Using the England football team to identify the education production function: Student effort, educational attainment and the World Cup

Robert Metcalfe
University of Oxford

Simon Burgess
CMPO, University of Bristol

Steven Proud
University of Bristol

July 2011

PRELIMINARY – NOT TO BE QUOTED

Abstract

We use a sharp, exogenous and repeated change in the value of leisure to identify the impact of student effort on educational performance. Performance is measured using the high-stakes tests that students in England take at the end of compulsory schooling. The treatment arises from the fact that the world's major international football tournaments overlap with the exam period in England. Our data enable a clean difference-in-difference design. We find a significant negative effect, substantial for some pupils. The deterioration in grades for subjects with exams during the tournament relative is 0.09 SDs of student performance for poor boys.

Correspondence: robert.metcalfe@merton.ox.ac.uk

1. Introduction

Understanding the education production function is a central goal of much research in education. Recent studies have analysed many components of this including the impact of class size, teacher quality, other school resources, peer groups, cognitive skills, family income, parental human capital and so on. One factor that has received comparatively little attention, however, is perhaps the one that the student focuses on most – her effort in studying is one of the few aspects of educational attainment actually under her own control. Does it matter, does studying hard pay off? Data and identification problems are undoubtedly a major constraint: causal analysis of the impact of student effort on educational attainment, requires an exogenous change in the marginal value of leisure or effort. This paper attempts to quantify how much does student effort matters to educational attainment.

We use a sharp, exogenous and repeated change in the value of leisure to identify the impact of student effort on educational performance. Performance is measured using the universal high-stakes tests that students in schools in England take at the end of compulsory schooling. The treatment arises from the fact that the world's major international football tournaments overlap with the exam period in schools in England, well known to be a nation obsessed with football. These tournaments are both attention-grabbing and highly salient for many students. They happen every other summer, so each year is sequentially either a treatment year or a control year. Because of the nature of the treatment and our data, we can implement a clean difference-in-difference design. We compare within-student variation in performance during the exam period between tournament and non-tournament years using seven years of student*subject data on practically all the students in England. This data allows us to bring out the heterogeneity of impact as well as quantifying the average effect.

In some ways, the treatment is ideal for a causal study. Exposure to the treatment is random: whether a particular student is born in an even year or an odd year. Neither students nor schools can affect the timing of the exams, which are always scheduled for the same period each year. The maximum potential treatment is very strong, as detailed below: the competition always completely dominates TV, radio and other media during the weeks it takes place. Because we do not observe actual hours spent thinking about the tournament, this is an intention to treat study, and actual treatment depends on an individual's interest in football[1]; our data allow us to estimate the heterogeneity of effects. Another benefit of this data relative to other studies is that student effort is

[1] And other parameters too, discussed below, but these can be differenced out.

not confounded with teacher effort. There is little teaching at this time, leaving the student time to revise.

The key high-stakes examinations in England (called the General Certificate in Secondary Education or GCSE) are taken at the end of secondary schooling (at age 16), and are always scheduled for May and June. We obtained data on exam timetables for each subject, and compare with the tournament dates. A proportion of exams overlap with these major football tournaments, and this generates within-student variation in tournament years.

We find a significant negative average effect, substantial for some groups. The mean deterioration in grades for subjects with exams during the tournament relative to earlier subjects is 0.063 SDs of student performance. For highly affected groups such as male students, from disadvantaged families and of Black Caribbean heritage, the effect size is 0.10 SDs. Thus the overlap of exams and the tournaments reduces average attainment and raises educational inequality. In fact, since on average the summer exams only count for around half of the grade on average, the impact of the effort reduction directly on the exam score is about double these numbers. We consider the implications of these results for the role of effort in educational production in the Conclusion.

Other studies have attempted to demonstrate how non-financial effects matter for effort. Stinebrickner & Stinebrickner (2008) found that randomly being assigned to a college roommate who has a video game console significantly reduces time allocated to studying (using self-completed surveys), which then negatively impacts on educational attainment. They use the roommate with the video console as an instrument to establish the causal impact of effort on attainment. The estimated effect may include peer effects as well as the changing marginal value of leisure. There is mixed evidence that class attendance in economics at higher education is positively related to student performance. The work by Schmidt (1983), Michaels and Miethe (1989), Douglas and Sulock (1995), Young et al (2003), and George et al (2008) found increased study time, and hence effort, to be associated with higher educational attainment. Conversely, Park and Kerr (1990) and Nonis and Hudson (2006) found study time to be unrelated to student performance, and Krohn and O'Connor (2005) found that effort in undergraduate degree courses, measured by study time, is negatively associated to performance. These studies do not have clear identification strategies to demonstrate the marginal causal impact of changing the incentives of either effort or leisure. Using a birth cohort, De Fraja et al (2010) provide a theoretical and empirical model to test the impact of student, parent and teacher effort on attainment in the United Kingdom. Using the National Child Database Survey, they find that the students' and parents' effort are complements, and both seem more important to attainment than teacher effort. Effort and educational attainment are self-reported, and the former is based on a host of subjective variables about schoolwork, parents' involvement in that schoolwork, and teacher involvement.

There is a small but growing literature within education on incentivising students. These studies increase the relative marginal value of effort over leisure. There have been some studies showing substantial positive effects of financial incentives on primary/elementary and secondary/high school students (Angrist et al, 2002; Henry & Rubinstein, 2002; Jackson, 2009; Kremer et al, 2009; Angrist & Lavy, 2009; Deardon et al, 2009; Dee, 2009; Pallais, 2009). More recent experimental evidence demonstrates a lack of positive effects of financial incentives in educational attainment (Bettinger, 2008; Sharma, 2010; and Fryer, 2010) and Rodriguez-Planas (2010) reports negative long-run effects of financial incentives for high school men. In a series of randomised trials in the U.S. tertiary sector, Angrist et al (2009; 2010), MacDonald et al (2009), and Barrow et al (2010) show that incentives (in terms of merit scholarships) have modest positive effects on educational attainment, especially when combined with extra academic support services, although these benefits were short-lived. In the Netherlands, Leuven et al (2003) find mixed effects of small financial incentives for university students.

More specifically, there is little evidence of how sporting events, that are not part of the traditional human capital production function, impact on attainment.[2]

Section 2 sets out the model underlying our approach, and section 3 describes our data. We present our results in Section 4 and section 5 discusses the implications of our results.

2. Model

We adopt a very simple model of student effort, a simplified version of that developed by de Fraja et al (2010). The level of educational attainment, as reflected in an exam result, is q. We assume that students value higher levels of attainment more highly because of the expected impact on future lifetime income, denoted $V(q)$. Students exert effort, e, when studying for their examinations, which has a utility cost, $\theta(e)$, increasing and convex in effort. Effort is to be understood in a broad sense as the number of hours spent preparing for the exams; this might include making the effort to ignore

[2] There is research that demonstrates that the impact of participation in sport is beneficial for educational attainment. There is some evidence from the United States that being involved in sport has a positive association with educational and labor market success (see for example Hanks, 1979; Long & Caudill, 1991; Maloney & McCormick, 1993; Barron et al., 2000; Eide & Ronan, 2001; Cornelißen & Pfeifer, 2010; Stevenson, 2010; Lechner, forthcoming). While there seems to be some positive impacts of participating in sport on education, the actual interest in sport on effort and educational attainment has received little attention.

distractions and to create an environment to concentrate on study. Students have an underlying level of ability, a. Attainment is produced by ability and effort: $q = \delta(e, a)$. Student *i* will choose e_i to maximise $V(\delta(e_i, a_i)) - \theta(e_i)$. Optimal effort is the value of *e* that solves: $V_i^{'}.\delta_i^e(e, a) = \theta_i^{'}(e)$, allowing all of the functions to vary by student.

The major cost of effort is the value of the leisure time forgone to undertake study. We assume that this may depend on observable individual characteristics such as gender, age, family background, denoted Z_i, and an unobservable individual component, μ_i. The key factor for this paper is that the value of leisure increases for some individuals when a major football tournament takes place. We allow this impact of the tournament on the value of leisure to vary by individual, ϕ_i. This taste for watching football will depend on cultural factors and an idiosyncratic component, which we expect to be substantial. The cultural factors may be associated with observable student characteristics, for example gender, ethnicity, social class, and location. An interest in football is by no means confined to men, but in England it remains a bigger part of male culture than female culture. Football has also been more associated with lower income and working class families (see Baker, 1979; Goldblatt, 2007). It may also matter whether a pupil lives in an urban or rural area since football is notoriously a city sport, and some cities have a tradition of a very strong interest in football.

We denote the incidence of the tournament by the indicator terms *I(t = T)* and *I(m = T)*; that is: it is a tournament year, *t*, and that the tournament is actually taking place on the time of the relevant exam, *m*. We might expect an increase in the value of leisure in a tournament year even before the matches begin because of all the media attention and build up (denoted ϕ_i^0), but we expect a larger impact once the tournament is actually under way (denoted ϕ_i^1).

So the overall cost of effort function for student *i* at time *t* is: $\theta_i^{'}(e) = (e; \mu_i, Z_i, \phi_i^0.\{I(t = T)\}, \phi_i^1.\{I(t = T).I(m = T)\})$. This is a very flexible formulation.

The attainment technology, the impact of revision time and effort on the qualification gained, will also likely vary by student observable and unobservable characteristics, γ_i, and possibly by school, s_i. We allow for the possibility that the exam setting and marking may vary year-by-year by including year effects, *t*; these will net out of within-student estimates. We also allow student performance to vary through the exam period. There are many possibilities: for example, it may be that students learn and improve their exam technique as time goes by; or it could be that students tire and do worse on later exams; or it could be that later exams provide less time for last-minute revision; or it could be that, anticipating this, students over-revise for the later exams. In any case, we allow for unrestricted, idiosyncratic within-period time dummies, *m*. This is all summarised in the attainment function: $\delta_i^e(e,a) = (e;\ a_i, Z_i, \gamma_i, s_i, t, m)$. Conditional on what the student writes down in her exams, there is nothing in the setting or marking of the exams that could vary between tournament and non-tournament years beyond general time trends.

The student's valuation of the qualifications, $V_i^{'}$, may also depend on the same observable individual characteristics, Z_i, and an unobservable factor, ϖ_i.

