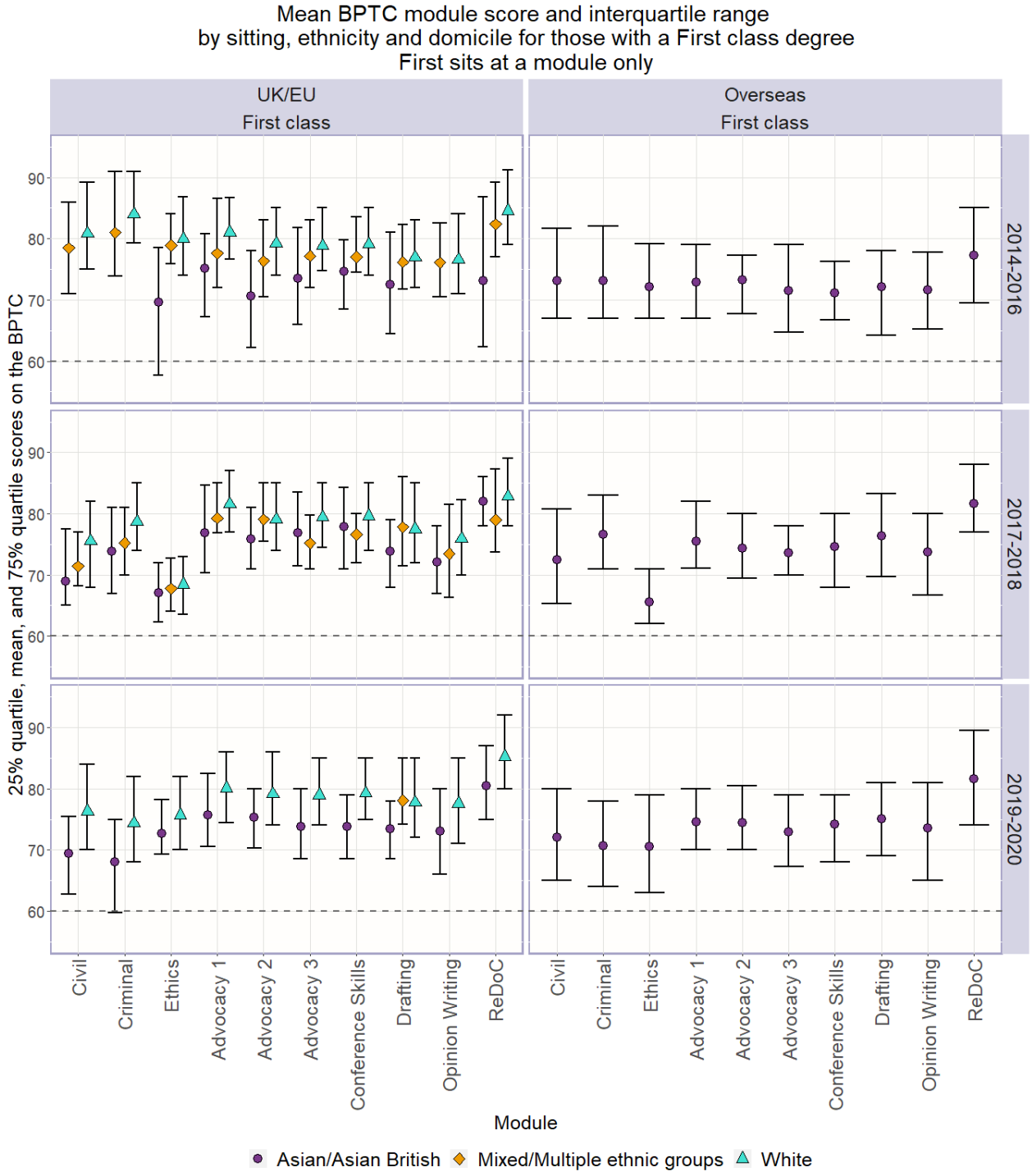
Chart 5. Mean BPTC module score and IQR – by sitting, module, age range and degree classification – first sits only



Ethnicity and domicile

76. When looking into the relationship between ethnicity and results by module it is useful to further disaggregate by ethnicity, as those who are domiciled outside of the UK prior to enrolment may display different patterns of results from those domiciled in the UK, and are also disproportionately from minority ethnic backgrounds. It is also useful to group sittings together, as the numbers in each ethnic and domicile group can be quite small at each individual sitting.

Chart 6. Mean BPTC module score and IQR by grouped sitting and domicile for those with a first class degree - first sits only[17]

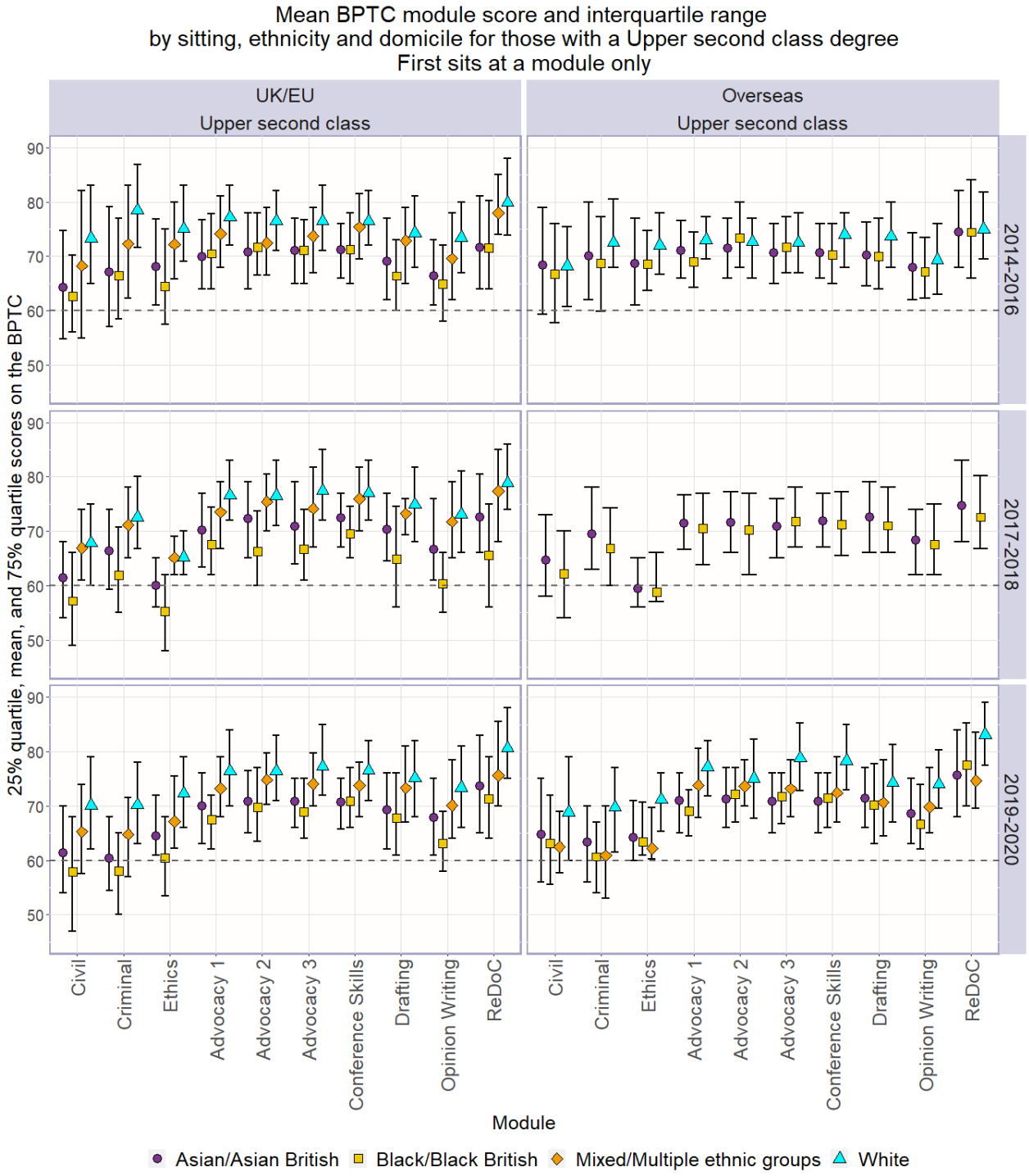


77. Chart 6 above shows the mean scores and interquartile range for the spring sittings from 2014-2019 as well as Summer 2020 by domicile and ethnic group for those with a first class degree. Chart 7

17 As with all of Charts 1-7 and Table 1, results are only presented for a group where there are more than 30 students in a group. This is why the results for Black/Black British students are not in Chart 6.

below that shows the picture for those with an upper second class degree.

