Erasmus University of Rotterdam

Erasmus School of Economics

BSc. Economics and Business Economics

Bachelor Thesis

The effect of Performance Pay on Teachers' Wage Scheme on Student Outcomes

Student: Germán Pons

Student number: 476639

Supervisor: dr. Anna Baiardi

Second Assessor: dr. Josse Delfgaauw

June 2020

Abstract:

This research studies the effect of performance pay in teachers' wage scheme on student outcomes. Theory and literature indicate a positive effect of such wage schemes on private sector workers, however, it is rare to find cases where this is applied in the public sector. To research this relationship two policies are studied, one implemented in Portugal in 2007 and one in England in the year 2000. Prior to those reforms, teachers in both countries perceived no element of incentives pay. In order to analyze student outcomes, national exams from each country taken by 15- and 16-year-old pupils are used. In the case of Portugal, the data includes the average grades obtained on Portuguese language and mathematics by region. The data on England contains a common threshold used to assess student performance. More specifically, the percentage of students achieving 5 or more grades between A* and C in the GCSE national exams by local education authority. To analyze the Portuguese case a difference-in-difference approach with regional fixed effects is used. For the English case, the chosen approach is a local education authority fixed effects regression. The results indicate a positive and statistically significant effect of both policies on student outcomes. In the case of Portugal, grades increased by 2,45% for both subjects and in England the percentage of students achieving the threshold increased by 5,02%.

TABLE OF CONTENTS

1 Introduction

Education is widely regarded as one of the best human capital investments children and young people can take on. Basic education helps people build a solid foundation for a good future, both as a professional and as an individual. The value that education provides to individuals and society as a whole is of such importance that the United Nations decided to incorporate the compulsory access to free elementary education as one of the 30 articles that compose the Universal Declaration of Human Rights. Moreover, providing access to high-quality education is one of the priorities of every responsible government around the World. In the words of Barack Obama, during his 2009 speech to American students "Every single one of you has something to offer. And you have a responsibility to yourself to discover what that is. That's the opportunity an education can provide. […] What you make of your education will decide nothing less than the future of this country." (Hairston-Ridgley, 2011).

Naturally, as the quality of education increases, so do the benefits obtained from it. Over the last decades profound research has been carried out on the effects of education on different fields of study, such as health or economics. The consensus in the literature is that the benefits of high education levels range from increasing the amount of future career opportunities and wage rates (Card 1999) to lowering crime rates, improving health and increasing citizens voting participation (Lochner 2011). Therefore, increasing the quality of public-access education is paramount for nations. However, finding ways to do so can be challenging as there are copious variables to consider. Most definitely, the role of teachers is fundamental in determining student outcomes. Previous research suggests that teacher motivation has a positive effect on students' interest (Keller et al 2016). Hence, enhancing teachers' motivation levels could positively affect the quality of students' learning outcomes. One possible way to increase teachers' motivation and make sure that students' performance is on their best interest is through the incorporation of an incentive-pay wage scheme. Principal-agent models show that adding an element of performance pay would lead to a positive effect on the levels of effort exerted (McMillan, 1992). However, in the public sector we mostly see wage schemes which fully consist of a fixed component. Overall, there are multiple reasons to believe that implementing this method in the education system could have an effect on teachers' motivation and effort exerted, which could affect students' performance. The aim of this research is to determine whether this is the case, hence, the research question is:

How does performance pay in teachers' salary affect students' learning outcomes?

To study the relationship between performance pay and student outcomes two policies implemented in different education systems will be considered. The first policy took place in Portugal in the year 2006 and the second one was implemented in England in the year 2000. Both policies as well as the education systems in which they were implemented will be discussed in more detail in later sections.

Regarding the methodology, a difference-in-difference approach with regional fixed effects is used for Portugal. In order to do so, this paper takes as a control group the autonomic region of the Madeira islands, which is not affected by the implementation of the analyzed policy. As for the English education system, the method used consists of a regional fixed effects regression, covering the 133 local education authorities existing during the research period.

The results from those regressions indicate a significant and positive effect of the introduction of performance pay on student outcomes in both cases. For Portugal, the effect is of approximately 2,45%. The results for England indicate a larger effect, to be precise, an increase of 5,02% in the fraction of students achieving the threshold of 5 or more subjects graded between A* to C. However, although the effect of the English policy is statistically significant, its interpretation is hindered by limitations affecting the method. Nevertheless, this research contributes to the field of personnel economics. More specifically to the existing yet limited literature on the effect of incentive pay on performance in the public sector. It is worth mentioning that the results are in line with those of previous research and principal-agent theory.

Given the obtained results and the previous literature on this topic, the introduction of incentive pay in the education system can be a great tool for policy makers who aim to improve student outcomes. Nevertheless, education entails more than student outcomes and it is unclear whether incentive pay could negatively affect the quality of other tasks that teachers are responsible for, such as counselling or creating a pleasant school atmosphere for students. Consequently, further research on those fields is encouraged and would be beneficial to get closer to understanding the full effect of incentive pay on students.

1.1 Relevance

As mentioned in the introduction, higher education levels are associated with a great range of benefits for both students and society as a whole. The effect of education on the economy has been thoroughly studied and literature indicates a direct positive effect of education on economic development. Generally speaking, the economic attainment of people and groups with more schooling is better than that of those with less education (Vila, 2000). Consequently,

the higher the rate of individuals with high education levels in a society, the higher the labour productivity will be and the more resources the government will have access to. Leigh (1998) presents two mechanisms through which education has a positive impact on a nation's economy; as education levels raise individuals improve their economic attainment, hence more money is collected through income taxes. Simultaneously, the number of individuals which depend on subsidies and other forms of economic support is reduced. Furthermore, education offers benefits which surpass the positive effects on individual income and on the economy. Increased education is positively and strongly correlated with measures of health, strong families, children's well-being, a clean environment and absence of violent crime (Leigh, 1998). Taking into consideration all the mentioned benefits it is indisputable that finding methods to improve the quality of education is relevant for society. Consequently, studying the effect of different policies on student outcomes is socially relevant.

Furthermore, even though performance pay has been researched in depth in the private sector it is not a commonly adopted wage scheme in public institutions, Therefore, not a lot of research has been done on its effect in the public sector, where employees are used to perceiving a salary completely based on a fixed amount. Hence, it is of scientific relevance to study whether performance pay affects public sector workers the way principal-agent theory predicts. It is worth mentioning that although limited, the current literature on performance pay in the public sector focuses on the fields of public administration, economics, education and health (Hasnain, Manning, & Pierskalla, 2012)

Moreover, finding effective methods to allocate economic resources in order to effectively improve the quality of education is of interest for governments. Consequently, understanding the relationship between performance pay in teachers' salary and student outcomes is relevant for policy makers.