Overall, optimal effort chosen by the student is $e_i^* = \xi(a_i, Z_i, s_i, t, m, \mu_i, \varpi_i, \gamma_i, \phi_i^0.\{I(t = T, \phi i1.It = T.Im= T$. Inserting this back into the attainment function, $q=\delta(e,a)$, gives the exam outcome, and we allow for interactions so that the impact on the exam of an hour of study effort may depend on ability. Assuming a linear form we arrive at our empirical formulation:

$$q_{itm} = \beta_0 + \beta_1.a_i + \beta_2.Z_i + \sum_\tau \alpha_\tau.I(t=\tau) + \sum_n \pi_{in}.I(m=n) + f(\phi_i^0, a_i).\{I(t = T)\} + f(\phi_i^1, a_i).\{I(t = T).I(m = T)\} + \upsilon_i + \eta_{itm} \quad (1)$$

Where υ_i combines μ_i, ϖ_i and γ_i, plus any unmeasured aspects of ability (and also absorbs the school effect, s_i), and η_{itm} is a noise term. In the results reported below, we also allow $f(\phi_i^1, a_i)$ to depend on Z_i.

In summary, the impact of the tournament on exam performance depends on the effect of the tournament on any reduction in study time (in turn depending on the student's taste for watching football), and the effectiveness of study time in raising exam scores (depending on the student's ability).

3. Data

a. Timing of Football Tournaments and Exams

Every four years (on even years) the FIFA World Cup takes place in June and July, and every other four years (on the different even years, so always two years apart) the UEFA European Championships also take place in June and July.[3] The FIFA World Cup attracts a massive worldwide audience. For instance, the 2006 World Cup in Germany had television coverage in 214 countries around the world, with 73,000 hours of dedicated programming, which generated a total cumulative television audience of 26.29 billion people (FIFA, 2007). The UEFA European Championships are not

[3] The history and background to the FIFA World Cup and European Championships can be found at http://www.fifa.com/worldcup/ and http://www.uefa.com/ respectively.

as large as the World Cup, although the 2008 Euro tournament was watched live by at least 155 million TV viewers, and the final round of the tournament was shown in a total of 231 countries (UEFA, 2008).

Appendix Table 1 documents the World Cups and European Championships from 2002 to 2008 (the years in which we have individual educational attainment scores in England). The English national football team qualified for the first three of these four international tournaments, but not for the 2008 European Championships[4]. We classify 2008 as a "non-tournament" year as England did not qualify for that tournament (see below for our support for this decision), although we also test the robustness of our results to this decision. This table also documents the time frame of the tournaments.

In each of these years, the tournament over-lapped with national UK examinations. We report the GCSE examination start dates and end dates in Appendix Table 2. There is no difference in exam dates between those years in which there is a football tournament and those years in which there is not. The proportion of exams during the football tournament ranges between 46% and 61%. The data on examination dates for each subject were obtained from Cambridge Examinations. Although different examination boards set their own exams, the exams of different boards for the same subject across the country are on the same day. The list of subjects used in our analysis is provided in Appendix Table 3. Some subjects have no exams during the tournament, others have a proportion of exams during the football period, and others have all the exams during the football period. By using the exact date of examination, we can calculate a variable that has the proportion of exams that are within the time period of the football tournament.

b. Pupil Data

The data on educational attainment is from the National Pupil Database (NPD). The NPD is an administrative dataset of all pupils in the state-maintained system, made available to researchers by the Department for Education. It includes a census of pupils, taken each year in January, from 2002 onwards. In each cohort there is approximately half a million pupils, and so over the seven year period we use, 2002 – 2008, we have some 3.5 m pupils. We have data on each student's gender, within-year age, ethnicity, an indicator of Special Educational Needs (SEN, which measures learning or behavioural difficulties), and eligibility for Free School Meals (FSM), which is dependent on eligibility for welfare benefits and is widely used as a crude measure of family poverty.

[4] As some readers may know, England did not progress very far through the knock-out stages of any of these tournaments. We considered whether we could differentiate between exams sat before and after England were eliminated, but in fact the team did manage to remain in the tournament for almost all the exam period.

This census of personal characteristics can be linked to each pupil's test score history. Our analysis uses the subset of pupils that are identifiable within the statesystem throughout this period, which amounts to 90% of the cohort[5].

Our dependent variable is the pupil's performance in the high-stakes exams at the end of compulsory schooling at age 16, called GCSEs. We have the pupil's grade for each subject taken. Each subject score will typically incorporate exams and coursework, in varying proportions, and we have the overall score, not separately exam by exam within each subject. We return to the implications of this for the interpretation of our results below. The GCSE scores are measured using the National Curriculum points system. We normalise the scores separately for each subject to remove any differences in subject difficulty; obviously the normalisation is done over all the years together as our focus is on across-year within-subject variation. Aggregate year to year variation is dealt with by common time effects. Some analyses below use the student's mean exam score; for those analyses, we normalise the total score to a mean zero and SD of one. In fact, it turns out that the SD of subject level scores and the SD of student averages over the (roughly) 7 subjects[6] they take are about the same. The latter is only about 7.5% lower because a student's scores on all the subjects s/he takes are so highly correlated. As a measure of prior attainment, we use data on tests taken at age 11, Keystage 2 tests (KS2) in English, Maths and Science. These are set and marked at a national level, remote from the school. The tests are always taken in early May so are unaffected by the tournament taking place.

Table 1 provides an overview of our data. The basic facts are that around 12% of students are eligible for FSM, and around 85% of the students are white. For some analyses we take a subset of students who take both "late" and "early" exams, and we these account for 82% of our overall data. This subset differs a little: those students taking only "early" exams are slightly more likely to be poor, and have slightly lower prior attainment. Since the compulsory subjects have "early" exams, there are no students only taking "late" exams. The lower panel of the Table shows that there is a slight shift away from late exams in tournament years: 24.9% of exams taken in non-tournament years are late, compared to 19.2% in tournament years. This suggests that exam authorities may be attempting to schedule exams a little earlier in tournament years. We consider below whether this induces a selection issue by differential sorting into early exams.

[5] Those that are excluded may have attended a private school for a period, may have spent time abroad (including Wales or Scotland), or may have been entirely educated in the English state system but their Unique Pupil Number was lost during a school transfer.
[6] Science entries are collapsed to one observation.

c. Television Viewing Data

TV viewing data provide useful support for our assumption that watching the major international football tournaments is a very widespread phenomenon. The Broadcasters' Audience Research Board (www.barb.co.uk) provides weekly data on the viewing figures of the top 30 programmes per channel.

First, we show in Appendix table 4 how football programmes dominate the list of top programmes watched in this window. Football programmes are by far the most popular thing on television across the tournament period for all years, and more people watch BBC and ITV during football periods. Second we show in Figure 1 the big changes in TV viewing habits between June and April over our sample period. People generally watch much less TV in June, but in football tournament years, June viewing is considerably higher. This difference–in–difference in viewing habits supports our interpretation of the difference-in-difference in educational outcomes below. It also supports our main assumption that 2008 (a tournament year, but for which England did not qualify) is more like a non-tournament year.

4. Results

We structure our results as follows. First, we present the simple aggregate time series story. Second, we look at simple differences between tournament and non-tournament years for matched groups of students. This gives an overall picture of the effect of tournaments but is vulnerable to being confounded with other year to year effects. So next we analyse within-individual differences and compare the distribution of these between tournament and non-tournament years. These provide our main results. This is an intention to treat study as we are estimating the effect of the potential treatment available. Fourth, we present robustness checks, and finally consider the effect size we have estimated relative to other education factors impacting on student performance.

Heterogeneity of response is likely to be very important in this context so in our results we emphasise the distribution of effects as much as the average. The effect we estimate depends on the change in the value of leisure time once the World Cup or European Championships is under way, and the impact of any reduction in study time on exam scores. The former at least is very likely to vary between students in a number of unobserved dimensions. We use observable characteristics of the students to attempt to proxy a high or low interest in tournament.

The most aggregate piece of evidence is the time series available on national average GCSE performance, which includes both state and private education. We focus on the percentage of students obtaining at least five good passes each year (grade C or higher) and Figure 2 plots the annual change in this from 1990 to 2008, with the tournament years highlighted. The visual impression that tournament years are associated with lower growth is supported by the means: a mean increase of 1.49 percentage points in tournament years against 1.63 in non-football years. However, this is just 18 observations.

a. Simple Differences across Years by Matched Groups within Schools

We now exploit the pupil-level data. We start with simple differences: how students perform in tournament years against a similar set of students in non-tournament years. Each student is only present in one year, so we must compare groups of students across years. Recalling our model, as set out in (1), and define for each group *W*, the mean of the observable and unobservable individual characteristics:

$$K_t^W = [\beta_1 . \bar{a} + \beta_2 . \bar{Z} + \bar{v}]_t^W$$

The subscript t means the individuals in group *W* in year t, noting that the school population can change slightly year to year. Using this we can define the mean attainment outcome for group *W* in tournament years:

$$\bar{q}_{t=T}^W = \beta_0 + K_{t=T}^W + \sum_{\tau=T} \alpha_\tau . I(t=T) + \overline{f(\phi^0, a)}_{t=T}^W . \{I(m \neq T)\} + \overline{f(\phi^1, a)}_{t=T}^W . \{I(m = T)\} \quad (2)$$

assuming that η_{itm} averages to zero within W. The within-period dummies are irrelevant for the total score (or rather the individual's sum of the π's just gets absorbed in the individual effect). The difference between tournament and non-tournament years is then:

$$\bar{q}_{t=T}^W - \bar{q}_{t=NT}^W = \overline{f(\phi^0, a)}_{t=T}^W . \{I(m \neq T)\} + \overline{f(\phi^1, a)}_{t=T}^W . \{I(m = T)\} + (K_{t=T}^W - K_{t=NT}^W) + (\sum_{\tau=T} \alpha_\tau . I(t=T) - \sum_{\tau=NT} \alpha_\tau . I(t=NT)) \quad (3)$$

We see that the simple difference incorporates the pre-tournament build-up effect and the effect during the tournament itself. It also includes factors from the possibly-differing populations in tournament and non-tournament years, and any differences in the general year dummies.