Chart 7. Mean BPTC module score and IQR by grouped sitting and domicile for those with a upper second class degree - first sits only



78. Results suggest that White students as a group achieve higher mean scores across modules when controlling for degree class and domicile. Differences by ethnic group appear to be relatively consistent across modules, but the lower average scores seen for the centrally assessed modules mean that in some cases, the mean score for those from a minority ethnic group with an upper second class degree is below, or very close to the pass mark.

Gender and disability status

79. When further disaggregating by gender, no clear further trends emerge: The mean scores and interquartile range are very similar across all modules and sittings by gender. The same can be said when

further disaggregating by disability status for those with a declared disability in comparison to those with no declared disability. The charts for both are given in the appendices.

Regression analysis: Change in differential outcomes on the centralised assessments

Civil Litigation

80. While the charts presented in the previous section can show us some broad trends, and differences across modules, they cannot tell us whether the trends seen remain when controlling for several other variables. For example, if female and male students have similar results across modules when controlling for ethnicity, type of university attended and study mode. This is the type of question that regression modelling is better suited to addressing.

81. The final model used for Civil Litigation included the standard variables used across all of the regression models for the centralised assessments and interaction terms for sitting with both ethnic group and first-degree classification.[18] This suggests that for Civil Litigation there is a statistically significant relationship between sitting and ethnic group, and between sitting and first-degree classification, of which more is detailed further on.

82. Other demographic variables did not show a significant interaction with sitting, suggesting no significant change over time in differential outcomes for the following variables:

- Gender; disability status; domicile; type of school attended; whether a parent/guardian attended university; and whether English was a student's first language.

83. Of the above variables, gender, disability status and parental degree status had a statistically significant main effect, suggesting some relationship with these variables and Civil Litigation score when controlling for the other variables in the model (these relationships are discussed in more detail below). However, type of school attended and English as a first language did not show significant effects, suggesting there was no relationship between these variables and Civil Litigation score once other variables are controlled for.

Statistically significant variables

Variables not interacted with sitting

84. There were several statistically significant variables that did not display a statistically significant change over different sittings. The following variables had statistically significant main effects/coefficients when controlling for the other variables in the model, but were not included with interaction terms with sitting:

- **Age:** Compared to those aged under 25:
  - **25-34** – Those aged 25-34 taking Civil Litigation on average would be predicted by the model to score 0.9 points higher on the assessment (p-value = 0.007).
  - **45 and older** – Those aged 45 and older taking Civil Litigation on average would be predicted to score 2.8 points lower on the assessment (p-value = 0.021).
- **Disability status** – Compared to those with no declared disability, those with a disability declared were predicted to score around 1.0 point lower on Civil Litigation (p-value = 0.043)

18 As noted in paragraph 51, interaction terms were included in the final models if the variable for the interaction between the demographic variable and the interaction variable had p-value below 0.10 on the first iteration of the model, and below 0.05 on the second iteration or resulted in a significant reduction in the sum of squares of the model when compared to the first iteration of the model (as determined by an ANOVA for model results).

- **Domicile** – Compared to those domiciled in the UK/EU, those domiciled overseas were predicted to score around 1.4 points higher on Civil Litigation (p-value < 0.001)
- **Gender –** Gender was a statistically significant variable but had a relatively small main effect. The model results suggested that male students on average score 0.9 points higher on Civil Litigation compared to female students (p-value = 0.002).
- **Parents' university status** – Those who did not have at least one parent attend university were predicted to score around 1.0 point lower on Civil Litigation on average compared to those who had at least one parent who attended university (p-value < 0.001).
- **University attended** – Compared to those who attended 'non-Oxbridge/Russell Group UK universities'':
    - those who attended an Oxbridge university were predicted to score 14.4 points higher; those who attended a non-Oxbridge Russell Group university were predicted to score 8.2 points higher; and those who attended a university overseas were predicted to score 7.1 points higher on Civil Litigation (p-values all < 0.001).
- **BPTC study mode** – Part-time students were predicted to score around 1.2 points higher on Civil Litigation in comparison to those studying the course full-time (p-value = 0.030).

Variables whose effect is related to sitting

85. The variables of ethnic group and first-degree classification both had statistically significant main effects, and statistically significant interaction terms with sitting. This indicates that these variables had a significant predictive relationship with exam results when controlling for other factors, and that the size of this relationship varied across sittings. Interpretation of interacted variables can be more complex, owing to the nature of the interaction terms.[19]

86. **Ethnicity** – In general, those from Asian/Asian British, Black/Black British, Mixed Multiple, and other ethnic backgrounds were predicted to perform worse on Civil Litigation than White students, a trend consistent with that seen in the previous research on differential outcomes.

- There were only two statistically significant interactions between ethnicity and sitting (Asian/Asian British Spring 2018; and Other Ethnic Group Spring 2019). In both cases, the interaction term was a positive value, suggesting a better performance than that seen in 2014 for the respective ethnic group relative to White students.
- Therefore, the results suggest that while there had been some variation year on year, there was not an overall pattern of differential outcomes by ethnic group widening following the introduction of the newer format assessments

87. The below table shows the predicted differences between White students and those from other ethnic groups by sitting. The table also indicates where these differences by sit represent a statistically significant change from Spring 2014.

19 It is more difficult to present results for interacted variables, as previously explained in the methodology section. The results for variables that were interacted are explained in more general terms.

Table 2. Civil Litigation - Predicted difference in module score by ethnicity compared to White students for each sitting

Sit	Predicted difference in score to White students for sitting by ethnic group			
	Asian/Asian British	Black/Black British	Mixed/ Multiple ethnic groups	Other ethnic group
Spring 2014	-4.83	-5.08	-4.65	-7.42
Spring 2015	-4.63	-3.54	-2.28	-2.63
Spring 2016	-5.23	-9.12	-5.01	-10.51
Spring 2017	-3.51	-6.15	-1.31	-4.68
Spring 2018	-2.04	-5.48	-4.12	-6.37
Spring 2019	-4.04	-4.73	-2.21	0.41
Summer 2020	-5.06	-7.10	-1.98	-3.98

= significant interaction (p-value < 0.05)
= close to significant interaction (p-value < 0.10)

88. **First degree classification** – The below table shows predicted differences between students with a first or upper second class degree when compared to students with a lower second class degree by sitting. There was only a single statistically significant difference for degree class, with those with a first class degree scoring worse in 2015 compared to 2014 relative to those with a lower second class degree. The model results therefore suggest that there has not been a consistent statistically significant change on a sit by sit basis in terms of differential outcomes by degree class following the introduction of the newer format assessments.

Table 3. Civil Litigation - Predicted difference in module score by degree class compared to those with a lower second class degree for each sitting

Sit	Predicted difference in score to those with a lower second class degree	
	First class	Upper second class
Spring 2014	13.29	8.37
Spring 2015	17.82	10.65
Spring 2016	15.71	7.75
Spring 2017	12.64	6.43
Spring 2018	12.55	5.92
Spring 2019	13.12	7.66
Summer 2020	13.91	6.03

Civil Litigation – Before/After model

89. While the above results provide detail on statistically significant interactions for each individual sitting, to get a more aggregate picture of the trends before and after the reforms to the assessments, we can use a grouping variable for sittings to get an average effect across sittings as they are grouped. This has the disadvantage of not taking into account variance between sittings that are grouped together but can give an overall picture.

90. When using a variable for grouped years (2014-2016, 2017-2020), for Civil Litigation, statistically significant interactions were found for the same variables as for the previously described model looking at individual sittings (ethnicity and first degree classification). The significant interactions were as follows:

- **Ethnicity:** The difference in score between White students and Asian/Asian British students sitting Civil Litigation in 2017-2020 was predicted to be around 2.3 points lower on Civil Litigation

than that seen for 2014-2016.

- **First degree classification:** Compared to 2014-2016, the difference in score between those with those with a lower second class degree and those with a first class degree in 2017-2019 was predicted to be around 3.2 points lower on Civil Litigation. The difference in score between those with those with a lower second class degree and those with an upper second class degree in 2017-2019 was predicted to be around 2.8 points lower.

91. The above suggests that for Civil Litigation the reform of the assessments had a statistically significant relationship with ethnicity (for those from Asian/Asian British ethnic backgrounds) and degree class when comparing pre- and post-reform sittings overall, rather than individual years. However, given that the previous model detailed that there was significant variation between sits for the differential outcomes observed for both these variables, this change may not be related to the change in the format in the exams, but instead reflect variation in exam results over time due to other factors.