2 Context and Policies

2.1 The Portuguese Education System

The Portuguese education system offers free and compulsory education to its students, starting at 6 and ending at 18 years of age or with the conclusion of upper secondary education (European Commission, 2020). The system is divided into three stages, pre-school education (which is not mandatory), basic education, from 6 to 15 years old, and upper secondary education, which covers the last three years of the compulsory education. The basic education stage is further divided into three cycles, which aim to build up on each other. The first one covers grades 1 to 4, the second cycle corresponds to grade 5 and 6 and together they make up primary education. The third and last cycle covers grades 7 to 9 and forms the lower secondary education. At the end of the third cycle, 9th grade students take written and oral examinations to test their knowledge on the subjects they have been cursing. Furthermore, students participate in the national examinations, where pupils from all Portuguese regions (with the exemption of the Azores islands) take the same exams to measure their abilities and knowledge on Portuguese language and mathematics

Once students successfully make it through basic education, they move on to the upper secondary education stage, which corresponds to grades 10 to 12. The organization of this stage varies depending on students' aims, either focusing on access to further studies or preparation for working life (European Commission, 2020). At the end of this cycle students are evaluated on their knowledge across all subjects, the acquired results are used to graduate secondary school and to access universities. Once again, students take part in national exams, which evaluate knowledge across a wide range of secondary education subjects instead of just Portuguese language and mathematics.

The Portuguese education system is very centralized in terms of organization and funding (European Commission, 2020). However, the Autonomous regions of Azores and Madeira have regional governments with legislative powers which include education. This difference will be exploited in the research method used in later sections in this paper.

2.2 The Portuguese Policy

In 2005 the recently elected Portuguese Socialist party had a clear goal as far as education is concerned: to raise the quality of the Portuguese education system to European standards. To understand this objective it is important to consider Portugal's participation in the 2001 PISA academic evaluation. Portuguese results were disappointing: performance was below the OECD

average in all three areas (reading, mathematics, and science). In reading and mathematics only four countries out of the 28 participants had a lower average. In the case of science, only three participating countries had a lower average (Crato, 2020).

As a result, small changes were made to the education system, however, it was not until the new party came into office in 2005 that significant action was taken and a reform in the education system was introduced. One of the vital points of this reform was breaking the, until then, single pay scale for teachers into two separate scales, which came into effect in January of 2007 (Martins, 2010). This policy included the introduction of performance-pay in teachers' salaries, given that in order to move up the pay scale teachers were measured on different variables such as the academic performance of the students taught by that teacher. Other less significant criteria considered included teacher's attendance, level of involvement in research projects and feedback from students' parents. This policy was aimed for teachers of both primary and secondary education and the decision of whether a teacher met the criteria to move up the pay scale was determined at a school level, following assessments provided by the national ministry of education. The economic benefits associated with this policy were generous, given that the gap between the last point in the lower scale and the first point of the higher scale was particularly large, at around 25%, from about 2000€ to 2500€ per month (Martins 2010). This policy entailed a drastic change in the wage scheme of teachers since previously it was only possible for teachers to increase their wage through tenure. Overall, this policy offers great possibilities for research since it introduced a large element of performance-pay in a system where previously performance was not rewarded.

2.3 The English Education System

The UK Government's Department for Education is responsible for the English education system, which is geographically divided into Local Education Authorities (from now onwards defined as LEAs). The English law states that parents are responsible for ensuring that children aged 5 to 16 receive efficient full-time education suitable for their age, ability, aptitude and to any special education needs they may have (European Commission, 2020). Even though education is mandatory, attendance to regular schools is not, which provides parents with the opportunity to educate their children at home. For young people aged 16 to 18 it is required that they either follow a full-time education, participate in an apprenticeship, or devote a minimum of 20 hours per week to work or volunteering while enrolled in part-time education.

The English education system is divided into four different stages, better known as key stages. Primary education consists of Key stage 1 and 2, for children aged 5 to 7 and 7 to 11 respectively.

Secondary education consists of key stage 3 and 4, for pupils aged 11 to 14 and 14 to 16 respectively. At the end of the fourth key stage, students take part in the General Certificate of Secondary Education (GCSE) evaluations. Those exams are pivotal for students given that it measures students' knowledge and skills in the subjects taken throughout the secondary education and play a key role in determining the pupils' progression into education, training or the labour market. The exams are taken at a national level and consequently provide a great opportunity to analyze performance differences across regions and LEAs.

2.4 The English Policy

In December of 1998 the secretary of state for education presented to the Parliament of the United Kingdom a project which aimed to adapt the education system in order to meet the challenge of change. Contrary to what one could imagine, it was a rather superficial proposal, with only a few actionable proposals. Out of those, the most significant one involved motivating teachers and the proposed way to do so was offering incentives for high-performance. As written in the policy proposal, the key features of the new system included: thorough annual assessment of performance against agreed targets, sustained levels of competence, achievement and commitment and higher pay ranges and new professional expectations for teachers (Department for Education and Employment, 1998). The Parliament ruled in favour of this project, which came into effect in September of 2000. This policy involved the introduction of the new upper pay spine for well performing teachers, who were rewarded with a £2000 increase in their yearly salary. Similarly to the Portuguese case, this policy affected both primary and secondary education. To make it through the threshold assessment and achieve the upper pay spine, teachers had to perform well in five different areas: pupil progress, teaching management, knowledge of the taught subject, professional effectiveness (other aspects unrelated to pupil progress) and professional characteristics (Department for Education and Employment, 2000). Prior to this policy the salary perceived by teachers depended on their position in a nine points scale, ranging from an annual salary of £14658 to £23193. An individual's position on this scale was determined by their qualifications and experience, and teachers mostly progressed through the scale on a yearly basis (Atkinson et al., 2004). Once the policy became effective, teachers who were on the ninth point of the scale had the possibility to apply for the upper pay spine. Teachers who had not yet reached the ninth point had to first do so, at a rate of a point increase per worked year, in order to be eligible for the upper pay spine.

