Cambridge Assessment

The link between subject choices and achievement at GCSE and performance in PISA 2015

Research Report

Matthew Carroll & Tom Benton

December 2018

Author contact details:

Matthew Carroll & Tom Benton
Assessment Research and Development,
Research Division
Cambridge Assessment
The Triangle Building
Shaftesbury Road
Cambridge
CB2 8EA
UK

Carroll.m@cambridgeassessment.org.uk
Benton.t@cambridgeassessment.org.uk

http://www.cambridgeassessment.org.uk

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

How to cite this publication:

Carroll, M. & Benton, T. (2018). *The link between subject choices and achievement at GCSE and performance in PISA 2015*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Contents

Introduction

The purpose of this report is to compare individual pupil achievement on two high profile sets of assessments in England: the Programme for International Student Assessment (PISA) and the General Certificate of Secondary Education (GCSE).

It should be noted to begin with that these two assessments serve entirely different purposes. GCSEs are focussed upon assessing the extent to which pupils have learnt and can apply the specific knowledge and skills taught within their particular programme of study, with implications for future educational and job opportunities. In contrast, the main focus of PISA is at the level of whole nations, with these being assessed against some proposed measures of how able their pupils are to apply skills[1] in reading, maths, science and (recently) collaborative problem solving. However, whilst there are obvious differences between the purposes of the two assessments, there are various reasons we might be interested in the association between these two assessments at an individual level.

1. In the past, data from PISA has been used to make decisions about the way GCSEs are administered and awarded. In particular, data from PISA was used by Ofqual as part of their justification for moving to a system of comparable outcomes to tackle "grade inflation". For example, Ofqual (2014) compared data showing rising proportions of candidates achieving the top grades at GCSE up until 2012 with the UK's fairly stable performance in PISA assessments over the same period to conclude that "there is no substantive evidence that suggests that increases in the proportions of higher grades in recent years can be justified". Similarly, using data from international studies including PISA (but also TIMSS[2] and PIRLS[3]), Heath et al (2013) concluded that "there can be little doubt that the official results showing dramatic improvements in standards at key stage 2, GCSE, and A level are grossly inflated." However, such interpretations depend on the correlation between PISA and GCSE at an individual level; if the association between the two forms of achievement was very weak, indicating that the assessments were measuring different skills, would it be reasonable to make conclusions regarding "grade inflation" in GCSEs, based on national performance in PISA?

2. Beyond arguments about GCSE grade inflation, PISA assessments are themselves intended to allow us to make meaningful statements about the quality of education in a country[4] and may be used in this manner by policymakers. For example, in 2010 the then Education Secretary Michael Gove used results from PISA 2009 to conclude that "our schools system is failing to fully develop the potential of many of our children" citing, for example, that an "alarming 18 per cent are failing to achieve a standard of literacy that will enable them to participate effectively and productively in life" (Gove, 2010). Judging the success or failure of a school system can, either

[1] The term "skills" is used throughout this report as shorthand for "the propensity for students to correctly answer questions of the type asked in PISA/GCSE". It is not intended to imply a dichotomy between "skills" and "knowledge and understanding".
[2] The Trends in International Mathematics and Science Study.
[3] The Progress in International Reading Literacy Study.
[4] The OECD states that the aim of PISA is to "evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students" (http://www.oecd.org/pisa/aboutpisa/ viewed on 4th May 2018).

directly or indirectly, imply judging the effectiveness of teaching. However, if PISA and GCSE achievement is only weakly related, indicating that the skills measured by GCSEs, and by implication the skills actually taught within schools, are very different to those measured by PISA, is it fair to judge the quality of teaching in the country on this basis? With such questions, knowing the correlation between GCSE and PISA achievement again informs the way we interpret results from PISA.

3. Some research studies, such as one published by the Department for Education (DfE) in 2011 (DfE, 2011), attempt to work out what would (theoretically) be required in terms of improvement in GCSE performance in order for England to catch up with "top-performing countries" in PISA. However, for such results to be truly meaningful it is necessary to establish whether PISA and GCSE are measuring similar skills. If not, improvements in GCSE performance may not lead to the expected improvements in PISA performance.

4. Results from PISA are sometimes criticised for being overly focussed on a few core skills and ignoring important 21st century skills such as creativity. For example, Villalba (2012) said that PISA "cannot be regarded as a measurement tool for creativity" and raised concerns that "certain very important aspects that education is supposed to provide will be overlooked". Similarly, Zhao (2012) raised concerns that countries' high performances in the core skills of reading, maths and science might be "masking important failures in developing innovators and entrepreneurs". With such concerns in mind it was of interest to look not only at the correlation between GCSE achievement and individuals' PISA performance, but also at how PISA performance relates to the selection of subjects that pupils choose to study. In particular, does PISA disadvantage students that choose to study more arts subjects at GCSE? Answering such questions affects the way in which we interpret findings from the PISA studies.

As Jerrim and Shure (2016) point out, PISA assessments differ from those taken as part of GCSEs in at least four respects:

- **Type of skills assessed**. GCSEs are intended to test pupils' knowledge and understanding of specific content as defined in the qualification specification (syllabus) whereas PISA tries to measure pupils' ability to apply knowledge to solve problems in real world situations.
- **Timing**. GCSE assessments are usually taken in May or June of year 11. Most GCSE assessments taken by England's PISA 2015 cohort would have been sat in May or June 2016. In contrast, the PISA 2015 tests themselves in England were taken around six months earlier in the first school term of year 11 (November or December 2015) .
- **Test administration mode**. GCSE tests are currently almost universally taken in paper-based formats whereas PISA 2015 made use of computerised tests.
- **Stakes**. Results from GCSE tests are supplied to each individual pupil and influence their future educational and job opportunities. In contrast, pupils are not even told how well they performed individually in PISA assessments.

Further to these differences, in contrast to the wide range of subjects assessed as part of GCSEs, PISA assesses students' skills in in four domains: Mathematics, Reading, Science and Collaborative Problem Solving. Individual students only complete relatively short assessments in each of these domains. Furthermore, in order to ensure that a broad range of questions are included in the PISA assessments, whilst at the same time ensuring the burden on individual students does not become too great, different students are presented with different sets of questions to respond to within each domain. For this reason, whilst for GCSEs performance can be summarised by simply adding up each student's score on each question, a more complex statistical model is used to create PISA scores (see OECD, 2017).

The research in this report will add to these well-documented differences between PISA and GCSEs to look at the strength of association between achievements on these two tests from an empirical perspective. A small amount of existing research has shown a relationship between PISA and GCSE within particular subjects (DfE, 2017). Specifically, the DfE's report compares achievement in PISA reading, maths and science with performance in the GCSE English, maths and science respectively. We will go beyond these published statistics by examining correlations of the PISA domains with a range of GCSE subjects, and also by benchmarking against correlations between GCSEs and scores from key stage 2 (KS2) tests taken at the end of primary education. As part of this process, the report will explore some of the technical issues that arise when attempting to link the various sources of data. Finally, and for the first time, this report will show how performances in the different PISA domains relate to the subject choices that pupils make at GCSE.

Preliminary steps

Data

For the purposes of this study the DfE provided a data set containing all available data from the PISA 2015 study for participants in England matched to a large number of key variables from the National Pupil Database (NPD). In particular, the NPD data gave details of each individual's performance across a large number of GCSE subjects, the vast majority of which would have been assessed in summer 2016, as well as performance at KS2 in summer 2011. This data set was anonymised so that no individual pupil could be personally identified. The full data set contained information on 5,194 pupils from a total of 206 schools.

Producing revised PISA ability estimates

In order to perform any of the analyses described in the introduction we need a measure of each pupil's performance in the PISA tests. However, in the original data set, the only formal estimates of each individual student's ability in each PISA domain come in the form of plausible values (hereafter, "PVs"). As will be described next, there are some problems with using these estimates directly in our analyses.

Plausible values are used within all existing international surveys to overcome the ubiquitous problem of measurement error in educational assessments. In high stakes tests (for example, GCSEs) it is necessary to produce a single number summarising the performance of each student – typically their total score across all items. However, it is generally recognised that the performance of individuals can vary depending upon the precise selection of items they are asked to respond to (see, for example, Winkley & Cresswell, 2012). Whilst this measurement error is fully acknowledged, it is rarely formally accounted for in educational research. Indeed, for high-stakes assessments, typically comprising large numbers of items, it may be safe to assume that the impact of measurement error on analyses is fairly limited.

In contrast to this, perhaps because the tests used to assess each subject are fairly short, ability estimates within the PISA data sets make measurement error fully explicit. Specifically, rather than providing a single number summarising each pupil's performance in maths, the PISA data set includes 10 PVs for each pupil that give a range of possibilities for their likely mathematical ability given everything that is known about them. These PVs are explicitly *not* intended to allow inferences to be made about individual students. Rather the idea is that by analysing these PVs as a whole across large groups of students we will be able to make accurate inferences about the distribution of abilities whilst accounting for the measurement error inherent in the test. In other words, the aim is to be able to make accurate inferences about populations without claiming such short tests can produce accurate ability estimates for individuals.

The official plausible values within the PISA data sets make use of all of the information that is collected about pupils as part of the study. For example, PVs relating to maths performance are generated not only using pupils' responses to the specific maths items but also:

- Responses to items in other PISA domains (science, reading, and collaborative problem solving),
- The number of items that were "not reached" in the PISA assessments (i.e. the number of items omitted in a row at the end of the assessments),
- Information regarding which school they attend,
- Responses to the student questionnaire including details of what possessions are in their home, parental occupation, the language they speak at home, their attitudes to learning and their aspirations for the future.

The inclusion of all of this information in the production of PVs is not an arbitrary decision. Rather it is absolutely necessary for subsequent analyses looking at relationships between ability and other student and school characteristics to produce unbiased estimates that fully account for measurement error. For example, if the PVs were generated without including gender in the generating model, subsequent analyses would (slightly) underestimate the difference in average abilities between boys and girls. Similarly, if PVs for different subjects (e.g. maths and reading) were generated entirely independently, subsequent analyses would again (slightly) underestimate the strength of the association between abilities in different domains (Von Davier, Gonzalez & Mislevy, 2009).

The aim of the present research is to explore the strength of the relationship between abilities in different PISA domains and subsequent performance in GCSEs. However, the only ability estimates in the original PISA data are in the form of PVs, and these PVs were generated from a model that did *not* include GCSE attainment. As such, if we analysed these variables as they are, we are likely to underestimate the strength of the association between PISA abilities and GCSE performance. This is of substantive importance, as our analysis will not only discuss which GCSE subjects are most strongly associated with PISA (i.e. the rank order of correlations) but will also compare the absolute size of these correlations with other analyses – for example, comparing the correlation between PISA and GCSE with the correlation between KS2 and GCSE. In order for us to make any inferences from these comparisons it is therefore crucial that the correlations are not biased downwards.

With this in mind, the first stage of analysis was to generate our own set of plausible values that specifically include performance in GCSE subjects within the generating model. The official methodology used to incorporate all of this information in the generation of PVs, and still retain comparability between results from different countries is fairly complicated and will not be repeated here. Further details are available in the PISA technical report (OECD 2017). As far as possible, the methodology used to produce the official PISA PVs was replicated using functions provided within the R package 'TAM' (Robitzsch et al. 2018). The main difference was that, rather than including all PISA questionnaire variables within the model, only a small selection were included, and, in their place, a large number of variables from the NPD such as performances in GCSE and KS2 subjects were included. Crucially, all of the variables to be used in subsequent analyses within this report were included at this stage. Except where explicitly stated otherwise, all references to plausible values (or PVs) for the remainder of the report will refer to the values generated specifically for our own analysis rather than those supplied with the original PISA data.

In order to check that our conclusions were robust to the procedure used to generate PVs, we also generated simpler ability estimates for each PISA participant based purely upon their performance in the items in the relevant subject. Specifically, EAP ability estimates[5] were produced based on upon a unidimensional partial credit IRT model underlying the items in each subject, with the item parameters set to equal those used in the main international analyses of England's students[6]. These estimates of ability are the closest we get to producing a simple PISA test score for each student – simple sum scores cannot be used because different students take different combinations of items. Note that, in contrast to the PVs, EAP ability estimates are only produced for students that answered at least one item in the domain of interest. Within the data set used for analysis, all but 2 students answered at least one science item. However, only 41 per cent of students answered any reading items, 41 per cent of students answered any maths items and 31 per cent answered any collaborative problem solving items.

To illustrate the importance of producing our own PVs and the substantive impact this has on conclusions, Table 1 shows the correlation between various methods of estimating PISA mathematics ability and achievement in maths in both GCSE and KS2. The left hand side of the table shows the correlations when the data is restricted to students who answered at least one PISA maths question, whereas the right hand side shows the correlation across all students. Note that with both the original and revised PVs, 10 values were created for each student and only the first of these is used in the top two rows of Table 1. As can be seen, the original PISA PVs result in a much lower estimated correlation with both external maths tests

[5] EAP stands for "Expected *A Posteriori*". These ability estimates are the most likely level of ability in each subject for each student taking account of the overall distribution of ability across the population and the individual student's responses to the items.

[6] Many thanks to John Jerrim from UCL for supplying a suitable file of item parameters.

(GCSE and KS2) than the newly-estimated PVs described above. This is particularly the case if the analysis is performed across all students, including those who did not take any PISA maths items (whose ability estimates are imputed based solely on performance in other domains and other variables included in the generating model). Similarly, although easier to understand, using EAP values also results in lower correlations, as this approach does not account for measurement error in the (fairly short) PISA tests. For this reason, and because the present analysis involves making substantive comparisons with correlations between KS2 and GCSE, it is important that we do not underestimate the strength of the association between PISA and GCSE. As such, the production of revised PVs for this analysis is crucial.

Table 1 also shows that the estimated correlation between the first of our revised PVs and the other achievement measures hardly changes when all students were included in analysis, rather than just those who answered some PISA maths items. This indicates that the approach to imputation was effective. Being able to include all students in subsequent analyses considerably simplifies things on a practical level.

Finally, Table 1 shows the effect of using the mean of PVs within analysis. Although this has been repeatedly warned against, it is still fairly common practice within exploratory analyses of PISA data (Rutkowski et al, 2010). Table 1 shows that even taking the mean across all of the original PVs fails to bring the correlation with GCSE and KS2 up to the level estimated from the first of our revised PVs. Finally, Table 1 shows that taking the mean of our own revised PVs leads to even higher estimated correlations between PISA and external achievement. However, such results are misleading because both GCSE and KS2 were used in the generation of revised PVs in the first place. As such, taking the mean of PVs generated in this way will result in estimated correlations that are too high.

Table 1. Weighted Pearson correlations between PISA maths ability estimates and performance in GCSE and KS2 .

	Pearson correlations			
	Students answering at least one PISA maths Item		All students (including those with no PISA maths items)	
PISA maths ability estimate	**GCSE maths grade**	**KS2 maths level**	**GCSE maths grade**	**KS2 maths level**
First PISA PV (original)	0.743	0.597	0.674	0.542
First PISA PV (revised)	0.795	0.668	0.783	0.654
EAP ability estimate	0.742	0.615	-	-
Mean of original PISA PVs	0.767	0.627	0.717	0.579
Mean of revised PISA PVs	0.833	0.705	0.826	0.692
N	**1,954**	**1,891**	**4,778**	**4,621**

Correlation between PISA abilities and GCSE performances

Method

Deriving variables

To begin with we considered the degree to which performance in individual GCSE subjects was correlated with performance in the PISA tests. To allow correlations to be calculated, GCSE grades were coded numerically, with A* = 8, A = 7, B = 6, etc., down to U = 0. The mean GCSE grade of each pupil was the average of these values across all GCSEs[7] and was also used within this analysis.

A similar process was carried out for teacher-assessed KS2 levels: awarded numerical levels (i.e. levels 1 – 6) were treated as that number, whilst "W" (working toward level 1) was assigned as 0; remaining values were assigned as missing data. KS2 test levels were available for English and maths, but it was considered that test marks would provide finer-level information, so these were used in preference. Teacher-assessed levels were also retained because they provided a measure of attainment in all of maths, English and science.

GCSE subjects taken by fewer than 100 students in the sample were removed to ensure adequate sample sizes for all correlations. This left 26 individual subjects to be considered: Additional Science, Art and Design, Biological Science, Business Studies, Chemistry, Core Science, D&T: Food Technology, D&T: Graphic Products, D&T: Resistant Materials Technology, D&T: Textiles Technology, Drama, English Language[8], English Literature, French, Geography, German, History, Home Economics: Child Development, Information Technology, Maths, Media, Film and Television Studies, Music, Physical Education, Physics, Religious Studies, Spanish, and Statistics. This, slightly restricted, list of GCSE subjects includes those that tend to be most popular in the overall GCSE population (see Carroll & Gill, 2017). Two subject-specific summary metrics were included: highest Science grade, and average science grade (only calculated for students who took three separate sciences). Finally, six KS2 levels and test marks were considered: reading test mark, writing test mark, total maths test mark, teacher-assessed English level, teacher-assessed maths level, and teacher-assessed science level.

Calculating correlations

Correlations were calculated between each of the 10 PVs from each PISA domain and all of the above variables. Following recommendations for the handling of PVs (Foy, 2017), for a given domain, Pearson correlations were calculated between each PV and the variable of

---

[7] Technically this was the NPD variable "average points per entry". It is calculated by awarding points to each grade achieved (as described) including some GCSE equivalent qualifications, summing these across all subjects, and dividing by the number of subjects for which the student was entered. For ease of the language this will be referred to as mean GCSE grade for the remainder of this report.

[8] Technically this variable recorded each student's "highest" English grade to account for the fact that candidates may have taken either English Language or a combined English Language and Literature GCSE. The vast majority of students will only have ever taken one of these options.

interest, weighted by the PISA final student weight[9]; and then the mean of the 10 correlation coefficients was calculated. To derive the standard error of the correlation, taking into account sampling variance and imputation variance, code was adapted from the 'intsvy' R package (Caro and Biecek, 2017). Further details are given in Appendix 2.

