Article

Panel Data Models for School Evaluation: The Case of High Schools' Results in University Entrance Examinations

Manuel Salas-Velasco

Department of Applied Economics, Faculty of Economics and Business, Campus Cartuja, University of Granada, 18071 Granada, Spain; msalas@ugr.es

Abstract: To what extent do high school students' course grades align with their scores on standardized college admission tests? People sometimes make the argument that grades are "inflated", but many school districts only use outcome-based descriptive methods for school evaluation. In order to answer that question, this paper proposes econometric models for panel data, which are less well-known in educational evaluation. In particular, fixed-effects and random-effects models are proposed for assessing student performance in university entrance examinations. School-level panel data analysis allows one knowing if results in college admission tests vary more between high schools than within a high school in different academic years. Another advantage of using panel data includes the ability to control for school-specific unobserved heterogeneity. For empirical implementation, official transcript data and university entrance test scores of Spanish secondary schools are used.

1. Introduction

This article introduces statistical tools for evaluating educational outcomes over time. Its main focus is on the evaluation of university admission test results. University or college admission is the process through which students enter higher education institutions (HEIs) at universities and colleges. Although systems vary widely from country to country, and sometimes from institution to institution, most countries or HEIs establish filters—such as standardized tests that students may need to take for admissions—to allocate students to limited university places. In Spain, upon completion of compulsory secondary education at the age of 16, students continue their education to upper secondary education. The upper secondary curriculum is structured in two itineraries in Spain: academic and vocational. Roughly 65 percent of the age cohort enrolls in the academic pathway called *Bachillerato*, which is required to attend university. The other 35 percent enter vocational education. The *Bachillerato* is usually studied between the ages of 16 and 18, and it is equivalent to 11th grade (junior year) and 12th grade (senior year) in the U.S. system. After finishing the *Bachillerato*, a high school graduate obtains a diploma and his/her high school grade point average (GPA) for the two years (i.e., an average of *Bachillerato* course grades). In the Spanish education system, the grading system is only decimal, with grades (marks) ranging from 0 to 10, passing with 5 points. However, a student can access the Spanish university system to study for an undergraduate degree as long as s/he has passed the university entrance examinations, called *Selectividad*. They are external (public) standardized admission tests set by each Spanish region (*comunidad autónoma*) taken every year in June by senior students, just after finishing high school. These tests assess the knowledge acquired by students in their *Bachillerato* courses (subjects). Each student obtains an average score in these standardized tests, which is averaged with his or her high school GPA to obtain their final university entry marks. In the data used here, the

Selectividad exams scored from 0 to 10 points. A student's final mark to enter university was a weighted mean between *Bachillerato* grades (60%) and *Selectividad* scores (40%). This final mark is essential for access to some university degrees where the number of student applications (demand) exceeds the number of available places (supply) (e.g., Medicine and Nursing studies). In Spain, a degree in Medicine is part of the bachelor's programs.

Therefore, due to the numerus clausus system to allocate places at Spanish universities, close attention should be paid by policymakers to the university entry mark formation to guarantee academic equity in access to higher education and make sure that high schools do not alter ("inflate") students' *Bachillerato* course grades. Indeed, the debate in Spain is whether or not private high schools really favor students by inflating their course grades in the *Bachillerato* because they represent an important percentage of the final mark of access to the university. However, in order for research results to adequately inform education policy and citizens, the statistical methods of analysis must be rigorous. Currently, many school districts only use outcome-based indicators for school evaluation, with econometric models for panel data being less well-known in educational evaluation. However, learning outcomes can also be empirically studied with the help of panel data. Approaches to the analysis of panel data such as change score models and fixed/random effect models constitute a powerful analytical tool for exploiting the time dimension of educational data. Panel data are obtained by observing the outcomes of the same school over several time periods. The *Selectividad* examination results (test scores) are also standardized to be comparable across the years. We can examine whether or not average test scores of high schools in *Selectividad* examinations can be explained by (are related to) their average course grades in the *Bachillerato* using evidence from panel data. Furthermore, panel data econometrics allow us to take into account omitted exogenous factors that are specific to each school and change over time, called random error (also noise) such as students' health status, even luck, when taking standardized exams. Additionally, school-level panel data analysis has the ability to control for school-specific unobserved heterogeneity, which refers to those school-specific time-invariant characteristics that are not directly observable to the econometrician but influence student learning outcomes (e.g., the school's intrinsic motivation, even prestige). A panel data approach to assessing student performance also allows us to determine whether educational outcomes in *Selectividad* tests differ more between high schools than within a high school over time.