Criminal Litigation

92. The final model used for Criminal Litigation included the standard variables used across all of the regression models for the centralised assessments and interaction terms for sitting with both Ethnic group and first degree classification. This suggests that for Criminal Litigation there is a statistically significant relationship between sitting and ethnic group, and between sitting and first degree classification, of which more is detailed below.

93. Other demographic variables did not show a significant interaction with sitting, suggesting no significant change over time in differential outcomes for the following variables:

- Disability status; domicile ; gender; type of school attended; whether a parent/guardian attended university; and whether English was a student's first language.

94. Of the above variables gender and whether a parent/guardian attended university had a statistically significant main effect, suggesting some relationship with these variables and Criminal Litigation score when controlling for the other variables in the model (these relationships are discussed in more detail below). However, disability status, domicile, type of school attended, study mode, and English as a first language did not show significant effects, suggesting there was no relationship between these variables and Criminal Litigation score once other variables are controlled for.

95. An interaction between domicile and sitting was included in the model given the criteria used for model selection. However, neither the main effect term or any interaction terms for domicile and sitting were statistically significant, and inclusion of this interaction did not result in an improvement in the fit of the model compared to a model without this interaction term, and so this term was dropped for ease of model interpretation.

Statistically significant variables

Variables not interacted with sitting

96. There were several statistically significant variables that did not display a statistically significant change over different sittings. The following variables had statistically significant main effects/coefficients when controlling for the other variables in the model, but were not included with interaction terms with sitting:

- **Age -** Compared to those aged under 25, those aged 45 and older on average would be predicted by the model to score 2.9 points lower on Criminal Litigation (p-value = 0.005).
- **Gender –** Male students were predicted to score 0.6 points higher on Criminal Litigation compared to female students (p-value = 0.015). This is a relatively small main effect.
- **Parents' university status** – Those who did not have at least one parent attend university were predicted to score around 0.7 points lower on Criminal Litigation on average than those

who had at least one parent who attended university (p-value = 0.006).

- **University attended** – Compared to those who attended 'non-Oxbridge/Russell Group UK universities':
  - those who attended an Oxbridge university were predicted to score 12.3 points higher; those who attended a non-Oxbridge Russell Group university were predicted to score 7.5 points higher; and those who attended a university overseas were predicted to score 5.9 points higher on Criminal Litigation (p-values all < 0.01).

Variables whose effect is related to sitting

97. The variables of ethnic group and first degree classification both had statistically significant main effects, and statistically significant interaction terms with sitting. This indicates that these variables had a significant predictive relationship with exam results when controlling for other factors, and that the size of this relationship varied across sittings.

98. **Ethnicity** – In general, those from Asian/Asian British, Black/Black British, Mixed/Multiple, and other ethnic backgrounds were predicted to perform worse on Criminal Litigation than White students, a trend consistent with that seen in the previous research on differential outcomes.

- There were a few statistically significant interactions between ethnicity and sitting (Asian/Asian British Spring 2015-2018; and Other Ethnic Group Spring 2019). For Asian/Asian British Spring 2015 and Spring 2016, the interaction term was a negative value, suggesting a worse performance than that seen in Spring 2014 for Asian/Asian British students relative to White students. For all other cases, the interaction term was a positive value, suggesting a better performance that that seen in 2014 for the respective ethnic group relative to White students.
- Therefore, the results suggest that while there had been some variation year on year, there was not an overall pattern of differential outcomes by ethnic group widening following the introduction of the newer format assessments.

99. The below table shows the predicted differences between White students and those from different ethnic groups by sitting. The table also indicates where these differences by sit represent a statistically significant change from Spring 2014.

Table 4. Criminal Litigation - Predicted difference in module score by ethnicity compared to White students for each sitting

Sit	Predicted difference in score to White students for ethnic group			
	Asian/Asian British	Black/Black British	Mixed/ Multiple ethnic groups	Other ethnic group
Spring 2014	-5.35	-6.77	-5.70	-7.58
Spring 2015	-8.36	-6.49	-3.79	-5.88
Spring 2016	-8.36	-9.75	-5.84	-9.15
Spring 2017	-2.53	-4.86	-1.51	-8.47
Spring 2018	-1.97	-5.53	-3.62	-3.36
Spring 2019	-4.26	-5.31	-3.21	0.25
Summer 2020	-6.10	-7.56	-2.94	-3.72

= significant interaction (p-value < 0.05)
= close to significant interaction (p-value < 0.10)

100. **First degree classification** – The below table shows predicted differences between students with a first or upper second class degree when compared to students with a lower second class degree by sitting. There were only two statistically significant differences for degree class, with those with a first class degree or a upper second class degree scoring better in 2015 compared to 2014. Model results therefore suggest that there has not been a consistent statistically significant change in terms of dif-

ferential outcomes by degree classification on a sit by sit basis following the introduction of the newer format assessment for Criminal Litigation.

Table 5. Criminal Litigation - Predicted difference in module score by degree class compared to those with a lower second class degree for each sitting

Sit	Predicted difference in score to those with a lower second class degree	
	First class	Upper second class
Spring 2014	11.15	6.55
Spring 2015	16.05	10.30
Spring 2016	12.43	6.69
Spring 2017	9.33	4.33
Spring 2018	11.45	5.59
Spring 2019	11.49	6.33
Summer 2020	10.56	4.25

Criminal Litigation – Before/After model

101. When using a variable for grouped years (2014-2016, 2017-2020), for Criminal Litigation, statistically significant interactions were also found for ethnicity and first degree classification. These were as follows, and are all compared to the 2014-2016 period for the relevant group:

- **Ethnicity:** The difference between White students and Asian/Asian British students sitting Criminal Litigation in 2017-2020 was predicted to be around 3.6 points lower than that seen in 2014-2016.
- **First degree classification:** The difference between those with a lower second class degree and those with a first class degree and those with an upper second class degree sitting Criminal Litigation in 2017-2020 were both predicted to be around 3.1 points lower on the exam respectively compared to 2014-2016.

102. The above suggests that for Criminal Litigation the reform of the assessments also had (in common with Civil Litigation) a statistically significant relationship with ethnicity and degree class when comparing pre- and post-reform sittings overall rather than individual years. However, given that the previous model detailed that there was significant variation between sits for the differential outcomes observed for both these variables, this change may not be related to the change in the format in the exams, but instead reflect variation in exam results over time due to other factors.

Professional Ethics

103. The final model for Professional Ethics included the standard variables used for the Civil and Criminal models as well as interaction terms for sitting with:

- Gender; domicile; ethnic group; degree class; whether a parent/guardian attended university; and English as a first language.

104. There were more significant interactions for the Ethics model than there were for the Civil or Criminal Litigation models. This suggests, that for Professional Ethics there are statistically significant relationships between sitting and several other variables.

105. All variables in the Professional Ethics model except for English as a first language, had either had a statistically significant main effect, or interaction, or both.

Statistically significant variables

Variables not interacted with sitting

106. There were several statistically significant variables that did not display a statistically significant change over different sittings. The following variables had statistically significant main effects/coefficients when controlling for the other variables in the model, but were not included with interaction terms with sitting:

- **Disability -** Compared to those with no declared disability, those with a declared disability were predicted to score around 0.8 points higher on Professional Ethics (p-value = 0.046).
- **Type of school attended –** Those who mainly attended a state school between the ages of 11-18 on average were predicted to score 0.6 points higher compared to those who attended a fee-paying school (p-value = 0.007). This was a relatively small main effect.
- **University attended –** Compared to those who attended 'non-Oxbridge/Russell Group UK universities':
  - those who attended an Oxbridge university were predicted to score 10.1 points higher; those who attended a non-Oxbridge Russell Group university were predicted to score 5.8 points higher; and those who attended a university overseas were predicted to score 4.9 points higher on Professional Ethics (p-values all < 0.001).
- **Mode of BPTC study –** Compared to full-time students, those that studied the BPTC part-time were predicted to score 1.0 point lower on Professional Ethics (p-value = 0.016).

Variables whose effect is related to sitting

107. The variables of ethnic group and first degree classification both had statistically significant main effects, and both ethnic group and first degree classification had statistically significant interaction terms with sitting. Age, domicile, gender, and whether a parent attended university did not have statistically significant main effects, but did have one or more significant interaction terms. English as a main language had no statistically significant main effects or interactions, but was included as an interaction term due to the model selection criteria used.

108. Age – There were only two statistically significant interaction terms for age, these being for those aged 25-34 for Spring 2018, and those aged 45 plus for Summer 2020. Those in these groups performed significantly worse than those aged under 25 for these sittings.