We first simply split our data up by gender, poverty status and quintiles of prior performance, and tabulate for each of those 20 groups the mean GCSE score in tournament-years minus the mean GCSE score in non-tournament-years, in student-level SD units. The results are in Table 2. For example, taking low ability non-poor boys, we see that this group typically score 0.044 SDs lower in tournament years than non-tournament years. Across the table, most of the entries are negative, other than for non-poor girls. Comparing the column averages, we see negative effects for three of the four groups, larger for boys than girls. There is generally a greater difference for poor boys than non-poor boys. The overall row averages show little pattern by ability. The metric in this table is student-level SDs of GCSE scores. The magnitude of these effects is therefore not negligible. We make no explicit provision for time trends here, so as can be seen from equation (3), any year-to-year aggregate effects on exam scores will also be part of these numbers.

b. Difference in Differences across Years by Matched Groups within Schools

A difference-in-difference analysis deals with this problem. Using our data on the timing of the exams and of the tournaments, we define 'late' subjects and 'early' subjects. In tournament years, late subjects are those in which at least two thirds of the exams are on dates overlapping the tournament, and early subjects are the rest. In non-tournament years, we simply take the same calendar dates in the tournament years to define late subjects. This allows us to compare performance in late and early exams within a year. As we noted above, it is likely that there will differences in performance on subjects late in the exam period versus early in the period for a number of reasons. We control for this and look for any differences in late - early performance gradients in tournament years.

There would be a potential selection problem if some students avoided taking subjects with late exams in tournament years. However, there are a number of reasons for believing this not to be a serious issue. First, subjects are chosen over two years in advance of the exam period. While obviously the occurrence of the football tournaments is fully predictable, differential overlap of this with the exams is probably not a major reason for subject choice. Second, exam dates change, so even if students were attempting to strategise this they would be unable to do so perfectly. The exams varied between early and late for around half of all subjects over our period. Thirdly, we compared the prior attainment and other characteristics of students picking late or early options in tournament and non-tournament years, and found no substantive differences, available from the authors.

For each pupil in each year, we define a late – early difference as the student’s mean score over her/his late subjects minus her/his mean score over the early subjects. This is obviously only defined for the 82% of students with both late and early subjects. From (1), the expected late – early difference in a non-tournament year is simply: $\bar{q}_{i,t=NT,m=late} - \bar{q}_{i,t=NT,m=early} = \pi_{i,late} - \pi_{i,early}$, all observed and unobserved individual characteristics drop out, the year effect drops out leaving only that student’s idiosyncratic performance change through the exam time. In tournament years, it is the same plus the impact of the tournament whilst it is in progress, $f\left(\phi_i^1, a_i\right)$.

We compare the distribution of this gradient in tournament and non-tournament years in Figure 3. We see a wide range of outcomes, but that the weight of the distribution is above zero, that is, that somewhat more students perform better in late subjects than early ones. The figure also shows a difference between tournament and non-tournament years: the latter curve is shifted to the right, that is, the typical late-early gradient is lower in years when football tournaments are staged.

We use a regression analysis to summarise the differences in the distribution across tournament and non-tournament years. We run a regression of the gradient on observable characteristics, a dummy for whether it is a tournament year, and the interaction of the tournament dummy with all the observable characteristics, all with and without school fixed effects. We cluster standard errors at school level. The results are in Table 3. In terms of reporting statistical significance, this table is the main exhibit of the paper.

Column 1 shows that the mean late-early difference is lower by 0.063 SDs in tournament years[7], and is unchanged by adding school fixed effects in column 3. Both these estimates are conditional on all our observable student characteristics (coefficients not reported). Both are very precisely estimated. We return to the quantitative significance below, but this is not trivial in magnitude.

Columns 2 and 4 introduce interactions of the observable student characteristics and the tournament indicator. Only the statistically significant interactions are reported in the Table. The strongest single interaction is with gender: the impact of the tournament is higher by 0.024 for males. Note, however, that the effect is statistically and quantitatively significant for female students too. Students from poor families also experience an additional negative impact of the tournament, adding 0.018 SDs to the base effect. Students of Black Caribbean or other Black heritage also on average have a significantly greater impact.

[7] The tournament dummy is a linear combination of year dummies. Simply including a tournament dummy and omitting two year dummies obviously means that the reported tournament effect depends on which years are omitted. So to avoid an ad hoc choice, we decided to include a full set of dummies (omitting the constant) and compute the tournament effect as: ((d2002+d2004+d2006)/3) – (d2003+d2005+d2007+d2008)/4) and test whether this is different from zero. This is the standard error and significance level reported in the table.

These effects are plausibly interpreted as mainly picking up cultural differences, as proxying a strong interest in watching the football tournament. But as we noted the impact of the tournament on exam scores in our model is the product of fewer hours of study and the effectiveness of those lost hours. The latter will depend on the student's ability, and the table shows results that confirm this view. Conditional on the other interactions, high ability students tend to suffer a higher penalty than the base effect, and low ability students less so. The extra hours of study lost due to the tournament are particularly costly to high ability students. Of course, ability is also likely to be correlated with cultural factors, but these effects appear to be swamped by the effectiveness factor.

A less parametric approach allows more of the response heterogeneity to be examined. We match groups at a very fine level, exploiting our very large and quite rich data. We do an exact match, forming groups for each student in a tournament year. We match within school, and using the key observables of student gender, FSM status, prior attainment group[8], ethnic group[9] and quarter of birth. So each student in a tournament year is matched with a student in a non-tournament year in the same school and defined by the same set of observables. We take the mean late – early difference within each school*observables group and difference this between tournament and non-tournament years.

Under our model (1), the within-individual difference removes all observable and unobservable individual characteristics, the year dummies and the effect of the general tournament build-up. The second difference across matched groups this difference yields the mean direct tournament impact for that group, $\overline{f(\phi^1, a)}^W$ plus any residual mean differences in unobservables across tournament years within our matched school*groups. The only threat to our identification would be if, for example, the mean unobservable characteristics for poor, white, middle ability boys born in the first quarter of the year and attending a specific school differed significantly between the years (2002, 2004, 2006) and the years (2003, 2005, 2007, 2008), and differed in a way correlated with tournament years across the 400,000 school*groups.

The difference-in-difference for each matched school*groups yields the mean impact of the tournament for that school*group. We examine the sign and size of these. We also examine the pattern across groups to see if there is a correlation with our views on the likely groups to have a stronger interest in football tournaments, plus the effectiveness of their lost study hours.

[8] We use three broad groups, working below the expected level (Keystage 2 score below 27), working at the expected level (KS2 of 27), or above the expected level.
[9] We use four aggregated groups: White, Black, Asian and Other.

We start in Table 4 with the results of this calculation for the same simple twenty sub-groups as in Table 2. This is the mean score in late exams in tournament years minus the mean score in early exams in tournament years, relative to the same calculation for non-tournament years. The Table shows that negative numbers are common: all the boys' groups are negative, and eight out of the ten girls' groups are. The patterns across groups are less pronounced: the effects are bigger for boys than girls, somewhat bigger for the poor then the non-poor, and a somewhat higher effect for higher ability students.

Our main matching results return to the exact matching within school*group. With such a lot of groups, and some small cell sizes, we trim these values[10]. Table 5 shows selected quantiles from the distribution of this difference-in-difference. The median school*group effect is 0.049 SDs. This is different to the equivalent regression value in Table 3 because that comes from an individual-level regression and the median is not weighted by school*group size. Focussing on the distribution across school*groups, the table reveals that some experience a very substantial effect. For example, 5% of school*groups have an effect more negative than -0.156 SDs. More than 75% of male school*groups suffer a negative effect. The difference between ability groups that we noted in the regression results is reflected here: a higher fraction of high ability school*groups face a negative impact than do average or low ability school*groups.

To focus attention on the heterogeneity of the effect by pupil characteristics, we take the weighted average across schools of the within-school difference for each group. The value for each of the 192 groups is plotted, ranked, in Figure 4, with the trimming of extreme values as described above. The Figure shows that the majority of groups see a reduction in the late-early performance gradient in tournament years. An important number of groups experience declines greater than 0.1 SDs. Equally, a smaller but non-trivial number of groups sees gains of the order of 0.05 SDs in tournament years.

The Figure displays in sub-panels the variation within gender groups, within poverty status, ability and ethnicity groups[11]. The groups ranked within gender show that in general, the most affected male groups are more affected than the most affected female groups. This is only just true for the most extremely affected male and female groups. At the other end, positively affected female groups are more positively affected than positively affected male groups. The sub-panel splitting by

[10] We drop any group that has fewer than 40 schools with data. Then, within groups, the bottom and top 2.5% of schools are winsorised. The weighted mean of the group types is then calculated, and this is winsorised for the top and bottom 2.5% again.

[11] We take out the otherwise-equivalent groups to those missing for the other subsets – so for example, if poor*females*high ability*white*first qtr birth is missing, we delete the same non-poor group.

poverty status shows that the most affected poor groups suffer a greater impact than the most affected non-poor groups, though for much of the range, the two groups experience broadly similar levels of impact. In the sub-panel for ability, the three lines are generally not far apart. What differences there are suggest that the middle ability group experiences the least impact among all but the most affected ten or so groups. The most affected low ability groups see a greater effect than the most affected high ability groups. But among the less affected 45 groups, the high ability groups are more affected than the low ability groups. This pattern is suggestive of the two countervailing effects we have proposed: the greater impact of the loss of an hour of study of a high ability student, and the greater likelihood of low ability being correlated with factors typically associated with a strong interest in football.

c. Robustness checks

We review the sensitivity of our results to a number of the data issues we have dealt with. First, instead of defining “late” subjects as those in which two thirds of exams overlapped with the tournament, we used a half as the cut-off. The average effect is still strongly negative, but as this definition is less sharp, employing this produces a lower estimated effect of the tournament: an overall average of -0.033 in the equivalent of Table 4 rather than -0.046. There is also less of a distinction between the genders.

Second, we counted the year 2008 as a non-tournament year for students in England as the national team did not qualify for the European Championships. Including it as a tournament year, on the grounds that there is still a lot of football acting as a distraction to effort, we get the results in column 2 of Table 6. We see that the overall average effect declines from -0.071 to -0.056. However, the interaction of the tournament with being male becomes more negative, so the impact for boys is barely diminished.

Third, we can extend the simple dichotomous early/late subject variable and construct a continuous variable from the exam timetable information. The variable P_{itz} is the proportion of examinations in subject *z* that are taken during the tournament period by student *i* in year *t*. We modify our main model (1) as follows, indexed now by subject *z* rather than by exam-month *m*, and collecting all the observed and unobserved student characteristics in S_i:

$$q_{itz} = \beta_0 + \beta_1 . S_i + \alpha_{t(i)} . t(i) + \phi_i . I(t = F) . P_{itz} + \varphi_i . P_{itz} + \zeta_{itz} \qquad (4)$$

Note that each student only appears in one year and so only one time dummy is relevant. The term φ_i represents the early – late gradient in exam scores in non-tournament years and $(\varphi_i + \phi_i)$ is the gradient in tournament years.