This paper, thus, aims to introduce the panel data methodology in the context of educational evaluation. By "panel data", we mean data that contain repeated measures of the same variable, taken from the same set of units over time. In our applications, the units are high schools. This paper does not provide details of specific software packages, although the statistical and econometric analyses were run using Stata®17 [1]. By using exemplars, we provide a guide for social scientists new to the area of panel data analysis. Due to data availability, the analysis of student performance in university admission tests was restricted to Andalusian high schools. The nine public universities in Andalusia (southern Spain) are part of the *Distrito Único Andaluz*, a regional admissions system. To apply for bachelor's degrees at these universities, students submit an ordered list of their preferred undergraduate programs to the centralized system, which fills the available places using the highest university access marks.

The remainder of the paper is organized as follows. Section 2 provides some background information. Section 3 gives an overview of panel data methods for assessing educational outcomes. In Section 4, in the evaluation of Andalusian high schools' results on the *Selectividad* exams, we present descriptive statistics along with panel data model estimations. The main conclusions are drawn in Section 5.

2. Background

Many school districts use performance indicators for school evaluation, often using longitudinal data to describe progress and highlight the needs for change [2]. One strategy for evaluating schools "typically aims to compare schools on standardised measures to

allow the benchmarking of their performance in relation to other schools, particular districts or regions, or national averages" [3]. In this regard, students' national examination results at the end of their upper secondary education have received much public attention in recent decades internationally, partly justified by the importance they have among the criteria for admission to higher education [4]. College admission tests are standardized exams that students may need to take to apply to four-year colleges and universities. Higher education institutions in the United States have used standardized test scores, particularly the ACT and SAT exams, along with high school GPA as predictors of academic performance and persistence for decades [5]. If college access were based solely on high school grades, high schools would be tempted to inflate their students' course grades. "Grade inflation refers to an increase in grade point average without a concomitant increase in achievement" [6]. Grade inflation can exacerbate socioeconomic inequities in educational outcomes when it varies systematically by the student or school background [7]. If more affluent high schools inflate their top students' grades more than less-affluent ones over time, grade inflation could also exacerbate socioeconomic stratification across universities. Thus, standardized tests are designed to be statistically fair, reliable, and valid assessments of the high school curriculum. "A standardized test is one where the method of administering the test, including the test conditions and system of scoring, is regulated and controlled so that it is consistently applied across multiple groups" [8]. It may be thought that standardized college admission tests promote equal access to higher education, but this is not always true. African American and Latino students in the United States have historically performed worse on standardized tests, such as the SATs, than their White and Asian American peers. Lower test scores remain a persistent barrier to pursuing postsecondary education, as well as a lower likelihood of admission to elite institutions, for African American and Latino students, particularly those from low-income urban areas [9]. However, the question that remains unanswered is the following: *Is there statistical evidence that high school grade inflation exists?* One way to answer the question is to compare high school course grades to an objective measure of student achievement in standardized tests that is stable over time by applying panel data econometrics. Below, we show the main models that could be used to assess school results.

3. Methodology

In this section, we introduce panel data analysis models that are used most frequently in the literature. For readers who want to gain an introductory level understanding of panel data analysis without using matrix algebra, Wooldridge's textbook is recommended [10]. In panel data applications for continuous dependent variables, linear models are still the most widely used. For ease of exposition, let us consider outcome Equation (1), which provides a linear specification of university entrance examination scores of high schools. The basic model for panel data can be written in matrix notation as follows:

$$y_{it} = x'_{it}\beta + \mu_i + e_{it} \tag{1}$$

$$i = 1, \ldots, N; t = 1, \ldots, T$$

Observations in panel data involve at least two dimensions: a cross-sectional dimension, indicated by subscript i, and a time series dimension, indicated by subscript t, where T is the number of time periods available for each unit. In Equation (1), in the context of school evaluation, the subscripts i and t refer to the i-th high school and the t-th time period, respectively; y_{it} is the dependent or the response variable (i.e., standardized test scores); x_{it} is a K × 1 vector of explanatory variables; and β is the K × 1 vector of coefficients on the set of explanatory variables. However, due to the limitations of the information contained in the administrative database used, our analysis only considers a single explanatory variable (average course grades of the last two years, 11th and 12th grades).