109. Domicile - The main effect for domicile was not statistically significant. There was one significant interaction term for Spring 2018. For this year, overseas candidates were predicted to score around 2.9 points lower than those domiciled in the UK/EU (p-values = 0.021).

110. Gender – The main effect for gender was not statistically significant. There were interaction terms that were significant for Spring 2018, and Spring 2019. For these years, male candidates were predicted to score around 2.4 points lower (Spring 2019) and 2.1 points lower (Spring 2019) than females at those sits (p-values < 0.05). At other sittings, the difference in predicted scores between male and female students was smaller.

111. Ethnicity – In general, those from Asian/Asian British, Black/Black British, Mixed/Multiple, and other ethnic backgrounds were predicted to perform worse on Professional Ethics than White students, a trend consistent with that seen in the previous research on differential outcomes.

112. The below table shows the predicted differences between White students and those from other ethnic groups by sitting. The table also indicates where these differences by sit represent a statistically significant change from Spring 2014. Results did not suggest that differential outcomes relating to ethnicity had widened for each ethnic group after the introduction of newer format assessments and suggested that it had narrowed for those from the 'other ethnic group' category. However, there was a large jump in differential outcomes for those from Black/Black British ethnic backgrounds for Summer

2020. This was also a statistically significant interaction.

Table 6. Professional Ethics - Predicted difference in module score by ethnicity compared to White students for each sitting

Sit	Predicted difference in score to White students sitting in same year for ethnic group			
	Asian/Asian British	Black/Black British	Mixed/ Multiple ethnic groups	Other ethnic group
Spring 2014	-5.51	-4.85	-4.04	-9.85
Spring 2015	-8.03	-5.60	-5.01	-6.03
Spring 2016	-2.61	-5.68	-0.16	-4.35
Spring 2017	-2.82	-3.34	-1.70	-3.41
Spring 2018	-2.68	-6.49	-1.92	-1.85
Spring 2019	-4.15	-4.69	-3.51	-2.22
Summer 2020	-4.44	-10.43	-0.51	-1.11

= significant interaction (p-value < 0.05)
= close to significant interaction (p-value < 0.10)

113. **First degree classification –** As with that seen for Civil and Criminal, the main effect terms for degree class for Professional Ethics were statistically significant (the values for these are given in the below table in the row for Spring 2014). Overall, the gap in predicted score between those with a first class degree and those with a lower second class degree, and between those with an upper second class degree and those with a lower second class degree has not widened following the introduction of the newer format assessments for Professional Ethics. However, it does fluctuate from sit to sit. This is shown in the below table.

Table 7. Professional Ethics - Predicted difference in module score by degree class compared to those with a lower second class degree for each sitting

Sit	Predicted difference in score to those with a lower second class degree	
	First class	Upper second class
Spring 2014	11.93	7.54
Spring 2015	11.57	7.28
Spring 2016	9.31	5.07
Spring 2017	5.31	2.38
Spring 2018	9.93	5.81
Spring 2019	11.46	7.41
Summer 2020	8.35	4.00

Professional Ethics – Before/After model

114. When using a variable for grouped years (2014-2016, 2017-2020), for Professional Ethics, statistically significant interactions were found for the following variables:

- **Age:** There was a narrowing of the difference between the under 25 and 25-35 age groups in 2017-20 compared to 2014-16. In 2017-20 those aged 25-34 were predicted to score 0.65 points lower than those aged under 25, whereas in 2014-16 they were predicted to score 0.85 points higher (a difference of 1.5 points between 2014-16 and 2017-20). Both of the differences were relatively small, but the interaction term was significant.
- **English as a first language:** There was a widening in the difference seen between those with and without English as a first language in 2017-20 compared to 2014-16, with those with English as an additional language predicted to score 1.9 points lower than those with English as

a first language for 2017-2020 sits, whereas in 2014-2016 they were predicted to score around 0.1 points higher (a difference of 2 points between 2014-16 and 2017-20).

- **Ethnicity:** There was a narrowing of the difference in score between White students and Asian/ Asian British students, and between White students and those from the 'Other ethnic group ' category, in 2017-20 compared to 2014-16. Asian/Asian British students were predicted to score 3.4 points lower than White students in 2017-20, whereas in 2014-16 they were predicted to score 4.6 points lower (a difference of 1.2 points between 2017-20 compared to 2014-16). Students from Other ethnic backgrounds were predicted to score 2.4 points lower than White students in 2017-20, whereas in 2014-16 they were predicted to score 6.7 points lower (a difference of 4.3 points between 2017-20 compared to 2014-16).
- **First degree classification:** There was a narrowing of the difference in score between students with a lower second class degree and a first class degree, and between those with a lower second class degree and an upper second class degree, in 2017-20 compared to 2014-16. In 2017-20, those with a lower second class degree were predicted to score 9.7 points lower than those with a first class degree, compared to 11.9 points lower in 2014-16 (a difference of 2.2 points between 2017-20 compared to 2014-16). In 2017-20, those with a lower second class degree were predicted to score 5.5 points lower than those with a upper second degree, compared to 7.4 points lower in 2014-16 (a difference of 1.9 points between 2017-20 compared to 2014-16).
- **Gender:** There was a widening in the difference in predicted score between female and male students in 2017-20 compared to 2014-16. Male students in 2017-20 were predicted to score 1.4 points lower than female students, compared to 0.3 points lower in 2014-16 (a difference of 1.1 points).
- **Parental university status:** There was a narrowing of the difference in score between those who did and who did not have a parent who attended university in 2017-2020 compared to 2014-16. Those who did not have a parent who attended university were predicted to score 0.2 points lower in 2017-20, compared to 1.4 points lower in 2014-16 (a difference of 1.2 points).

115. The above largely supports the findings of the model with individual sitting as an interaction term, and suggests that for Professional Ethics the reform of the assessments has a statistically significant relationship with more variables than that seen for Civil and Criminal.

Regression analysis: Differential outcomes across modules

116. The regression model developed consisted of the same demographic variables in the regression models for the centralised assessments, and the outcome variable of module result (as a percentage out of 100). The data covered those sitting each module for the first time over the Spring 2019 and Summer 2020 sits. Module was included as an additional variable, and sitting was removed as a variable, as the inclusion of sit as a variable did not significantly improve the fit of the model.

117. The reference module was the first Advocacy module taken by BPTC students, Advocacy 1, this was simply because it was the first module in the list when sorted alphabetically.

118. Interaction terms were included in the model between module and the following variables:

- Age range
- Disability status
- Domicile
- English as a first language
- Ethnic group

- First degree classification
- Gender
- University attended

119. Type of school attended, BPTC study mode, and whether a parent/guardian attended university did not show any significant interactions with module, suggesting no notable differences between Advocacy 1 and other modules for these variables.

120. We are principally interested in the interaction terms for this section of the research as these show the difference in predicted score for the module and demographic variable (eg Males sitting Opinion Writing) in comparison to the predicted score for this group for Advocacy 1. There may be a significant main effect for Advocacy 1 for the category, but we are principally concerned with differences seen between modules in differential outcomes, particularly for the centralised assessments.

Significant interaction terms with larger main effects

121. Results are only presented below where there was a statistically significant interaction between module and a given variable, and only where the effect size (the coefficient for the interaction term) is greater than 1 or below -1 (representing a difference of one mark either way), as anything between these two figures is a relatively small difference in predicted score compared to that seen for the same category for the Advocacy 1 module. As already noted, the data used covered only the Spring 2019 and Summer 2020 sits.

122. **Age**: Those aged 25-34 were predicted to score 0.7 points higher than those aged under 25 for Advocacy 1. However, for Advocacy 2 and Opinion Writing, they were predicted to score 0.8 points lower than those aged under 25 (a difference of 1.5 points when comparing Advocacy 1 to Advocacy 2 and Opinion Writing). Those aged 45 and over were predicted to score around 1.6 points lower than those aged under 25 in Advocacy 1, and were predicted to score around 6.7 points lower than those aged under 25 in Ethics (a difference of around 5.2 points when comparing Advocacy 1 to Ethics).

123. **Disability**: Those with a declared disability were predicted to score 1 point higher than those without a declared disability on Advocacy 1. However, for Ethics, they were predicted to score 1 point lower (a difference of 2.1 points when comparing Advocacy 1 to Ethics).

124. **Domicile**: Overseas domiciled students were predicted to score around 1.0 point higher on Advocacy 1 than UK domiciled students. In comparison, those sitting Professional Ethics who were domiciled overseas were predicted to score 1.0 point lower on the assessment in comparison to those domiciled in the UK or EU (a difference of 2.0 points when comparing Advocacy 1 to Ethics).