For the main analysis, it was considered that correlations could be affected by the range of abilities of pupils taking different subjects. For example, if the range of PV values was restricted in smaller subjects, the correlation could artificially appear weaker than the true value. Hence, Thorndike corrections (Thorndike, 1947) were applied to estimated correlation coefficients. To do this, the standard deviations of the PV in the full population and in the restricted population (i.e. only for students with a value for the variable of interest) were calculated. An adjustment for the difference between these two standard deviations was made using the equation stated in Appendix 2. For example, if a particular GCSE subject was taken by a group of pupils within a relatively narrow range of ability (i.e. a low standard deviation) then the initially calculated correlations would be adjusted upwards so that correlations could be meaningfully compared across different subjects.

Checking correlation robustness

To test the robustness of correlations, analyses were repeated with varying conditions. First, analyses were repeated without the Thorndike correction, to identify whether the correction substantially affected results. Next, analyses were repeated with original PISA PVs, to determine the extent to which the use of updated PVs changed conclusions. Analyses were also repeated using different correlation methods: weighted Spearman correlations were calculated for all variables, and weighted polyserial correlations were calculated for any variables that could be considered to be ordinal categorical variables (i.e. all grades and levels). Weighted Spearman and polyserial correlations were calculated using the 'wCorr' R package (Emad and Bailey, 2017). Analyses were then repeated using the EAP ability estimates, which provided the most straightforward measure of each student's achievement within the PISA tests without attempting to adjust for measurement error. These last estimates also restricted the sample for each correlation only to students who had actually taken items in the relevant domain.

As PISA is a low-stakes test, there is the potential that student motivation could substantially influence performance. To test for an effort effect, correlations were re-estimated separately for 'high effort' and 'low effort' groups. An effort indicator was derived by fitting a linear model with a response of 'total number of items attempted' and 27 categorical predictor variables, each indicating whether or not a student was assigned a particular cluster of PISA questions (each student was assigned four clusters in total); the fitted model had an R^2 value of 0.955. Residuals were extracted from the model: students who had answered more items than expected given the clusters of items they took were placed into the 'high effort' group (n = 4,053), whilst those who had answered fewer items than expected were placed into the 'low effort' group (n = 1,141). Analyses using updated PVs were then repeated separately for each of these groups.

[9] These were supplied along with the original PISA data sets. They are designed to help ensure that the analyses are based upon samples of students that are representative of the overall population of students in England.

Providing context for correlations
To provide context for interpreting the observed correlations in the main analyses, two further analyses were carried out. First, correlations were calculated between GCSE grades achieved in June and those achieved in resits in November. This was done to examine the correlation in performance over a similar time period to that between PISA and GCSEs. To do this, data on grades of OCR candidates sitting maths in both June and November 2016, and those sitting English in both June and November 2017 were acquired[10]. Grades were recoded numerically as described above. Unweighted Pearson correlations were then calculated between first attempt and the resit. Thorndike corrections were applied, using the standard deviation of grades in the entire population of candidates from the June session; this was particularly important because the range of candidates taking resits is likely to be highly restricted relative to the full population[11].

Finally, correlations between GCSE and KS2 variables considered in the main analysis were examined. Whilst performance at KS2 and GCSE is correlated, it is widely understood that the two key stages are quite different, and pupils may mature at different rates, so the correlation is not extremely strong. Therefore, if correlations between KS2 and GCSE were similar to those between GCSE and PISA, it could also be assumed that PISA and GCSE are themselves quite different. To test this, Pearson correlations were calculated between GCSE and KS2 variables, with final student weights still applied to make results comparable to those from main analyses. However, for these correlations, the Thorndike correction was not applied, as range restriction could feasibly occur on both variables.

Results

The results of the correlation analysis are shown in Table 2. Further details, including standard errors are given in Table A1 of Appendix 1. Red-blue shading has been added to Table 2 to help show the patterns more clearly, with red shading indicating stronger correlations and blue shading indicates weaker correlation (scaled relative to observed values). The GCSE subjects within Table 2 are sorted by their correlation with PISA science (the major PISA domain).

The strongest correlations observed were for PISA maths and science. Across the entire table, the largest correlation was between PISA maths and maths GCSE ($r = 0.777$), whilst the next strongest was between PISA science and highest science grade ($r = 0.760$). Hence, for maths and science, the strongest correlation was with the most relevant GCSE subject. However, in absolute terms, correlations were only moderately strong. To provide context for this finding we note that the correlation between KS2 and GCSE performance in maths for

[10] Ideally both sets of analyses would have been based on 2016 as this was the year in which PISA pupils would have taken their GCSEs. However, the analysis of English Language was based on 2017 as it was the first year where assessment was entirely exam based so that controlled assessment scores were not carried forward from the first attempt to the resit.

[11] Candidates are most likely to resit if they were close to achieving a higher grade, particularly so at the boundary of a ‘pass’, i.e. C or above in old GCSEs, or 4 and above in reformed GCSEs. Candidates achieving the highest grades would be unlikely to resit at all, whilst those achieving the lowest grades may seek alternative qualifications. Hence, the population taking resits is likely to be dominated by candidates in a small part of the overall grade distribution.

this same group of pupils was 0.748 (Appendix 1, Table A7). In other words, despite being separated by 5 years, the correlation between KS2 maths and GCSE maths is nearly as high as that between PISA and GCSE.

To provide further context for this finding, correlations between GCSE maths grades achieved in June 2016 and in resits the following November were examined using data from those candidates taking their maths GCSE with OCR. The time difference between the first and second attempts at GCSE is similar to that between when students take PISA tests and when they sit GCSE exams, so these correlations indicate expected variation in performance over time. For 2016 maths grades, the Pearson correlation (with Thorndike correction) was 0.903[12]. This implies that the correlation that we observe between PISA and GCSE maths would be considerably higher if time was the only thing that changed between the two assessments.

After this, the strongest maths and science correlations were with the mean GCSE grade (both r = 0.753). For PISA reading and collaborative problem solving (CPS), the strongest correlation was with mean GCSE grade (reading r = 0.741; CPS r = 0.611), indicating that performance in these domains was more strongly correlated with overall attainment than with any particular subject.

For PISA maths, relatively strong correlations were also seen with science subjects, notably with the highest science grade (r = 0.746) and physics (r = 0.732). This reflects a relatively high correlation between these subjects within GCSEs themselves[13]. As may be expected, PISA science also showed relatively strong correlations with science subjects: the strongest was with core science (r = 0.748), followed by additional science (r = 0.734). However, PISA science showed stronger correlations with maths (r = 0.728) and geography (r = 0.714) than with any separate science, of which physics was the strongest (r = 0.699). It must be noted though that when these estimates of correlation are considered in the context of their associated precision (see the standard errors in Appendix 1, Table A1), these differences are not large. For example, for PISA science, the relatively weak correlation with GCSE chemistry (0.659) displayed an estimated 95 per cent confidence interval of 0.732-0.586, thus overlapping the coefficients for most higher-ranked subjects. Hence, subject rankings should be interpreted with some caution.

Correlations with PISA reading were slightly weaker, and there was no clear pattern to which subjects were most strongly correlated. Most notably, subjects that may have been expected to correlate with reading scores were actually relatively weak: English (r = 0.680) and English literature (r = 0.637) showed weaker correlations than highest science grade (r = 0.708), history (r = 0.696), core science (r = 0.692) and geography (r = 0.687). This indicates that, despite the association shown by previous analyses (DfE, 2017), PISA reading clearly measures very different skills to those assessed by GCSE English. In particular, this may

[12] Without Thorndike correction, r = 0.750. Note that Thorndike corrections made a large difference because resits are most often attempted by candidates in a narrow attainment range just below the grade C boundary, leading to extreme range restriction. For 2017 English grades, the corrected correlation between original grade and resit grade was 0.911 (uncorrected, r = 0.717).

[13] Exploratory analysis indicated that the correlations of GCSE maths grade with highest science grade and physics were each just above 0.85 (after applying a Thorndike correction).

relate to the fact that none of the PISA domains measure important skills such as essay writing, that are both taught and assessed within GCSE English. Further consideration of these results and the way they relate to the nature of the items used in PISA reading will be given later.

One of the most striking features of Table 2 is the fact that CPS correlations with GCSE performance were substantially lower than those from other domains. Maths (r = 0.593) showed the strongest correlation of individual subjects, but this was still relatively weak. Hence, none of the GCSE subjects considered in this study particularly measure the skills assessed by PISA collaborative problem solving. Furthermore, even the correlation with mean GCSE grade (r = 0.611) is relatively low. For example, this is roughly the same size as the correlation between KS2 reading and PISA maths (r = 0.601). This indicates that, overall, assessment at GCSE is not particularly strongly associated with pupils' skills in collaborative problem solving. This finding is of interest as collaborative problem solving is the PISA domain in which England's performance is the strongest relative to the OECD average (Jerrim and Shure, 2017). In other words, England's strongest performance comes in a domain that does not appear to relate to what is actually taught and assessed as part of GCSEs. We will return to this finding later in the discussion.

The strongest correlation with any KS2 metric was between PISA maths and KS2 maths marks (r = 0.740), showing that attainment at age 11 correlated reasonably strongly with PISA maths scores, just slightly less so than GCSE maths. In general, KS2 correlations were weaker than equivalent GCSE subjects. However, there were some notable exceptions: PISA science showed a slightly stronger correlation with KS2 English reading marks (r = 0.648) than with GCSE English (r = 0.628), whilst PISA reading displayed as high a correlation with KS2 English reading (r = 0.638) as with GCSE English literature (r = 0.637).

For maths, science and reading, the weakest correlations were with Spanish and Art & Design, with coefficients typically at or below 0.5. These subjects also showed weak correlations with CPS but coefficients were under 0.4. Further subjects showed similarly weak correlations with CPS, including resistant materials, French and German.

Further analysis to verify that the above results were robust to different methods of calculating correlations, and to restricting analysis to pupils that appear to have made a reasonable level of effort in the PISA tests, was also undertaken; details are given in Appendix 1. Overall, these checks did not reveal any differences with the major conclusions listed above.

Table 2. Weighted Pearson correlations with updated plausible values, with Thorndike correction applied. Variable gives the variable against which PVs were correlated; where only a subject name is given, this indicates the GCSE grade in that subject. N gives the number of students included. Table is sorted with mean GCSE attainment first, then GCSE subjects, and then KS2 metrics; within each, table is sorted in order of Science correlation strength. Red shading indicates stronger correlation; blue shading indicates weaker correlation (scaled relative to observed values). For standard errors of correlations, see Appendix 1, Table A1.

Variable	N	Science	Maths	Reading	CPS
Mean GCSE grade	4,912	0.753	0.753	0.741	0.611
Highest Science	4,677	0.760	0.746	0.708	0.587
Core Science	3,037	0.748	0.714	0.692	0.550
Additional Science	2,779	0.734	0.706	0.660	0.525
Maths	4,778	0.728	0.777	0.672	0.593
Geography	2,232	0.714	0.698	0.687	0.583
D&T: Textiles Technology	199	0.700	0.677	0.656	0.512
Physics	1,563	0.699	0.732	0.604	0.492
Average across separate sciences	1,544	0.698	0.729	0.626	0.499
History	2,373	0.696	0.675	0.696	0.561
Statistics	390	0.682	0.694	0.672	0.548
Biological Science	1,580	0.681	0.696	0.624	0.494
Business Studies	746	0.672	0.701	0.674	0.531
Chemistry	1,566	0.659	0.699	0.607	0.479
English	4,735	0.628	0.625	0.680	0.534
Home Economics: Child Development	152	0.624	0.583	0.604	0.469
English Literature	4,287	0.613	0.592	0.637	0.534
Music	363	0.601	0.594	0.567	0.417
D&T: Food Technology	311	0.589	0.606	0.635	0.494
Media, Film and Television Studies	450	0.588	0.577	0.609	0.472
German	530	0.587	0.631	0.579	0.397
Religious Studies	2,447	0.575	0.564	0.595	0.481
Physical Education	1,102	0.571	0.579	0.538	0.454
D&T: Graphic Products	217	0.559	0.614	0.547	0.466
D&T: Resistant Materials Technology	446	0.552	0.582	0.531	0.371
French	1,387	0.535	0.551	0.538	0.376
Drama	552	0.534	0.509	0.557	0.400
Information Technology	1,167	0.524	0.541	0.532	0.448
Art and Design	1,334	0.500	0.475	0.507	0.328
Spanish	930	0.476	0.484	0.495	0.379
KS2 English: marks in reading test	4,564	0.648	0.601	0.638	0.572
KS2 maths: total test marks	4,575	0.645	0.740	0.553	0.512
KS2 maths: teacher-assessed NC level	4,628	0.619	0.685	0.550	0.503
KS2 science: teacher-assessed NC level	4,626	0.610	0.642	0.533	0.504
KS2 English: teacher-assessed NC level	4,628	0.603	0.610	0.592	0.524
KS2 English: marks in writing test	4,564	0.519	0.524	0.549	0.441

Relationships with mean GCSE

To further explore some of the patterns identified above, relationships between PISA scores and mean GCSE attainment were explored graphically. First, the relationship between mean GCSE (rounded to the nearest grade) and scores in each PISA domain was plotted (Figure 1). Across all domains, PISA scores increased as mean GCSE grade increased. For students achieving A* or A on average, median PISA science abilities were larger than those for all other domains, indicating that students who attained the highest GCSE grades did particularly well on PISA science, relative to performance on other domains. Conversely, for students achieving D to F on average, median abilities on collaborative problem solving ("C" within the Figure legend) were higher than those in other domains. Hence, lower attaining students appeared to perform disproportionately well on PISA CPS. Further, the spread of abilities observed for CPS was larger than that of other domains: students attaining C on average could achieve scores as high as the highest achieved by A* average students, whilst even those attaining F on average could have abilities higher than the median score of A* average students.

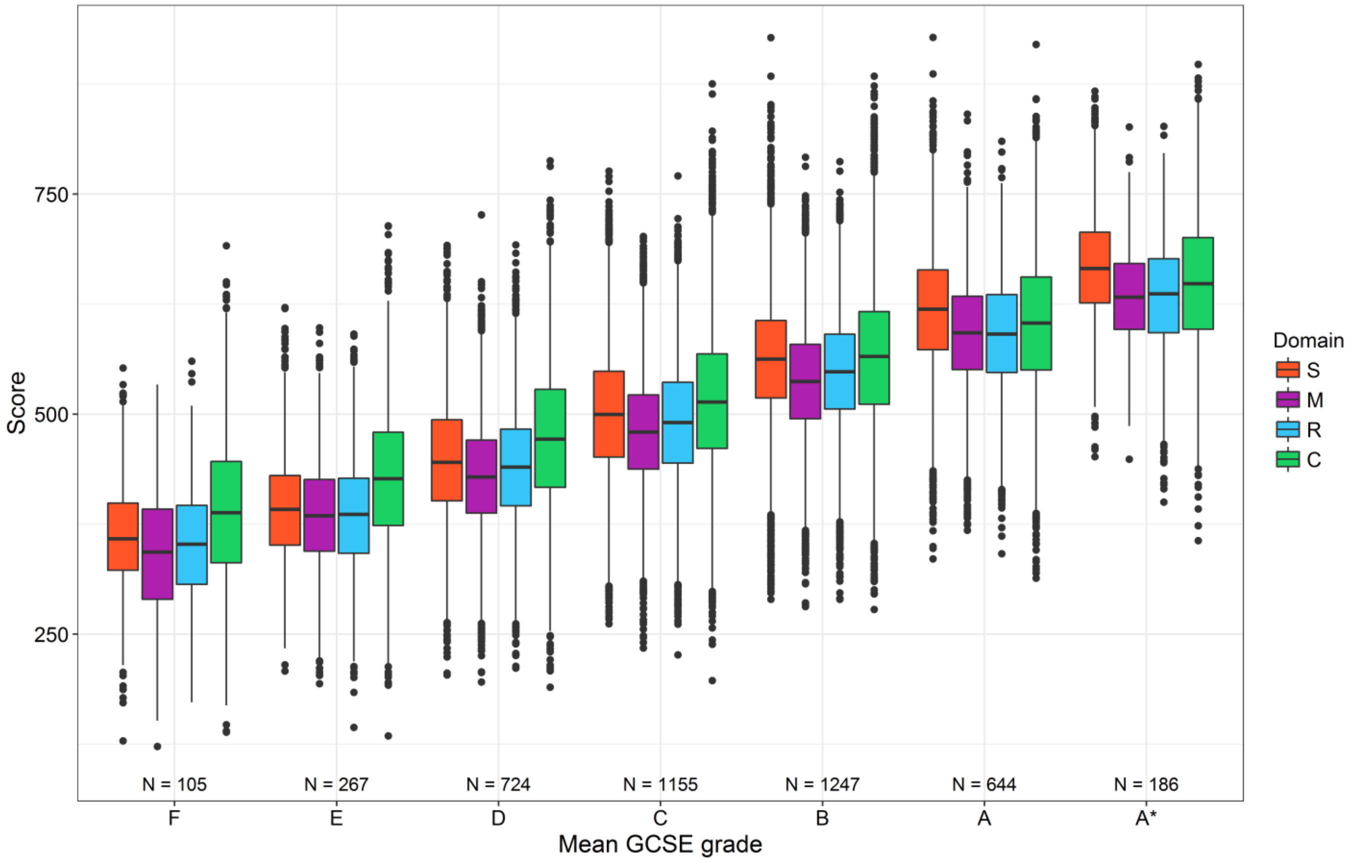

Figure 1. Boxplots of PISA scores for each mean GCSE grade. Scores on all 10 PVs for each domain are plotted. Mean GCSE grade is the average KS4 points per entry, rounded to the nearest full grade. Boxes indicate interquartile range and median; whiskers indicate 1.5*interquartile range; points are outliers. N values listed indicate the number of students achieving that grade. Note that due to small sample sizes, students attaining U (N=1) or G (N=19) on average are not included on this graph.