The composite error term in Equation (1), $\mu_i + e_{it}$, is an important feature of panel data models. The stochastic part of the model specifies the effects of all other variables that

affect student learning outcomes but are not explicitly included as explanatory variables. It distinguishes between two components: (i) e_{it}, omitted exogenous variables (including measurement errors) that affect educational outcomes, which vary across schools and change over time (the so-called random error term); these are the time-varying unobserved factors that affect y_{it}, and they are often called "idiosyncratic errors", and (ii) μ_i, school-specific time-invariant characteristics that are not directly observable to the econometrician but influence student learning outcomes (termed "unobserved heterogeneity" in panel data econometrics). Such factors can be regarded as time-invariant, and at the same time, it is extremely hard to measure them. The fact that we have repeated measurements of the same units allows us to control for their unknown characteristics that are constant over time. The failure to account for these unobserved individual differences leads to bias in the resulting estimates [11]. Depending on our assumptions about this latter term, different estimation procedures are available [12]. If the assumption is that the unobserved heterogeneity is uncorrelated with the explanatory variables in the model, random effects (REs) estimation is used to assess the relationship between the explanatory variables and educational outcomes (*mu* can be taken as random). Additionally, the standard assumption is that *e* behaves like a random variable and is uncorrelated with *x*. Otherwise, with correlated heterogeneity, we have to use other techniques that have become known as *fixed effects (FEs) estimation*. This variant is called the fixed-effects model, as early treatments modeled these effects as parameters $\mu_1, \ldots, \mu_N$ to be estimated [13]. FEs remove the effect of those time-invariant characteristics from the predictor variables, so we can assess the predictors' net effect. The key insight is that if the unobserved variable does not change over time, then any changes in the dependent variable must be due to influences other than these fixed characteristics [14]. So, the estimated coefficients of the fixed-effects models cannot be biased because of omitted time-invariant characteristics. Unlike the fixed-effects model, the rationale behind the random-effects model is that the variation across units is assumed to be random and uncorrelated with the predictors or independent variables included in the model. If we believe that differences across entities have some influence on the dependent variable, then we should use random effects. Because the estimate of the slope parameters (β) differs across the different estimation methods, a frequently asked question in empirical research is which model to use: the fixed-effects model or the random-effects model. Although sometimes researchers prefer random-effects models merely because they simply want to obtain the effects of time-invariant variables, this is not a sufficient justification. A formal Hausman test can be used to test whether or not the school-specific heterogeneity is fixed. Whether μ_i is assumed to be fixed or random is crucial to obtaining unbiased parameter estimates. The null hypothesis is that H_0: *random effects (REs) are consistent and efficient*. If the test fails to reject the null hypothesis, as the *p*-value is greater than 0.05, we would select the RE model: the school-specific heterogeneity, though present in the data, is not correlated with the explanatory variables and can very well be taken as random; the RE estimators will be consistent and efficient. Otherwise, if we reject H_0, then $\text{Cov}(x_{it}, \mu_i) \neq 0$, and it would be wiser to use the fixed-effects (FEs) estimator to obtain unbiased estimates [12].

4. Results

4.1. Data and Results

For the empirical implementation, this study used administrative panel data on the population of high schools that participated in standardized university admission tests (*Selectividad*) in Andalusia (Spain) over four consecutive years (2005, 2006, 2007, and 2008). The data set was provided at the time by the regional statistical institute. We could not obtain updated data. Although, the database is adequate to run the panel data models proposed in this methodological paper. However, the data set only contained, averaged by high school, the following information: (i) *Selectividad* test scores of senior students who took the exam in June 2005, 2006, 2007, and 2008, and (ii) *Bachillerato* course grades for those years. Table 1 shows the descriptive statistics of both high schools' average GPA in the

Bachillerato and high schools' average test scores in the *Selectividad*. The statistical analysis is performed by using the three main types of *Bachillerato* (B.): Technological B., Health Sciences B., and Social Sciences B. The different, upper-secondary academic pathways provide high school students with early exposure to the subjects related to their future educational and career options. Some courses are shared (such as Spanish and English); others are specific to each *Bachillerato* (e.g., Physics for the Technological B., Biology for the Health Sciences B., or Economics for the Social Sciences B.). Additionally, we split the population of high schools by ownership structure to take into account that the educational production process may be different in private and public schools [15]. Moreover, "equity would require that the achievement levels in schools be compared with achievement in schools of similar socioeconomic status" [16]. We included subsidized private high schools (called *institutos concertados*) in the category of private high schools. In Andalusia, an important percentage of private high schools receive public funding.