125. **English as First Language**: Those with English as an additional language were predicted to score around 1.2 points lower than those with English as a first language on Advocacy 1. The predicted difference in score between those with and without English as a first language was 2.4 points, and 2.9 points on Advocacy 2 and Opinion Writing respectively (differences of around 1.2 and 1.7 points greater than that seen for Advocacy 1)..

126. **Ethnicity**: Asian/Asian British students were predicted to score around 4.5 points lower than White students on Advocacy 1. For Drafting, and Opinion Writing, the difference was narrower, with Asian/Asian British students scoring 2.6 and 2.7 points lower respectively. Those from Other ethnic backgrounds sitting Drafting were predicted to score 1.5 points higher on Drafting in comparison to White students, whereas for Advocacy 1 they were predicted to score around 3.1 points lower.

127. **Gender**: Male students sitting Advocacy 1 were predicted to score around 0.3 points higher than female students. For Drafting, and Professional Ethics, male students were predicted to score 0.7 and 1.3 points lower on the assessments respectively, in comparison to females. In both cases, this represents a widening in the differences seen by gender, although the effect sizes are quite small overall.

128. First degree classification:

- **First class**: Those with a first class degree sitting Advocacy 1 were predicted to score around 7.5 points higher than those with a lower second class degree. There was a greater difference seen between these degree classes for Civil, Criminal, Ethics, and ReDoC in comparison to Advocacy 1 – the gaps were around 6.4; 3.9; 2.5, and 4.9 points wider respectively. In comparison the gap was narrower for those sitting Advocacy 2 and Advocacy 3 by around 2.6, and 3.6 points respectively.
- **Upper second class**: Those with an upper second class degree sitting Advocacy 1 were predicted to score around 3.6 points higher than those with a lower second class degree There was a greater difference seen between these degree classes for Civil, Criminal, Ethics, and ReDoC compared to Advocacy 1 – the gaps were around 3.6, 2.0, 2.1; and 3.7 points wider on these assessments respectively. In comparison the gap was narrower for those sitting Advocacy 2 and Advocacy 3 by around 1.6, and 1.5 points respectively.

129. University attended:

- **Oxbridge:** Those who attended Oxbridge were predicted to score around 7.1 points higher than those who attended a non-Russell Group UK university on Advocacy 1. There was a greater difference between these university groupings for Civil, Criminal, Ethics, and ReDoC compared to Advocacy 1 – the gaps were predicted to be 6.2, 4.4, 3.7; and 4.6 points wider on these assessments respectively. In comparison the gap was narrower for Advocacy 2, Advocacy 3, and Conference Skills by around 1.8, 3.0 and 2.5 points respectively.
- **Non-Oxbridge Russell Group:** Those who attended a non-Oxbridge Russell Group university were predicted to score around 3.4 points higher than those who attended a non-Russell Group UK university on Advocacy 1. There was a greater difference between these university groupings for Civil, Criminal, Ethics, and ReDoC compared to Advocacy 1 – the gaps were predicted to be 4.3, 3.1, 3.0; and 3.6 points wider on these assessments respectively. In comparison the gap was around 1.3 points narrower for Advocacy 2.
- **Overseas University:** Those who attended university overseas were predicted to score around 3.4 points higher than those who attended a non-Russell Group UK university on Advocacy 1. There was a greater difference between these university groupings for Civil, Criminal; Ethics, and ReDoC compared to Advocacy 1 - the gaps were predicted to be around 5.6; 3.8; 4.3; and 4.9 points wider on these assessments respectively.

Key findings for this model

130. For most variables, there were only one or two significant differences seen compared to Advocacy 1, but as described above, for first degree classification, and university attended there were more seen, particularly for the centrally assessed models and ReDoC.

131. Overall, results from this model do not suggest that the centralised assessments exacerbate differential outcomes seen for the protected characteristics of ethnicity or gender in comparison to other modules. In other words, all modules show similar patterns in the level of differential outcomes for ethnicity and gender, meaning that the issue of differential outcomes related to ethnicity is not unique to the centralised assessments.

132. Compared to those with a declared disability sitting Advocacy 1, those with a declared disability sitting Ethics would be predicted to score an additional 2.2 points higher compared to those with no declared disability. This would represent a narrowing in differential outcomes between the two groups, meaning that the significant result seen for disability for Ethics led to a reduction in the magnitude of differential outcomes between those with and without a declared disability. There was a significant result seen for age for Ethics that suggested a widening in differential outcomes for those aged 45 and over. Such a result may be partly due to smaller numbers of such students.

133. Overall, the centralised assessments are linked with greater differential outcomes related to academ-

ic history, as defined by the variables of first degree classification and university attended. This also appears to be the case for ReDoC and may possibly be related to the format of the assessments, as ReDoC was also assessed using MCQs (and SAQs).

Conclusions

Descriptive statistics across modules

134. Overall, Charts 1 to 4 and Table 1 clearly show that Civil Litigation, Criminal Litigation and Professional Ethics were more difficult to pass from 2014 to 2020 than the other BPTC modules analysed. This is particularly the case for those with lower second and upper second class degrees in comparison to those with a first class degree.

135. Over the period analysed it has often been the case that for those with a lower second class degree the mean score on the centrally assessed modules has been lower than the pass mark of 60 per cent. This is not the case for any of the other modules analysed, although Opinion Writing comes the closest in this regard.

136. It appears that the introduction of the newer format assessments for the centralised assessments in Spring 2017 led to a reduction in average scores on the modules, and this was seen across degree classification. However, the introduction did not lead to a notable change in failure rates on the Civil and Criminal modules in the year they were introduced (the previous format exams required a student to pass both the MCQ and SAQ section of the exam). For Professional Ethics, there was an increase in failure rate in the year the newer format exams were introduced, but the failure rate returned to the previous level in subsequent years.

137. The introduction of the newer format assessments also saw the lowest scoring 25% of those with an upper second class degree scoring below the pass mark of 60 per cent in some sittings of the centralised assessments (Civil: Spring 2017-2019 and Summer 2020; Criminal: Spring 2019 and Summer 2020; and Professional Ethics: Spring 2017-2018).

138. It does appear that from Spring 2014 to Summer 2020, failure rates and mean scores fluctuated more from year to year on the centralised assessments than that seen for other modules, including when controlling for degree class.

139. This may be related to a number of factors, including the way the modules are taught, and how examinations are prepared for following changes in the assessment format. In addition, the BSB has taken steps to change its focus in re-selecting questions - we use those that have evidenced good discrimination between weaker and stronger candidates, and have largely eliminated questions that were a choice between two options rather than four. We have also almost eliminated all negative questions (eg 'Which of the following is wrong?'). This may have resulted in examinations with questions that might be harder, as we no longer use questions that have performed poorly in the past, and improve those that haven't worked as intended.

140. We believe that the standard setting processes will be less of a factor in determining fluctuation in passing rates on the centralised assessments in the future, as they are there to ensure that there is parity in the level of difficulty between different sits of each exam.[20] The BSB takes steps to ensure that the methodology for standard setting is the same for each sit of an exam, and also tries to use the same people as standard setters (although we do have a wider pool of standard setters available).[21]

20 This is something that does not happen for the non-centralised exams.

21 We are also bringing in 'anchor questions' from the December 2022 sit that will allow us to better monitor performance between cohorts.

141. Whatever the factors involved in determining different pass rates by sit may be, it is recommended that year to year fluctuations in the proportion passing an exam for all modules be more closely monitored going forwards.

142. There also does appear to be some differential outcomes present by ethnicity as highlighted in Charts 6-7, with those from minority ethnic backgrounds achieving lower scores across BPTC modules even when controlling for degree class and domicile. There may also be some differential outcomes on some modules for older students (those aged 35+) in comparison to those aged under 35. The discrepancy by age does not appear to be as large as that seen by ethnicity. These trends were explored further in the regression models.

143. A summary of each regression model follows.

Civil Litigation

144. Results suggest that the introduction of the newer format assessments did not lead to a widening of differential outcomes for the majority of the variables included in the model. For most variables that exhibited differential outcomes, the differential outcomes observed remained relatively constant over time. However, for some variables, there was a significant change in the level of differential outcomes for certain sits.

145. There were several variables that show a strong predictive relationship (ie a statistically significant relationship) with score on Civil Litigation, some with larger predicted effect sizes than others. However, most of the variables that had a significant predictive relationship with exam score did not show a strong interaction with sitting. This indicates that the level of differential outcomes associated with these variables remained relatively constant over time, and did not vary significantly over different sittings. The variables that showed a statistically significant relationship with exam score, but did not vary significantly by sit, were;

    - Age (for those aged 25-34, and 45+, compared to those aged < 25); Disability status; Domicile; Gender; Parents university status; University attended; and BPTC study mode.