In order to illustrate these findings, PISA abilities were divided into quintiles for each domain (this was done separately for each of the ten PVs in each domain). Students were then classified as occurring in the top quintile or not. The proportion of students in each mean

GCSE grade who were in the top score quintile was plotted (Figure 2). As expected, higher average grades had higher proportions of students in the top quintile: over 80% of A* average students had PISA science and maths abilities in the top quintile, whilst over 75% had PISA reading scores in the top quintile. At grade A average, proportions declined to around 55% for science and maths, and 50% for reading, whilst by grade B, proportions declined to around 20-25%. However, CPS showed a different pattern: only around 65% of A* average students and only around 45% of A average students were in the top quintile, but around 10% of C average, 5% of D average and 1% of E and F average students were in the top quintile. Hence, for CPS, it was possible that pupils with low GCSE grades on average could have high abilities in collaborative problem solving, whereas for the other domains the tendency for pupils with the highest GCSE grades to also have higher PISA abilities was stronger. To put this another way, pupils may have a high ability in collaborative problem solving that is not captured within their GCSE assessments.

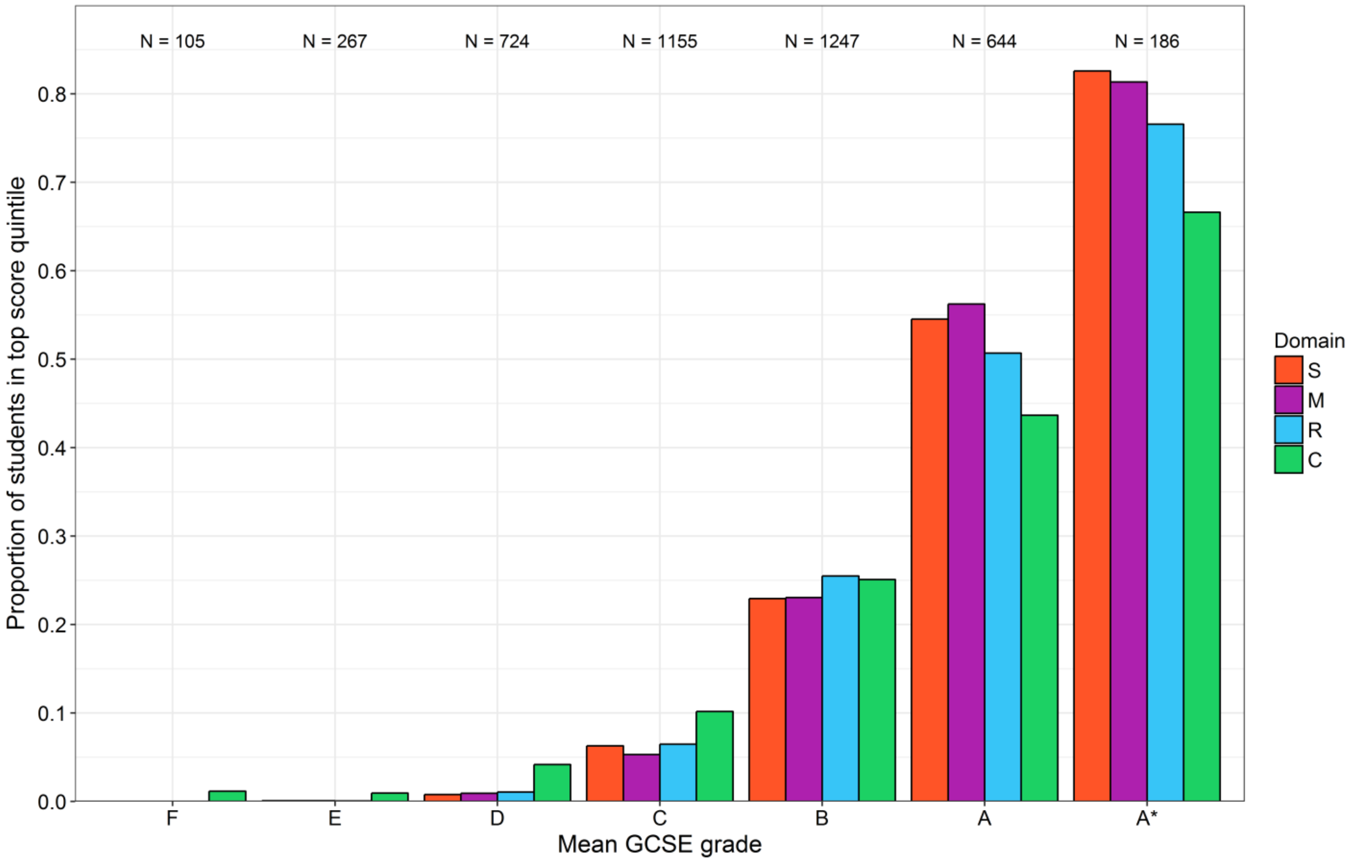

Figure 2. The proportion of students achieving each mean GCSE grade who also achieved PISA scores in the top quintile. Quintiles were calculated separately for each PV within each domain, but all are taken into account on this plot. N values listed indicate the number of students achieving that grade. Note that due to small sample sizes, students attaining U (N=1) or G (N=19) on average are not included on this graph.

Modelling effects of subject choice on PISA performance

The results in the previous section considered the correlation between achievement in various GCSE subjects and pupil ability as measured by the PISA tests. As such, they could be considered to be concerned with whether the PISA tests are measuring the same skills as are *assessed* by different GCSEs. This next section will consider the extent to which abilities in the PISA tests relate to which subjects pupils choose to study during KS4. As such, this new analysis could be seen as exploring the extent to which the PISA tests measure skills that are *learnt* (or taught) as part of different GCSE subjects. As we will see, such a simplistic interpretation of the relationship between subject choices and PISA scores is subject to a number of caveats. However, the results may still provide a prompt for reflection on the different skills that are acquired by pupils with differing GCSE subject choices. A second aim of this analysis is allow us to explore whether taking GCSE subjects, such as Art & Design, that teach skills that are self-evidently *not* measured by the PISA assessments, has a detrimental impact on the pupil abilities that are reported by the OECD.

Method

Deriving variables

To enable the investigation, variables were derived from the NPD data indicating whether a student had taken a particular subject: for each of the subjects considered in the correlation analysis, a variable was coded as 1 if the student had taken it (that is, a grade between A* and U was recorded in the NPD) or 0 if they had not.

As indicators of taking the various science subjects were strongly collinear[14], a single 'separate sciences' variable was derived: if the number of separate sciences taken (i.e. biology, chemistry and physics) exceeded the number of other sciences taken (i.e. core or additional science), the student was classed as taking separate sciences as opposed to core and additional science.

Only 8.8% of students had no grade recorded for GCSE English, only 8.0% had no grade recorded in GCSE maths, and only 10% of students had no grade recorded for any science GCSEs. Furthermore, it was considered likely that in these cases pupils were making use of qualifications other than GCSEs to study these subjects rather than not studying them at all. Consequently, rather than including participation in these GCSE subjects as binary variables, which would not be particularly informative, the sample was restricted to the 4,348 pupils with GCSE grades in all of English, maths and science. Then, the actual grades achieved by pupils in these subjects were included in one model formulation to indicate concurrent attainment (see below). This left 21 variables indicating specific subject choices.

In some models, variables indicating concurrent or prior attainment were included (see below), along with the PISA variable indicating economic, social and cultural status (hereafter, "ESCS"). These continuous variables were centred before use.

[14] Of the 1,593 students who took any of the separate science GCSEs, 96.9% took all three separate sciences, and 98.7% took at least two. Hence, any student taking any separate science was highly likely to be taking all three, and almost certain to be taking at least one other. Of the 3,044 students taking core or additional science, 91.1% were taking both core *and* additional science, so these subjects were also strongly collinear.

Modelling approach
Models were fitted by adapting code from the ‘intsvy’ R package (Caro and Biecek, 2017). The regression in this package fits simple linear models, accounting for imputation and sampling variance. However, simple linear models may not adequately account for clustering within schools, even with PISA weights included. Indeed, Akaike Information Criterion (AIC) values for models fitted with ‘school’ as a random factor were substantially superior to those for standard linear models. Consequently, the function was updated to fit mixed models using the ‘lme4’ R package (Bates et al., 2015). These models included ‘school’ as a random factor, and thus, at least partially, accounted for the influence of individual schools on pupils’ PISA abilities. See Appendix 3 for more information on the models fitted.

The first models only included binary variables indicating whether or not a student took a subject. Model coefficients from this approach therefore indicated the raw effect of taking each subject for each PISA domain, not controlling for any student background or attainment effects.

The second models included binary subject variables, ESCS, student gender, and a range of variables indicating concurrent attainment: highest GCSE science grade, GCSE maths grade, GCSE English grade, mean GCSE grade, and the number of GCSE and equivalent entries. Highest science grade, maths grade, English grade and mean GCSE grade were entered as quadratic terms to account for non-linear relationships with PISA scores[15]. These models therefore indicated differential effects of taking each subject for a fixed level of overall KS4 attainment.

A final model was run that included variables indicating *prior* attainment at KS2 in place of those measuring concurrent attainment at GCSE. There was less evidence of a non-linear relationship between KS2 attainment and PISA scores, so these were entered as linear terms. These models indicated differential effects of taking each subject after accounting for KS2 attainment, so could be used to explore the ‘progress’ in PISA produced by taking particular subjects. Due to some missing cases in KS2 attainment variables, these models were further restricted to a sample of 4,053 students. Details of the subject choices included and the number of pupils taking them that were included in the different models are given in Table 3.

Model coefficients and standard errors returned were used to produce a t value, which was used to test whether the coefficient was significantly different from 0. To reduce the risk of spurious findings due to the very large number of coefficients being examined, significance testing was carried out at the 1% level.

[15] Preliminary investigations suggested that linear forms did not fit the relationship between attainment and PISA scores well at high and low values, whereas including a quadratic term improved the fit; cubic terms provided little further benefit over the quadratic term.

Table 3. Details of the GCSE subject choices including in modelling along with the number of pupil taking each subject within the data set. The number when the data is restricted to those with available KS2 data is also shown.

GCSE choice	Number taking GCSE choice in analysis	Number taking GCSE choice in analysis with KS2 available
Art & Design	1,158	1,064
Business Studies	683	653
D&T: Food Technology	286	272
D&T: Graphic Products	206	198
D&T: Resistant Materials Technology	387	357
D&T: Textiles Technology	187	179
Drama	497	469
English Literature	3,875	3,623
French	1,258	1,176
Geography	2,010	1,884
German	494	474
History	2,120	2,004
Home Economics	140	131
Information Technology	1,052	978
Media, Film and TV	395	374
Music	325	303
Physical Education	998	933
Religious Studies	2,214	2,069
Separate sciences	1,483	1,406
Spanish	849	801
Statistics	341	332

Results

Exploring score differences

Before looking at the results of regression analyses, some descriptive analysis of the differences between pupils taking different GCSE subjects is shown in Figure 3. For this Figure, the differences between each student's abilities and the sample means within each PISA domain were calculated, such that if a student performed better than average in a particular domain the score was positive. The quantiles of these differences are shown in Figure 3 for those pupils that took each GCSE subject.

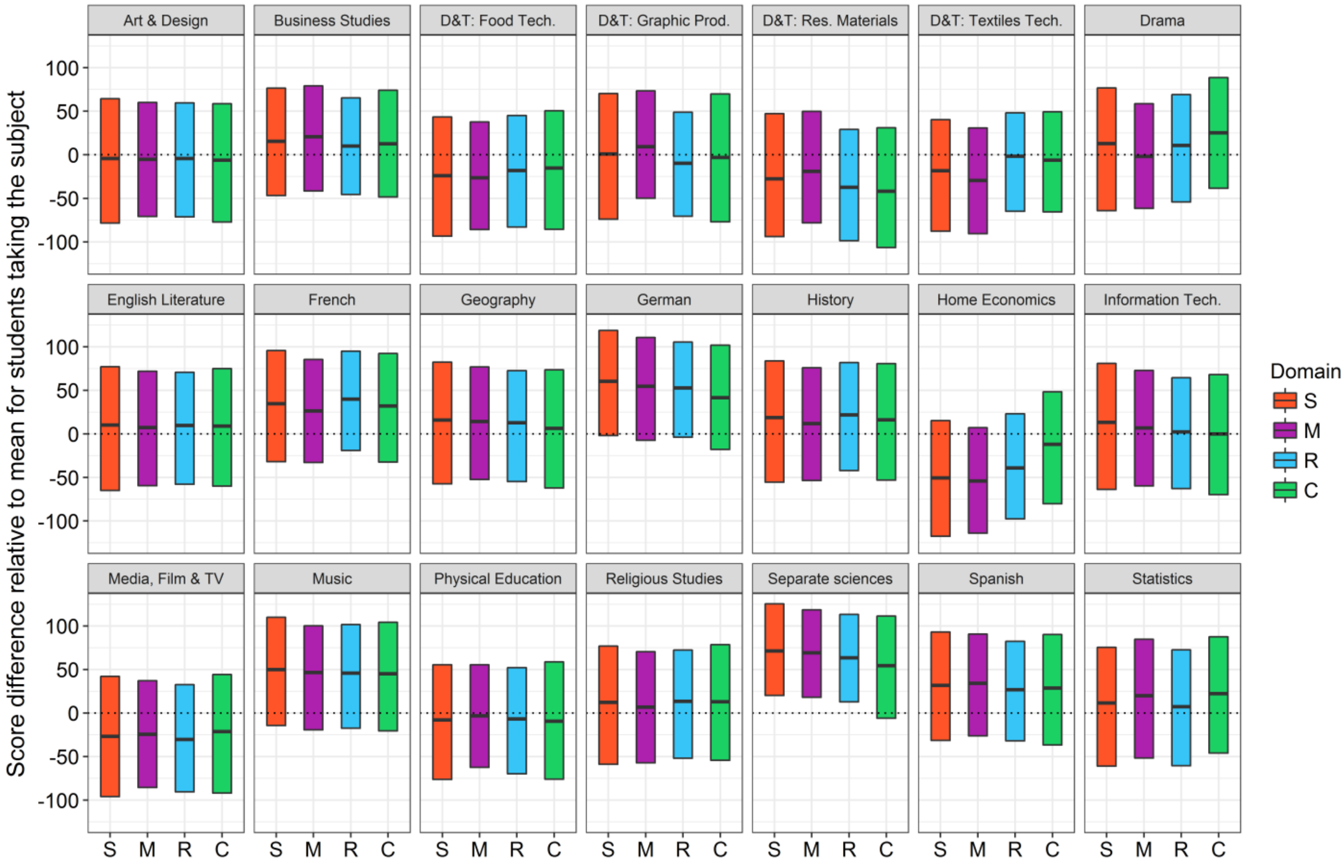

Figure 3. PISA scores for students who took each GCSE subject considered relative to the overall mean. Differences were calculated separately for each PV, but all are plotted here. Boxes indicate 25th percentile, median and 75th percentile; full ranges are not plotted for clarity of interpretation. The dotted lines indicate 0 difference, i.e. the mean score in England overall.

There are two aspects of Figure 3 that are of interest. The first is the general level of difference between pupils taking each GCSE subject and the overall average across the four PISA domains. For example, for separate sciences the median difference is well above zero across all four domains. This simply indicates that the pupils who choose to take separate sciences tend to be of generally higher ability (as measured by PISA) than those that do not. In fact, in the case of separate sciences, the large gap between those taking the subject and others is of similar magnitude to the difference in national means between the UK as a whole and Singapore (the top performer in PISA). However, given that we have not controlled for any form of attainment (prior or concurrent), this may simply reflect pre-existing differences in the characteristics of pupils who take different subjects (see Carroll & Gill (2017) for examples). Of slightly more interest are differences between domains within particular GCSE subjects. For example, this allows us to see that within those taking separate sciences at GCSE, the PISA domain showing the largest difference is itself science. One possible explanation for this result is that the extra time spent studying science at GCSE translates into particularly strong performance in this PISA domain. Of course, alternative explanations are possible. For example, it may be that those pupils who choose to study separate sciences are those with a particular aptitude for, or interest in, the subject and that it is this, rather than the additional time they have spent being taught the subject, that leads to the observed differences seen in Figure 3. Again, it may be that this particularly large difference in science performance can be explained by the prior attainment of pupils or by some other measured characteristic. This last explanation will be explored further in the regression modelling.

Models not accounting for attainment

One weakness of the descriptive analysis above is that no attempt is made to calculate the statistical significance of difference between subjects or between domains. This means that it is difficult to know which differences may simply be the result of the relatively small numbers of pupils studying particular subjects combined with normal levels of variation in ability between pupils. To address this issue, initial regression models simply considered whether students took each subject, but did not account for any background characteristics or attainment (Appendix 1, Table A8). The coefficients from these models are also shown in Figure 4. These coefficients reflect the mean PISA score difference between those taking each subject and those not, after controlling for whether or not pupils entered the other subjects included in this model. As can be seen, most subjects tended to have positive coefficients. This probably relates to the fact that higher ability pupils tend to take greater numbers of GCSEs so that GCSE entry in any subject will, in most cases, be associated with higher ability. As with Figure 3 earlier, we note that there are several subjects, such as separate sciences or modern languages, with large positive coefficients across all domains. However, this may well be explained by the background characteristics of students entering the subjects and is not of particular interest. Of perhaps more interest in Figure 4 are the differences between coefficients for different PISA domains within a given GCSE subject. In particular, we might be interested in the 5 GCSE subjects where the 99 per cent confidence intervals for the coefficients do not all overlap with an average value (indicated by the dotted line). In these cases the results suggest that taking the GCSE given subject has a particular effect on a given PISA domain. Specifically we might note that:

- The estimated difference between those taking separate sciences and those not is particularly large when looking at PISA science (86 points) compared to the coefficients for other domains (80, 69 and 60 for Maths, Reading and CPS respectively).
- The coefficient for the effect of taking French GCSE on PISA reading (51 points) is larger than the average coefficient for this subject (41 points), whereas the coefficient for PISA maths (31 points) is noticeably smaller.
- The coefficient for the effect of taking History GCSE on PISA reading (38 points) is larger than the average coefficient for this subject (28 points).
- The coefficient for the effect of taking D&T: Resistant Materials GCSE on PISA maths (8 points) is larger than the average coefficient for this subject (-5 points).
- The coefficient for the effect of taking Information Technology on PISA science (33 points) is larger than the average coefficient for this subject (22 points).