When accounting for differences in outcomes by school ownership structure, students in private high schools perform better than those in public ones. In Table 1, "Yes" means that private high schools scored significantly higher than public high schools (the null hypothesis is rejected at the 5% level of significance). In general, for the three types of *Bachillerato*, the GPAs of private high schools are higher than those of public schools, but their performance in the standardized tests (*Selectividad*) is also better. People sometimes make the argument that grades are "inflated". A priori, however, the belief that private high schools "inflate" grades does not seem to be supported by the descriptive statistics analysis. Student performance in the *Selectividad* is based on previous academic endowments. What should concern us are inequalities in access to higher education; that is, if students' *Bachillerato* course grades, as well as *Selectividad* test scores, are superior in private high schools, their final university entry marks will also be superior, and they will have a greater probability of accessing highly requested degrees such as Medicine. Yet, a basic question that has to be empirically confirmed is whether this result creates inequalities in access to Spanish HEIs.

Nevertheless, the descriptive analysis shown in Table 1 yields limited information for school evaluation. *Selectividad* scores, for example, increased over time in the *Bachillerato* of Social Sciences both in private and public schools. How do we test whether these test scores increased significantly from 2005 to 2008? The increase from 5.82 (5.61) in 2005 to 6.17 (5.80) in 2008 could be due to random error (in parentheses, test scores for public schools). Focusing on the change of the dependent variable (Δy_{it} is also called a *change score*), school-level panel data analysis enables a relatively higher level of statistical validity in policy analysis and school evaluation.

Table 1. High school grades and university entrance examination scores: descriptive statistics. Andalusian high schools, 2005–2008.

	Total Observations [a]	Percentage	High Schools' Average GPA (*Bachillerato*)	High Schools' Average Test Scores (*Selectividad*)	High Schools' Average GPA (*Bachillerato*)				High Schools' Average Test Scores (*Selectividad*)			
			2005–2008	2005–2008	2005	2006	2007	2008	2005	2006	2007	2008
Technological B.												
Private high school	627	23.93	7.58	6.15	7.57	7.51	7.57	7.68	6.36	5.90	6.20	6.12
Public high school	1993	76.07	7.47	5.90	7.40	7.45	7.46	7.54	6.09	5.73	5.94	5.84
Difference of means (private–public)			0.12	0.25	0.17	0.06	0.11	0.14	0.27	0.17	0.26	0.28
Is the diff. statistically significant at 5%? §			Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Health Sciences B.												
Private high school	638	23.85	7.54	6.19	7.39	7.48	7.54	7.73	6.15	6.25	6.13	6.24
Public high school	2037	76.15	7.41	5.96	7.32	7.35	7.42	7.55	5.93	6.00	5.95	5.97
Difference of means (private–public)			*0.13*	*0.23*	0.07	*0.12*	*0.12*	*0.18*	*0.22*	*0.25*	*0.19*	*0.27*
Is the diff. statistically significant at 5%? §			Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Social Sciences B.												
Private high school	643	23.86	7.11	6.01	7.08	7.11	7.10	7.15	5.82	6.02	6.01	6.17
Public high school	2052	76.14	6.99	5.69	7.03	6.98	6.97	7.00	5.61	5.66	5.68	5.80
Difference of means (private–public)			*0.12*	*0.32*	0.05	*0.13*	*0.13*	*0.15*	*0.21*	*0.36*	*0.33*	*0.37*
Is the diff. statistically significant at 5%? §			Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

§ Mean comparison test. diff.—mean(private)–mean(public); H_0—diff. = 0. [a] There are around 160 private high schools and 520 public high schools. For each high school (group), we have about four observations (an average score per year). Source: author's calculations.

Table 2 presents the estimation results. In most panel applications, a choice has to be made between FE and RE estimation. We specify the fixed-effects model, Equation (2), and the random-effects model, Equation (3), as follows:

$$Y_{it} = \mu_i + \beta_1 X_{it} + e_{it} \tag{2}$$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + (\mu_i + e_{it}) \tag{3}$$

where $i = 1, 2, \ldots, N$ (# of high schools) and $t = 2005, 2006, 2007, 2008$; Y_{it} is the dependent variable (average test scores in *Selectividad* of the i-th Andalusian high school in the t-th year); X_{it} is the explanatory variable (average GPA of the i-th Andalusian high school in the t-th year); β_1 is the coefficient for this independent variable; and e_{it} is the error term.