146. Of the above listed variables, it was only the variables for age and university attended that had a large effect size (the presence of the category related to a predicted difference in score of more than ±2 points), with the university attended variable being a particularly important one.

    - For age, those aged 45 and older taking Civil Litigation on average would be predicted by the model to score 2.8 points lower on the assessment than those aged under 25.
    - For university attended, compared to those who attended a non-Oxbridge/Russell Group UK based university; those who attended an Oxbridge university; those who attended a non-Oxbridge Russell Group university; and those who attended an overseas university were predicted to score 14.4; 8.2; and 7.1 points higher on Civil Litigation respectively.

147. There were two variables that did show a statistically significant interaction with sitting, and these were ethnicity and first-degree classification. For the model exploring the effect of each sitting, there did not appear to be a consistent trend of differential outcomes improving or worsening following the introduction of the new format assessments for either of these variables, as it fluctuated over time from sit to sit.

148. For the model that grouped all old format exam results and all new format exam results together rather than looking at individual sits, it appears that when controlling for other factors, the outcomes of British/British Asian students improved relative to White students under the newer examinations compared to their performance under the old format examinations. In contrast, the outcomes of those with a first class and upper second class degree declined relative to those with a lower second class degree under the newer examinations compared to their performance under the old format examinations. However, given that there was significant variation between sits for the differential outcomes

observed for both of these variables, this change may not be related to the change in the format in the exams, but instead reflect variation in exam results over time.

Criminal Litigation

149. The results for Criminal Litigation also suggest that the introduction of the newer format assessments did not lead to a widening of differential outcomes for the majority of the variables included in the model. For most variables that exhibited differential outcomes, the differential outcomes observed remained relatively constant over time. However, for some variables, there was a significant change in the level of differential outcomes for certain sits.

150. There were several variables that displayed a statistically significant relationship with score on Criminal Litigation, some with larger predicted effect sizes than others. However, most of the variables that had a significant predictive relationship with exam score did not show a strong interaction with sitting. This indicates that the level of differential outcomes associated with these variables remained relatively constant over time, and did not vary significantly over different sittings. The variables that showed a statistically significant relationship with exam score, but did not vary significantly by sit, were:

    - Age (for those aged over 45 compared to those under 25); Disability status; Domicile; Gender; Parents' university status; University attended; and mode of BPTC study.

151. Of the above listed variables, it was only the variables for age and university attended that had a large effect size (the presence of the category related to a predicted difference in score of more than ±2 points), with the university attended variable being a particularly important one.

    - For age, those aged 45 plus taking Criminal Litigation on average would be predicted by the model to score 2.9 points lower on the assessment than those aged under 25.
    - Compared to those that attended a non-Oxbridge/Russell Group UK based university, those who attended an Oxbridge university; those who attended a non-Oxbridge Russell Group university; and those who attended an overseas university were predicted to score 12.4, 7.5 and 5.9 points higher on Criminal Litigation respectively.

152. As with that seen for Civil Litigation, the same two variables of ethnicity and first degree classification displayed a statistically significant interaction with sitting. However, for the model exploring the effect of each sitting there did not appear to be a consistent trend of differential outcomes improving or getting worse following the introduction of the new format assessments for either of these variables, as it fluctuated over time from sit to sit.

153. For the model that grouped all old format exam results and all new format exam results together rather than looking at individual sits, it appears that when controlling for other factors, compared to those in the same demographic group the outcomes of Asian/Asian British improved relative to White students, and the outcomes of those with a first class and upper second class degree declined following the introduction of the new format examinations relative to those with a lower second class degree. However, as with the other centrally assessed examinations, given that there was significant variation between sits for the differential outcomes observed for both these variables, this change may not be related to the change in the format in the exams, but instead reflect variation in exam results over time.

Professional Ethics

154. As with Criminal Litigation and Civil Litigation, the results for Professional Ethics suggest that the introduction of the newer format assessments did not lead to a widening of differential outcomes for the majority of the variables included in the model. For most variables that exhibited differential outcomes, the differential outcomes observed remained relatively constant over time. However, for some

variables, there was a significant change in the level of differential outcomes for certain sits.

155. There were several variables that displayed a strong predictive relationship (statistically significant relationship) with score on Professional Ethics, these were disability; type of school attended; university attended; and mode of BPTC study. Of these variables, it was only university attended that had a relatively large effect size:

    - Compared to those who attended 'non-Oxbridge/Russell Group UK universities', those who attended an Oxbridge university; those who attended a non-Oxbridge Russell Group university; and those who attended an overseas university were predicted to score 10.1; 5.8; and 4.9 points higher on Professional Ethics respectively.

156. There were more variables with statistically significant interaction terms with sitting for Professional Ethics in comparison to that seen for Civil Litigation and Criminal Litigation. The same two variables of ethnicity and first degree classification did show a statistically significant interaction with sitting, along with age, domicile and gender. However, once again, there was no consistent trend of differential outcomes improving or getting worse following the introduction of the new format assessments, as it fluctuated over time from sit to sit.

157. For the model that grouped all old format exam results and all new format exam results together rather than looking at individual sits, on average ;

    - Compared to the same group sitting the exams in 2014-2016, outcomes for Asian/Asian British students, and those from Other ethnic groups improved relative to White students.
    - In common with Criminal and Civil, outcomes for those with a first class or upper second class degree declined relative to those with a lower second class degree after the newer format assessments were introduced. This was also the case for male students (relative to female students), those with English as an additional language (relative to those with English as a first language), and those aged 25-34 (relative to those aged under 25).
    - However, as with the other centrally assessed examinations, given that there was significant variation between sits for the differential outcomes observed for these variables, this change may not be related to the change in the format in the exams, but instead reflect variation in exam results over time.

Differential outcomes across modules

158. The model comparing differential outcomes across modules only covered the Spring 2019 and Summer 2020 sits. For the sake of brevity how the modules compared before and after the introduction of the newer format assessments was not looked at.

159. Similar levels of differential outcomes were seen across modules for most demographic variables (in comparison to the reference group of those sitting the first advocacy module, Advocacy 1). There were a few significant differences, but no clear pattern for most variables.

160. However, for the variables of first degree classification, and university attended, there was a clearer trend, with differences in predicted score between those with different degree classifications, and between those who attended different universities being greater for the centralised assessments and ReDoC, and lesser for Advocacy 2, and Advocacy 3.

161. Overall, results from this model do not suggest that the centralised assessments exacerbate differential outcomes seen for the demographic variables of ethnicity, gender, parental university status, and type of school attended in comparison to other modules. This means that that the issue of differential outcomes related to ethnicity on the Bar training course is not unique to the centralised assessments.

162. A significant difference seen for disability for Ethics led to a narrowing in the level of differential outcomes between those with and those without a declared disability. There was a significant result seen

for age for Ethics that suggested a widening in differential outcomes for those aged 45 and over. Such a result may be partly due to smaller numbers of such students.

163. Overall, the centralised assessments appear to be linked with greater differential outcomes related to academic history as indicated by first degree classification and university attended. This also appears to be the case for ReDoc and may be related to the format of the assessments, as multiple-choice questions were also utilised for the ReDoC assessmentt.

Summary

164. The aims of this research were to:

- Compare differential outcomes on the centralised assessments under system brought in for the 2017 sits with the system in place until 2015-16;
- investigate differential outcomes on other modules on the BPTC to highlight any trends that may be occurring and how other modules compare to the centralised assessments; and
- report on other trends that may be of note following analysis of the data.

165. Upon exploration of the data , it was found that the centralised assessments were the modules that consistently displayed the lowest mean scores for those sitting them. This was particularly the case for those with an upper second class and lower second class degree.

166. There also appears to have been a drop in mean scores seen for the centrally assessed modules for the majority of sits following the introduction of the newer format examinations in 2017 (compared to 2014-2016 sittings), particularly for those with a first class or upper second class degree.

167. The mean score for the centralised assessments also appears to vary more widely between years than that seen for other BPTC modules, as does the failure rate for those sitting the centrally assessed modules for the first time.

168. Results from the regression models found that, as with that seen in previous research on differential outcomes, ethnicity was found to be a statistically significant variable with a relatively large main effect size in relation to predicted score on Civil Litigation, Criminal Litigation, and Professional Ethics. Those from Asian/Asian British, Black/Black British, Mixed/Multiple ethnic backgrounds, and from other ethnic backgrounds were all predicted to do worse on the assessments than White students on each centralised assessment, even when controlling for other variables such as prior academic attainment.