Overall, this policy is very similar to the Portuguese one given that they both involve the introduction of performance pay schemes into systems where, prior to those policies, wage was mostly or fully determined by tenure. Even though they are similar policies it is interesting to study both of them given that they are applied in different education systems and the degree of incentives offered is different for each of them.

3 Previous Research

Extensive research has been conducted on the effects of performance pay on productivity. The existing literature focuses on the private sector, where such wage schemes are given more frequently than in the public sector. In the case of Britain, only 7% of the public sector employees are paid for their performance, almost 4 times less than the 27% of their private sector counterparts (Bryson et Al, 2017). As far as the private sector is concerned, there is a consensus on the existing literature that indicates a positive effect of economic incentives on productivity. Naturally, this is supported by a significant number of studies carried out in different settings throughout the last decades. To illustrate this, it is worth mentioning the work of Edward Lazear, commonly referred to as the founder of personnel economics. From 1994 to 1995 Lazear worked closely with a production firm which decided to switch its wage scheme from hourly wages to paying piece rates. The results of this research indicate a large increase on productivity, estimated between 20% and 36% (Lazear, 1996). Similarly, research carried out on a textile factory which used performance-pay to replace hourly wages found an increase in productivity of 29% and 26% for low and high skilled workers respectively (Franceschelli, Galiani, Gulmez, 2010). It is important to note that although those studies focus on the impact of incentive pay on productivity, research also covers other effects it has on the workforce. For example, Heywood and Wei found that performance pay increases job satisfaction given that it enables employees to optimize their productivity and reach higher wages (Heywood, Wei, 2006).

Although performance pay is implemented less frequently in the public sector its effect on performance has also been researched. There is evidence that employees in different areas of the public sector respond positively to incentive pay. However, those responses are in most cases small (Prentice, 2007). It is important to keep in mind that the public sector involves multiple fields which vary largely between each other. Therefore, it is unrealistic to assume that all public sector fields will react similarly to performance pay wage schemes. Hence, the following research analysis will focus on the education sector for relevance reasons.

A common challenge researchers face when studying the effects of teacher incentive schemes on academic outcomes is finding measurable proxies and accessing data for key variables (Atkinson et al., 2009). Nonetheless, reliable studies have been conducted and the existing literature suggests a positive effect of such policies on students' performance. One of the earliest studies available on incentive pay and student outcomes available was conducted by Ladd, who researched an incentive program implemented in Dallas from 1991 to 1995. This policy was characterized by its sophistication measuring student improvement, given that multiple regressions on individual and school level data were ran and compared to the state average in order to determine the magnitude of teachers' contributions. The staff of the awarded schools, which typically represented about 20% of the schools in Dallas, received bonuses of up to $1000, depending on their position. The findings suggest an overall positive and relatively large effect on student outcomes when compared to schools in Texas that did not implement this policy (Ladd, 1997). Figlio and Kenny approach this research question by analyzing school level programs which incorporate merit raises and bonuses as a reward for good performance. Once again, their findings suggest a positive correlation between incentive pay and student outcomes. Moreover, their results indicate that the effect is stronger in schools which offer higher rewards (Figlio, Kenny, 2006) However, their research is marked by a crucial limitation: the analyzed sample is not representative of schools across the US. Consequently, their results lack external validity and cannot be translated to different settings.

Incentive pay has also been used in other education systems worldwide. The Israeli government implemented a policy which rewarded teachers with monetary bonuses for improvements in their students' performance. To measure student performance both the passing rate and the average score of students' high-school matriculation exams were used. Similarly to the program implemented in Dallas, relative performance was used to assign rewards. Therefore, this policy created a tournament among teachers, given that only the teachers ranked highest for each subject were rewarded. Lavy researched this program through two different strategies; a regression discontinuity and propensity score matching. Once again, his results suggest that incentive pay has a significant effect on student outcomes (Lavy, 2004). An important conclusion drawn from this paper is the fact that this program was more cost-effective than school-level bonuses, which is likely to be explained by reduced free-riding. Furthermore, Lavy conducted a survey among teachers that participated in the program to find how they reacted to it. The results suggest that teachers changed their teaching methods, became more responsive to student needs and overall exerted more effort.

The policies that will be analyzed in this paper have also been researched before. Martins studied the effect of the Portuguese program on student achievement by using a difference-in-difference approach. His results indicate a significant decline in student achievement, which contradicts the previously mentioned literature. Martins argues that this could be caused by the disruption of teacher cooperation created by the competition for promotions and the tournament characteristics of the program (Martins, 2010). It is important to note that Martins uses high-school national exams as a proxy for student outcomes, which are required for graduating high-school and to enter university. Hence, it could be argued that the students participating in those exams are mature and highly motivated in the first place and could therefore be less affected by teachers' effort levels. This paper will instead measure the student outcomes of 15-year-old pupils finishing their lower secondary education, which are arguably less independent and more prone to be affected by student effort. Hence, this research will complement Martin's work and help determine whether the incentives pay program was successful or not.

As for the English policy, researchers worked with schools during its implementation to study its effect on student outcomes. Given that this was a pilot study, researchers were able to analyze data from students taught by teachers eligible for the incentive pay program and use other pupils as their control group. This enabled researchers to use a difference-in-difference regression as their methodology, which included teacher and student individual effects. To measure student outcomes, GCSE exams from 1997 and 1999 are used. The results suggest that teachers eligible for the incentive payment increased student outcomes by approximately half a GCSE grade per pupil more than ineligible teachers (Atkinson et al., 2009). However, the results are limited by the sample size; given the difficulties to store and access such data only 18 schools in England participated in this research. Furthermore, the schools that did participate are more likely to have above average management and IT systems. Hence, the sample cannot be considered as representative of all English secondary schools. This paper will build up on this research in two ways: first, sample representativeness will be improved given that secondary schools from all England regions will be analyzed. Second, this paper will also study the effect the policy had on student outcomes up to 4 years after its implementation, whereas Atkinson et al. (2009) focus on the short-term effect.