In each of the cases above, the results may possibly indicate the specified GCSE subject has a particular benefit in terms of acquiring the skills in the given PISA domains. In other words, studying separate sciences or Information Technology builds ability in PISA science, studying French or history helps acquire reading skills, and studying resistant materials supports ability in PISA mathematics. However, it remains possible that these findings could also be explained by differences in the particular aptitudes or background characteristics of students choosing to study these subjects. For example, it is interesting that the particularly strong effect of French on PISA reading is not repeated in either of the other modern languages (German and Spanish). Regarding the finding relating to Information Technology, one possible explanation is that the computer-based format of the PISA science tests means that familiarity with using computers built up whilst studying the Information Technology GCSE may support achievement in this domain, although it's difficult to understand why science would benefit more than other domains.

Two further subjects where the 99 per cent confidence intervals for the coefficients on different PISA domains only just overlapped with the average are worth mentioning. Specifically, Figure 4 shows that the coefficient for GCSE drama on PISA collaborative problem solving was larger (30 points) than its average coefficient (16 points). This may indicate that studying drama is particularly beneficial for collaborative skills – an intuitively plausible possibility. We can also see that for GCSE geography the coefficient for collaborative problem solving is smaller (10 points) than for the average coefficient (19 points). This may indicate that studying this GCSE is not particularly helpful in developing collaborative skills, or, perhaps more likely, that studying geography is rather more useful in developing skills in science, maths and reading.

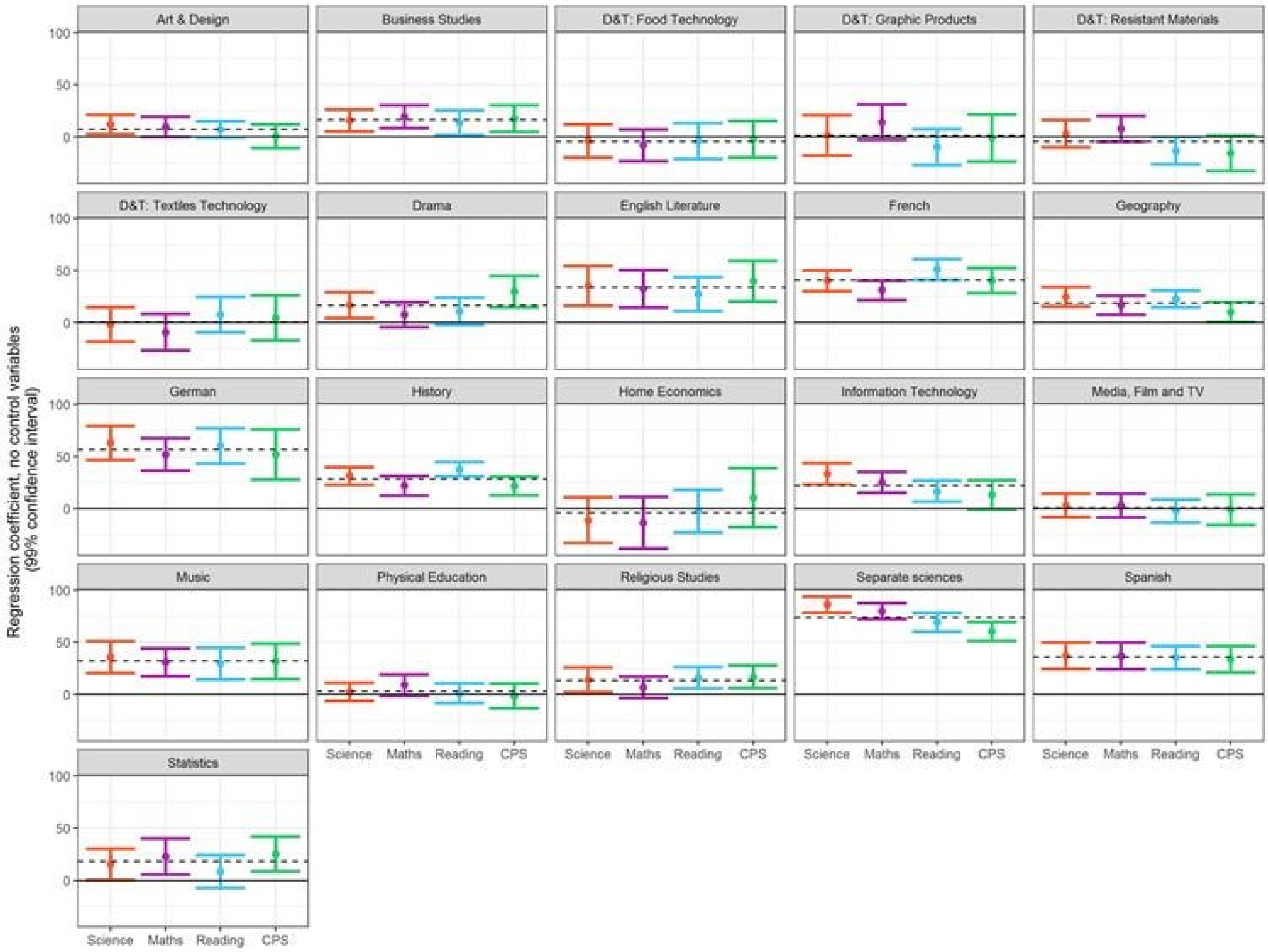

Figure 4. Coefficients from linear mixed models with PISA score as a function of taking/not taking GCSE subjects, and no control variables. Coefficients are displayed with 99 per cent confidence intervals. The average coefficient size for each GCSE subject is shown by a dotted line.

Models accounting for concurrent attainment

Whilst the coefficients in Figure 4 could indicate the contribution of each subject to PISA scores, the models did not account for attainment or background. The next models controlled for concurrent attainment by including variables indicating performance in GCSE English, maths and science, overall GCSE performance, and the total number of GCSEs taken. Variables indicating gender and economic, social and cultural status were also included to help control for student background. These models therefore indicated the effect of taking each subject relative to other students with equivalent levels of achievement overall at GCSE. In other words, they explored whether there were any subjects where the students who studied them performed disproportionately well in PISA given their overall attainment at GCSE.

Note that concurrent attainment had a highly significant effect on PISA scores (Table 4; standard errors in Appendix Table 8). All of the various main measures of concurrent attainment had positive coefficients with each of the PISA domains[16]. Although it is not the main focus of our research, it is worth noting in passing that male pupils performed better in PISA than might be expected from their GCSE performances in both science and maths. This result fits with previous research showing that whilst male pupils tend to outperform females in PISA maths no such male advantage is found for GCSE assessments (Bramley, Vidal Rodeiro and Vitello, 2015). In contrast to results in maths, female pupils tended to display the highest achievement in collaborative problem solving relative to their GCSE achievement. Socio-economic status (as measured by ESCS) had no significant effect for any of the domains once we accounted for concurrent attainment.

Unlike the earlier models without controls, few subject coefficients were significantly different from 0 (Table 4). The biggest effects were seen for separate sciences, which had a large positive effect for both science (25 points) and reading (25 points), but also significant positive effects for maths (19 points) and CPS (16 points). This indicates that, even after accounting for their high achievement in their GCSEs, as well as for gender and socio-economic status, students studying separate sciences performed particularly well across all of the PISA domains. Interestingly the size of these coefficients was similar across the different domains. If we interpreted these coefficients in a purely causal manner they would suggest that, without even needing to improve GCSE results, if the two-thirds of pupils studying combined science moved to studying separate sciences, the performance of the UK in PISA could rise to being close to that of Korea (ranked between 4th and 9th across different domains). However, whilst this coefficient may indicate that studying separate sciences causes students to particularly develop the skills measured by PISA, it is also possible these effects may be explained by unmeasured aptitudes of the pupils choosing to study separate sciences in the first place.

Some checking of alternative explanations for the large and positive science coefficients was performed. Firstly, as shown in Table A9 of Appendix 1, the analysis was rerun including extra variables derived from the PISA pupil questionnaire controlling for pupils' enjoyment of, broad interest in, self-efficacy in, and instrumental motivation in science. However, inclusion of these variables in the model was found to hardly alter the coefficients for separate

[16] Some of the quadratic terms had negative coefficients indicating a slightly curved relationship. However, in all cases a positive relationship was observed overall.

sciences from those reported in Table 4. Other analysis (not shown in this report) investigated the impact of including explicit controls for whether pupils attended selective or independent schools. Again, this was found to make no difference to the reported results. Having said this, whilst some effort has been made to explore alternative, non-causal explanations for the science coefficient, we cannot completely exclude the possibility that the reported effects are the result of some pupil characteristics not captured by the data.

Table 4. Coefficients from linear mixed models with PISA score as a function of taking/not taking GCSE subjects, concurrent attainment and student background. Models were fitted with 'school' as a random effect. Concurrent attainment variables were entered in quadratic form, so both linear and quadratic components are reported. Coefficients are the mean across all 10 PVs, and standard errors take into account sampling and imputation variance. Coefficients significantly different from 0 at P < 0.01 are indicated in bold; standard errors are presented in Appendix 1, Table A8.

	Science	Maths	Reading	CPS
(Intercept)	**496.6**	**482.4**	**491.5**	**525.92**
Art & Design	4.5	5.6	-3.4	-11.0
Business Studies	-1.9	2.6	0.6	3.0
D&T: Food Technology	-2.3	-3.6	-4.7	-5.3
D&T: Graphic Products	-5.7	5.6	-11.3	-5.8
D&T: Resistant Materials	-5.9	-2.6	-9.0	-14.6
D&T: Textiles Technology	-8.4	-12.1	-7.4	-13.2
Drama	8.5	2.1	0.1	**18.7**
English Literature	0.8	-0.3	0.2	14.5
French	-1.1	-7.2	**15.0**	3.8
Geography	0.2	-4.4	3.8	-9.0
German	**12.6**	3.8	**19.2**	11.1
History	4.0	-1.6	**13.1**	-1.1
Home Economics	-3.3	-2.7	-6.9	2.72
Information Technology	6.9	-1.6	3.3	-2.5
Media, Film and TV	-3.4	-2.2	-5.5	-6.0
Music	9.9	7.7	7.9	9.7
Physical Education	**-9.3**	-2.5	-4.9	-8.9
Religious Studies	-3.4	-7.8	-0.1	1.1
Separate sciences	**25.0**	**19.5**	**25.3**	**16.2**
Spanish	-3.7	-2.2	-0.1	-2.1
Statistics	-2.6	3.8	-3.2	10.4
Mean GCSE grade	**9.2**	6.4	**11.5**	3.3
Mean GCSE grade squared	-1.3	-2.5	-1.1	-0.3
GCSE English grade	**5.8**	**6.0**	**15.4**	7.5
GCSE English grade squared	1.1	0.5	**1.8**	1.2
GCSE Maths grade	**14.0**	**27.1**	**9.0**	**17.7**
GCSE Maths grade squared	**1.0**	**2.2**	1.0	0.8
Highest GCSE science grade	**23.6**	**11.7**	**10.3**	**10.8**
Highest GCSE science grade squared	1.2	**1.5**	-1.5	-1.6
Number of GCSE and equivalent entries	4.0	2.5	-0.4	2.9
ESCS	-1.7	-0.1	-0.7	-0.1
Male	**16.4**	**20.8**	-8.5	**-16.8**

Studying GCSE German had a positive association with PISA science (13 points), whilst PE had a significant negative association (-9 points). The reasons for these patterns are unclear. Given that the effect for German was not repeated for either French or Spanish it may be that this relates to skills or aptitudes of students taking German that were not accounted for in the model. A possible explanation for the PE effect is that it may reflect an impact of spending less time on classroom-based learning (note also that although non-significant, all D&T subjects also showed negative coefficients for PISA science). Note that we make no value judgement as to whether this is a good or a bad thing; we are merely noting that such learning may not be actively beneficial for developing the abilities deemed as important by the PISA tests. Clearer patterns were seen for PISA reading, with positive effects of French (15 points), German (19 points) and history (13 points). Given that all of these subjects have strong elements of reading and comprehension this may relate to the skills that are acquired by studying them, although, as with earlier findings, alternative explanations are possible. Finally, drama showed a positive effect for CPS (19 points), which may possibly be linked to the nature of CPS testing: it intends to test collaboration by presenting a sequence of events that students must respond to, and studying drama could help to develop such skills.

One of the main purposes of conducting this analysis was to see whether PISA disadvantages students who take large number of arts subjects. From the results in Table 4 there is no sign of this being the case. Indeed, although falling short of statistical significance the majority of the coefficients for both Art & Design and for Music were positive. This indicates that, given equally good performance at GCSE overall, students studying these subjects perform no worse within PISA tests than those that had made alternative, perhaps less artistic, choices.

Models accounting for prior attainment

The final models fitted included prior attainment, indicated by KS2 performance variables, rather than concurrent attainment (Appendix 1, Table A8). All of the various measures of prior attainment at KS2, whether from tests or from teacher assessments, were positively associated with PISA abilities across all domains; with the exception of scores on the KS2 writing test. This finding is interesting, since it supports the suggestion, noted earlier, that ability in essay writing is not captured at all by the PISA tests. As such, having controlled for key stage 2 performances in maths, reading and science, pupils' scores on the writing test are not at all predictive of later performance in PISA[17].

The effect of gender on ability in the PISA domains was similar to that shown in the models accounting for concurrent attainment. Male pupils tended to outperform females with similar prior attainment in the PISA domains of maths and science. On the other hand, females tended to outperform similar males in reading and collaborative problem solving.

More subjects showed significant effects in these models, perhaps indicating that the effects of student characteristics were not fully accounted for by considering only KS2 attainment. However, the coefficients can be taken as indicating the 'progress' in PISA scores produced by taking each subject, relative to a student with the same level of KS2 attainment who did not take the given GCSE subject. Again, separate sciences had the largest effect across all domains (science 45; maths 35; reading 37; CPS 26). After this, German also displayed large, positive coefficients for science (31), maths (20) and reading (31), but showed no significant coefficients for collaborative problem solving. Considering other languages,

[17] Note that published reliability coefficients for KS2 writing tests (Newton, 2009, page 201) indicate that they are almost as reliable as the KS2 reading tests. This means that differences in reliability cannot be used as an explanation for these results.

French showed significant positive coefficients for science (14), collaborative problem solving (14) and, in particular, for reading (25) but Spanish showed no significant effects. Of the humanities, history and geography showed significantly positive coefficients for science and reading. Of the arts, English literature and drama showed positive effects for collaborative problem solving (17 and 19 respectively), whilst music showed positive coefficients for science and maths (19 and 14 respectively). Information Technology showed positive coefficients for science, maths and reading, which may reflect the shift toward computer-based testing in PISA (although CPS is also assessed via computer-based testing, and was not significant at the 1% level). Finally, the negative association between studying PE and PISA science scores seen in the concurrent attainment model was also seen here.

Graphical illustration of the "impact" of taking separate sciences

The mixed effect modelling described in the previous section showed that the largest relationship between GCSE subject choice and PISA abilities related to whether pupils studied separate sciences, as opposed to combined sciences. In order to illustrate this relationship, Figure 5 shows how the distribution of plausible values in PISA maths varied according to pupils' grades in GCSE maths and whether they studied combined or separate sciences at GCSE. The figure shows how, even amongst pupils with equally good performance in GCSE maths, those who studied separate sciences tended to achieve higher scores in PISA maths than those who studied combined sciences. There is some indication that this difference was more prominent amongst those with lower grades at GCSE. Specifically the differences in medians between those taking separate and combined sciences was around 40 PISA points amongst those with grade D, around 25 points amongst those with grade A, and 10 points amongst those with grade A*. This may suggest that the separate science effect is partially caused by the impact of students being entered and prepared for different tier exams as the difference is smaller amongst students with grades B and above who were definitely entered for higher tier GCSE papers than amongst lower grades[18]. Having said this, it is clear that some gap persists across all grades (and hence both tiers). Also, given the restricted sample sizes amongst those with different grades, these comments should be treated with some caution[19]. Similar charts could be produced for PISA science and PISA reading.

[18] For maths GCSE exams in England students could either enter higher tier examinations where grades A*-E were available or foundation tier exams where grades C-G were available.

[19] Amongst those taking combined science there were 498, 1124, 634, 263 and 89 pupils included in this particular analysis with grades D-A* respectively. Amongst those taking separate sciences there were 43, 238, 409, 452 and 404 pupils with grades D-A* respectively. Very few candidates taking separate sciences had GCSE maths grades of E or below and so these grades were not included in this chart.