169. The gap in results seen for ethnic minorities is not unique to the training for the Bar. There is a substantial body of research that highlights similar differences in other disciplines and academic stages. In an analysis of Higher Education Statistics Agency data from the late 1990s and early 2000s, it was found that British students from minority ethnic backgrounds were less likely to get good degrees than White students.[22] In their analysis of trainee doctors, Woolf et al (2011) determined that differential outcomes were both 'consistent and persistent'; but while it was clear that ethnicity was related to exam performance the reasons for this were opaque.[23] Studies that have controlled for other factors associated with academic performance have shown that differences in outcomes for ethnicity cannot be fully accounted for.[24]

170. Richardson (2015) noted the 'phenomenon of an attainment gap… is correlational rather than causal

22 Connor, H. et al (2004) Why the difference? A closer look at Higher Education minority students and graduates. Department for Education and Skills. Research report RR552

23 Woolf, K., Potts, H.W.W., McManus, I. C. (2011) Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. BMJ (Clinical Research Ed.) 342:d901-d

24 Degree Attainment, Ethnicity and Gender – Interactions and the modifications of effects (Fielding et al, 2008)

in nature'.[25] Ethnicity per se is most probably not the effective variable affecting students' success on the BPTC. Instead, it is a proxy for other factors correlated with ethnicity that are not controlled for in the analyses that have been undertaken; these are likely to relate to socio-economic status and psychosocial-cultural experience (including family and other support networks), and differing behaviour towards those of different ethnicities. Overt racism does still occur in Higher Education in the UK,[26] and there is often more subtle racial discrimination against specific groups of people in Higher Education.[27] [28]

171. In other research commissioned for the BSB,[29] BPTC students from minority ethnic backgrounds described racial microaggressions and differing treatment towards them from lecturers in some cases, and inappropriate comments in general. From the same research, examples of factors related to socio-economic background (which in turn has some relationship with ethnicity in the UK[30]) and success on the BPTC included the impact of needing to work part time when studying[31], and that the BPTC required extensive self-directed learning which was highlighted as putting those who had attended more 'elite' academic institutions at some advantage compared to others. It has also been shown that irrespective of background pupils from all minority ethnic backgrounds are significantly more likely to go to university than their White British counterparts.[32] This may translate to the socio-economic background of UK domiciled students on the BPTC differing by ethnic background, which may also explain some of the gap in differential outcomes. We did not have the data to control for such factors in this study.

172. First degree classification and university attended were also statistically significant variables with relatively large effect sizes across the Civil, Criminal and Professional Ethics models.

173. The effect sizes of other variables in the centralised assessments models were generally smaller (with the exception of those in older age groups to an extent), and while this does not mean the results seen are not of concern, it does mean that differences in score are explained less by these variables than that seen for ethnicity and variables relating to academic history.

174. Overall results on a sit by sit basis suggest that the introduction of the newer format assessments did not appear to lead to a consistent change in differential outcomes on the centralised assessments for the demographic variables of age, disability, domicile, ethnicity, gender, parental university status, and type of school attended.

175. When looking more broadly at before/after type models, there are some variables for which differential outcomes may have widened for Professional Ethics. These were age, English as a first language, gender, and whether a parent attended university.

176. The level of differential outcomes on the centralised assessments was, in general, found to be in line with that seen for other BPTC modules for the demographic variables analysed, including ethnicity.

25 Richardson, J. T.E. (2015) The under-attainment of ethnic minority students in UK higher education: what we know and what we don't know, Journal of Further and Higher Education, 39:2, 278-291

26 Wong, B., Elmorally, R., Copsey-Blake, M., Highwood, E., & Singarayer, J. (2021). Is race still relevant? Student perceptions and experiences of racism in higher education. Cambridge Journal of Education, 51(3), 359-375.

27 Savas, G. (2014) Understanding critical race theory as a framework in higher educational research, British Journal of Sociology of Education, 35:4, 506-522

28 Singh, G. (2009). Black and minority ethnic (BME) students' participation in higher education: Improving retention and success. A synthesis of research evidence. York: Higher Education Academy.

29 Bar Standards Board (2017) Barriers to Training for the Bar. Research undertaken by NatCen on behalf of the BSB

30 House of Commons Library (2020). Which ethnic groups are most affected by income inequality?. Can be found at: https://commonslibrary.parliament.uk/income-inequality-by-ethnic-group/

31 Scales, J and Whitehead, J M (2006) found a strong relationship between needing to work part time at university and overall course results. See Scales, J and Whitehead, J M (2006) 'The undergraduate experience of Cambridge among three ethnic minority groups'. Cambridge University Reporter, 2005-06

32 Crawford, C., & Greaves, E. (2015). Socio-economic, ethnic and gender differences in HE participation.

However, compared to other modules, the centralised assessments were linked with a higher level of differential outcomes related to academic history as defined by the variables of first degree classification and university attended.

177. The differences by ethnicity in differential outcomes between the centralised assessments and other BPTC modules were broadly similar. However, as the centrally assessed modules were more difficult to pass on average, the differences in outcomes by ethnicity had a larger impact on pass rates for the centralised assessments than for other modules - with higher proportions of students from minority ethnic backgrounds failing to pass the centralised examinations than other modules on the course.

Appendices

Charts

Chart A1. Mean BPTC module scores and IQR by sitting, gender and degree class – first sits only