4 Data

4.1 Portugal

The data used to measure student outcomes consists of the average grades on the lower secondary education exams of pupils from across all Portuguese regions, with the exemption of the Azores island. This data is released yearly by the Portuguese National Exams Committee (JNE, Jurí Nacional de Exames), a branch of the Ministry of Education, and can be accessed through their website (JNE, 2020). As mentioned in the context section, those exams are taken by students at the end of their lower secondary education, more precisely at the age of 14 or 15 years old. Although students are evaluated in all the subjects they pursue, the national exams only test Portuguese language and mathematics skills. As for the grading, it is done using a five-point grading scale, where 5 is the best possible grade, 4 is considered good, 3 represents average performance, a 2 is considered unsatisfactory and 1 is the lowest grade attainable. As for the coverage of the data, it includes average grades in Portuguese language and mathematics for 286 of the 308 municipalities that form Portugal. Out of the 22 municipalities not included, 19 belong to the Azores region, which does not participate in those exams, and the other 3 where not included in the official reports given their limited population. Given that the Madeira region will be used as the counterfactual, the control group consists of its 10 municipalities, whereas the other 276 municipalities form the control group. The data covers the exams from the school year 2004/2005 until 2007/2008. Even though the policy was implemented in the school year 2006/2007, it is useful to have data points on previous years to study the behaviour of the trends before the policy took place. This is especially relevant when implementing a difference-in-difference regression, given that this allows to test whether the parallel trends assumption holds. This will be discussed in more detail in later sections. Table A1 and A2 in the appendix contain statistical information on the number of participants per year and the overall yearly averages for each subject.

4.2 England

GCSE results will be used in order to measure student performance in the English education system. More specifically, the percentage of students achieving 5 or more grades between A* and C. This data is released yearly by the English Department for Education and Employment through their Statistics of Education reports, (Department for Education, 2020). The data collected covers 133 out of the 152 currently existing LEAs in England. It is important to note that the number of LEAs has not been constant over the years, given that some highly populated ones have split whereas others have joined to facilitate management. Nevertheless, the

examined data covers all LEAs that have existed throughout the researched years, 1997/98 until 2003/04, with the exemption of one region. Namely the isles of Scilly, for which data covering all years is not available given its reduced population.

As mentioned earlier, GCSE exams are taken by 15 and 16 years-old students at the end of their compulsory education. The results from those evaluations play a significant role in students' progression into higher stages of education and are also considered by employers. Hence, they carry great importance in individuals' professional careers. Contrary to their Portuguese counterparts, English students have a lot of freedom to decide how many and which subjects they take exams on. Nevertheless, there are some criteria students must meet in order to acquire their secondary school diploma. Pupils are required to take exams on a minimum of 5 subjects, however, most schools advice a minimum of 8. This way students avoid the existing risk of failing and not making it through to the next cycle and acquire knowledge necessary in higher education. Furthermore, although students are free to choose among subjects, there are four mandatory ones. Namely, mathematics, English literature, English language and science.

The assessment has changed over the last decade, but during the researched years each subject was graded from A* to G, being A* the maximum grade attainable and E the lowest passing grade. A very common threshold students, employers, and the education system itself use to define a "good pass" is achieving 5 or more grades between A* and C. Therefore, school performance is oftentimes studied by measuring the percentage of students achieving 5 or more grades from A* to C. This paper will use this variable as a proxy to measure student outcomes, given that it provides information on the fraction of students doing well in their examinations by national standards and is resistant to outliers. Furthermore, it is more constant and reliable than other variables such as GCSE point grade, which is largely determined by the number of average subjects taken by students, which significantly rose from 1997 to 2003. Hence, it is not a reliable proxy of changes in student outcomes. Tables A3 to 11 in the appendix contain descriptive statistics on the studied variable across the 8 regions of England from 1997 to 2003. As it can be observed, the percentage of students reaching the threshold is increasing over time. It is important to consider this, given that it could pose a limitation and affect the interpretation of the results.

5. Theoretical Framework

The fundamental idea behind incentives is as simple as rewarding someone for behaving and acting following your interests. McMillan (1992) defines incentives as "the uncontroversial, even bland proposition that people must be rewarded if they are to be induced to do something they would prefer not to do". The role incentive pay has on effort levels is covered by principal-agent theory. However, before going into more detail on principal-agent models, it is necessary to first understand how the effort exerted by a teacher can impact student outcomes. On top of the existing literature supporting this idea, a theoretical approach can be taken. To do so, we can use the multi-tasking model of Holmstrom and Milgrom (1991) adapted to value added to student outcomes from teacher's effort from Neal (2011):

Consider a scenario where a teacher is hired to teach a student. There are two tasks amongst which the teacher allocates her effort, f_1 and f_2 . Let t_1 and t_2 denote the time the teacher allocates to f_1 and $f_2$respectively. The human capital production model is as follows:

$$(1)\quad h = f_1 t_1 + f_2 t_2 + g$$

Where h denotes the human capital the student acquires from the teacher efforts and g represents a random shock that affects the learning process. This random shock captures factors that affect the student's learning that are unrelated to effort exerted by the teacher. A way to understand f_1 and f_2 coefficients is as the ability level of a teacher. Some individuals are more skillful than others when it comes to teaching. For those, less effort is required to improve student outcomes when compared to individuals with low ability. Therefore, ability A takes the form of a function of f_1 and f_2 (2). Effort exerted by a teacher could involve a variety of actions such as adapting the teaching method to students learning process amongst others. In this case, in order to simplify the model, effort will be understood as the time spent on each task. Hence, effort e can be seen as a function of t_1 and t_2 (3).

$$(2)\quad A = f(f_1, f_2)$$
$$(3)\quad e = f(t_1, t_2)$$

As depicted in (1), given a fixed level of ability A for an individual, the more effort e a teacher exerts, the higher the human capital acquisition from the student. Hence, it is beneficial to find ways to increase teachers' levels of effort. Principal-agent theory explains this through individual utility maximizing models such as the following:

$$(4)\quad Utility = U(Y, e)$$

Where Y denotes workers income and e the effort exerted by said worker. Naturally, exerting effort is associated with a cost, and consequently has a negative effect on teachers' utility. However, this is compensated by an increase in income, as shown in the following function:

$$(5)\quad Y = a + B(e, A, g)$$

Where income Y is formed by a fixed part a and a bonus or incentive pay B. The amount of income a teacher receives from the incentive part of the model depends on the coefficient of B (so the strength of the incentives) and her performance, which depends on effort exerted, ability level and the external shock g. It is also necessary to consider the cost of effort, which represents the decrease in utility an individual experiences as a result of exerting effort:

$$(6)\quad C(e) = f(e)$$

Given the income and the cost of effort equations, it is now possible to show the utility maximizing equation of a teacher:

$$(7)\quad Max\ U(a + B(e, A, g) - C(e))$$

Optimal levels of effort differ amongst people, given that each individual maximizes their own utility and cost of effort functions. Nonetheless, from this equation we can deduce that an individual with a relatively low cost of effort – ceteris paribus – will exert more effort than someone with a high cost of effort. The utility function also plays a key role, given that it determines the valuation individuals have for income. This equation indicates that in the absence of incentives individuals exert no effort. Hence, theory predicts that if only a fixed wage is offered, individuals maximize their utility by doing the bare minimum to keep their job. On the other hand, if income is fully depended on incentives pay individuals would exert effort until the marginal benefit from doing so was equal to the marginal cost of exerting it. In the cases analyzed in this paper B is positive and larger than 0. Therefore, principal-agent theory predicts that teachers would on average exert more effort than before the introduction of the policy. As argued in (1), effort exerted by teachers positively affects student outcomes. Hence, linking the research question with the presented theory, the first hypothesis proposed is:

H_1: The implementation of the policies had a positive effect on student outcomes.