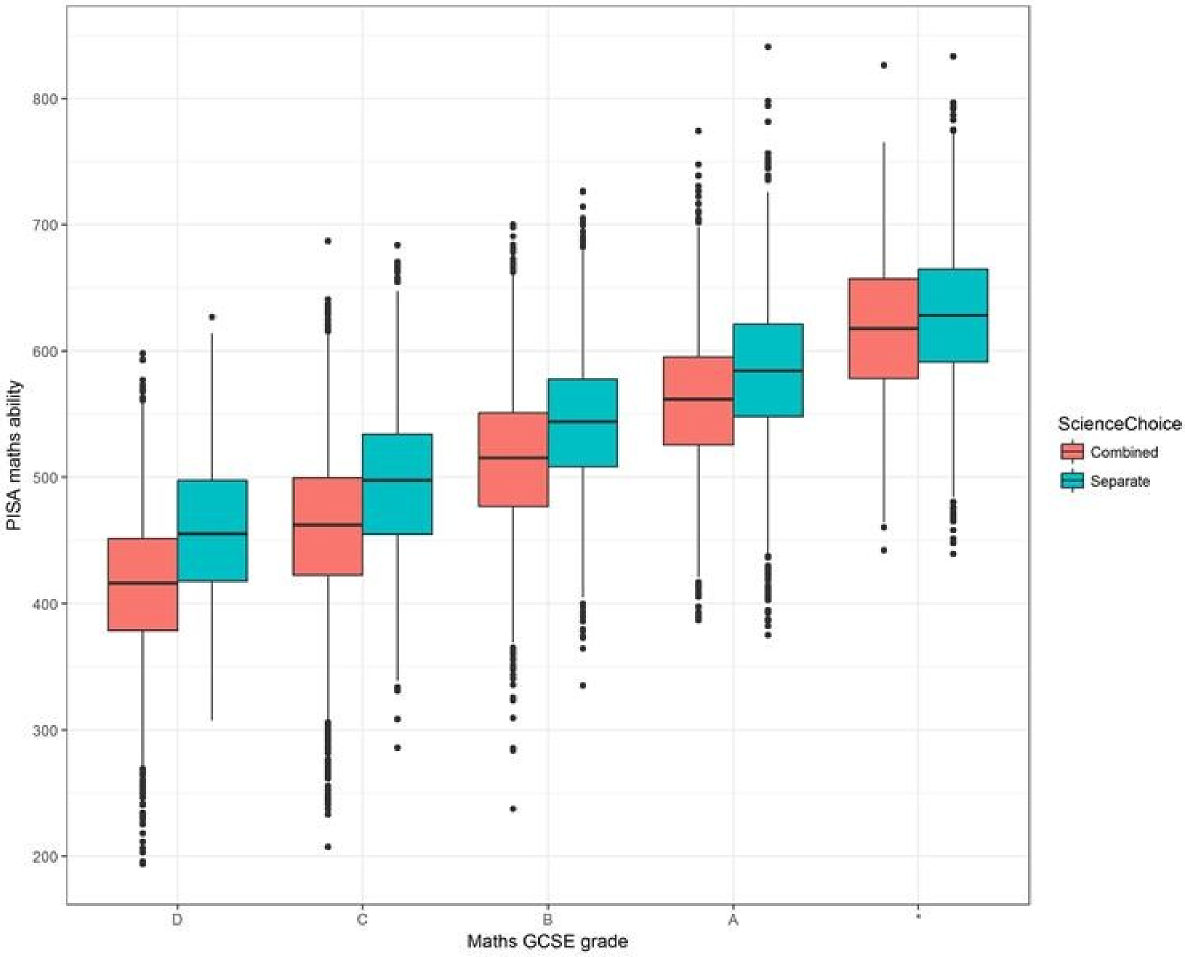

Figure 5: The relationship between Maths GCSE grade and PISA plausible values, split by whether students studied separate or combined sciences at GCSE

Discussion

This report has considered the links between pupils' abilities in the PISA tests and their performances and subject choices at GCSE. Understanding these links is important as it affects the way we interpret the findings from the PISA studies. Specifically, knowing how the PISA tests relate to what is actually taught and assessed as part of the GCSE curriculum may inform the extent to which we believe they provide a reasonable evaluation of our school system, as well as the extent to which it is justifiable to make decisions about the administration of GCSEs based upon results from the PISA tests.

Our research has shown that the correlation between PISA abilities and GCSE performance is not particularly high. For example, the correlation between PISA maths ability and GCSE maths grade (0.78) is only slightly higher than the correlation between KS2 maths and GCSE maths grade (0.75). Furthermore, as shown by data from GCSE resits, the relatively low correlation between PISA and GCSE cannot be explained simply by the time interval between the two tests. Rather the results indicate that the PISA tests measure something rather different to GCSEs. This fact need not solely relate to differences in content but may also be due to other factors, such as the fact that PISA tests are low-stakes for the pupil (thus pupils may not apply full effort, and will not have prepared as thoroughly). Alternatively, it may relate to the fact that, in contrast to GCSEs, the tests are fully computer-based. Although differences in content between GCSEs and PISA have been noted before (see, for example, Jerrim and Shure, 2016), this research confirms the impact of these differences empirically. As such, the skills that are explicitly measured by PISA cannot be assumed to act as a proxy for all the skills that are not.

The results noted above essentially relate to the differing rank orders of individual students depending upon whether they are assessed by PISA or by GCSEs. Although the aim of the PISA tests is explicitly *not* to rank or assess individuals, these results still raise important questions. Specifically, given that the rank order of students is surely affected by at least some of the differences in content, stakes and format of the PISA tests, how can we be sure that such factors do not also influence the average performance levels that are reported for whole countries? For example, existing research has already suggested that the switch to computer-based testing may have affected countries' performances (Jerrim, 2016; Jerrim et al., 2018). It is therefore important to question whether we can be sure that other elements of the testing procedure are not also important in determining countries' rankings.

Although the results comparing GCSE and PISA maths are interesting, those comparing GCSE English and PISA reading are even more striking. Naively, one might assume that these two tests measure similar skills. However, our analysis of correlations shows that performance on the PISA reading test is at least as closely aligned with achievement in GCSE science as it is with GCSE English. This, initially surprising, finding becomes considerably easier to understand once a few of the items used to assess PISA reading have been explored. For example, if we look through a sample of released PISA items[20] we find that the several of the reading tasks actually ask students to read and interpret tables or figures of scientific information, a skill that is assessed in GCSE science tests rather than in

[20] Available from https://www.oecd.org/pisa/pisaproducts/Take%20the%20test%20e%20book.pdf. The "Lake Chad" item gives an example of a PISA reading item that requires looking at scientific information.

GCSE English. The substantive importance of this finding comes in that GCSE English forms a fundamental part of the way that school performance is judged in England, meaning that schools devote substantial time to teaching skills such as essay writing that are fundamental to success in this subject. However, it is clear that when it comes to judging the performance of our education system as a whole using PISA, nothing similar is measured. Thus the performance of our country's education system is judged whilst ignoring some of the key skills that schools are trying to teach.

The PISA domain that shows by far the biggest difference from those skills assessed within GCSEs is that of collaborative problem solving. This is interesting for two reasons. Firstly, this domain is the one where England displays its strongest performance relative to the OECD average. As such, we may ask how this is being achieved given that the skills being assessed show little similarity with those tested (and, thereby, presumably taught) within any mainstream GCSE. One possibility is that these skills are not widely taught in other countries either, so that the performances of countries may be more affected by activities out of school than is the case for the other PISA domains.

Secondly, the relatively low correlation between GCSE performance and PISA collaborative problem solving raises the question of whether all of our assessment at GCSE fails to recognise certain pupils' skills. For example, our analysis has shown that a sizeable minority of students with relatively low performance at GCSE (i.e. averaging at grade C or below) may be amongst those with the highest ability when it comes to collaborative problem solving. Whether we see this as a genuine deficiency in the current GCSE system will depend upon our view of the importance of this skill in its own right and the validity of the OECD's approach to assessing it. For a further discussion of merits of the OECD's collaborative problem solving assessment see Shaw and Child (2017).

This report has also considered the association between GCSE subject choices and PISA abilities. Although assigning causality is problematic, some clear associations were found. In particular those students who had chosen to study separate sciences displayed higher ability in the PISA tests than students with equally good achievement at GCSE, but who had chosen to study combined science. Within PISA reading, students who had chosen to study history, German or French displayed higher abilities than those, with equally good performance at GCSE overall, who had not studied these subjects. Although alternative explanations are possible, this may indicate that PISA performance in reading is associated with the extent to which students are taking subjects that encourage these skills (e.g. comprehension of texts for PISA reading) as part of their GCSEs. Similarly, the results suggested a positive association between taking GCSE drama and performance in collaborative problem solving - an understandable finding given that the collaborative problem solving tasks are themselves a form of role play. Conversely, a negative association was found between taking physical education GCSE and performance in PISA. This may potentially indicate that PISA favours students taking subjects that are fully classroom based. Having said this, there was no indication that students taking arts subjects such as Art & Design or music performed worse in PISA than those choosing less artistic options.

It is worth noting that, even if all the results relating pupil subject choices to PISA scores were genuine causal effects (and we provide no guarantee that they are), this does not necessarily imply those subjects associated with higher PISA scores are necessarily "better".

For example, would it necessarily be right to force all pupils to study separate sciences up until the age of 16, regardless of their interest in the subject, *purely* to boost results in PISA? Likewise, would it be correct to discourage the uptake of the GCSE in physical education, regardless of pupils' interest or aptitudes, just because taking this subject is associated with lower scores in PISA science? Rather than such leading to conclusions, we hope that our work will encourage reflection on what exactly is (and isn't) measured by PISA and that this will help inform the way results from these studies are interpreted.

This research report has highlighted the fact that PISA and GCSEs measure different skills, and has also shown, perhaps as might be expected, that an individual's performance in PISA is likely to depend upon their choices with regard to what they study. These findings warn against an uncritical use of the PISA results without careful consideration of exactly what is being measured. They also give the opportunity to reflect upon the skills that are currently assessed both by PISA and by our own GCSEs and the extent to which each of these fit with wider societal goals.

References

Bates, D., M. Maechler, B. Bolker and S. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67(1)*, 1-48.

Bramley, T., Vidal Rodeiro, C.L., and Vitello, S. (2015). *Gender differences in GCSE*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment. Available from http://www.cambridgeassessment.org.uk/Images/gender-differences-in-gcse.pdf.

Caro, D. H. and P. Biecek (2017). intsvy: An R Package for Analyzing International Large-Scale Assessment Data. *Journal of Statistical Software 1(7)*.

Carroll, M & Gill, T. (2017). *Uptake of GCSE subjects 2016*. Statistics Report Series No. 114. Cambridge, UK: Cambridge Assessment. Available from http://www.cambridgeassessment.org.uk/Images/420406-uptake-of-gcse-subjects-2016.pdf.

DfE (2011). *PISA 2009 Study: How big is the gap? A comparison of pupil attainment in England with the top-performing countries*. London: Department for Education DFE-RR630. Available from https://www.gov.uk/government/publications/pisa-2009-study-how-big-is-the-gap-a-comparison-of-pupil-attainment-in-england-with-the-top-performing-countries.

DfE (2017). *Achievement of 15-Year-Olds in England: PISA 2015 Further Analysis Report*. London: Department for Education DFE-RB766. Available from https://www.gov.uk/government/publications/pisa-2015-national-report-for-england.

Emad, A. and P. Bailey (2017). wCorr: Weighted Correlations. R package version 1.9.1. https://CRAN.R-project.org/package=wCorr.

Foy, P. (2017). *TIMSS 2015 User Guide for the International Database*. Boston: IEA.

Gove, M. (2010). PISA slip should put a rocket under our world-class ambitions and drive us to win the education space race. TES, 17 December 2010. Available from http://www.tes.co.uk/article.aspx?storycode=6066185.

Heath, A., Sullivan, A., Boliver, V., & Zimdars, A. (2013). Education under new labour, 1997–2010. *Oxford Review of Economic Policy*, *29*(1), 227-247.

Jerrim, J. (2016). PISA 2012: How do results for the paper and computer tests compare?. *Assessment in Education: Principles, Policy & Practice*, *23*(4), 495-518.

Jerrim, J., & Shure, N. (2016). *Achievement of 15-Year-Olds in England: PISA 2015 National Report*. London: Department for Education DFE-RR630. Available from https://www.gov.uk/government/publications/pisa-2015-national-report-for-england.

Jerrim, J., & Shure, N. (2017). *Achievement of 15-Year-Olds in England: PISA 2015 Collaborative Problem Solving National Report*. London: Department for Education DFE-RR765. Available from https://www.gov.uk/government/publications/pisa-2015-national-report-for-england.

Jerrim, J., Micklewright, J., Heine, J. H., Salzer, C., & McKeown, C. (2018). PISA 2015: how big is the 'mode effect'and what has been done about it?. *Oxford Review of Education*, 1-18.

OECD (2017). *PISA 2015 Technical Report*, Paris, France. Available from http://www.oecd.org/pisa/data/2015-technical-report/.

Ofqual (2014). *Setting GCSE, AS and A level grade standards in summer 2014*. Coventry: Ofqual/14/5428. Avaiable from http://webarchive.nationalarchives.gov.uk/20141031163546/http://ofqual.gov.uk/documents/setting-gcse-level-grade-standards-summer-2014/ .

Newton, P. (2009). The reliability of results from national curriculum testing in England. *Educational Research*, 51 (2), 181-212.

Revelle, W. (2018). psych: Procedures for Personality and Psychological Research. R package version 1.8.3. https://CRAN.R-project.org/package=psych.

Robitzsch, A., T. Kiefer and M. Wu (2018). TAM: Test analysis modules. R package version 2.10-24. https://CRAN.R-project.org/package=TAM.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, *39*(2), 142-151.

Shaw, S. and Child, S. (2017). Utilising technology in the assessment of collaboration: A critique of PISA's collaborative problem-solving tasks. *Research Matters: A Cambridge Assessment publication*, 24, 17-22. Available from http://www.cambridgeassessment.org.uk/Images/426176-research-matters-24-autumn-2017.pdf.

Thorndike, R. L. (1947). *Research problems and techniques. (Report No. 3)*. Washington DC: U.S. Government Printing Office.

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI monograph series*, *2*, 9-36. http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_01.pdf.

Villalba, E. (2012). Searching for the holy grail of measuring creativity. *Creativity Research Journal*, *24*(1), 1-2.

Winkley, J. & Cresswell, M. (2012). *Introduction to the concept of reliability in the context of educational assessment*. Chap. 1 in Ofqual's Reliability Compendium, edited by D. Opposs and Q. He, 9-66. Coventry: Ofqual/12/5117. Also available from https://www.gov.uk/government/publications/reliability-of-assessment-compendium/introduction-to-the-concept-of-reliability.

Zhao, Y. (2012). Flunking innovation and creativity. *Phi Delta Kappan*, *94*(1), 56-61.

Appendix 1: Robustness checks and tables of further results

Standard errors of the Pearson correlations between PISA PVs in each domain and GCSE and KS2 achievements are given in Table A1.

Correlations were repeated without the Thorndike correction for restricted range and results are shown in Table A2. Most often, uncorrected correlations were slightly weaker, but differences were greater for subjects with fewer students. For all PISA domains, the variable with strongest correlation was identical to the corrected version, and ranks were broadly similar. The biggest impacts appeared to be on sciences, with core science, additional science and each separate science showing smaller coefficients if the Thorndike correction was not applied. This was due to the restricted ability ranges of pupils taking these subjects: those taking separate sciences were skewed towards higher abilities (as measured by PISA), whilst those taking core and additional science were skewed towards lower abilities. The relatively strong correlations with GCSE geography (particularly with PISA science) were robust to the lack of Thorndike correction, despite only 43% of students in the sample taking it; this indicates that the students taking GCSE geography covered a wider ability range.

Analyses were repeated with original PISA PVs used in the place of the newly calculated ones (Table A3). Relatively small impacts were seen for PISA science and reading. For science, correlations were typically reduced by 0.01 – 0.03. For reading, some coefficients were slightly larger (but only by 0.009) but most were smaller; some by as much 0.127. Despite this, the ranking of GCSE subjects by correlation with PISA ability was similar. For PISA maths, more changes were observed: on average, coefficients were 0.071 smaller and more rank changes occurred. Notably, the correlation with maths GCSE weakened substantially ($r = 0.681$), becoming only the second-strongest correlation and a similar decline was seen for KS2 maths. The fact that such changes were seen justifies the decision to recreate our own plausible values. For CPS, coefficient changes were relatively evenly distributed between increases and decreases (mean change 0.010 smaller).

To examine whether correlation method affected results, analyses were repeated with Spearman and, where appropriate (i.e. when the variable could be an ordinal categorical variable), polyserial correlations (Table A4). Spearman correlations were tested as they should be more robust to the exact distributions of the variables being correlated. As such, it is reassuring to note that the results were very similar to those from Pearson correlations; both in coefficient and rank. Polyserial correlations were used to see the effect of accounting for the fact that GCSE grades and KS2 teacher assessment levels are reported on a coarse scale with a small number of available categories rather than a fully continuous one. In theory, this may attenuate the size of observed correlations. The polyserial correlations with GCSE subject grades were almost universally larger than the Pearson coefficients, but ranks were usually similar and, furthermore, the differences were small; averaging just below 0.02. However, the polyserial correlations between PISA abilities and KS2 teacher assessment levels were all substantially larger than original Pearson correlations, with an average increase in excess of 0.06. For example, the correlation of PISA maths with KS2 maths teacher assessment level increased substantially ($\rho = 0.749$), becoming the joint third strongest. Hence, polyserial correlations made relationships with KS2 teacher assessment levels appear stronger. However, given that for teacher assessments, unlike GCSEs, measurement is never made on a continuous scale (e.g. a mark scale), whether such observations are meaningful is debatable.

Relationships were examined with EAP ability estimates (Table A5); these were again tested with Pearson, Spearman and polyserial correlations, but only Pearson correlations are discussed as results were broadly similar. Note that to have an EAP ability estimate generated for a particular domain, a student had to have taken at least one relevant cluster of items. Across all three domains the rank order of correlations was similar to in the original analysis. However, since no effort was made to adjust for measurement error in the PISA tests, most correlations were slightly weaker than those shown in Table 2 of the main report (and repeated with standard errors in Table A1). Hence, although some small changes in the rank orders of subjects were observed, the key results discussed in the main report appeared robust.

As a final check, correlations were calculated separately for students making ‘high’ or ‘low’ effort (Table A6). Across all PISA domains, the ‘low effort’ coefficients tended to be slightly lower, on average, by just 0.03. Although some larger differences were visible for some GCSE subjects, these were generally amongst those with relatively small numbers of observations from the ‘low’ effort group and, as such, cannot be taken to imply meaningful differences in the rank orders correlations. One larger difference between the ‘low’ and ‘high’ effort groups that was supported by larger sample sizes was seen for the correlation between PISA reading and KS2 reading marks (high effort r = 0.656, low effort r = 0.545, high effort N = 3,588, low effort N = 976). This might imply that the association of PISA reading scores with external performance measures was more susceptible to variation in effort. Generally, however, results appeared to be reasonably robust to variation in effort.

Tables A7, A8 and A9 provide full results for other analyses discussed in the main report.

Table A1. Weighted Pearson correlations (with standard error) with updated plausible values, with Thorndike correction applied. Variable gives the variable against which PVs were correlated; where only a subject name is given, this indicates the GCSE grade in that subject. N gives the number of students included. Table is sorted with mean GCSE attainment first, then GCSE subjects, and then KS2 metrics; within each, table is sorted in order of Science correlation strength. Red indicates stronger correlation; blue indicates weaker correlation (scaled relative to observed values).