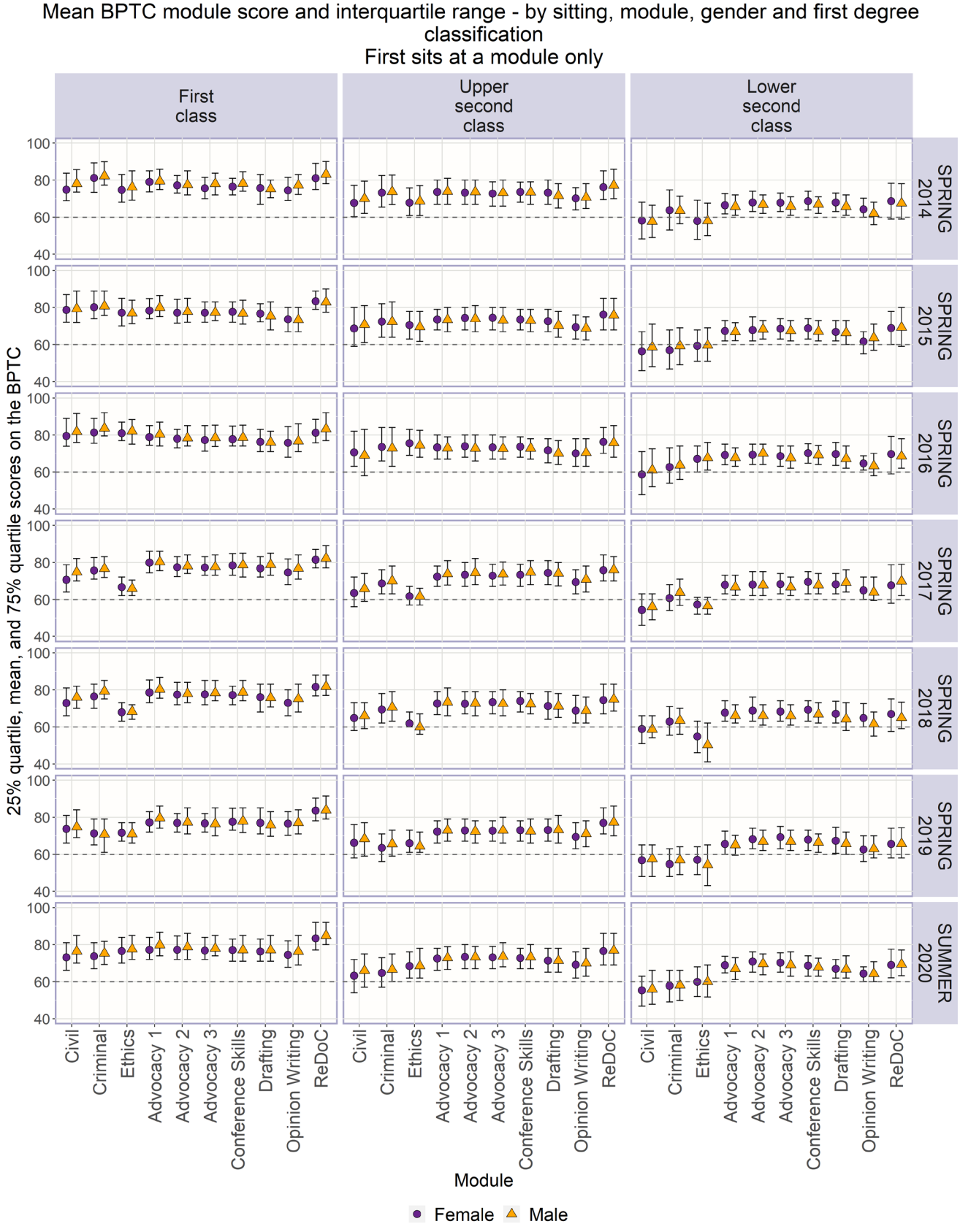
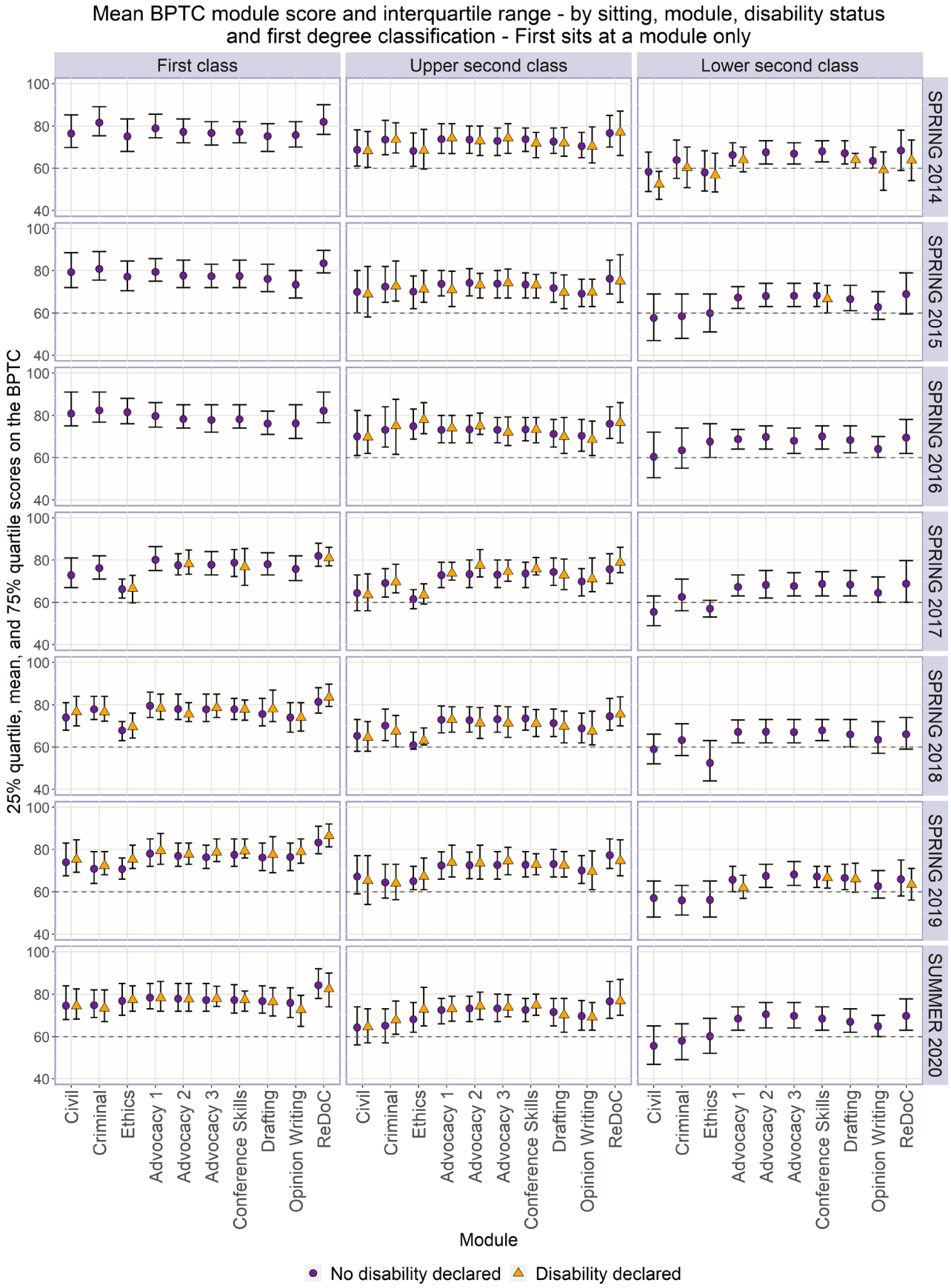


Chart A2. Mean BPTC module scores and IQR by sitting, disability status and degree class – first sits only



Interpretation of interaction effects

- The intercept in a model only including categorical predictors that includes interactions has the same interpretation as in a model without interactions. It is the predicted outcome when all of the variables are at their reference level.
- As already noted, the main effects of a categorical variable in a model without interaction terms can be interpreted as the effect of that variable compared to a reference group for that specific variable when controlling for all of the other variables in the model.
- The main effect coefficient of an interacted variable however, has a different meaning. It is the effect of that coefficient at the reference level of the interaction variable. This can be quite difficult to grasp, particularly if there are multiple interaction terms in a model.
- The interpretation of coefficients for the interaction terms can also be quite difficult to grasp and so to provide elucidation, an example model with an interaction term is highlighted below. The below contains the main effects and interaction terms for an interaction between gender and year in a relatively simple regression model used to predict student score on an assessment. The figures are made up.

Table. A1.

Type of term	Variable	Term	Coefficient	P-value
Intercept			60.05	
Main Effect	**Gender** (reference group: Female, and due to the interaction term, 2015)	2015:Male	2.10	0.001
	Year (reference group: 2015, and due to the interaction term, female)	F:2016	1.20	0.042
		F:2017	1.70	0.008
		F:2018	2.20	0.001
Interaction	**Gender:Year** (Comparison group is males sitting the assessment in 2015)	Male:2016	1.10	0.048
		Male:2017	1.40	0.031
		Male:2018	1.50	0.027

- Interpreting the above table:
    - The intercept in the above table shows the predicted value for females sitting the assessment in 2015: this is 60.05.
    - The main effect for gender shows the predicted effect for males sitting the assessment in 2015 in comparison to females sitting the assessment in 2015. This is a difference of +2.10, and this is statistically significant at the $p < 0.05$ level, suggesting that there was a statistically significant difference in the results of males sitting the assessment in 2015 compared to females sitting the assessment in the same year.
    - The main effect for year shows the predicted effect of year for females only, in comparison to a reference group of females sitting the assessment in 2015 (1.6 for 2016; 1.7 for 2017; and 2.2 for 2018). These main effect terms are all statistically significant at the $p < 0.05$ level, suggesting that each year has a statistically significant difference in predicted score for female students in comparison to the reference category of females sitting the assessment in 2015.
    - For the interaction between gender and year, the coefficients show the additive effect of male and year compared to males sitting the assessment in 2015 (1.1 for 2016; 1.4 for 2017; and 1.5 for 2018). These terms are all statistically significant at the $p < 0.05$ level, suggesting that the effect of each interaction term is likely to be significantly different to the reference category of males sitting the assessment in 2015.
- If you wanted to predict the score of a male sitting the assessment in 2016 using this model, then you

would use the following calculation:

- Intercept + male + male:2016 = 60.05 + 2.10 + 1.10 = 63.25
- To break this down;
- intercept + male; would give the predicted value for males sitting the assessment in 2015 = 62.15;
- then adding male:2016 onto the above would give the predicted value for 2016 for males, which is 63.25.

- The interpretation of models with more interaction effects becomes even more complicated, but follows the same premise, i.e. interpretation of coefficients is based on what the reference categories for the main effects are, given the interacted terms in the model.

Full results to the regression models are provided in a separate spreadsheet available with this report.

Contacting us

We are committed to providing a high standard of service and dealing with everyone in a way that is fair, transparent, and proportionate. We welcome your feedback on our services, particularly where the level of service has exceeded or fallen below your expectations. Your comments and suggestions are important to us as they will help us to meet our obligations to you and to improve our performance.

Write to us

Bar Standards Board

289-293 High Holborn

London WC1V 7HZ

Tel: 020 7611 1444

contactus@barstandardsboard.org.uk

www.barstandardsboard.org.uk

Twitter: @barstandards

www.linkedin.com/company/the-bar-standards-board

In addition, If you would like this report in an alternative format, please contact the BSB Research Team on 020 7611 1467 or email Research@BarStandardsBoard.org.uk