From (7) it is also possible to deduct that the larger the value of B – ceteris paribus – the more effort will be exerted by individuals. Given that the Portuguese case involved a larger incentive pay than the English case, the second hypothesis proposed is:

H_2: The effect of the policy implemented in Portugal is larger than that of the English policy.

6. Methodology

To approach this research question two different models will be used. Naturally, each model will focus on one of the studied policies and adapt to the available data. Namely, the first model will focus on the policy implemented in the Portuguese education system, whereas the second model will analyze the impact of the English policy.

6.1 First model

The first model uses a difference-in-difference with regional fixed effects approach. For this method, the control group consists of the student outcomes in the different regions that form the Madeira islands. The reasoning behind this is that students across the country participate in the same examinations however teachers in the Madeira islands are not affected by the policy. Therefore, as long as the parallel trends assumption holds, it is fair to assume that in the absence of treatment the change in student outcomes in the mainland regions would have followed a similar trend to that in the islands. On top of this, allowing the constant to differ between regions is a realistic approach given that regions within a country differ in multiple unobservable characteristics. Hence, adding fixed effects to this model offers an advantage over a classical difference-in-difference approach given that there is no need to control for time invariant variables in our model. The estimated model is as follows:

$$Y_{it} = \alpha_i + \rho D_g + y_t + \delta T_{it} + \varepsilon_{gt} \qquad , t = 0 \; or \; t = 1$$

Where Y_{it} represents the average student results in region i and period t for the chosen academic subject, α_i accounts for region fixed effects, D_g is the treatment group dummy, which takes the value 1 for mainland Portugal and 0 for the Madeira islands. y_t is a dummy variable indicating whether the policy has been implemented, t=1 refers to the treatment period and t=0 to all pretreatment periods. The effect of the policy is captured by the coefficient δ, which multiplies T_{it} , the interaction term between the treatment group and the treatment period.

An important concern when performing a difference-in-difference regression is the parallel trends assumption, which assumes that the outcomes of treated and non-treated would have changed in the same way in the absence of treatment. Given that it is a fundamental assumption of the model it is necessary to assess whether it holds, otherwise the results could not be interpreted and a different approach would be required. This will be tested in the results section.

6.2 Second model

Contrary to the Portuguese case, there is not a region where GCSE exams are taken that did not incorporate performance that can act as the control group. Therefore, a difference-in-difference approach is not feasible. Nevertheless, given the variation in treatment over the researched years, it is possible to use each LEA before the implementation of the policy as its own counterfactual. Hence, an individual fixed effects approach will be used. Just as in the first model, this model captures all time-invariant characteristics individual to each LEA. Once again, this is a realistic approach given the innate differences between regions depicted in tables 1-7 in the appendix. The estimated model is the following:

$$Y_{it} = \alpha_i + \delta T_{it} + \varepsilon_{gt} \qquad , t = 0 \; or \; t = 1$$

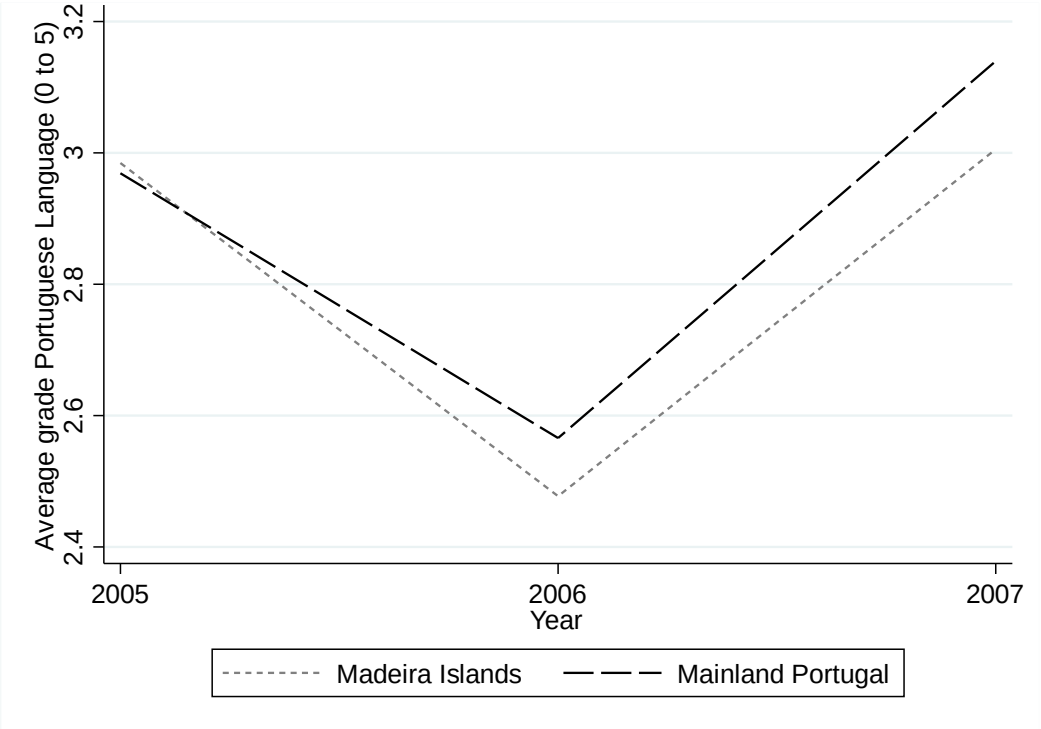
Where Y_{it} refers to the percentage of students in LEA i at time period t achieving 5 or more grades between A* and C. The α_i controls for LEA fixed effects, δ refers to the treatment effect and T_{it} to the treatment dummy. Just as in the first model, t=1 refers to post-treatment periods whereas t=0 refers to all the pretreatment periods.