The within-estimator simply relates the exam scores across subjects within-student with the timetabling of those exams[12]. The source of variation is simply how many of the exams were timetabled later in the period, and differencing the impact of that timing across tournament and non-tournament years. We also estimate a pooled version of (4) including observable student characteristics and no student fixed effects. In this, there is again variation across subjects and across years for a given subject.

The results are in Table 7, the three columns reporting the two key coefficients of interest, $\bar{\varphi}$ and $\bar{\phi}$, for models with (1) just year dummies, (2) year dummies plus individual characteristics, and (3) year dummies and student fixed effects. As the residual error term, ζ_{itz} is likely to be correlated across each pupils' exam results, we cluster standard deviations at the individual level. The results suggest that typically students do better in later exams than earlier ones. However, this difference is significantly less in tournament years, the coefficient $\bar{\phi}$ being negative and precisely determined. This closely mirrors the results above from the simple late/early dichotomy.

Fourthly, it is clear that the identifying variation we have is between years. While our dataset gives us millions of observations, we only have a very small number of years, and so the statistical precision of our estimation is over-stated. One way round this is to run a simple regression of mean GCSE score against a time trend and a tournament dummy, within the 192 student-type groups we defined above, including school fixed effects and clustering standard errors at school level. We find negative coefficients far more common than positive on the tournament dummy. We find p-values well below 5% for the most populous groups, but much less well determined for the smaller groups[13]. Unsurprisingly given the earlier results, these regressions tend to be those for male groups, poor groups and Black groups.

e. Quantifying the effect

We have two main sets of results: the overall effect in Table 2 and the difference in difference analysis. The latter are cleanly identified and seem unlikely to be confounded, but only tell part of the story. The former gives the full effect but may be confounded by other common year effects more complex than a simple trend. We illustrate how the two different results relate to one another

[12] Recall that potential differences in the difficulty of the subject are dealt with by normalising the scores by subject.
[13] The full results are available from the authors.

in Figure 5; in principle, either one could be larger depending on the degree of impact of the pre-tournament build up, and the fraction of late exams. We focus on the difference-in-difference results.

We present the average effect and a representative high effect. The latter serves two purposes: it shows how affected the most affected groups are, and it allows us to address the importance of effort in general by asking how much do scores fall when students significantly reduce their effort. We present the results in terms of effect sizes (standard deviation units for student average test scores) and in terms of GCSE grades (counting the difference between A and B as one) for local policy interest.

The results are in Table 8, column 1 simply repeating the coefficients from the tables specified. To get to the numbers in column 2, note that the dependent variable is {(pupil mean score in late exams) – (pupil mean score in early exams)} measured in subject-SD units. The dummy on "tournament year" means that the mean late exam score is lower by the coefficient in those years. The effect on the overall average is this number times the fraction of exams that are scheduled late, which is (1.75/7.80); we also convert from subject-SD units to pupil-SD units.

These results are the impact of the tournament on the overall GCSE score. However, it is important to note that the summer exams only form part of the basis for this final outcome score, the rest being made up of coursework and continual assessment. For instance, in English, coursework can typically account for 40% of the overall GCSE grade.

The fraction of the mark from the final summer exams varies by course and by year[14]. Assembling information from a range of sources:

Year 11 summer exam contribution

Subject	%	Subject	%	Subject	%
English Language	60	English Literature	70	Maths	55 – 70
Science	50	Child Development	50	IT	40 – 50
Languages	50	Geography	75	History	75
Design & technology	40	Law	80	Media Studies	50
Music	25	Physical Ed'n (gym)	50	Religious Education	75 – 100

The three compulsory subjects average out to about 60% of the final mark coming from the summer exams, and the other subjects not that different. This suggests that the direct impact of a reduction

[14] Since 2009, coursework has been phased out in favour of "controlled assessment", and the system has changed.

in student effort on exam performance is about double the regression results: 0.126 SDs mean effect from Table 3, col 3, and 0.19 SDs for poor boys. At the very extremes, table 5 shows the most affected five percent have an impact on their grade of 0.15 SDs, so an impact on their exam score of 0.3 SDs.

To put these numbers in perspective, we can compare to a range of policy effects (see Jacob and Ludwig, 2008, and Dobbie and Fryer, 2010). For example, the "Literacy Hour" intervention in England raised reading attainment by 0.06 SDs (Machin & McNally, 2008). A unit SD increase in teacher quality raises test scores by around 0.15 to 0.24 SDs per year, 0.27 in England (Rockoff, 2004; Hanushek and Rivkin, 2005; Aaronson, Barrow, and Sander, 2007; Kane and Staiger, 2008; Slater, Davies and Burgess, 2009). The effect of major "early years" programmes such as Head Start is 0.147 SDs in applied problems and 0.319 in letter identification (Currie and Thomas, 1995; Ludwig and Phillips, 2007). Dearden et al (2007) have shown that a student's month of birth has effects on GCSE outcomes: students who have spent their entire school careers as the youngest in the class (August-borns) score on average 0.116 SDs (girls) or 0.131 SDs (boys) lower than the oldest in the class (September-born students). Substantial effects on pupil progress have been found in "No Excuses" Charter schools, of between 0.10 - 0.40 standard deviations increase per year in mathematics and reading (Abdulkadiroglu et al, 2009; Angrist et al, 2010). More closely related to our focus on effort, Fryer (2011) shows that incentivising students to raise their effort (input-based student incentives) have an effect size of about 0.15 SDs.

In this context, our results suggest that effort clearly has a non-trivial impact on student performance.

5. Conclusion

We use a sharp, exogenous and repeated change in the value of leisure to identify the impact of student effort on educational performance. This arises from the fact that the world's major international football tournaments overlap with the main high-stakes exam period in schools in England, well known to be a nation obsessed with football. We compare within-student variation in exam performance during the exam period between tournament and non-tournament years using seven years of student*subject data on practically all the students in England. This data allows us to bring out the heterogeneity of impact as well as quantifying the average effect.

We find a significant negative average effect, substantial for some groups. The mean deterioration in grades for subjects with exams during the tournament relative to earlier subjects is 0.063 SDs of student performance. For highly affected groups such as male students, from disadvantaged families and of Black Caribbean heritage, the mean effect size is 0.10 SDs. Our analysis of the heterogeneity shows that the most affected 5% of students see an impact of 0.15 SDs. In the context of the range of effect sizes reported by Jacob and Ludwig (2008), these are non-trivial amounts.

We think of this impact arising through a reduction in student effort, with that time being spent instead on watching the tournament. The variation in impact arises because of differing tastes for football, arising in turn from cultural norms and idiosyncratic factors, and from the differing effectiveness of an hour of study on learning. There are a number of contributory factors, worthy of further research. Firstly, the allocation of time by children is a largely ignored topic by economists (Larson and Verma, 1999), but could end up being crucial to educational attainment. Understanding this, along with their rate of time preference, and whether they are maximising their utility by watching the football as opposed to revising for their examinations, will be important for unpicking this relationship. Secondly, students could be involved in associated behaviours that slow cognition, such as drinking too much alcohol (Clark, 2004), which is associated with lower human capital (Chatterji, 2005; Renna, 2007). There is little research on this during football tournaments, but future research could rely on time-use data to examine this possibility. Thirdly, it could be that students' parents are watching the football tournament and pay less attention to their child revising for their examinations, which lowers their encouragement or confidence to the child during their examinations (Barnard, 2004).

Our results relate to two issues: a broad debate on the nature of educational attainment function, and a policy question specific to England about the timing of summer exams.

Taking the second first, our results show that having important exams in the tournament period reduces educational attainment at the median by 0.02 SDs of pupil average scores. We estimate much greater negative effects for male students, students from disadvantaged families and Black Caribbean students. These are groups that have lower performance anyway, and our results show that the tournament has a substantial negative effect on their performance. This in turn will affect their likelihood of progression through the educational system and their lifetime income. Given this, the benefits of moving the exams just a few weeks earlier in tournament years are significant. What are the costs? If the start of the school year was unchanged, then this would mean fewer weeks of teaching and learning, or more bunching of exams, which might be costly. But if the start was also brought forward the same number of weeks, then it would simply be readjusting the whole school calendar by a few weeks. Since we no longer have to time the long school holiday to harvest times, this is not obviously very costly in terms of ongoing costs (clearly there would be transitional costs).

The pattern of the effects we have shown means that this change would particularly benefit students from disadvantaged families, male students and students of black Caribbean heritage. There would be losers too, but the net effect of the policy would be positive and would also reduce educational inequalities. More generally, scheduling GCSE examinations during football tournaments lowers overall human capital in the UK, with implications for future economic growth.

Our results carry a number of implications for our understanding of the educational production function. The first is simply to note that student effort matters a lot. The coefficients directly show the impact on subject grades for late subjects. For strongly affected groups, this is as high as 0.1 SDs. As we noted above, this is the impact on the overall GCSE score, and given that exam performance is typically worth about half of the overall score, the impact of reduced effort in the exam period on the exam score will be roughly twice that, 0.2SDs. This is a very substantial effect. Obviously, it would be wrong to extrapolate from this number to a longer-term reduction in effort, as we have captured the effect of a reduction in high-value effort just before the exam. Nevertheless, comparing two otherwise identical students of average ability, the one putting in considerably less effort to their school work will perform substantially less well, at least 0.2 SDs worse, and conceivably worse still.

This matters for a number of reasons. First, unlike genetic characteristics, innate ability or soft skills, effort is almost immediately changeable. Our results suggest that this could have a big effect. This ties in with recent results on policies aimed at raising attainment. Fryer's (2011) and Jackson's (2010) results suggest that directly paying students for greater effort has an impact on test scores. Furthermore, the dramatic test score gains cited for "No Excuses" schools in the KIPP and HCZ or some Charter schools (Abdulkadiroglu et al. (2009), Angrist et al.(2010), Fryer and Dobbie (2010)) can plausibly be interpreted as those environments eliciting greater effort from the students. The fact that we find changes in student effort to be very potent in affecting test scores suggests that policy levers to raise effort through incentives or changing school ethos are worth considering seriously. Such interventions would be justified if the low effort resulted from market failures due to lack of information on the returns to schooling, or time-inconsistent discounting.

Secondly, the importance of a manipulable factor such as effort for adolescents' educational performance provides evidence of potentially high value policy interventions much later than "early years" policies. This is encouraging, offering some hope that low performing individuals' trajectories in life are not so heavily and irredeemably determined early in life.