Table 2. Andalusian high schools' results in university entrance examinations: panel data estimates.

	Technological B.			Health Sciences B.			Social Sciences B.		
	Private High Schools			**Private High Schools**			**Private High Schools**		
	Coef.		Std. Err.	Coef.		Std. Err.	Coef.		Std. Err.
High schools' average GPA in the *Bachillerato*	0.6864	**	0.0385	0.6792	**	0.0350	0.6964	**	0.0400
Constant	0.9364	**	0.2942	1.0670	**	0.2661	1.0506	**	0.2870
sigma_μ	0.4332			0.4002			0.4154		
sigma_e	0.4706			0.3838			0.3774		
rho	0.4588			0.5209			0.5478		
Wald chi2(1)	318.05			376.76			302.45		
Prob > chi2	$p < 0.001$			$p < 0.001$			$p < 0.001$		
Number of obs.	625			638			642		
Number of groups	160			164			166		
Dep. var. = high schools' average test scores (*Selectividad*)									
	Random-effects GLS regression			Random-effects GLS regression			Random-effects GLS regression		
Hausman test	chi2(1) = 0.50; Prob > chi2 = 0.4795			chi2(1) = 2.59; Prob > chi2 = 0.1078			chi2(1) = 0.15; Prob > chi2 = 0.6941		
	Public high schools			Public high schools			Public high schools		
	Coef.		Std. Err.	Coef.		Std. Err.	Coef.		Std. Err.
High schools' average GPA in the *Bachillerato*	0.6763	**	0.0224	0.6967	**	0.0215	0.6197	**	0.0228
Constant	0.8503	**	0.1679	0.7972	**	0.1596	1.3550	**	0.1596
sigma_μ	0.4839			0.4176			0.3894		
sigma_e	0.5087			0.4311			0.3939		
rho	0.4750			0.4840			0.4943		
F(1;1464)	908.98	**							
F(1;1510)				1050.95	**				
F(1;1528)							740.10	**	
Number of obs.	1988			2033			2051		
Number of groups	523			522			522		
Dep. var. = high schools' average test scores (*Selectividad*)									
	Fixed-effects (within) regression			Fixed-effects (within) regression			Fixed-effects (within) regression		
Hausman test	chi2(1) = 6.90; Prob > chi2 = 0.0086			chi2(1) = 17.51; Prob > chi2 = 0.0000			chi2(1) = 32.62; Prob > chi2 = 0.0000		

** represents 5% levels of significance. Source: author's calculations.

Our goal is to estimate the slope parameter β_1 that indicates how much Y changes over time, on average per school, when X increases by one unit in Equation (2) or the average