7. Results

7.1 First Model

Before running the regression, it is necessary to check whether the fundamental parallel trends assumption holds. This implies determining whether the student outcomes in the treatment and control group followed a similar trend in the periods prior to the implementation of the performance pau policy. This can be done through a graphic approach.

Figure 1: Parallel trends analysis for average grade in Portuguese language. Covering pre-treatment school years 2004/2005 to 2006/2007



As shown in Figure 1, the assumption holds for the Portuguese Language National exams, given that from the 2004/05 to 2005/6 the trends are almost identical. Parallel trends assumption also holds for the mathematics exams. Taking a look at figure A1 in the appendix it can be seen how in the years prior to the policy implementation both groups follow very similar trends.

Having tested the parallel trends assumption, it is now possible to interpret the results from the difference-in-difference with region fixed effects regression shown in the first model. The results for Portuguese language and Mathematics, shown in tables 1 and 2 below, suggest a positive and significant effect of the performance pay policy on student outcomes. However, the effect is relatively small. In the case of the Portuguese language subject, the results indicate an increase of 0.0978 on the average grade obtained by students. Similarly, the results indicate an increase in mathematics grades of 0.0980. Considering assessment is done in a scale of 1 to 5, this implies an increase in average national grades for both subjects of approximately 2,45%.

Table 1: Portuguese Language results using a Difference-in-Difference with region fixed

Variables	Coefficient	Robust Std. Error	P-value	95% Confidence Interval	
Treatment	0.0978	0.0372	0.008***	0.0249	0.1706
Treatment Group	0.0365	0.0510	0.474	-0.0635	0.1365
Treatment Period	0.2745	0.0357	0.000***	0.2046	0.3444
Constant	2.7309	0.0500	0.000***	2.6328	2.8290

*Significant at 10%. ** Significant at 5%. *** Significant at 1%

Table 2: Mathematics results using a Difference-in-Difference with region fixed effects regression

Variables	Coefficient	Robust Std. Error	P-value	95% Confidence Interval	
Treatment	0.0980	0.0400	0.014**	0.0195	0.1765
Treatment Group	0.1512	0.0582	0.009***	0.0370	0.2653
Treatment Period	-0.1705	0.0377	0.000***	-0.2443	-0.0967
Constant	2.0105	0.0569	0.000***	1.8988	2.1222

*Significant at 10%. ** Significant at 5%. *** Significant at 1%

7.2 Second model

Similarly, the second model also suggest a positive and significant effect of performance-pay on student outcomes. As shown in table 3 below, the results indicate an increase of 5,02% in the fraction of students achieving the threshold of 5 or more subjects graded between A* and C.

Table 3: individual fixed effects: % of students achieving 5+ A*-C GCSE. 0-1 scale (1=100%, 0=0%)

Variables	Coefficient	Robust Std. Error	P-value	95% Confidence Interval	
Treatment	0.0502	0.0021	0.000***	0.0461	0.0544
Constant	0.4376	0.0012	0.000***	0.43528	0.4402

*Significant at 10%. ** Significant at 5%. *** Significant at 1%

8. Discussion

8.1 Limitations

Some limitations should be considered with regards to the interpretation of the presented results. Regarding data, the effect of the Portuguese policy is only researched for two subjects, given that national exams are only taken for Portuguese language and mathematics. Ideally, this research would study the impact of performance pay on student outcomes across all or most subjects. Furthermore, it is necessary to consider the difference in sizes between the control and the treatment group, given the first one is formed by 276 regions whereas the later only consists of 10. This is far from optimal and poses a limitation to the interpretation of the results. As for the English data, it is important to consider that the studied measure of performance was increasing in the years previous to the policy. Hence, it is difficult to determine to what extent the increase in student outcomes is a consequence of the policy, limiting the causal interpretation of the policy. Furthermore, the data analyzed only covers the percentage of students achieving five or more grades between A* and C. This percentage does not consider the effect the policy had on pupils who would have achieved this threshold regardless of the policy, nor on those who did not reach it but may have benefited from the policy. To illustrate this; if a student would have only achieved one grade between A* and C in a scenario without the policy and four with it, this student's performance would have been drastically improved. However, this improvement would not be captured by the data given that the threshold was not met. Nevertheless, this is still the best proxy for student outcomes available for two reasons. Firstly, it is the only variable that has been disclosed yearly in official reports. Secondly, it is resistant to outliers, contrary to other variables such as point score GCSE results, which is affected by students taking a large number of subjects.

Regarding the methodology, it is important to mention some limitations affecting the first model. First and foremost, it is necessary to consider omitted variable bias, given that throughout the analyzed time period there could have been changes that have not been considered. This is especially important considering the observations are at a yearly level and hence multiple events could have taken place during this period. Nevertheless, no major policies affecting the education system took place during this time period. Therefore, it seems reasonable to consider that the potential effect of any omitted variable bias would not largely affect the causal interpretation of the results. Secondly, although the parallel trends assumption seems to strongly hold, it would be beneficial for its assessment to test whether this is the case for a larger number of previous periods. As for the second model, it would benefit from having

a control group formed by a region not affected by the policy, similarly to Portugal. Nevertheless, given that no major education reforms other than the studied one took place in the researched years and the results obtained are significant, it seems reasonable to consider the effect as causal. Furthermore, the model is not affected by the common concerns associated with individual fixed effects. Namely not being able to estimate the effect of time-invariant variables, issues with measurement errors and not knowing the cause behind variation in treatment.

An important limitation of this research is assuming that increased effort exerted by teachers leads to an increase in student outcomes. Although improving students' knowledge is a crucial task for any teacher, the job involves much more. Hence, it is possible that teachers' increased effort could be directed to unobservable variables such as controlling for the atmosphere in the class or taking a more significant role as a mentor. Nonetheless, it could be argued that those unobservable variables are mechanisms through which teacher effort affects student outcomes.