Variable	N	Science	Maths	Reading	CPS
Mean GCSE grade	4,912	0.753 (0.010)	0.753 (0.011)	0.741 (0.010)	0.611 (0.014)
Highest Science	4,677	0.760 (0.009)	0.746 (0.010)	0.708 (0.010)	0.587 (0.017)
Core Science	3,037	0.748 (0.011)	0.714 (0.013)	0.692 (0.013)	0.550 (0.017)
Additional Science	2,779	0.734 (0.014)	0.706 (0.015)	0.660 (0.017)	0.525 (0.020)
Maths	4,778	0.728 (0.010)	0.777 (0.010)	0.672 (0.012)	0.593 (0.013)
Geography	2,232	0.714 (0.016)	0.698 (0.018)	0.687 (0.017)	0.583 (0.024)
D&T: Textiles Technology	199	0.700 (0.039)	0.677 (0.040)	0.656 (0.046)	0.512 (0.076)
Physics	1,563	0.699 (0.030)	0.732 (0.023)	0.604 (0.047)	0.492 (0.059)
Average across separate sciences	1,544	0.698 (0.030)	0.729 (0.024)	0.626 (0.041)	0.499 (0.056)
History	2,373	0.696 (0.016)	0.675 (0.016)	0.696 (0.017)	0.561 (0.021)
Statistics	390	0.682 (0.036)	0.694 (0.051)	0.672 (0.046)	0.548 (0.047)
Biological Science	1,580	0.681 (0.029)	0.696 (0.024)	0.624 (0.036)	0.494 (0.054)
Business Studies	746	0.672 (0.030)	0.701 (0.030)	0.674 (0.029)	0.531 (0.043)
Chemistry	1,566	0.659 (0.037)	0.699 (0.028)	0.607 (0.041)	0.479 (0.054)
English	4,735	0.628 (0.012)	0.625 (0.015)	0.680 (0.011)	0.534 (0.017)
Home Economics: Child Development	152	0.624 (0.053)	0.583 (0.085)	0.604 (0.084)	0.469 (0.081)
English Literature	4,287	0.613 (0.015)	0.592 (0.017)	0.637 (0.016)	0.534 (0.019)
Music	363	0.601 (0.049)	0.594 (0.049)	0.567 (0.042)	0.417 (0.057)
D&T: Food Technology	311	0.589 (0.051)	0.606 (0.048)	0.635 (0.033)	0.494 (0.052)
Media, Film and Television Studies	450	0.588 (0.041)	0.577 (0.048)	0.609 (0.043)	0.472 (0.045)
German	530	0.587 (0.056)	0.631 (0.043)	0.579 (0.066)	0.397 (0.091)
Religious Studies	2,447	0.575 (0.023)	0.564 (0.023)	0.595 (0.020)	0.481 (0.026)
Physical Education	1,102	0.571 (0.027)	0.579 (0.029)	0.538 (0.031)	0.454 (0.030)
D&T: Graphic Products	217	0.559 (0.063)	0.614 (0.066)	0.547 (0.067)	0.466 (0.071)
D&T: Resistant Materials Technology	446	0.552 (0.040)	0.582 (0.041)	0.531 (0.044)	0.371 (0.056)
French	1,387	0.535 (0.026)	0.551 (0.025)	0.538 (0.028)	0.376 (0.036)
Drama	552	0.534 (0.038)	0.509 (0.047)	0.557 (0.039)	0.400 (0.050)
Information Technology	1,167	0.524 (0.026)	0.541 (0.030)	0.532 (0.025)	0.448 (0.028)
Art and Design	1,334	0.500 (0.023)	0.475 (0.024)	0.507 (0.023)	0.328 (0.031)
Spanish	930	0.476 (0.045)	0.484 (0.042)	0.495 (0.048)	0.379 (0.053)
KS2 English: marks in reading test	4,564	0.648 (0.009)	0.601 (0.012)	0.638 (0.011)	0.572 (0.015)
KS2 maths: total test marks	4,575	0.645 (0.012)	0.740 (0.010)	0.553 (0.014)	0.512 (0.015)
KS2 maths: teacher-assessed NC level	4,628	0.619 (0.011)	0.685 (0.010)	0.550 (0.012)	0.503 (0.014)
KS2 science: teacher-assessed NC level	4,626	0.610 (0.011)	0.642 (0.011)	0.533 (0.014)	0.504 (0.013)
KS2 English: teacher-assessed NC level	4,628	0.603 (0.009)	0.610 (0.011)	0.592 (0.011)	0.524 (0.013)
KS2 English: marks in writing test	4,564	0.519 (0.013)	0.524 (0.014)	0.549 (0.013)	0.441 (0.016)

Table A2. Weighted Pearson correlations (with standard error) with updated plausible values, with no corrections applied. Variable gives the variable against which PVs were correlated; where only a subject name is given, this indicates the GCSE grade in that subject. N gives the number of students included. Table is sorted with mean GCSE attainment first, then GCSE subjects, and then KS2 metrics; within each, table is sorted in order of Science correlation strength. Red indicates stronger correlation; blue indicates weaker correlation (scaled relative to observed values).

Variable	N	Science	Maths	Reading	CPS
Mean GCSE grade	4,912	0.754 (0.010)	0.755 (0.011)	0.741 (0.010)	0.614 (0.015)
Highest Science	4,677	0.757 (0.009)	0.741 (0.010)	0.705 (0.010)	0.588 (0.017)
Maths	4,778	0.727 (0.009)	0.777 (0.010)	0.671 (0.012)	0.595 (0.013)
Core Science	3,037	0.710 (0.012)	0.667 (0.014)	0.658 (0.014)	0.532 (0.017)
Geography	2,232	0.703 (0.016)	0.692 (0.019)	0.678 (0.018)	0.578 (0.023)
History	2,373	0.688 (0.016)	0.666 (0.016)	0.685 (0.017)	0.556 (0.022)
Additional Science	2,779	0.683 (0.015)	0.646 (0.016)	0.611 (0.018)	0.495 (0.020)
D&T: Textiles Technology	199	0.659 (0.040)	0.644 (0.043)	0.613 (0.046)	0.472 (0.075)
Statistics	390	0.658 (0.036)	0.689 (0.050)	0.673 (0.045)	0.529 (0.044)
English	4,735	0.627 (0.012)	0.625 (0.015)	0.679 (0.011)	0.536 (0.017)
Business Studies	746	0.626 (0.031)	0.677 (0.032)	0.625 (0.032)	0.497 (0.043)
English Literature	4,287	0.611 (0.015)	0.589 (0.017)	0.635 (0.016)	0.532 (0.019)
Physics	1,563	0.602 (0.032)	0.636 (0.027)	0.514 (0.046)	0.443 (0.057)
Average across separate sciences	1,544	0.595 (0.032)	0.625 (0.027)	0.529 (0.042)	0.444 (0.054)
Biological Science	1,580	0.581 (0.031)	0.596 (0.025)	0.531 (0.035)	0.441 (0.052)
Home Economics: Child Development	152	0.580 (0.054)	0.547 (0.086)	0.558 (0.079)	0.468 (0.081)
Media, Film and Television Studies	450	0.573 (0.041)	0.557 (0.053)	0.586 (0.042)	0.468 (0.046)
Music	363	0.571 (0.047)	0.561 (0.049)	0.546 (0.041)	0.411 (0.053)
D&T: Food Technology	311	0.561 (0.051)	0.592 (0.048)	0.634 (0.036)	0.497 (0.052)
Religious Studies	2,447	0.560 (0.023)	0.550 (0.023)	0.580 (0.020)	0.474 (0.026)
Chemistry	1,566	0.560 (0.037)	0.600 (0.031)	0.515 (0.041)	0.430 (0.052)
D&T: Resistant Materials Technology	446	0.559 (0.040)	0.600 (0.043)	0.532 (0.045)	0.373 (0.060)
German	530	0.553 (0.056)	0.604 (0.045)	0.519 (0.065)	0.364 (0.086)
Physical Education	1,102	0.537 (0.026)	0.537 (0.028)	0.511 (0.030)	0.438 (0.029)
D&T: Graphic Products	217	0.533 (0.063)	0.592 (0.066)	0.529 (0.066)	0.493 (0.070)
Drama	552	0.523 (0.038)	0.473 (0.045)	0.541 (0.040)	0.391 (0.049)
Information Technology	1,167	0.520 (0.026)	0.537 (0.030)	0.521 (0.024)	0.448 (0.029)
Art and Design	1,334	0.501 (0.024)	0.480 (0.025)	0.515 (0.023)	0.327 (0.030)
French	1,387	0.498 (0.025)	0.514 (0.024)	0.489 (0.028)	0.345 (0.034)
Spanish	930	0.433 (0.042)	0.440 (0.040)	0.458 (0.048)	0.363 (0.052)
KS2 English: marks in reading test	4,564	0.643 (0.010)	0.598 (0.012)	0.630 (0.011)	0.570 (0.015)
KS2 maths: total test marks	4,575	0.639 (0.012)	0.737 (0.010)	0.544 (0.014)	0.509 (0.015)
KS2 maths: teacher-assessed NC level	4,628	0.617 (0.011)	0.687 (0.010)	0.546 (0.012)	0.506 (0.014)
KS2 science: teacher-assessed NC level	4,626	0.608 (0.011)	0.644 (0.010)	0.529 (0.014)	0.506 (0.013)
KS2 English: teacher-assessed NC level	4,628	0.602 (0.009)	0.612 (0.011)	0.588 (0.011)	0.526 (0.013)
KS2 English: marks in writing test	4,564	0.513 (0.013)	0.520 (0.014)	0.540 (0.013)	0.440 (0.016)

Table A3. Weighted Pearson correlations (with standard error) with original PISA plausible values, with Thorndike correction applied. Variable gives the variable against which PVs were correlated; where only a subject name is given, this indicates the GCSE grade in that subject. N gives the number of students included. Table is sorted with mean GCSE attainment first, then GCSE subjects, then KS2 metrics; within each, table is sorted in order of Science correlation strength. Red indicates stronger correlation; blue indicates weaker correlation (scaled relative to observed values).

Variable	N	Science	Maths	Reading	CPS
Mean GCSE grade	4,912	0.730 (0.011)	0.681 (0.014)	0.699 (0.012)	0.606 (0.012)
Highest Science	4,677	0.736 (0.011)	0.683 (0.013)	0.678 (0.013)	0.579 (0.014)
Core Science	3,037	0.720 (0.014)	0.649 (0.017)	0.653 (0.017)	0.534 (0.020)
Maths	4,778	0.706 (0.010)	0.681 (0.013)	0.650 (0.011)	0.562 (0.013)
Additional Science	2,779	0.700 (0.016)	0.623 (0.022)	0.633 (0.020)	0.507 (0.022)
Geography	2,232	0.692 (0.016)	0.637 (0.023)	0.652 (0.016)	0.553 (0.021)
History	2,373	0.675 (0.016)	0.630 (0.018)	0.657 (0.017)	0.554 (0.021)
Physics	1,563	0.668 (0.032)	0.639 (0.035)	0.577 (0.042)	0.473 (0.036)
D&T: Textiles Technology	199	0.668 (0.050)	0.634 (0.048)	0.622 (0.058)	0.480 (0.073)
Statistics	390	0.663 (0.027)	0.610 (0.041)	0.652 (0.032)	0.576 (0.045)
Average across separate sciences	1,544	0.663 (0.034)	0.628 (0.036)	0.600 (0.040)	0.488 (0.036)
Biological Science	1,580	0.647 (0.034)	0.609 (0.038)	0.598 (0.038)	0.494 (0.036)
Business Studies	746	0.647 (0.032)	0.590 (0.036)	0.604 (0.036)	0.481 (0.038)
Chemistry	1,566	0.627 (0.041)	0.596 (0.042)	0.575 (0.040)	0.475 (0.041)
English	4,735	0.613 (0.013)	0.573 (0.016)	0.609 (0.013)	0.528 (0.013)
English Literature	4,287	0.600 (0.016)	0.555 (0.018)	0.602 (0.015)	0.523 (0.017)
Home Economics: Child Development	152	0.597 (0.060)	0.496 (0.076)	0.601 (0.068)	0.516 (0.068)
Music	363	0.577 (0.042)	0.530 (0.049)	0.569 (0.048)	0.447 (0.059)
German	530	0.577 (0.054)	0.556 (0.056)	0.533 (0.057)	0.452 (0.069)
Media, Film and Television Studies	450	0.573 (0.044)	0.541 (0.044)	0.592 (0.046)	0.470 (0.040)
D&T: Food Technology	311	0.571 (0.055)	0.545 (0.058)	0.508 (0.056)	0.428 (0.055)
Religious Studies	2,447	0.559 (0.024)	0.526 (0.022)	0.562 (0.021)	0.463 (0.024)
Physical Education	1,102	0.545 (0.026)	0.517 (0.028)	0.525 (0.028)	0.423 (0.033)
D&T: Resistant Materials Technology	446	0.527 (0.049)	0.490 (0.044)	0.488 (0.048)	0.418 (0.057)
Drama	552	0.525 (0.043)	0.504 (0.048)	0.529 (0.044)	0.436 (0.053)
French	1,387	0.522 (0.024)	0.503 (0.025)	0.516 (0.029)	0.407 (0.035)
D&T: Graphic Products	217	0.522 (0.069)	0.495 (0.074)	0.484 (0.069)	0.492 (0.059)
Information Technology	1,167	0.493 (0.030)	0.446 (0.037)	0.471 (0.030)	0.431 (0.030)
Art and Design	1,334	0.480 (0.025)	0.430 (0.025)	0.463 (0.026)	0.361 (0.028)
Spanish	930	0.460 (0.038)	0.439 (0.040)	0.480 (0.037)	0.419 (0.037)
KS2 English: marks in reading test	4,564	0.625 (0.011)	0.576 (0.012)	0.610 (0.011)	0.526 (0.013)
KS2 maths: total test marks	4,575	0.623 (0.012)	0.620 (0.012)	0.559 (0.013)	0.474 (0.015)
KS2 maths: teacher-assessed NC level	4,628	0.603 (0.011)	0.594 (0.011)	0.546 (0.013)	0.469 (0.015)
KS2 science: teacher-assessed NC level	4,626	0.589 (0.012)	0.565 (0.012)	0.534 (0.014)	0.458 (0.014)
KS2 English: teacher-assessed NC level	4,628	0.587 (0.010)	0.551 (0.011)	0.570 (0.014)	0.494 (0.013)
KS2 English: marks in writing test	4,564	0.513 (0.014)	0.479 (0.015)	0.517 (0.016)	0.448 (0.013)

Table A4. Weighted Pearson, Spearman and polyserial correlations with updated PVs, with Thorndike correction applied. Variable gives the variable against which PVs were correlated; where only a subject name is given, this indicates the GCSE grade. N gives the number of students. Table is sorted by Science Pearson correlation (as in Table 1). Red indicates stronger correlation; blue indicates weaker correlation (scaled relative to observed values).

Variable	N	Science			Maths			Reading			CPS		
		Pearson	Spearman	Polyserial	Pearson	Spearman	Polyserial	Pearson	Spearman	Polyserial	Pearson	Spearman	Polyserial
Mean GCSE grade	4,912	0.753	0.753	–	0.753	0.752	–	0.741	0.737	–	0.611	0.598	–
Highest Science	4,677	0.760	0.758	0.780	0.746	0.743	0.765	0.708	0.698	0.723	0.587	0.577	0.600
Core Science	3,037	0.748	0.743	0.769	0.714	0.698	0.732	0.692	0.675	0.707	0.550	0.536	0.561
Additional Science	2,779	0.734	0.725	0.755	0.706	0.701	0.728	0.660	0.640	0.674	0.525	0.521	0.540
Maths	4,778	0.728	0.737	0.759	0.777	0.789	0.809	0.672	0.676	0.696	0.593	0.587	0.611
Geography	2,232	0.714	0.714	0.727	0.698	0.688	0.706	0.687	0.685	0.702	0.583	0.579	0.598
D&T: Textiles Technology	199	0.700	0.704	0.709	0.677	0.695	0.692	0.656	0.629	0.657	0.512	0.476	0.513
Physics	1,563	0.699	0.670	0.718	0.732	0.722	0.749	0.604	0.557	0.620	0.492	0.455	0.506
Average across separate sciences	1,544	0.698	0.669	0.704	0.729	0.724	0.734	0.626	0.581	0.629	0.499	0.459	0.503
History	2,373	0.696	0.688	0.707	0.675	0.668	0.683	0.696	0.682	0.703	0.561	0.546	0.567
Statistics	390	0.682	0.675	0.704	0.694	0.716	0.721	0.672	0.699	0.700	0.548	0.552	0.559
Biological Science	1,580	0.681	0.659	0.699	0.696	0.689	0.714	0.624	0.585	0.640	0.494	0.463	0.510
Business Studies	746	0.672	0.654	0.684	0.701	0.684	0.711	0.674	0.671	0.689	0.531	0.493	0.539
Chemistry	1,566	0.659	0.624	0.677	0.699	0.688	0.715	0.607	0.569	0.620	0.479	0.436	0.489
English	4,735	0.628	0.628	0.647	0.625	0.629	0.643	0.680	0.681	0.702	0.534	0.518	0.548
Home Economics: Child Development	152	0.624	0.581	0.638	0.583	0.567	0.595	0.604	0.581	0.617	0.469	0.402	0.475
English Literature	4,287	0.613	0.617	0.632	0.592	0.602	0.610	0.637	0.637	0.654	0.534	0.513	0.543
Music	363	0.601	0.580	0.603	0.594	0.545	0.585	0.567	0.526	0.560	0.417	0.407	0.404
D&T: Food Technology	311	0.589	0.599	0.611	0.606	0.620	0.631	0.635	0.640	0.651	0.494	0.483	0.512
Media, Film and Television Studies	450	0.588	0.582	0.608	0.577	0.559	0.595	0.609	0.602	0.631	0.472	0.461	0.479
German	530	0.587	0.573	0.603	0.631	0.629	0.647	0.579	0.565	0.592	0.397	0.360	0.408
Religious Studies	2,447	0.575	0.559	0.582	0.564	0.548	0.573	0.595	0.582	0.601	0.481	0.438	0.480
Physical Education	1,102	0.571	0.573	0.583	0.579	0.572	0.591	0.538	0.537	0.547	0.454	0.466	0.465
D&T: Graphic Products	217	0.559	0.571	0.578	0.614	0.619	0.628	0.547	0.537	0.559	0.466	0.492	0.484
D&T: Resistant Materials Technology	446	0.552	0.561	0.580	0.582	0.593	0.609	0.531	0.545	0.558	0.371	0.366	0.390
French	1,387	0.535	0.545	0.547	0.551	0.540	0.561	0.538	0.540	0.551	0.376	0.390	0.385
Drama	552	0.534	0.543	0.544	0.509	0.516	0.519	0.557	0.561	0.562	0.400	0.384	0.399
Information Technology	1,167	0.524	0.535	0.550	0.541	0.541	0.567	0.532	0.542	0.554	0.448	0.443	0.470
Art and Design	1,334	0.500	0.520	0.519	0.475	0.487	0.489	0.507	0.514	0.524	0.328	0.336	0.340
Spanish	930	0.476	0.468	0.484	0.484	0.468	0.493	0.495	0.482	0.507	0.379	0.364	0.386
KS2 English: marks in reading test	4,564	0.648	0.651	–	0.601	0.600	–	0.639	0.645	–	0.572	0.571	–
KS2 maths: total test marks	4,575	0.645	0.649	–	0.740	0.745	–	0.553	0.566	–	0.512	0.515	–
KS2 maths: teacher-assessed NC level	4,628	0.619	0.622	0.679	0.685	0.685	0.749	0.550	0.555	0.599	0.503	0.494	0.548
KS2 science: teacher-assessed NC level	4,626	0.610	0.611	0.687	0.642	0.639	0.719	0.533	0.533	0.598	0.504	0.500	0.566
KS2 English: teacher-assessed NC level	4,628	0.603	0.611	0.671	0.610	0.613	0.675	0.592	0.596	0.656	0.524	0.524	0.581
KS2 English: marks in writing test	4,564	0.519	0.517	–	0.524	0.521	–	0.549	0.548	–	0.441	0.441	–

Table A5. Weighted Pearson, Spearman and polyserial correlations with EAP ability scores, with Thorndike correction applied. Variable gives the variable against which PVs were correlated; where only a subject name is given, this indicates the GCSE grade. N gives the number of students. Table is sorted by Science Pearson correlation; note that the order is different from previous tables, which were sorted based on correlation strength with science PVs. Red indicates stronger correlation; blue indicates weaker correlation (scaled relative to observed values).