effect of X over Y when X changes across time and between schools by one unit in Equation (3). However, not controlling for unobserved school-specific effects leads to bias in the resulting estimates. The variable μ_i captures all unobserved factors affecting Y that do not change over time. In Equation (2), μ_i ($i = 1 \ldots$ N) is an unknown intercept for each high school (N school-specific intercepts). This implies that all behavioral differences between high schools are fixed over time and are represented as parametric shifts of the regression function. The intercept is allowed to vary from high school to high school, while the slope parameter is assumed to be constant in both the school and time dimensions. In Equation (3), intercepts vary randomly among schools, and μ_i is a random disturbance term that is assumed to be constant over time. The term μ_i is a stochastic variable that embodies the unobservable or non-measurable school differences. Essentially, the effect is thought to be a random school effect rather than a fixed parameter. In short, we might try to discern whether there is a difference in achievement between high schools in the *comunidad autónoma* of Andalucía. Instead of including every high school in the equation (as we would have in the fixed-effects model using dummy variables) one can randomly sample high schools and assume that the effect is randomly distributed across high schools but constant through time. The first step in our analysis is, therefore, deciding whether to estimate a fixed- or random-effects panel model. The results of the Hausman specification test supported the fixed-effects (FEs) model for public high schools and the random-effects (REs) model for private high schools (Table 2). Serial correlation is not a problem in micro panels (with very few years), and taking group means can remove heteroskedasticity. In all cases, for both public and private high schools, the results indicate that students' scores on the *Selectividad* exams are explained by their grades in the *Bachillerato*. These latter grades are a summary indicator of the academic performance during two school years, which reflects, among other factors, student effort and socioeconomic characteristics. The grade point average of private high schools has a larger effect on the *Selectividad* examination results in the Technological B. and Social Sciences B. compared to public secondary schools. For example, a one-unit increase in high school GPAs increases the *Selectividad* exam scores by almost 0.70 points in the Social Sciences B. in private schools versus 0.62 in public schools. However, in the *Bachillerato* of Health Sciences, the value of the estimated β_1 is greater for public high schools compared to private ones: as GPAs increase by 1 unit, standardized test scores increase by 0.70 and 0.68 points, respectively. Nevertheless, the results in the *Selectividad* exams vary between high schools not only by the observed student characteristics (GPAs) but also by unobserved characteristics (time-invariant unobserved heterogeneity). Analyzing the "rho" value in Table 2, an important portion of the variance of the outcomes on the *Selectividad* exams is due to unobserved characteristics that differ between high schools. "rho" is known as the intraclass correlation (% of the variance due to differences across panels). We should highlight the results obtained for the *Bachillerato* of Health Sciences and the *Bachillerato* of Social Sciences in private high schools (RE estimation). If μ_i is conceptualized as a random variable in RE estimation, it can be interpreted as a randomly varying intercept in Equation (3) that captures unmodeled unit-specific heterogeneity of Y's level (e.g., the heterogeneity of *Selectividad* test scores). The above 50 percent variation in high school test scores is explained by the constant over time school effect. That is, educational outcomes in the *Selectividad* vary more among high schools than within a high school on different occasions. This greater dispersion (variability) in the test scores across private high schools would indicate that some exceptional schools set a very high benchmark for the group of private schools. However, how does that help to infer grade inflation? From a short panel, we cannot infer that *Bachillerato* course grades are inflated in those private high schools with poor performance on the *Selectividad* exams. They would be inflated if this pattern holds up over several years.

4.2. Discussion

In light of the results shown in this section, at least two relevant policy questions arise. What do the results tell us about school performance? Do they tell us anything

about the relative performance of public and private schools? Let us examine the results of the *Bachillerato* of Social Sciences to find answers to these questions. The private–public gap in school performance increased over time, both in course grades and test scores (Table 1). This result is verified by the higher marginal effect estimated in Table 2. Private high school GPAs have a larger effect on *Selectividad* examination results (0.70) than public secondary school GPAs (0.62). The usual interpretation of the achievement gap between students from private schools and those from public ones is that private schools have higher standards, more rigorous requirements, and more motivated teachers. However, the "rho" value (known as the intra-class correlation) is 0.548 in private high schools, meaning that between-school differences account for 54.8% of the variation in students' performance in *Selectividad* (i.e., 54.8% of the total variance is due to cross-school variability). The "rho" value is less than 0.50 in public schools. Hence, a key result seems to be that test scores are more dispersed for the private schools, perhaps due to a few exceptional schools. In other words, there are private high schools that have consistently performed well in both the *Bachillerato* and the *Selectividad*. However, there are other private high schools that have very good grades in the *Bachillerato*, but their *Selectividad* results are mediocre. If we assume that the *Selectividad* tests were similarly difficult during the study period, this result could indicate "grade inflation" in GPAs in some private schools. By "grade inflation", we do not necessarily mean that teachers automatically raise students' grades but that they are likely lowering academic standards and students performing worse on standardized tests. In this regard, panel evidence can be very useful to inform those responsible for education (public authorities, for example) about possible "irregular" grading practices in high schools, although panels spanning many more years would be needed. The education policy should focus on the pre-university stages, trying to equalize the educational achievement of students regardless of their origin and the type of high school (private vs. public). This is the only way to guarantee equity in access to the different bachelor's programs at the university.