8.2 Conclusion

Evaluating the results, it is possible to approach the research question. The results indicate a positive and significant effect of performance-pay in teachers' wage scheme on student outcomes. In the case of the policy implemented in Portugal the results suggest an increase in student outcomes of 2,45% for both Portuguese language and mathematics. As for England, the research indicates a large increase of 5,02% in the proportion of students achieving the well-known threshold of 5 or more subjects graded from A* to C. Consequently, the first hypothesis presented holds and cannot be rejected. It is important to note that the variables used to measure student outcome in each education system are different and hence cannot be compared. In the case of Portugal, average grades are measured, whereas for England a threshold of "good" performance is assessed. Nevertheless, it seems sensible to interpret the effect of the English policy as larger than that of the Portuguese one. This would imply that, contrary to what the theoretical framework predicted, the policy which incorporated a smaller element of performance pay achieved better results. Hence, the second hypothesis does not hold and can be rejected.

The results for the English case are larger than those found by Atkinson et al. (2009), which, as mentioned in the previous research section, found an increase of half a GCSE grade per pupil. Once again, this research offers a larger and more representative sample and analyzes the effect of the policy over a longer time frame. As for the Portuguese policy, the results differ from those of Martins (2010), who found a small (virtually zero) significant decline in student performance.

However, this research studies the effect on pupils from a different age group which seems to be more reactive to changes in teacher effort levels.

In order to assess the internal and external validity of this research it is necessary to analyze the policies separately. Although the Portuguese policy faces some limitations which could pose a threat to internal validity, they do not significantly affect the causal interpretation of the policy. As previously mentioned, omitted variable bias is unlikely, and even though the parallel trends assumption could be assessed in more detail and the control group could be larger it is still possible to consider the assumption to hold and the sample size as representative of the Portuguese population. Hence, taking this into account and given that the results are statistically significant, the first method and its results have a strong internal validity. As for the external validity, considering the strong internal validity and since sample sizes are representative of the Portuguese population, it is expected to obtain similar results in countries with comparable education systems. However, those results can only be translated to very similar settings, it could be the case that students of different ages react differently to those policies, which could explain the findings of Martins (2010). Furthermore, this research does not explore in detail the differences in results between all subjects taken by students, which could vary.

Similarly, in the case of the English policy the necessary assumptions hold and the sample is representative of the studied population. However, the constant increase in student performance which took place before the implementation of the policy poses too strong of a limitation for the causal interpretation of the results. Therefore, in this case, the internal validity is weak, and consequently so is the external validity.

Overall, this research contributes to the literature on personnel economics in the public sector and on the field of education economics by studying the effect of performance pay on productivity in the education system. The results indicate a significant positive effect of policies which implemented incentives into teachers' wage scheme on student outcomes. Considering those results and the existing literature, policy makers and governments could take into consideration the implementation of similar programs in order to increase the student outcomes of young students. However, studying the full effect of policies implemented in education systems is complex given the presence of important unobservable characteristics associated with teaching. It remains unclear whether performance pay would affect other aspects of schooling, such as the mentoring role of teachers or the classroom ambiance. This should therefore be further researched to the extent that it is possible to determine whether this is a potential benefit or a potential drawback of performance pay in the education system.

In one of the reviewed papers, Martins (2010) argues that the Portuguese policy increased competition between teachers at a school level, hence reducing their job satisfaction although their wages were increased. This is a potential drawback of such policies, mostly determined by the structure in which they are implemented. Taking into account those factors, further research is needed and encouraged. Some ideas include studying how students from different age groups react to such policies and whether the effects of performance pay affects differ between subjects. However, the study of such policies faces multiple challenges, such as the different assessment methods across education systems. A solution to this issue could be using a measure of performance used in different countries, such as the PISA exams carried out by the OECD members. Another aspect on which future research should focus is long term effects. Once again, assessing the long-term effects of policies in the education system is complex given the constant changes and the implementation of new policies, which pose a threat to the causal interpretation of the long term effect of policies.

9. Bibliography

Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H., & Wilson, D. (2009). Evaluating the impact of performance-related pay for teachers in England. *Labour Economics, 16*(3), 251-261.

Bryson, A., Forth, J., & Stokes, L. (2017). How much performance pay is there in the public sector and what are its effects? *Human Resource Management Journal, 27*(4), 581-597.

Card, D. (1999). The Causal Effect of Education on Earnings. *Handbook of Labor Economics,* 1801-1863.

Crato, N. (2020). Curriculum and Educational Reforms in Portugal: An analysis on Why and How Students' Knowledge and Skills Improved. In F. Reimers (Author), *Audacious Education Purposes* (pp. 209-233). Cambridge, Massachusets: Springer International Publishing.

Department for Education: Statistics of Education reports. (n.d.). Retrieved June 05, 2020, from https://webarchive.nationalarchives.gov.uk/20130320104839/https://www.education.gov.uk/researchandstatistics/statistics/allstatistics

European Commission. (n.d.). The education system in Portugal. Retrieved from https://eacea.ec.europa.eu/national-policies/eurydice/content/portugal_en

European Commission (n.d.). The education system in the United Kingdom - England. Retrieved from https://eacea.ec.europa.eu/national-policies/eurydice/content/united-kingdom-england_en

Figlio, D., & Kenny, L. (2006). Individual Teacher Incentives And Student Performance. *National Bureau of Economic Research*.

Franceschelli, I., Galiani, S., & Gulmez, E. (2010). Performance Pay and Productivity of Low- and High-Ability Workers. *Labour Economics, 17*(2), 317-322.

GCSE/GNVQ and GCE A/AS Level and Advanced GNVQ Examination Results 1997/98 - England (Rep.). (1999). Statistical Bulletin.

Hairston-Ridgley, H., Jr. (2011). A message from President Obama [Preface]. In *You can't drop out of highschool and drop into a job: Avoiding the ocean of economic and social instability*. Bloomington, Indiana: AuthorHouse.

Hasnain, Z., Manning, N., & Pierskalla, J. H. (2012). *Performance-related pay in the public sector: A review of theory and evidence*. Washington, D.C.: World Bank.

Heywood, J. S., & Wei, X. (2006). Performance Pay and Job Satisfaction. *Journal of Industrial Relations*.

Holmstrom, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization, 7*(Special), 24-52.

Jurí Nacional de Exames: Portuguese National Exams. (n.d.). Retrieved June 05, 2020, from https://www.dge.mec.pt/relatoriosestatisticas-0

Keller, M. M., Neumann, K., & Fischer, H. E. (2016). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching, 54*(5), 586-614.

Ladd, H. F. (1997). The Dallas School Accountability and Incentive Program: An Evaluation of Its Impacts on Student Outcomes. *Economics of Education Review, 18*.

Lavy, V. (2004). Performance Pay and Teachers' Effort, Productivity and Grading Ethics. *National Bureau of Economic Research*.