Variable	Science				Maths				Reading				CPS			
	N	Pearson	Spearman	Polyserial	N	Pearson	Spearman	Polyserial	N	Pearson	Spearman	Polyserial	N	Pearson	Spearman	Polyserial
Mean GCSE grade	4,910	0.720	0.722	–	2,006	0.713	0.721	–	2,006	0.702	0.703	–	1,493	0.610	0.596	–
Highest Science	4,675	0.729	0.731	0.747	1,909	0.700	0.710	0.721	1,900	0.664	0.661	0.679	1,431	0.595	0.579	0.607
Core Science	3,035	0.709	0.710	0.730	1,230	0.654	0.652	0.676	1,226	0.624	0.616	0.641	944	0.531	0.513	0.541
Maths	4,776	0.695	0.709	0.726	1,954	0.746	0.765	0.781	1,941	0.643	0.645	0.663	1,456	0.585	0.588	0.609
Additional Science	2,778	0.690	0.686	0.712	1,125	0.637	0.648	0.661	1,135	0.593	0.579	0.608	848	0.519	0.506	0.531
Geography	2,231	0.686	0.691	0.697	894	0.648	0.656	0.659	905	0.644	0.641	0.655	703	0.566	0.552	0.575
D&T: Textiles Technology	199	0.684	0.697	0.690	88	0.650	0.659	0.654	82	0.630	0.596	0.636	54	0.404	0.288	0.410
Physics	1,563	0.661	0.633	0.677	654	0.712	0.690	0.727	643	0.567	0.545	0.586	463	0.559	0.507	0.564
History	2,373	0.660	0.657	0.669	1,000	0.634	0.635	0.639	969	0.661	0.649	0.667	691	0.540	0.535	0.547
Average across separate sciences	1,544	0.659	0.634	0.662	645	0.684	0.669	0.687	637	0.606	0.573	0.611	455	0.542	0.492	0.541
Biological Science	1,580	0.645	0.629	0.659	658	0.655	0.626	0.672	651	0.607	0.581	0.619	468	0.528	0.485	0.540
Statistics	390	0.638	0.641	0.662	159	0.667	0.709	0.703	151	0.616	0.656	0.652	125	0.587	0.562	0.583
Business Studies	745	0.624	0.612	0.636	311	0.621	0.616	0.634	292	0.669	0.667	0.679	228	0.500	0.459	0.510
Chemistry	1,566	0.622	0.594	0.636	651	0.643	0.629	0.658	648	0.591	0.563	0.603	461	0.513	0.472	0.520
Home Economics: Child Development	152	0.606	0.574	0.618	61	0.513	0.494	0.535	56	0.646	0.618	0.659	49	0.576	0.503	0.600
English	4,733	0.598	0.602	0.617	1,935	0.599	0.610	0.619	1,924	0.636	0.634	0.655	1,443	0.530	0.528	0.545
English Literature	4,287	0.586	0.591	0.604	1,759	0.557	0.571	0.574	1,734	0.624	0.624	0.639	1,314	0.520	0.498	0.528
Media, Film and Television Studies	450	0.572	0.575	0.591	181	0.589	0.553	0.602	174	0.545	0.583	0.564	147	0.530	0.487	0.538
D&T: Food Technology	311	0.567	0.582	0.585	132	0.630	0.655	0.646	124	0.525	0.522	0.523	97	0.444	0.459	0.472
Music	362	0.562	0.560	0.566	152	0.563	0.482	0.554	153	0.555	0.527	0.525	97	0.470	0.470	0.489
German	530	0.555	0.549	0.569	224	0.541	0.553	0.563	206	0.503	0.461	0.517	159	0.469	0.436	0.485
Religious Studies	2,446	0.544	0.529	0.549	1,006	0.534	0.529	0.543	1,009	0.570	0.557	0.575	740	0.481	0.438	0.477
Physical Education	1,101	0.525	0.528	0.536	437	0.496	0.500	0.506	444	0.565	0.560	0.573	346	0.462	0.467	0.477
D&T: Graphic Products	217	0.520	0.550	0.539	78	0.519	0.573	0.533	92	0.493	0.494	0.513	64	0.528	0.478	0.541
D&T: Resistant Materials Technology	446	0.517	0.531	0.544	186	0.527	0.541	0.560	174	0.486	0.489	0.506	136	0.410	0.417	0.435
French	1,387	0.511	0.520	0.521	563	0.504	0.501	0.511	577	0.488	0.488	0.500	411	0.378	0.382	0.387
Drama	552	0.503	0.511	0.513	219	0.551	0.573	0.557	229	0.496	0.487	0.502	177	0.417	0.427	0.422
Information Technology	1,167	0.497	0.511	0.520	455	0.535	0.533	0.558	489	0.501	0.507	0.516	362	0.432	0.427	0.444
Art and Design	1,333	0.492	0.510	0.510	534	0.443	0.470	0.458	533	0.445	0.454	0.460	422	0.316	0.324	0.328
Spanish	930	0.453	0.447	0.459	377	0.438	0.428	0.446	387	0.516	0.511	0.528	278	0.309	0.324	0.321
KS2 English: marks in reading test	4,563	0.621	0.627	–	1,875	0.575	0.580	–	1,841	0.614	0.625	–	1,404	0.582	0.584	–
KS2 maths: total test marks	4,574	0.617	0.625	–	1,876	0.701	0.714	–	1,847	0.546	0.557	–	1,407	0.505	0.512	–
KS2 maths: teacher-assessed NC level	4,626	0.591	0.599	0.647	1,895	0.660	0.671	0.724	1,874	0.538	0.546	0.585	1,422	0.495	0.490	0.542
KS2 science: teacher-assessed NC level	4,624	0.582	0.589	0.653	1,894	0.604	0.615	0.677	1,873	0.526	0.528	0.591	1,422	0.496	0.496	0.559
KS2 English: teacher-assessed NC level	4,626	0.579	0.589	0.643	1,895	0.573	0.588	0.638	1,874	0.574	0.580	0.634	1,422	0.524	0.529	0.584
KS2 English: marks in writing test	4,563	0.500	0.498	–	1,874	0.511	0.517	–	1,844	0.513	0.513	–	1,402	0.465	0.454	–

Table A6. Weighted Pearson correlations with updated plausible values, split by effort: 'high' indicates students attempted more items than expected; 'low' effort indicates students attempted fewer items than expected. Variable gives the variable against which PVs were correlated; where only a subject name is given, this indicates the GCSE grade in that subject. N gives the number of students included. Table is sorted in order of Science correlation strength in 'high' group. Red indicates stronger correlation; blue indicates weaker correlation (scaled relative to observed values).

Variable	N		Science		Maths		Reading		CPS	
	high	low	high	low	high	low	high	low	high	low
Mean GCSE grade	3,823	1,089	0.754	0.729	0.756	0.727	0.738	0.725	0.597	0.609
Highest Science	3,646	1,031	0.764	0.730	0.752	0.709	0.708	0.681	0.575	0.574
Core Science	2,240	797	0.746	0.740	0.714	0.693	0.688	0.678	0.529	0.557
Additional Science	2,090	689	0.735	0.715	0.710	0.675	0.659	0.638	0.515	0.513
Maths	3,710	1,068	0.732	0.699	0.784	0.745	0.672	0.642	0.580	0.584
D&T: Textiles Technology	156	43	0.716	0.617	0.696	0.588	0.656	0.625	0.512	0.433
Geography	1,795	437	0.713	0.697	0.694	0.690	0.690	0.649	0.575	0.563
Physics	1,350	213	0.706	0.625	0.731	0.710	0.600	0.586	0.474	0.533
History	1,882	491	0.706	0.639	0.678	0.637	0.698	0.658	0.547	0.546
Average across separate sciences	1,334	210	0.703	0.620	0.728	0.699	0.623	0.595	0.480	0.533
Statistics	290	100	0.699	0.607	0.731	0.569	0.700	0.580	0.521	0.562
Biological Science	1,364	216	0.685	0.606	0.692	0.669	0.624	0.576	0.482	0.499
Business Studies	615	131	0.672	0.624	0.697	0.684	0.674	0.622	0.525	0.499
Chemistry	1,353	213	0.672	0.544	0.704	0.636	0.610	0.545	0.472	0.466
English	3,689	1,046	0.633	0.577	0.626	0.586	0.678	0.659	0.522	0.513
Home Economics: Child Development	105	47	0.618	0.641	0.585	0.558	0.576	0.663	0.478	0.436
English Literature	3,372	915	0.612	0.574	0.592	0.546	0.630	0.620	0.517	0.523
Media, Film and Television Studies	341	109	0.595	0.582	0.574	0.599	0.621	0.570	0.468	0.463
German	461	69	0.589	0.524	0.635	0.562	0.580	0.499	0.387	0.325
D&T: Graphic Products	172	45	0.583	0.483	0.619	0.595	0.563	0.501	0.506	0.353
D&T: Food Technology	230	81	0.580	0.568	0.621	0.524	0.626	0.637	0.485	0.457
Music	295	68	0.571	0.590	0.541	0.613	0.533	0.546	0.371	0.434
Religious Studies	1,923	524	0.571	0.553	0.557	0.550	0.585	0.589	0.462	0.481
French	1,144	243	0.562	0.418	0.571	0.458	0.555	0.453	0.385	0.312
Physical Education	878	224	0.561	0.583	0.568	0.595	0.535	0.521	0.438	0.468
D&T: Resistant Materials Technology	342	104	0.540	0.565	0.557	0.643	0.513	0.555	0.358	0.348
Information Technology	927	240	0.537	0.445	0.545	0.500	0.540	0.470	0.443	0.412
Drama	451	101	0.506	0.533	0.482	0.502	0.527	0.570	0.336	0.477
Art and Design	990	344	0.486	0.545	0.468	0.489	0.499	0.528	0.308	0.345
Spanish	763	167	0.477	0.456	0.497	0.415	0.505	0.439	0.368	0.378
KS2 English: marks in reading test	3,588	976	0.661	0.574	0.603	0.555	0.656	0.545	0.574	0.519
KS2 maths: total test marks	3,596	979	0.653	0.602	0.749	0.708	0.565	0.477	0.512	0.469
KS2 maths: teacher-assessed NC level	3,618	1,010	0.619	0.596	0.682	0.684	0.554	0.498	0.488	0.498
KS2 science: teacher-assessed NC level	3,616	1,010	0.614	0.570	0.644	0.616	0.536	0.479	0.496	0.474
KS2 English: teacher-assessed NC level	3,618	1,010	0.607	0.559	0.608	0.588	0.597	0.543	0.516	0.497
KS2 English: marks in writing test	3,589	975	0.522	0.480	0.525	0.489	0.553	0.507	0.442	0.391

Table A7. Weighted Pearson correlations between KS2 and KS4 variables, with no Thorndike correction applied. Table sorted by strength of correlation with KS2 English marks. Red indicates stronger correlation; blue indicates weaker correlation (scaled relative to observed values).

	KS2 English: marks in reading test	KS2 English: marks in writing test	KS2 English: teacher-assessed NC level	KS2 maths: total test marks	KS2 maths: teacher-assessed NC level	KS2 science: teacher-assessed NC level
Mean GCSE grade	0.657	0.632	0.664	0.652	0.637	0.615
English	0.627	0.617	0.633	0.541	0.547	0.555
Geography	0.618	0.570	0.623	0.596	0.575	0.569
History	0.612	0.572	0.619	0.563	0.544	0.558
Highest Science	0.608	0.554	0.608	0.637	0.623	0.604
Statistics	0.607	0.517	0.623	0.675	0.585	0.565
English Literature	0.593	0.575	0.608	0.494	0.503	0.516
Maths	0.581	0.529	0.594	0.748	0.703	0.614
Media, Film and Television Studies	0.564	0.549	0.606	0.447	0.483	0.464
Core Science	0.559	0.507	0.574	0.568	0.565	0.557
Music	0.548	0.534	0.612	0.484	0.520	0.498
D&T: Food Technology	0.548	0.501	0.530	0.490	0.507	0.525
Drama	0.539	0.503	0.563	0.432	0.418	0.458
D&T: Textiles Technology	0.539	0.410	0.510	0.539	0.519	0.460
Religious Studies	0.537	0.530	0.554	0.475	0.472	0.470
Home Economics: Child Development	0.516	0.551	0.469	0.429	0.432	0.461
Business Studies	0.513	0.501	0.521	0.551	0.510	0.500
German	0.497	0.515	0.464	0.548	0.493	0.470
Additional Science	0.491	0.441	0.503	0.524	0.512	0.491
Biological Science	0.488	0.434	0.438	0.529	0.461	0.421
Average across separate sciences	0.474	0.417	0.427	0.542	0.464	0.415
Physics	0.473	0.393	0.438	0.565	0.476	0.438
D&T: Graphic Products	0.471	0.458	0.455	0.445	0.466	0.414
Information Technology	0.463	0.455	0.471	0.453	0.431	0.437
Chemistry	0.457	0.395	0.391	0.508	0.423	0.382
French	0.455	0.508	0.459	0.424	0.413	0.375
Spanish	0.436	0.447	0.423	0.436	0.387	0.403
Physical Education	0.428	0.405	0.493	0.496	0.457	0.423
Art and Design	0.395	0.405	0.398	0.388	0.397	0.367
D&T: Resistant Materials Technology	0.343	0.385	0.406	0.442	0.482	0.377

Table A8. Coefficients (standard error in parentheses) from linear mixed models with PISA score as a function of taking/not taking GCSE subjects. Models were fitted with 'school' as a random effect. Coefficients are the mean across all 10 PVs, and standard errors take into account sampling and imputation variance. Coefficients significantly different from 0 at P ***< 0.01 are indicated in bold and highlighted yellow.***