Finally, there are suggestions from neuroscience that activation of the brain's motivational circuitry directly affects cognitive learning processes involving many different regions throughout the brain (Adcock,2006; Bavelier et al. (2010); Howard-Jones (2011). If so, this means that students devoting a lot of time to their studies are likely to have long-lasting effects on their ability in addition to the immediate effects on their test scores.

References

Aaronson, Daniel, Lisa Barrow & William Sander (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25: 95-135.

Abdulkadiroglu, Atila, Joshua Angrist, Susan Dynarski, Thomas J. Kane, Parag Pathak. (2009). Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots. NBER Working Paper No. 15549.

Adcock, R. A. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron,* 50: 507-517.

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. (2002). Vouchers for Private Schooling in Columbia: Evidence from a Randomized Natural Experiment. *American Economic Review*, 92: 1535-1558.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, 1: 136-63.

Angrist, Joshua D., and Victor Lavy (2009). The Effect of High School Matriculation Awards: Evidence from a Randomized Trial. *American Economic Review*, 99: 1384-1414.

Angrist, Joshua D., Phil Oreopoulos and Tyler Williams (2010). When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards. NBER working Paper No. 16643.

Baker, William J. (1979). The Making of a Working-Class Football Culture in Victorian England. *Journal of Social History*, 13: 241-251.

Barnard, Wendy M. (2004). Parent involvement in elementary school and educational attainment. *Children and Youth Services Review*, 26: 39-62.

Barron, John M., Ewing, Bradley T., and Wadell, Glen R. (2000). The Effects of High School Athletic Participation on Education and Labor Market Outcomes. *Review of Economics and Statistics* 82: 409-421.

Barrow, Lisa, Lashawn Richburg-Hayes, Cecilia Elena Rouse, and Thomas Brock. (2010). Paying for Perfomance: The Educational Impacts of a Community College Scholarship Program for Low-Income Adults. Federal Reserve Bank of Chicago Working Paper No. 2009-13.

Bavelier, D., D. M. Levi, et al. (2010). Removing Brakes on Adult Brain Plasticity: From Molecular to Behavioral Interventions. *Journal of Neuroscience*, 30: 14964-14971.

Bettinger, Eric P. (2008). Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores. Program on Education Policy and Governance Working Paper 08-03.

Chatterji, Pinka. (2005). Does alcohol use during high school affect educational attainment?: Evidence from the National Education Longitudinal Study. *Economics of Education Review*, 25: 482-497.

Clark, Duncan B. (2004). The natural history of adolescent alcohol use disorders. *Addiction*, 99: 5–22.

Corneließen, Thomas and Pfeifer, Christian (2010). The Impact of Participation in Sports on Educational Attainment: New Evidence from Germany. *Economics of Education Review* 29, 94-103.

Currie, Janet and Duncan Thomas. 1995. "Does Head Start Make a Difference?" American Economic Review. 85:3, 341-364.

Dearden, Lorraine, Claire Crawford and Costas Meghir. (2007). *When you are born matters: the impact of date of birth on child cognitive outcomes in England*. IFS Commentary.

Deardon, Lorraine, Carl Emmerson, Christine Frayne, and Costas Meghir (2009). Conditional Cash Transfers and School Dropout Rates. *Journal of Human Resources*, 44: 827-857.

Dee, Thomas (2009). Conditional Cash Penalties in Education: Evidence from the Learnfare Experiment. NBER Working Paper No. 15126.

De Fraja, Gianni, Oliveira, Tania and Zanchi, Luisa. (2010). Must try harder: Evaluating the role of effort in educational attainment. *Review of Economics and Statistics*, 92: 577-597.

Dobbie, Will and Roland Fryer Jr. (2010). Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*.

Douglas, S., and Sulock, J (1995). Estimating educational production functions with correction for drops. *Journal of Economic Education,* 26: 101–12

Eide, Eric R. and Ronan, Nick (2001). Is Participation in High School Athletics an Investment or a Consumption Good? Evidence from High School and Beyond. *Economics of Education Review* 20: 431-442.

Federation Internationale de Football Association (FIFA). (2007). FIFA World Cup and Television. http://www.fifa.com/mm/document/fifafacts/ffprojects/ip-401_06e_tv_2658.pdf

Fryer, Roland G., Jr. (2011). Financial Incentives and Student Achievement: Evidence From Randomized Trials. *Quarterly Journal of Economics*.

George, D., Dixon, S., Stansal, E., Gelb, S., & Pheri, T. (2008). Time diary and questionnaire assessment of factors associated with academic and personal success among university undergraduates. *Journal of American College Health*, 56: 706–715.

Goldblatt, David. (2007). *The Ball is Round: A Global History of Football*. London: Penguin Books.

Hanks, Michael (1979). Race, Sexual Status and Athletics in the Process of Educational Achievement. *Social Science Quarterly* 60: 482-496.

Henry, Gary T., and Ross Rubinstein (2002). Paying for Grades: Impact of Merit-Based Financial Aid on Education Quality. *Journal of Policy Analysis and Management*, 21: 93-109.

Howard-Jones, P. A. (2011). *The impact of digital technologies on human wellbeing*. Oxford, Nominet Trust

Jackson, Clement (2009). A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program. Cornell University ILR School Working Paper.

Jacob, Brian, and Jens Ludwig (2008). Improving Educational Outcomes for Poor Children. NBER Working Paper 14550.

Kane, Thomas J. and Douglas O. Staiger (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. NBER Working Paper No. 14607.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. (2009). Incentives to Learn. *The Review of Economics and Statistics*, 91: 437-456.

Krohn, Gregory A. and O'Connor, Catherine M. (2005). Student Effort and Performance over the Semester. *Journal of Economic Education*, 36: 3-28.

Larson, Reed W. and Verma, Suman (1999). How children and adolescents spend time across the world: Work, play, and developmental opportunities. *Psychological Bulletin,* 125: 701-736.

Lechner, Michael (forthcoming). Long-Run Labour Market Effects of Individual Sports Activities. *Journal of Health Economics*, forthcoming.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw (2003). The Effect of Financial Rewards on Students' Achievements: Evidence from a Randomized Experiment. CEPR Discussion Paper No. 3921.

Long, James E. and Caudill, Steven B. (1991). The Impact of Participation in Intercollegiate Athletics on Income and Graduation. *Review of Economics and Statistics* 73: 525-531.

Ludwig, Jens and Deborah Phillips (2007). The benefits and costs of Head Start. NBER Working Paper No. 12973.

Machin, Stephen and McNally, Sandra. (2005). Gender and Student Achievement in English Schools. *Oxford Review of Economic Policy*, 21: 357-372.

MacDonald, Heather, Lawrence Bernstein, and Cristofer Price (2009). Foundations for Success: Short-Term Impacts Report. Report to the Canada Millennium Scholarship Foundation.

Maloney, Michael T. and McCormick, Robert E. (1993). An Examination of the Role that Intercollegiate Athletic Participation Plays in Academic Achievement: Athletes' Feats in the Classroom. *Journal of Human Resources* 28: 555-570.

Michaels, J. and Miethe, T. (1989). Academic effort and college grades. *Social Forces*, 68: 309–319.

Nonis, S. and Hudson, G. (2006). Academic performance of college students: Influence of time spent studying and working. *Journal of Education for Business*, 81: 151–159.

Pallais, Amanda (2009). Taking a Chance on College: Is the Tennessee Education Lottery Scholarship Program a Winner? *The Journal of Human Resources*, 44: 199-222.

Park, K.H. and Kerr, P.M. (1990). Determinants of academic performance: A multinominal logit approach. *Journal of Economic Education*, 21: 101-111.

Renna, Francesco. (2007). The economic cost of teen drinking: late graduation and lowered earnings. *Health Economics*, 16: 407-419.

Rivkin, Steven G., Eric A. Hanushek and John F. Kahn. (2005). Teachers, Schools and Academic Achievement. Econometrica, 73: 417-458.

Rockoff, Jonah E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review,* 94: 247-52.

Rodriguez-Planas, Nuria (2010). Longer-Term Impacts of Mentoring, Educational Incentives, and Incentives to Learn: Evidence from a Randomized Trial. IZA Discussion Paper No. 4754.

Schmidt, R.M. (1983). Who maximizes what? A study in student time allocation. *American Economic Review*, 73: 23-28.

Sharma, Dhiraj (2010). Incentives for Academic Achievement: The Impacts on Exam Grades and Household Behavior of a Randomized Trial in Nepal. Working Paper.

Slater (2008) ???

Slater, Helen, Simon Burgess & Neil Davies (2009). Do teachers matter? Measuring the variation in teacher effectiveness in England. The Centre for Market and Public Organisation 09/212, Department of Economics, University of Bristol, UK.

Stevenson, Betsey (2010). Beyond the Classroom: Using Title IX to Measure the Return to High School Sports. *Review of Economics and Statistics* 92: 284-301.

Stinebrickner, Ralph, and Stinebrickner, Todd R. (2008). The Causal Effect of Studying on Academic Performance. *B.E. Journal of Economic Analysis and Policy*, 8: article 14.

UEFA (2004). UEFA EURO 2008™ review.

Young, M., Klemz, B. and Murphy, W. (2003). Enhancing learning outcomes: The effects of instructional technology, learning styles, instructional methods, and student behavior. *Journal of Marketing Education*, 25: 130–142.

Figure 1: Major Football Tournaments and TV viewing figures

The difference in monthly TV viewing figures, June minus April.