5. Conclusions

This article proposes the use of econometric analysis for school evaluation. When we have repeated observations from each school over time, we can use panel data methods to control for the school-specific time-invariant characteristics that are not directly observable to the econometrician but influence student learning outcomes (called "unobserved heterogeneity"). The estimation methods are classified by how to treat that heterogeneity. We used the fixed-effects model to eliminate the unobserved heterogeneity because it is assumed to be correlated with any of the explanatory variables. The random-effects model, also known as the variance components model, regards the unobserved heterogeneity as random variables rather than fixed ones. Between the fixed-effects and random-effects models, which model we should use is a critical issue. Essentially, the debate lies in how to treat the unobserved heterogeneity and which model is more efficient. The Hausman test is generally used to choose between a fixed-effects and a random-effects model. If the null hypothesis is not rejected, then it is preferred to use random effects because it produces more efficient estimators. If it is rejected, the fixed-effects model outperforms the random-effects model. For the empirical implementation, we used administrative panel data on the total number of high schools that participated in the university entrance exams in Andalusia (southern Spain), called *Selectividad*. The data covered the years 2005–2008. Our approach implies explaining, at the school level, the mathematical relationship that links *Selectividad* scores with *Bachillerato* course grades. For each type of upper-secondary curriculum in the academic pathway (i.e., *Bachillerato*) and ownership structure, the results indicate that the GPAs of high schools have a positive and statistically significant influence on student performance on the *Selectividad* exams. Compared to public secondary schools, GPAs of private high school had a larger effect on the *Selectividad* examination results in the Technological *Bachillerato* and the *Bachillerato* of Social Sciences. Quite the opposite was observed in the *Bachillerato* of Health Sciences. Is there statistical evidence that grade

inflation exists in high school? Analyzing the "rho" value, educational outcomes in *Selectividad* vary more across private high schools than within a private high school on different occasions. However, since we worked with a short panel, we must be cautious in stating that *Bachillerato* grades are inflated in those private high schools with poor performance on the *Selectividad* exams. Finally, we can highlight that this methodological paper can be useful within the community of educational evaluators, including superintendents, principals, and teachers.

References

1. StataCorp. *Stata Statistical Software: Release 17*; StataCorp LP: College Station, TX, USA, 2021.
2. Sanders, J.R.; Davidson, E.J. A model for school evaluation. In *International Handbook of Educational Evaluation*; Kellaghan, T., Stufflebeam, D.L., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 807–826.
3. OECD. *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes*; OECD Publishing: Washington, DC, USA, 2013.
4. Coutinho-Pereira, M.; Moreira, S. Efficiency of secondary schools in Portugal: A stochastic frontier analysis. *Banco Port. Econ. Bull.* **2007**, *13*, 101–117.
5. Zwick, R. Higher education admission testing. In *Educational Measurement*, 4th ed.; Brennan, R.L., Ed.; American Council on Education/Praeger Series on Higher Education: Washington, DC, USA, 2006; pp. 647–679.
6. Bejar, I.I.; Blew, E.O. Grade inflation and the validity of the Scholastic Aptitude Test. *Am. Educ. Res. J.* **1981**, *18*, 143–156. [CrossRef]
7. Gershenson, S. *Grade Inflation in High Schools (2005–2016)*; Thomas, B., Ed.; Fordham Institute: Washington, DC, USA, 2018.
8. Dowling, A. Output Measurement in Education (Australia Council for Educational Research, December 2008). Available online: https://research.acer.edu.au/cgi/viewcontent.cgi?article=1001&context=policy_analysis_misc (accessed on 19 November 2022).
9. Walpole, M.; McDonough, P.M.; Bauer, C.J.; Gibson, C.; Kanyi, K.; Toliver, R. This test is unfair: Urban African American and Latino high school students' perceptions of standardized college admission tests. *Urban Educ.* **2005**, *40*, 321–349. [CrossRef]
10. Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*; Cengage Learning: Mason, OH, USA, 2015.
11. Baltagi, B.H. Panel data methods. In *Handbook of Applied Economic Statistics*; Ullah, A., Giles, D.E.A., Eds.; CRC Press: Boca Raton, FL, USA, 1998; pp. 311–323.
12. Andreß, H.J.; Golsch, K.; Schmidt, A.W. *Applied Panel Data Analysis for Economic and Social Surveys*; Springer: Berlin/Heidelberg, Germany, 2013.
13. Cameron, A.C.; Trivedi, P.K. *Microeconometrics: Methods and Applications*; Cambridge University Press: Cambridge, UK, 2005.
14. Stock, J.H.; Watson, M.W. *Introduction to Econometrics*; Addison Wesley: Boston, MA, USA, 2003.
15. Lazear, E.P. Educational production. *Q. J. Econ.* **2001**, *116*, 777–803. [CrossRef]
16. Johnson, R.L. The development and use of school profiles. In *International Handbook of Educational Evaluation*; Kellaghan, T., Stufflebeam, D.L., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 827–842.