	No control variables				Concurrent attainment, gender and ESCS				Prior attainment, gender and ESCS			
Covariate	**Science**	**Maths**	**Reading**	**CPS**	**Science**	**Maths**	**Reading**	**CPS**	**Science**	**Maths**	**Reading**	**CPS**
(Intercept)	**378.9 (7.3)**	**377.7 (7.0)**	**382.1 (7.3)**	**412.7 (6.7)**	**496.6 (9.2)**	**482.4 (8.8)**	**491.5 (9.7)**	**525.9 (12.2)**	**466.3 (7.5)**	**463.7 (6.3)**	**473.4 (7.8)**	**506.5 (7.6)**
Art & Design	**12.1 (3.5)**	9.8 (3.8)	7.3 (3.1)	0.6 (4.4)	4.5 (2.9)	5.6 (3.1)	-3.4 (3.1)	-11.0 (5.3)	**9.5 (3.3)**	8.0 (3.4)	0.1 (3.4)	-7.5 (4.5)
Business Studies	**15.8 (4.1)**	**19.4 (4.2)**	**13.8 (4.6)**	**17.6 (4.9)**	-1.9 (3.3)	2.6 (4.0)	0.6 (3.6)	3.0 (4.3)	2.0 (3.4)	4.5 (3.9)	1.7 (4.1)	5.9 (4.3)
D&T: Food Technology	-3.8 (6.2)	-7.9 (5.9)	-3.8 (6.7)	-2.2 (6.8)	-2.3 (4.7)	-3.6 (4.4)	-4.7 (5.0)	-5.3 (6.0)	0.6 (4.5)	-2.3 (4.3)	-4.9 (4.9)	-5.1 (5.6)
D&T: Graphic Products	1.5 (7.5)	14.2 (6.6)	-9.5 (6.7)	-1.0 (8.7)	-5.7 (5.6)	5.6 (4.4)	-11.3 (5.2)	-5.8 (7.6)	-5.9 (5.9)	4.4 (4.9)	-13.3 (5.4)	-5.0 (7.7)
D&T: Resistant Materials	3.2 (5.1)	7.9 (4.8)	**-13.3 (5.0)**	-15.8 (6.6)	-5.9 (4.5)	-2.6 (3.9)	-9.0 (4.5)	-14.6 (6.4)	-0.2 (4.7)	0.3 (4.5)	-7.3 (4.7)	-10.6 (6.0)
D&T: Textiles Technology	-1.6 (6.5)	-8.9 (6.8)	7.9 (6.6)	4.7 (8.3)	-8.4 (5.4)	-12.1 (5.4)	-7.4 (6.3)	-13.2 (7.6)	-0.4 (6.0)	-4.1 (6.1)	-1.9 (6.5)	-6.6 (8.2)
Drama	**17.0 (4.7)**	7.9 (4.6)	11.1 (5.1)	**29.9 (5.8)**	8.5 (3.4)	2.1 (3.4)	0.1 (4.5)	**18.7 (4.9)**	9.5 (4.0)	2.7 (3.8)	-0.8 (4.5)	**19.1 (5.4)**
English Literature	**35.5 (7.4)**	**32.6 (7.0)**	**27.6 (6.3)**	**39.9 (7.6)**	0.8 (5.4)	-0.3 (4.5)	0.2 (4.8)	14.5 (6.9)	5.6 (5.3)	3.7 (4.5)	2.1 (4.7)	**16.5 (6.0)**
French	**40.4 (3.9)**	**31.2 (3.6)**	**51.2 (3.9)**	**40.6 (4.7)**	-1.1 (3.1)	-7.2 (3.3)	**15.0 (3.6)**	3.8 (4.4)	**14.4 (3.6)**	6.2 (3.3)	**24.6 (3.9)**	**13.7 (4.5)**
Geography	**25.0 (3.6)**	**16.9 (3.5)**	**22.9 (3.1)**	**10.4 (3.7)**	0.2 (3.4)	-4.4 (3.2)	3.8 (3.8)	-9.0 (4.4)	**10.5 (2.8)**	3.4 (2.8)	**9.9 (3.1)**	-3.0 (3.5)
German	**63.0 (6.3)**	**52.0 (6.0)**	**60.2 (6.5)**	**51.9 (9.3)**	**12.6 (4.8)**	3.8 (4.1)	**19.2 (6.1)**	11.1 (9.7)	**31.4 (4.9)**	**20.1 (4.6)**	**30.7 (6.1)**	21.6 (9.7)
History	**31.4 (3.4)**	**22.0 (3.6)**	**37.7 (2.7)**	**21.7 (3.4)**	4.0 (2.9)	-1.6 (3.6)	**13.1 (3.1)**	-1.1 (4.5)	**12.3 (3.0)**	5.1 (3.1)	**18.7 (2.7)**	2.4 (3.5)
Home Economics	-11.0 (8.5)	-13.6 (9.6)	-2.2 (7.9)	10.7 (11.3)	-3.3 (6.1)	-2.7 (7.4)	-6.9 (6.6)	2.7 (10.2)	2.5 (7.0)	1.9 (7.7)	-4.1 (6.8)	9.9 (10.7)
Information Technology	**33.4 (4.0)**	**25.2 (3.9)**	**16.7 (3.9)**	13.1 (5.4)	6.9 (3.3)	-1.6 (3.3)	3.3 (3.7)	-2.5 (5.3)	**20.4 (3.2)**	**8.7 (3.0)**	**11.7 (3.3)**	7.2 (5.0)
Media, Film and TV	3.1 (4.3)	3.0 (4.4)	-2.1 (4.3)	-1.0 (5.7)	-3.4 (4.0)	-2.2 (3.3)	-5.5 (4.4)	-6.0 (5.7)	0.3 (3.8)	1.2 (4.0)	-5.6 (4.6)	-4.5 (5.3)
Music	**35.8 (5.9)**	**31.0 (5.2)**	**29.7 (5.9)**	**31.7 (6.6)**	9.9 (4.4)	7.7 (4.4)	7.9 (5.1)	9.7 (6.5)	**18.6 (5.2)**	**13.7 (4.7)**	14.0 (5.5)	15.2 (6.4)
Physical Education	2.5 (3.3)	9.1 (3.9)	1.3 (3.7)	-1.3 (4.6)	**-9.3 (2.6)**	-2.5 (3.5)	-4.9 (3.2)	-8.9 (4.3)	**-8.1 (2.9)**	-4.0 (3.3)	-4.9 (3.5)	-7.4 (4.3)
Religious Studies	**14.3 (4.6)**	6.8 (4.0)	**15.9 (4.0)**	**17.0 (4.3)**	-3.4 (3.3)	-7.8 (3.0)	-0.1 (3.3)	1.1 (4.0)	2.4 (3.1)	-2.3 (2.4)	2.4 (2.8)	4.2 (3.6)
Separate sciences	**86.2 (3.0)**	**79.9 (3.0)**	**69.3 (3.5)**	**60.4 (3.5)**	**25.0 (3.2)**	**19.5 (3.4)**	**25.3 (3.6)**	**16.2 (4.6)**	**44.5 (2.7)**	**35.1 (2.8)**	**37.4 (3.5)**	**26.0 (4.4)**
Spanish	**37.3 (4.9)**	**37.0 (4.9)**	**35.4 (4.3)**	**33.9 (4.9)**	-3.7 (4.0)	-2.2 (3.7)	-0.1 (3.6)	-2.1 (4.7)	9.5 (4.3)	7.1 (4.0)	8.8 (4.1)	6.8 (4.7)
Statistics	**15.5 (5.8)**	**22.8 (6.7)**	8.4 (6.1)	**25.3 (6.4)**	-2.6 (4.4)	3.8 (5.1)	-3.2 (4.6)	10.4 (5.6)	2.9 (4.6)	7.6 (4.5)	-1.4 (4.9)	**14.3 (5.3)**
Mean GCSE grade	-	-	-	-	**9.2 (3.0)**	6.4 (5.0)	**11.5 (3.5)**	3.3 (4.5)	-	-	-	-
Mean GCSE grade^2	-	-	-	-	-1.3 (1.0)	-2.5 (1.0)	-1.1 (1.2)	-0.3 (1.5)	-	-	-	-
GCSE English grade	-	-	-	-	**5.8 (1.6)**	**6.0 (1.8)**	**15.4 (1.6)**	7.5 (3.1)	-	-	-	-
GCSE English grade^2	-	-	-	-	1.1 (0.4)	0.5 (0.5)	**1.8 (0.6)**	1.2 (1.0)	-	-	-	-
GCSE Maths grade	-	-	-	-	**14.0 (1.9)**	**27.1 (2.2)**	**9.0 (1.8)**	**17.7 (2.1)**	-	-	-	-
GCSE Maths grade^2	-	-	-	-	**1.0 (0.4)**	**2.2 (0.4)**	1.0 (0.4)	0.8 (0.5)	-	-	-	-
Highest GCSE science grade	-	-	-	-	**23.6 (2.0)**	**11.7 (2.6)**	**10.3 (2.3)**	**10.8 (3.9)**	-	-	-	-
Highest GCSE science grade^2	-	-	-	-	1.2 (0.6)	**1.5 (0.6)**	-1.5 (0.9)	-1.6 (1.4)	-	-	-	-
No. GCSE and equiv. entries	-	-	-	-	4.0 (1.6)	2.5 (1.8)	-0.4 (2.0)	2.9 (2.6)	-	-	-	-
ESCS	-	-	-	-	-1.7 (1.7)	-0.1 (1.6)	-0.7 (1.8)	-0.1 (2.0)	2.7 (1.9)	3.4 (1.7)	3.8 (1.9)	2.9 (2.1)
Male	-	-	-	-	**16.4 (2.7)**	**20.8 (2.7)**	-8.5 (3.5)	**-16.8 (3.7)**	**10.1 (3.1)**	**10.9 (3.3)**	**-14.0 (3.4)**	**-19.4 (3.9)**
KS2 English TA level	-	-	-	-	-	-	-	-	**9.9 (3.0)**	**11.8 (2.7)**	**11.1 (3.3)**	**11.2 (4.2)**
KS2 Maths TA level	-	-	-	-	-	-	-	-	8.0 (3.4)	5.8 (3.1)	**10.4 (3.5)**	5.3 (4.4)
KS2 Science TA level	-	-	-	-	-	-	-	-	**13.8 (2.8)**	**16.4 (3.3)**	1.6 (3.1)	**12.7 (3.8)**
KS2 Maths Test Total	-	-	-	-	-	-	-	-	**0.9 (0.1)**	**1.8 (0.1)**	0.3 (0.2)	**0.6 (0.2)**
KS2 English Reading Test Total	-	-	-	-	-	-	-	-	**2.7 (0.3)**	**0.6 (0.2)**	**2.7 (0.2)**	**2.9 (0.3)**
KS2 English Writing Test Total	-	-	-	-	-	-	-	-	-0.2 (0.2)	0.1 (0.2)	0.6 (0.3)	-0.8 (0.3)
N	**4,348**				**4,348**				**4,053**			

Table A9. Coefficients (standard errors in parentheses) from linear mixed models with PISA score as a function of taking/not taking GCSE subjects, concurrent attainment and attitudes to science. Models were fitted with 'school' as a random effect. Coefficients are the mean across all 10 PVs, and standard errors take into account sampling and imputation variance. Coefficients significantly different from 0 at P *< 0.01 are indicated in bold and highlighted yellow.*

	Controlling for concurrent attainment, gender, ESCS and science attitudes			
Covariate	**Science**	**Maths**	**Reading**	**CPS**
(Intercept)	**501.9 (9.3)**	**485.7 (9.8)**	**500.4 (9.8)**	**529.4 (12.8)**
Art & Design	4.1 (3.0)	5.8 (3.3)	-5.1 (3.3)	-12.3 (5.4)
Business Studies	-1.5 (3.3)	2.4 (4.4)	-1.1 (3.9)	3.6 (4.8)
D&T: Food Technology	-4.7 (5.0)	-4.7 (4.7)	-5.4 (5.3)	-5.8 (6.4)
D&T: Graphic Products	-5.8 (5.5)	5.2 (4.5)	-11.1 (5.5)	-6.0 (7.9)
D&T: Resistant Materials Technology	-5.6 (4.5)	-2.6 (4.1)	-7.0 (4.7)	-13.4 (6.8)
D&T: Textiles Technology	-8.1 (5.8)	-11.4 (5.8)	-7.0 (6.9)	-12.4 (8.8)
Drama	8.0 (3.5)	1.4 (3.5)	-1.5 (4.4)	**18.9 (5.0)**
English Literature	1.7 (5.6)	-0.6 (5.2)	-1.0 (5.0)	15.0 (6.7)
French	-0.5 (3.1)	-6.7 (3.3)	**13.5 (3.6)**	2.2 (4.7)
Geography	-1.0 (3.7)	-3.8 (3.4)	2.6 (3.8)	-9.4 (4.7)
German	**12.9 (4.8)**	4.2 (4.3)	**18.2 (6.3)**	10.5 (9.7)
History	3.2 (3.1)	-1.6 (3.6)	**12.3 (3.2)**	-2.0 (4.8)
Home Economics	-3.0 (6.3)	-2.7 (7.4)	-8.2 (6.7)	3.0 (10.7)
Information Technology	5.7 (3.4)	-2.4 (3.7)	2.5 (3.9)	-4.9 (5.8)
Media, Film and TV	-4.7 (4.0)	-4.0 (3.6)	-7.5 (4.5)	-6.8 (6.1)
Music	8.2 (4.3)	6.4 (4.7)	5.5 (5.1)	8.6 (6.5)
Physical Education	**-8.7 (2.7)**	-1.7 (3.4)	-5.5 (3.2)	-7.6 (4.4)
Religious Studies	-5.5 (3.5)	**-9.0 (3.1)**	-1.3 (3.4)	0.3 (4.3)
Separate sciences	**21.8 (3.6)**	**18.4 (3.6)**	**23.4 (3.6)**	**13.7 (5.0)**
Spanish	-3.8 (4.1)	-3.3 (3.7)	-1.8 (3.7)	-1.8 (4.9)
Statistics	-4.6 (4.7)	3.0 (5.2)	-4.9 (4.9)	9.3 (5.6)
Mean GCSE grade	**8.7 (3.1)**	6.4 (5.0)	**11.2 (3.7)**	1.7 (4.7)
Mean GCSE grade squared	-1.2 (1.2)	-2.6 (1.0)	-1.9 (1.5)	-0.1 (1.5)
GCSE English grade	**6.5 (1.7)**	**6.0 (1.8)**	**15.9 (1.7)**	**8.3 (3.2)**
GCSE English grade squared	1.0 (0.5)	0.5 (0.5)	**1.7 (0.6)**	1.5 (1.0)
GCSE Maths grade	**13.6 (2.0)**	**26.8 (2.2)**	**8.8 (1.8)**	**18.2 (2.2)**
GCSE Maths grade squared	1.0 (0.4)	**2.3 (0.4)**	1.1 (0.5)	0.8 (0.5)
Highest GCSE science grade	**22.6 (2.1)**	**11.6 (2.7)**	**9.8 (2.6)**	9.8 (4.1)
Highest GCSE science grade squared	0.9 (0.7)	1.4 (0.6)	-1.2 (0.9)	-1.7 (1.5)
Number of GCSE and equivalent entries	4.0 (1.7)	2.7 (2.0)	0.0 (1.9)	2.9 (2.9)
ESCS	-2.8 (1.8)	0.2 (1.6)	-1.1 (1.9)	0.0 (2.1)
Male	**15.3 (3.0)**	**20.7 (3.0)**	-9.4 (3.7)	**-18.3 (4.1)**
Enjoyment of science	1.9 (1.7)	-0.5 (1.3)	2.6 (1.6)	3.6 (2.1)
Interest in broad science	**7.5 (1.8)**	4.5 (1.9)	4.7 (2.0)	5.7 (2.3)
Instrumental motivation	**-4.1 (1.5)**	-2.1 (1.6)	**-4.9 (1.5)**	-3.2 (1.8)
Science self-efficacy	2.6 (1.3)	0.3 (1.3)	0.6 (1.4)	1.2 (1.7)
Index of science activities	-2.0 (1.4)	-1.0 (1.4)	-1.3 (1.5)	-2.0 (1.9)
N	**3,945**			

Appendix 2: Notes on calculation of standard errors for PISA data

Sampling variances for the estimates of correlation were calculated using the following procedure. To begin with sampling variance was calculated using only the first PV, and used the 80 Fay balanced repeated replication weights provided in PISA data. The variance was calculated as in equation 1:

$$sampling\ variance = \frac{\sum_{i=1}^{80}\left(r_1^i - r_1\right)^2}{80(1-0.5)^2} \qquad (1).$$

Here, r_1^i represents the Pearson correlation between the first PV and the variable of interest, weighted by Fay weight *i*; r_1 represents the Pearson correlation between the first PV and the variable of interest, weighted by the final student weight.

Imputation variance was calculated using all PVs and final student weights, as in equation 2:

$$imputation\ variance = \left(1+\frac{1}{10}\right)\frac{\sum_{j=1}^{10}\left(r_j - \bar{r}\right)^2}{10-1} \qquad (2).$$

Here, r_j represents the Pearson correlation between PV *j*, weighted by final student weights, and $\bar{r}$ represents the mean correlation across all 10 PVs. The standard error was then defined as (sampling variance + imputation variance)$^{1/2}$.

The Thorndike correction was applied using the 'psych' R package (Revelle 2018), using the following equation:

$$\hat{r} = \frac{r\left(\frac{S}{s}\right)}{\sqrt{1-r^2+r^2\left(\frac{S^2}{s^2}\right)}} \qquad (3).$$

Here, $\hat{r}$ is the corrected correlation, r is the correlation calculated on the restricted dataset, upper-case S is the standard deviation of the PV in the full population, and lower-case s is the standard deviation of the PV in the restricted population (i.e. only for those students with a data value for the variable of interest). When the correction was applied, this was applied before the estimation of standard error described above, such that the standard error took into account the correction.

Sampling and imputation variance for the mixed models were also taken into account using equations 1 and 2, but with regression coefficients used in place of correlation coefficients. Further, whereas correlation sampling variance was estimated using only the first PV, here, sampling variance was estimated using all PVs[21]. Imputation variance was calculated as in equation 2.

[21] Equation 1 was applied for each PV, rather than just for the first PV, producing 10 separate estimates of sampling variance. The final sampling variance estimate was simply the mean of these 10 values.

Appendix 3: Model specifications

Mixed models were fitted to assess the effect of taking specific subjects at GCSE on PISA scores. All of the models took the following general form:

$$Y_{ij} = \beta_0 + \beta_1 X1_{ij} + \beta_2 X2_{ij} + \cdots + \beta_k Xk_{ij} + u_j + e_{ij} \qquad (4).$$

Here, Y_{ij} is the score in the PISA domain being modelled for student *i* in school *j*; $X1$ to Xk are the k independent variables included as predictors; β_0 to β_k are the regression coefficients associated with each predictor variable; u_j is a random variable at school level; and e_{ij} is individual-level residual variation.

In the first model set, the only predictors included were binary variables indicating whether or not student *i* took that subject at GCSE. Hence, there were 21 X variables (i.e. Art & Design through to Statistics; see Table 4 in the main report for the full listing).

The second set of models included student background characteristics (gender and socio-economic status), and concurrent attainment. Hence, the models took the same form as the first set, but one X variable indicated gender (taking a value of 0 if student *i* was female and 1 if they were male) and one indicated socio-economic status (using the ESCS variable in the PISA dataset). Nine further X variables indicated concurrent attainment; see Table 4 in the main report for the full listing. Hence, there were 32 predictors in total in this model set.

The third set of models again included gender and ESCS, but included 6 variables indicating KS2 attainment. Hence, there were 29 predictors in total in this model set.

Models were fitted to PISA scores from a single PV at a time (i.e. Y_{ij} could only represent the student's score in one PV within a given domain). Consequently, to provide final regression coefficients, models were run separately for each of the 10 PVs for each domain, and the mean value across all 10 PVs was reported. This was carried out using code adapted from the 'intsvy' R package (Caro and Biecek, 2017) and available from https://github.com/CambridgeAssessmentResearch/intsvyExtras. For more information on the calculation of standard errors of the coefficients, see Appendix 2.