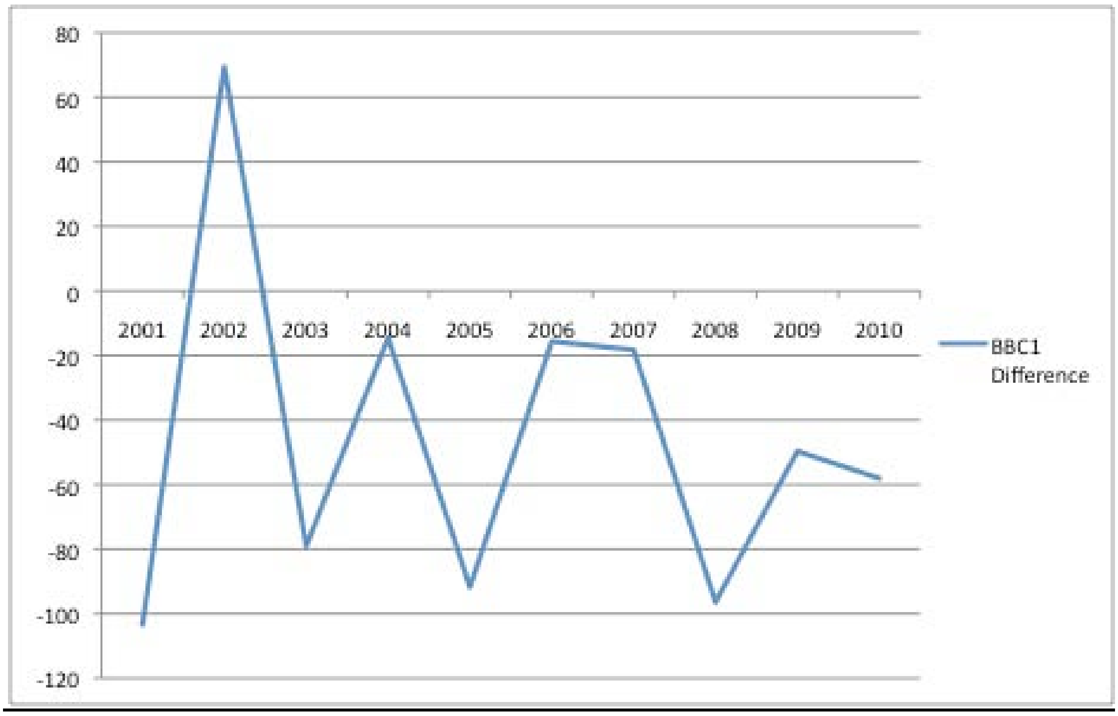


Figure 2: Annual change in percentage of pupils obtaining five good GCSEs

Tournament years highlighted.

Figure 3: Density functions for (late-early) subject score difference

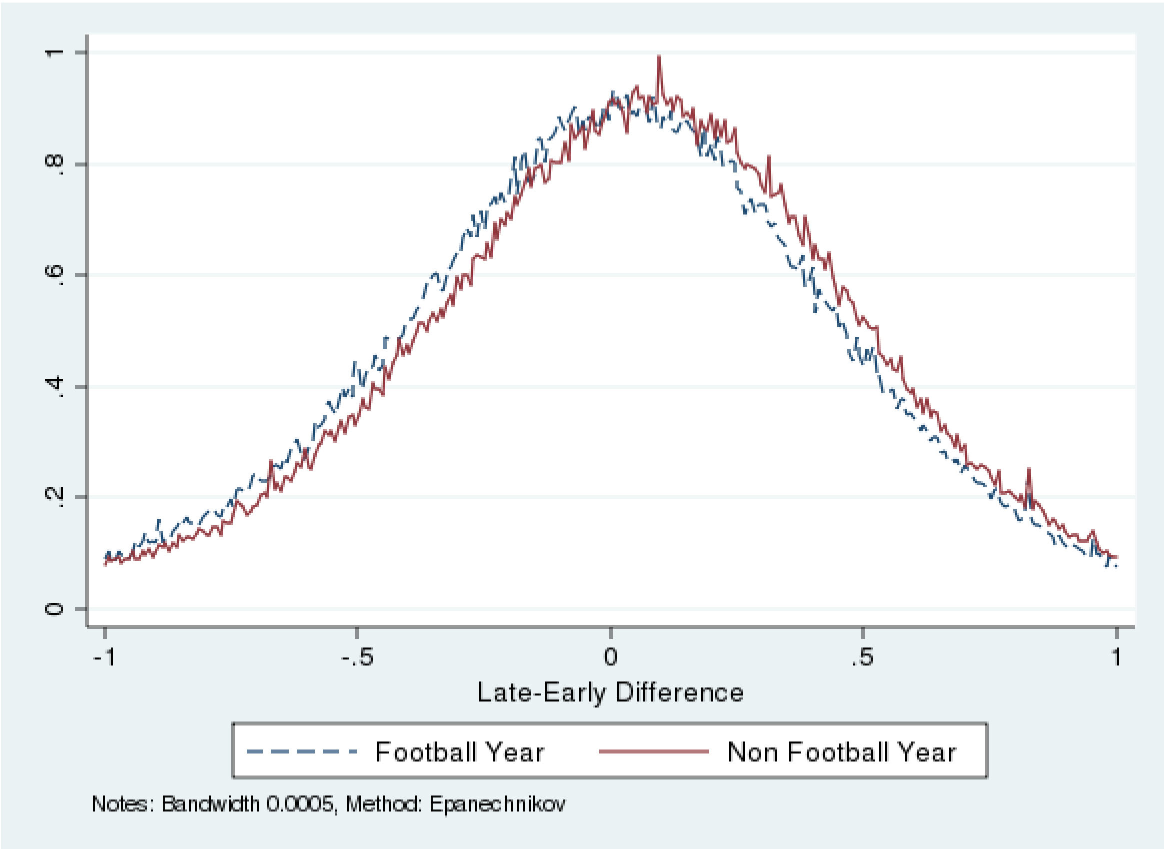
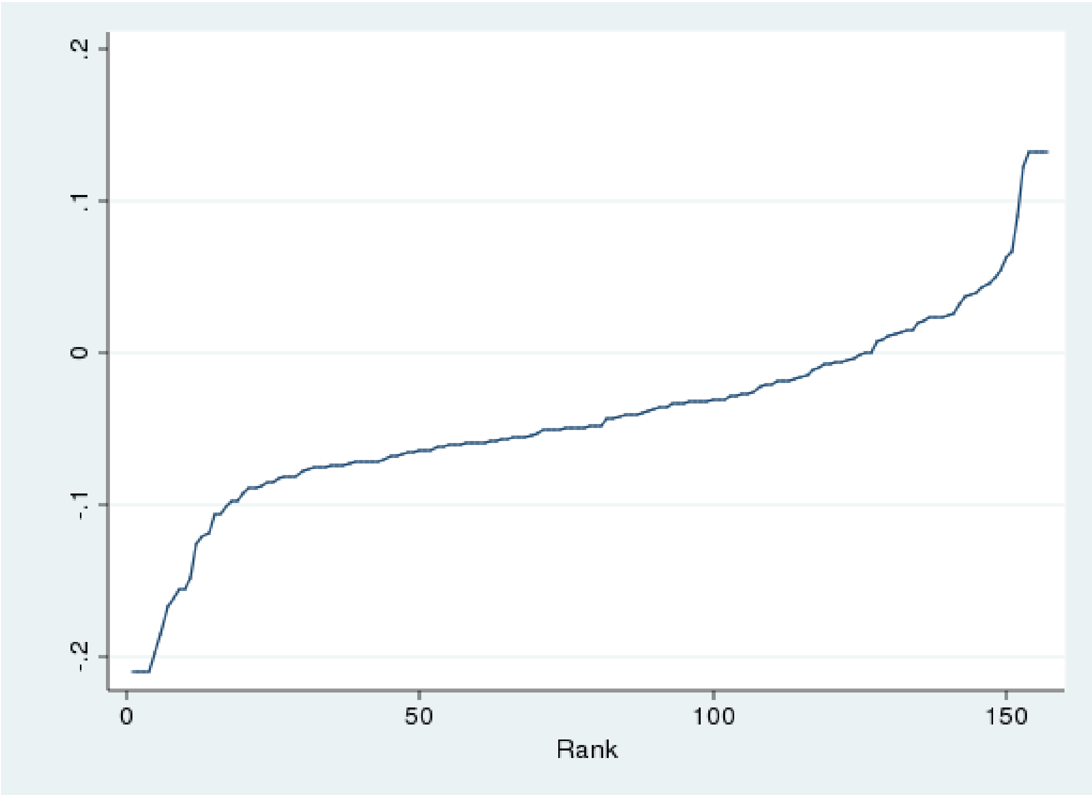
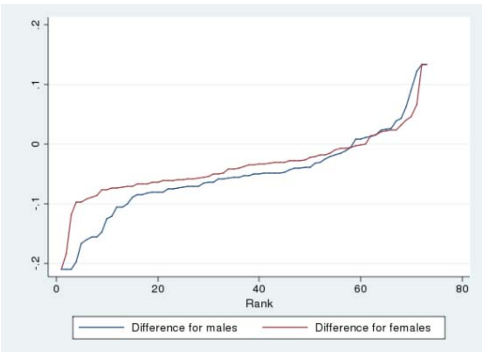