Lazear, E. P. (1996). Performance Pay and Productivity. *National Bureau of Economic Research*.

Leigh, J. (1998). Education and Development: Measuring the Social Benefits. *Economics of Education Review, 17*(3), 363-368.

Lochner, L. (2011). Non-Production Benefits of Education: Crime, Health, and Good Citizenship. *National Bureau of Economic Research*.

Martins, P. S. (2010). Individual Teacher Incentives, Student Achievement and Grade Inflation. *Centre for the Economics of Education, London School of Economics*.

McMillan, J. (1992). *Games, strategies, and managers: How managers can use game theory to make better business decisions*. New York: Oxford Univ. Press.

Neal, D. (2011). The Design of Performance Pay in Education. *Handbook of The Economics of Education Handbook of the Economics of Education, 4*, 495-550.

Prentice, G. (2007). Performance pay in the public sector: A review of the issues and evidence. *Office of Manpower Economics*.

Teachers: Meeting the challenge of change. (1998). London: Department for Education and Employment.

Vila, L. E. (2000). The Non-monetary Benefits of Education. *European Journal of Education, 35*(1), 21-32.

10. Appendix

10.1 Abbreviations

LEA: Local Education Authority

GCSE: General Certificate of Secondary Education

JNE: Jurí Nacional de Exames

10.2 Tables

Descriptive Statistics on Portuguese National Exams per subject

Table A1: Portuguese Language Results 2004/05 to 2007/08. Summary Statistics

Year	Participants	Mean	Min	Max
2004/05	85 100	2,97	2,42	3,69
2005/06	93 085	2,56	2	3,25
2006/07	96 941	3,13	2,52	3,68
2007/08	94 397	3,19	2,5	3,82

Table A2: Mathematics Results 2004/05 to 2007/08. Summary Statistics

Year	Participants	Mean	Min	Max
2004/05	84 987	2,07	1,49	2,66
2005/06	92 896	2,23	1,56	3,05
2006/07	96 829	2,08	1,38	2,88
2007/08	94 838	2,83	2,08	3,61

Descriptive statistics on GCSE results in England per region: A3-A11

Table A3: North East Region: % of students achieving 5+A*-C grades. Summary Statistics

Year	Mean	Min	Max
1997/98	37.88%	27.52%	49.47%
1998/99	40.76%	30.97%	54.54%
1999/00	42.66%	34.61%	56.96%
2000/01	43.65%	35.00%	55.80%
2001/02	45.66%	35.80%	58.20%
2002/03	47.92%	38.80%	59.80%
2003/04	49.64%	40.70%	62.20%

Table A4: North West Region: % of students achieving 5+A*-C grades. Summary Statistics

Year	Mean	Min	Max
1997/98	40.56%	22.92%	54.26%
1998/99	43.92%	23.62%	54.54%
1999/00	44.39%	24.99%	56.96%
2000/01	45.18%	27.10%	57.80%
2001/02	46.80 %	30.10%	60.10%
2002/03	48.16%	33.50%	62.30%
2003/04	49.52%	38.10%	64.00%

Table A5: Yorkshire & Humber Region: % of students achieving 5+A*-C grades. Summary Statistics

Year	Mean	Min	Max
1997/98	38.28%	22.84%	56.35%
1998/99	40.03%	23.35%	56.77%
1999/00	41.61%	24.42%	59.37%
2000/01	42.71%	27.50%	58.30%
2001/02	43.79%	28.90%	60.20%
2002/03	45.28%	32.10%	58.90%
2003/04	46.87%	34.70%	60.70%

Table A6: East Midlands Region: % of students achieving 5+A*-C grades. Summary Statistics

Year	Mean	Min	Max
1997/98	42.53%	34.76%	48.42%
1998/99	46.82%	37.10%	58.85%
1999/00	47.48%	33.76%	61.60%
2000/01	48.35%	37.00%	57.20%
2001/02	50.53%	40.50%	61.10%
2002/03	52.15%	42.80%	62.40%
2003/04	51.68%	42.90%	58.20%

Table A7: West Midlands: % of students achieving 5+A*-C grades. Summary Statistics

Year	Mean	Min	Max
1997/98	40.44%	28.73%	52.22%
1998/99	43.03%	29.69%	55.37%
1999/00	44.06%	31.74%	56.35%
2000/01	45.37%	34.30%	55.90%
2001/02	47.35%	34.20%	58.40%
2002/03	49.63%	35.90%	59.90%
2003/04	50.12 %	38.00%	60.20%

Table A8: East of England: % of students achieving 5+A*-C grades. Summary Statistics

Year	Mean	Min	Max
1997/98	46.73%	35.05%	54.01%
1998/99	48.77%	35.79%	54.90%
1999/00	49.75%	39.46%	55.36%
2000/01	50.47%	38.20%	56.30%
2001/02	51.71%	39.50%	56.60%
2002/03	52.49%	41.50%	58.10%
2003/04	53.26%	43.80%	58.50%

Table A9: London region: % of students achieving 5+A*-C grades. Summary Statistics

Year	Mean	Min	Max
1997/98	41.13%	23.26%	58.55%
1998/99	42.62%	27.40%	60.14%
1999/00	44.09%	25.66%	63.32%
2000/01	45.48%	28.70%	63.00%
2001/02	47.60%	31.10%	64.80%
2002/03	50.04%	35.70%	67.00%
2003/04	52.08%	40.20%	68.10%

Table A10: South East region: % of students achieving 5+A*-C grades. Summary Statistics

Year	Mean	Min	Max
1997/98	45.49%	30.66%	59.48%
1998/99	47.64%	32.15%	60.97%
1999/00	48.68%	33.96%	61.56%
2000/01	49.29%	34.30%	63.40%
2001/02	50.39%	36.50%	64.40%
2002/03	52.24%	44.30%	65.50%
2003/04	52.11%	41.80%	65.90%

Table A11: South West: % of students achieving 5+A*-C grades. Summary Statistics

Year	Mean	Min	Max
1997/98	48.18%	28.96%	54.66%
1998/99	49.65%	31.01%	56.12%
1999/00	50.85%	31.13%	59.54%
2000/01	51.59%	31.80%	58.50%
2001/02	53.21%	31.00%	60.90%
2002/03	54.15%	35.30%	61.20%
2003/04	53.96%	35.10%	62.00%

10.3 Figures

Figure A1: Parallel trends analysis for average grade in Mathematics.