Figure 4: Difference-in-differences by matched groups

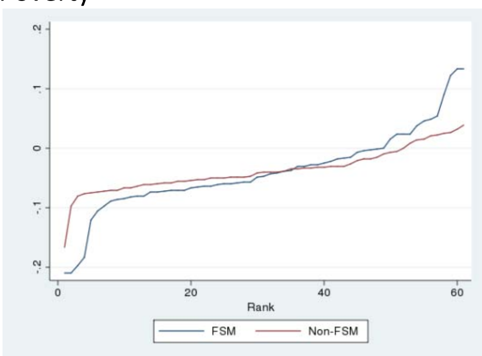
Metric is subject-level SD units



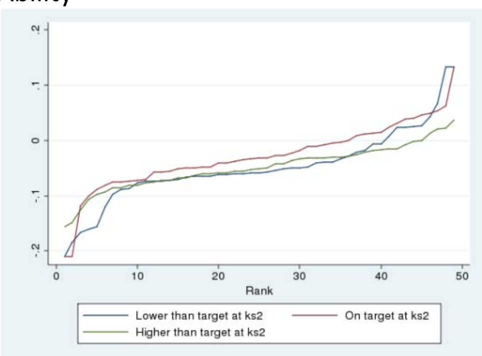
Gender



Poverty



Ability



Ethnicity

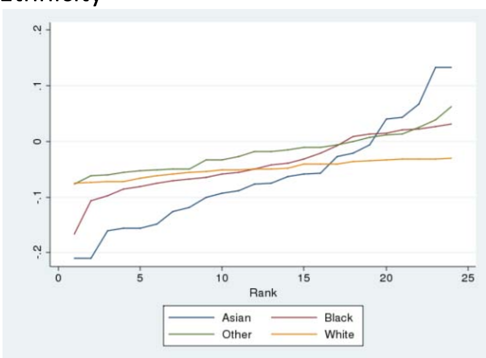


Figure 5: Comparing the Difference in difference and the total effect

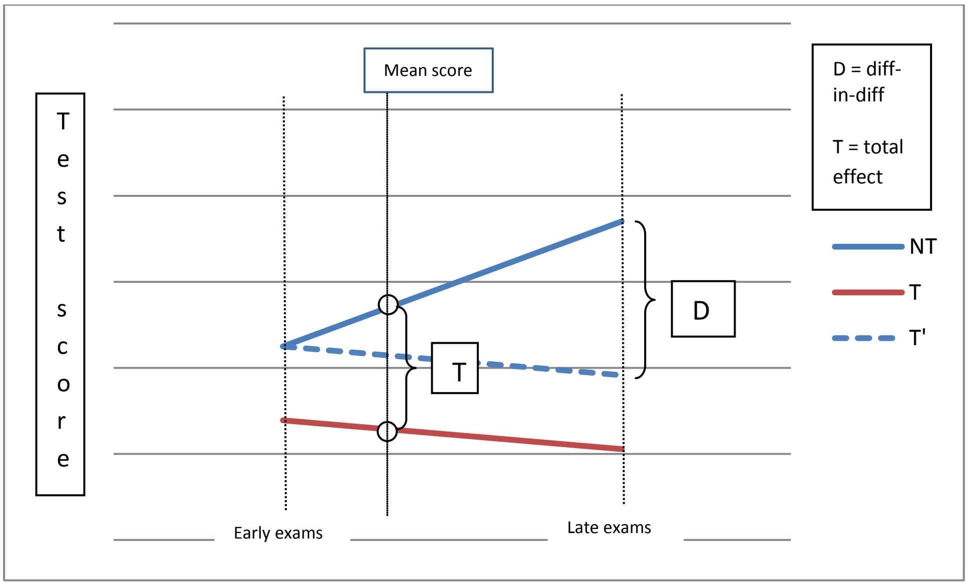


Table 1: Data Descriptives

a. Students	All	With both "late" and "early" subjects	Exam year	Students
	%			
Male	50.15	49.27	2002	478,277
FSM Eligible	12.05	11.03	2003	503,758
SEN – non-statemented	13.48	11.40	2004	528,493
SEN – statemented	2.03	1.53	2005	524,904
Selected ethnicities[2]			2006	535,033
White	84.64	84.06	2007	538,991
Black Caribbean	1.34	1.38	2008	541,229
Indian	2.33	2.47		
Pakistani	2.28	2.37	Total	3,650,685
GCSE score, normalised	-0.0405	0.0139		
Keystage 2 score	27.02761	27.34254		
Number of students	3,650,685	2,969,824		
Total observations (subjects*students)	25,739,948	21,978,566		

Notes: 1. One cohort of students per year; final year cohort.
2. Full set used in regressions.

b. Subjects	Total	Late Subjects	Early Subjects
All Years	7.80	1.76	6.33
Non-Tournament Years	7.71	1.92	6.05
Tournament Years	7.93	1.52	6.73

Table 2: Simple Average Differences

(Mean GCSE scores in football-years) – (Mean GCSE scores in non-football-years)
Metric is SD of student average score

Quintiles of KS2:	Not eligible for FSM		Eligible for FSM		All Pupils
	Girls	Boys	Girls	Boys	
Lowest	0.0137	-0.0444	0.0041	0.0051	-0.0137
2	0.0871	-0.0633	-0.0195	-0.0447	0.0149
3	0.0733	-0.0739	-0.0735	-0.0916	-0.0196
4	0.0369	-0.0691	-0.0937	-0.0978	-0.0268
Highest	0.0370	-0.0370	-0.0644	-0.0888	-0.0082
All	0.0494	-0.0564	-0.0433	-0.0552	-0.0093

Notes: The number in each cell is: {mean (over pupils in that cell) of the pupil-mean of (normalised GCSE scores) in football years} – {mean (over pupils in that cell) of the pupil-mean of (normalised GCSE scores) in non-football years}. The normalisation is by subject.

The smallest cell size is 30823

The quintiles are group and year specific.

Table 3: Regression analysis of (late – early) difference

Unit is Individual student
Metric is subject level SD

	(1)	(2)	(3)	(4)
Tournament year	-0.063***	-0.054***	-0.063***	-0.050***
	(0.001)	(0.002)	(0.001)	(0.002)
Tournament year interacted with:				
Low prior attainment		-0.008***		-0.008***
		(0.002)		(0.002)
High prior attainment		-0.013***		-0.013***
		(0.002)		(0.002)
FSM		-0.019***		-0.018***
		(0.002)		(0.002)
Male		-0.024***		-0.024***
		(0.001)		(0.001)
Black Caribbean		-0.011*		-0.011**
		(0.006)		(0.006)
Other Black Heritage		-0.025***		-0.020**
		(0.010)		(0.009)
Pakistani		-0.021***		-0.018***
		(0.006)		(0.005)
Other Asian Origin		-0.031***		-0.033***
		(0.010)		(0.010)
Mixed White and Black Caribbean heritage		0.041***		0.042***
		(0.008)		(0.008)
School Fixed effects			Y	Y
Observations	2969824	2969824	2969824	2969824
Number of schools			3283	3283
R-squared	0.03	0.03	0.03	0.03

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Dependent variable is the difference between the individual's mean score in late subjects and early subjects
Variables also included as main effects are: pupil gender, ethnicity, FSM status, SEN status, prior attainment, and location, plus year dummies. Only a selection of interactions are shown.

Table 4: Simple Difference in differences

Metric is subject-level SD

Quintiles of KS2:	Not eligible for FSM		Eligible for FSM		All Pupils
	Girls	Boys	Girls	Boys	
Lowest	-0.0018	-0.0569	-0.0436	-0.0741	-0.0336
2	-0.0422	-0.0698	0.0050	-0.0835	-0.0523
3	0.0134	-0.0669	-0.0104	-0.0828	-0.0338
4	-0.0383	-0.0919	-0.0500	-0.0825	-0.0653
Highest	-0.0484	-0.0519	-0.0286	-0.0760	-0.0504
All	-0.0233	-0.0658	-0.0245	-0.0796	-0.0455

Notes: The normalisation is by subject.
The smallest cell size is 30823.
The quintiles are group and year specific.

Table 5: Quantiles of Differences-in-differences for matched school*groups

Metric is subject-level SD

	{(Late – early) in tournament} - {(Late – early) in non- tournament}						
	p5	**p10**	**p25**	**p50**	**p75**	**p90**	**p95**
All Pupils	-0.1558	-0.0978	-0.0716	-0.0486	-0.0109	0.0244	0.0455
Male	-0.1610	-0.1483	-0.0815	-0.0536	-0.0209	0.0231	0.0383
FSM	-0.1207	-0.0872	-0.0716	-0.0481	-0.0075	0.0239	0.0491
Low KS2	-0.1666	-0.1558	-0.0732	-0.0589	-0.0210	0.0267	0.0667
Middle KS2	-0.1188	-0.0888	-0.0565	-0.0315	0.0086	0.0455	0.0540
High KS2	-0.1261	-0.0978	-0.0720	-0.0506	-0.0259	-0.0004	0.0201

Notes: The underlying data are the mean (late – early) difference in tournament years for a given group in a given school minus the same for non-tournament years. Groups are defined by gender*FSM status*prior ability group*broad ethnic category*quarter of birth. These are quantiles of the distribution of this number across all school*groups..
The normalisation is by subject.

Table 6: Robustness Checks

	Base	Treat 2008 as Tournament year
Tournament year	-0.071***	-0.056***
	(0.002)	(0.002)
Tournament year interacted with:		
Low prior attainment	-0.008***	-0.008***
	(0.002)	(0.002)
High prior attainment	-0.013***	0.020***
	(0.002)	(0.002)
FSM	-0.018***	-0.008***
	(0.002)	(0.002)
Male	-0.024***	-0.037***
	(0.001)	(0.001)
More ….		
Student characteristics	Y	Y
School fixed effects	Y	Y
Observations	2969824	2969824
Number of schools	3283	3283
R-squared	0.03	0.03

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Dependent variable is the difference between the individual's mean score in late subjects and early subjects
Variables also included as main effects are: pupil gender, ethnicity, FSM status, SEN status, prior attainment, and location, plus year dummies.
Only statistically significant interactions are shown.

Table 7: Student*subject fixed effect regression results

Observation = student*subject
Metric is subject-level SD

	(1)	(2)	(3)
Proportion of exams within subject which	0.067***	0.102***	0.126***
are “late”	(0.001)	(0.001)	(0.001)
Proportion of exams within subject which	-0.007***	-0.009***	-0.009***
are “late” * Year is a tournament year	(0.001)	(0.001)	(0.001)
Year dummies	Y	Y	
Student Characteristics		Y	
Student fixed effects			Y
Number of observations			
Number of pupils	3,650,516	3,650,516	3,650,516

Notes: 1. Standard errors in parentheses; standard errors clustered at student level.
2. * significant at 10%; ** significant at 5%; *** significant at 1%
3. “Late” is defined by calendar date for all years, and coincides with the tournament dates; see text for details
4. Student characteristics are: gender, ethnicity, month of birth, poverty status, SEN status, English as additional language, prior ability measures (Keystage 2 English score, Keystage 2 maths score, Keystage 2 Science score)

Table 8: Quantifying the Results: Other Metrics

	Coefficient	Effect size Metric: SD of pupil mean score	Effect Metric: GCSE grades
Difference in difference			
Table 3			
Mean (col. 3)	-0.063	-0.015	-0.213
Poor, male, white (col. 4)	-0.113	-0.024	-0.335
Table 5			
All pupils, (median)	-0.049	-0.012	-0.167
All pupils, (p10)	-0.098	-0.024	-0.333
Male pupils, (median)	-0.054	-0.013	-0.180
Male pupils, (p10)	-0.148	-0.036	-0.500

Notes: Column 2 = column 1* (1.76/7.80)*(11.54/10.68) {share of late exams}*{converting subject sd to pupil sd}

Column 3 = column 2*(10.68/6)*7.80 {converting to gcse points} {converting to letter grades} multiplying by the number of exams.

Appendix Tables

Appendix Table 1: The football tournaments from 2002 to 2008

Year	Host country	Tournament	Did England qualify?	Start date	End date
2002	South Korea and Japan	World Cup	Yes	31st May	30th June
2004	Portugal	European championships	Yes	12th June	4th July
2006	Germany	World Cup	Yes	9th June	9th July
2008	Austria and Switzerland	European championships	No	7th June	29th June

Appendix Table 2: The examination dates from 2002 to 2008

Year	'Football' year	Examination start date	Examination end date	% of exams during football
2002	Yes	13th May	28th June	61%
2003	No	12th May	27th June	-
2004	Yes	17th May	30th June	49%
2005	No	16th May	30th June	-
2006	Yes	15th May	28th June	48%
2007	No	14th May	27th June	-
2008	Yes	13th May	25th June	46%

Appendix Table 3: Subjects that do and do not have overlap with the football tournaments in 2002, 2004 and 2006

2002 Subjects that have no exams during the football	2002 Subjects that have some exams during the football	2002 Subjects that have all their exams during the football
D&T automobile studies	Biology	Economics
Drama	Business Studies	Greek
English Lit	Chemistry	History
Geography	Classical civilisation	Home economics
German	D&T Electronics	Information Technology
Health studies	D&T Food	Physics
Humanities	D&T Graphics	Psychology
Information Studies	D&T Industrial	Social Science
Music	D&T Resistant Materials	Sociology
Persian	D&T Systems and Control Technology	
Physical Education	D&T Textiles Technology	
	French	
	English Language	
	Latin	
	Mathematics	
	Portuguese	
	Religious Studies	
	Religious Education	
	Science (both double and single)	
	Turkish	

2004 Subjects that have no exams during the football	2004 Subjects that have some exams during the football	2004 Subjects that have all their exams during the football
Applied ICT	Biblical Hebrew	Applied Business
Applied Science	Biology	Chemistry
Business and communication systems	Business Studies	Classical Civilisation
Citizenship studies	D&T Electronics	Classical Greek

D&T Systems and Control technology	D&T Food	Economics
Drama	D&T Graphics	History
Engineering	D&T Industrial	Home economics
Geography	D&T Resistant Materials	Leisure and Tourism
German	D&T Textiles Technology	Media Studies
Health and Social Care	English Language	Physics
Humanities	English Lit	Psychology
ICT	French	Sociology
Latin	Mathematics	
Manufacturing	Religious Studies	
Music	Science (both double and single)	
Persian		
Physical Education		
Portuguese		
Rural and Agricultural Science		
Spanish		
Turkish		

2006 **Subjects that have no exams during the football**	**2006** **Subjects that have some exams during the football**	**2006** **Subjects that have all their exams during the football**
Applied ICT	Biblical Hebrew	Applied Business
Applied Science	Biology	Chemistry
Business and communication systems	Business Studies	Classical Civilisation
Citizenship studies	D&T Electronics	Classical Greek
D&T Systems and Control technology	D&T Food	Economics
Drama	D&T Graphics	History
Engineering	D&T Industrial	Home economics
Geography	D&T Resistant Materials	Leisure and Tourism
German	D&T Textiles Technology	Media Studies
Health and Social Care	English Language	Physics
Humanities	English Lit	Psychology
ICT	French	Sociology
Latin	Mathematics	
Manufacturing	Religious Studies	
Music	Science (both double and single)	

Persian
Physical Education
Portuguese
Rural and Agricultural Science
Spanish
Turkish

Appendix Table 4: The list of football programmes that reached the top 15 most viewed programmes

Channel and programme rank for that week		Week ending and programme	Viewers (millions)
2002			
BBC1		w/e 9th June 2002	
	2	WORLD CUP 2002: ARGENTINA V ENGLAND (FRI 1230)	12
	5	WORLD CUP 2002: POST-MATCH (FRI 1420)	10.49
BBC1		w/e 16th June 2002	
	2	WORLD CUP 2002: ENGLAND V DENMARK (SAT 1230)	12.47
	4	WORLD CUP 2002: ENGLAND V NIGERIA (WED 0730)	12.22
	7	WORLD CUP 2002: POSTMATCH (SAT 1420)	8.85
	9	WORLD CUP 2002: SPAIN V IRELAND (SUN 1230)	7.77
	13	WORLD CUP 2002: POST-MATCH (WED 0920)	6.65
	15	MATCH OF THE DAY (TUE 1859)	5.74
BBC1		w/e 23rd June 2002	
	1	WORLD CUP 2002: ENGLAND V BRAZIL (FRI 0730)	12.46
	6	WORLD CUP 2002: POST-MATCH (FRI 0920)	9.77
BBC1		w/e 30th June 2002	
	4	WORLD CUP 2002: GERMANY V BRAZIL (SUN 1200)	10.08
	7	WORLD CUP 2002: POST MATCH (SUN 1350)	8.95
2004			
BBC1		w/e 13th June 2004	
	7	EURO 2004: SPA V RUS (SAT 1945)	6.4
	8	EURO 2004: PORT V GRC (SAT 1700)	6.19
	15	EURO 2004: PREMATCH (SAT 1929)	5.02
ITV1		13th June 2004	
	1	EURO 2004 FRA V ENG (SUN 1944)	17.8

	12	EURO 2004 PRE MATCH (SUN 1915)	7.21
BBC1		w/e 20th June 2004	
	4	EURO 2004: SPA V PORT (SUN 1945)	8.78
	5	EURO 2004: GER V NETH (TUE 1945)	7.95
	6	EURO 2004: CRO V FRA (THU 1946)	7.55
	7	EURO 2004: POST-MATCH (THU 2135)	7.23
	8	EURO 2004: POST-MATCH (SUN 2135)	6.85
	9	EURO 2004: NETH V CZECH (SAT 1945)	6.74
ITV1		w/e 20th June 2004	
	1	EURO 2004 ENG V SWI (THU 1659)	14.31
	11	EURO 2004 RUS V POR (WED 1943)	6.51
	12	EURO 2004 ITA V SWE (FRI 1944)	5.78
	13	EURO 2004 SWE V BUL (MON 1944)	5.73
BBC1		w/e 27th June 2004	
	1	EURO 2004: POR V ENG (THU 1945)	20.66
	2	EURO 2004: CRO V ENG (MON 1945)	18.28
	3	EURO 2004: POST-MATCH (MON 2136)	14.48
	4	EURO 2004: POST-MATCH (THU 2229)	14.22
	5	EURO 2004: PREMATCH (THU 1929)	11.71
	7	EURO 2004: PREMATCH (MON 1929)	9.83
	14	EURO 2004: CZE V DEN (SUN 1945)	6.72
ITV1		w/e 27th June 2004	
	4	EURO 2004 GER V CZE (WED 1944)	8.28
	9	EURO 2004 SWE V NETH (SAT 1945)	7.04
	13	EURO 2004 ITA V BUL (TUE 1944)	6.08
	14	EURO 2004 FRA V GRE (FRI 1944)	5.69
2006			
BBC1		w/e 4th June 2006	
	1	MATCH OF THE DAY LIVE (TUE 1958)	9.29
	13	MATCH OF THE DAY LIVE (SAT 1330)	4.49
BBC1		w/e 11th June 2006	
	1	WORLD CUP 2006: ENG V PAR (SAT 1400)	12
	2	WORLD CUP 2006: POST-MATCH (SAT 1551)	9.29
	10	WORLD CUP 2006: GER V CRI (FRI 1701)	5.65
	15	WORLD CUP 2006: MEX V IRN (SUN 1702)	4.46

ITV1		w/e 11th June 2006	
	11	WORLD CUP 06 (SUN 1958)	6.02
	12	WORLD CUP 06 (FRI 1958)	5.71
	13	WORLD CUP 06 (SAT 1959)	5.44
BBC1		w/e 18th June 2006	
	1	WORLD CUP 2006: BRA V CRO (TUE 2000)	9.64
	2	WORLD CUP 2006: GER V POL (WED 2000)	8.11
	4	WORLD CUP 2006: POST-MATCH (WED 2149)	6.74
	5	WORLD CUP 2006: ITA V GHA (MON 2000)	6.69
	8	WORLD CUP 2006: POST-MATCH (SUN 2151)	6.39
	9	WORLD CUP 2006: POST-MATCH (TUE 2151)	6.38
	10	WORLD CUP 2006: FRA V KOR (SUN 2000)	6.17
	12	WORLD CUP 2006: POST-MATCH (MON 2150)	5.78
	13	WORLD CUP 2006: PREMATCH (TUE 1930)	5.29
	15	WORLD CUP 2006: FRA V SWI (TUE 1700)	5.14
ITV1		w/e 18th June 2006	
	1	WORLD CUP 06: ENG V TRI (THU 1650)	13.67
	5	WORLD CUP 06: BRA V AUS (SUN 1658)	8.08
	10	WORLD CUP 06: SWE V PAR (THU 1959)	6.63
	12	WORLD CUP 06: ITA V USA (SAT 1959)	5.76
BBC1		w/e 25th June 2006	
	1	WORLD CUP 2006: ENG V ECU (SUN 1600)	16.29
	2	WORLD CUP 2006: POST-MATCH (SUN 1750)	13.45
	3	WORLD CUP 2006: ARG V MEX (SAT 2000)	8.46
	4	WORLD CUP 2006: JAP V BRA (THU 2000)	7.81
	10	WORLD CUP 2006: PREMATCH (SUN 1529)	7.44
	11	WORLD CUP 2006: SPA V TUN (MON 2000)	7.21
	13	WORLD CUP 2006: POST-MATCH (THU 2149)	7.12
	15	WORLD CUP 2006: POST-MATCH (MON 2149)	6.25
ITV1		w/e 25th June 2006	
	1	WORLD CUP 06 (TUE 1950)	18.46
	3	WORLD CUP 06 (WED 1958)	8.74
	7	WORLD CUP 06 (SUN 1958)	7.43
	9	WORLD CUP 06: PREMATCH (TUE 1903)	6.7

2008

BBC1		w/e 8th June 2008	
	15	EURO 2008: MATCH OF THE DAY LIVE (SUN 1926)	4.53
BBC1		w/e 15th June 2008	
	10	EURO 2008: MATCH OF THE DAY LIVE (FRI 1929)	5.58
ITV1		w/e 15th June 2008	
	10	EURO 2008 LIVE (MON 1929)	5.74
	12	EURO 2008 LIVE (SUN 1929)	4.72
BBC1		w/e 22nd June 2008	
	4	EURO 2008: MATCH OF THE DAY LIVE (SUN 1930)	7.21
	5	EURO 2008: MATCH OF THE DAY LIVE (TUE 1929)	6.29
	7	EURO 2008: MATCH OF THE DAY LIVE (FRI 1929)	5.64
	14	EURO 2008: MATCH OF THE DAY LIVE (MON 1930)	4.64
ITV1		w/e 22nd June 2008	
	3	EURO 2008 LIVE (SAT 1929)	7.37
	5	EURO 2008 LIVE (THU 1929)	6.89
	13	EURO 2008 LIVE (WED 1929)	4.45
BBC1		w/e 29th June 2008	
	1	EURO 2008: MATCH OF THE DAY LIVE (SUN 1856)	8.84
	6	EURO 2008: MATCH OF THE DAY LIVE (WED 1929)	6.95
ITV1		w/e 29th June 2008	
	6	EURO 2008 LIVE (THU 1929)	6.77