COMPARING MEANS CAN BE MEAN: QUANTIFYING THE PERFORMANCE OF STUDENTS AT STANDARDIZED EXIT EXAMS[*]

Silvia C. Gómez Soler[**]

[**] PhD Fellow, Maastricht University, The Netherlands; Master of Arts in International Development (Development Economics), American University, Washington, DC., U.S.A., 2012; Master's degree in Economics, Pontificia *Universidad Javeriana*, Bogota, Colombia, 2009; Bachelor's degree in Economics, Pontificia *Universidad Javeriana*, Bogota, Colombia, 2008. Instructor at the School of Economics, Pontificia *Universidad Javeriana*, Bogota, Colombia. E-mail: sg8590a@american.edu; silvia.gomez@javeriana.edu.co

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

201

SILVIA C. GÓMEZ SOLER

Comparing means can be mean: quantifying the performance of students at standardized exit exams

ABSTRACT

The current work introduces an alternative analysis approach to standardized exam results, which have been so far scrutinized by means of ranking systems based on arithmetic mean comparisons. The Propensity Score Matching technique is applied to compare the results of equivalent students in the Colombian field-specific college exit exam (ECAES), which was introduced by the National Education Ministry in 2003. The specific case of the students enrolled at *Universidad Javeriana*'s Business Administration Program is addressed, but the methodology can be easily applied to other programs and universities. The results show a strong treatment effect of attending *Universidad Javeriana* on the performance in this exam. In contrast to the results of previous ranking studies based on simple average data, students from *Universidad Javeriana* were found to perform better than equivalent students. This shows that the construction and interpretation of those rankings might be flawed.

Keywords: higher education, exit exams, economics of education, impact evaluation

JEL classification: I23, I25

Comparar medias no es el medio: cuantificando el desempeño de los estudiantes en exámenes estandarizados

RESUMEN

El trabajo propone una metodología alternativa para analizar los resultados de exámenes estandarizados, que hasta la fecha se han analizado mediante *rankings* construidos a partir de promedios aritméticos. Utilizando la técnica *Propensity Score Matching* se comparan los resultados de estudiantes con características semejantes en los exámenes de Estado colombianos a nivel universitario (ECAES), implementado por el Ministerio de Educación en 2003. Se analiza el caso de los estudiantes de Administración de Empresas de la Universidad Javeriana, pero la metodología puede replicarse en otros programas y universidades. Los resultados muestran un impacto significativo del tratamiento (ser estudiante de la Javeriana) sobre el desempeño en el ECAES. A diferencia de los estudios que usan *rankings* a partir de promedios, los estudiantes de la Javeriana tienen mejor desempeño que los estudiantes comparables, demostrándose que la construcción e interpretación de *rankings* es inapropiada.

Palabras Claves: educación superior, exámenes de Estado, economía de la educación, evaluación de impacto

Clasificación JEL: I23, I25

Comparar as médias não é o meio: quantificando o desempenho dos estudantes em exames padronizados

RESUMO

Neste trabalho, propõe-se uma metodologia alternativa para analisar os resultados de exames padronizados que até o momento são analisados mediante rankings construídos a partir de médias aritméticas. Ao utilizar a técnica Propensity Score Matching, comparam-se os resultados de estudantes com características semelhantes nos exames de estado colombianos em nível universitário (ECAES), implementado pelo Ministério da Educação em 2003. Analisa-se o caso dos estudantes de Administração de Empresas da Universidade Javeriana, mas a metodologia pode ser aplicada em outros programas e universidades. Os resultados mostram um impacto significativo do tratamento (ser estudante da Javeriana) sobre o desempenho no ECAES. À diferença dos estudos que usam rankings a partir de médias, os estudantes da Javeriana têm melhor desempenho que os estudantes comparáveis, o que demonstra que a construção e a interpretação desses rankings não são apropriadas.

Palavras-chave: educação superior, exames de Estado, economia da educação, avaliação de impacto.

Classificação JEL: I23, I25

Introduction

In 2003, as part of an initiative to improve quality and accountability in Colombian higher education, the National Education Ministry introduced a field-specific college exit exam known as ECAES. The results of the ECAES have been used in recent years by the Colombian media to establish rankings of what they consider to be the best academic programs. The purpose of this paper is to develop an alternative analysis approach to the one that, based on arithmetic mean comparisons, is currently being used by the media to quantify and rank the performance of students in standardized exams. Although the study focuses on the specific case of the students enrolled at *Universidad Javeriana*'s Business Administration Program, the methodology can be easily applied to other universities and academic programs.

This paper examines and builds on the problems that might come up in attempting to analyze the results of the ECAES exam using the rankings that have been constructed by the Colombian media as a measure of educational quality. It is argued that, although the ECAES results are not a good proxy of education quality, they are a good source of information to compare student specific skills at the national level (Popham, 1999). Comparing and ranking average results misses the individual dimension of students' performance in this exam. In this respect, it has been stated that "the construction of indices by which institutions or departments are ranked is arbitrary, inconsistent and based on convenience measures" (Harvey, 2008). Ranking according to average results is a

naïve approach because it assumes that the change in outcome [ECAES score] is attributable exclusively to the program being evaluated. The main contribution of this paper is to develop an alternative econometric and analytical approach to the interpretation of standardized exam results.

Propensity Score Matching is used to compare the results of equivalent students and to deal with methodological problems related to self-selection bias. The results show a strong treatment effect of attending *Universidad Javeriana* on the performance in the ECAES exam. The average treatment effect on the treated varies from 5.863 to 9.051 points. If we compare students that are similar in observable characteristics, *Universidad Javeriana* students have significantly higher scores than other students in the ECAES exam, which indicates that the ranking approach might be misleading.

The robustness of the average treatment effect estimated in this paper is checked by comparing the results from propensity score matching with the results obtained using other methodologies. Specifically, the results of least squares regressions and fixed effects regressions were used to test the consistency of the propensity score matching results.

Due to the nature of the data analyzed in this paper, it is particularly interesting to examine the heterogeneity of the impact. Given its important policy implications, the specific effect of learning at *Universidad Javeriana* on certain variables like gender and socioeconomic stratum is particularly appealing. Students from lower socioeconomic strata

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

203

seem to benefit more from an education at this institution. Attending *Universidad Javeriana* also has a positive effect of around 1.728 points on the performance of female students in the ECAES exam, which suggests a positive impact on gender inequality reduction.

The rest of this paper is organized as follows: section 2 presents the conceptual background; section 3 describes flawed rankings and the media; section 4 presents background information about higher education and standardized exams in Colombia; section 5 describes the educational data from Colombia used in the analysis; section 6 reviews the econometric methodology; section 7 reports and analyzes the main results; section 8 presents the heterogeneous impact of the treatment; and section 9 presents conclusions and policy implications.

1. Conceptual Framework

1.1 Higher Education Rankings

Higher education rankings have been criticized on many fronts, but their publication keeps attracting a lot of attention and generating both academic and non-academic discussion. As mentioned by Tofallis (2012), there is a considerable amount of ranking related issues, which have been highlighted primarily by Educationalists. Tambi *et al.* (2008) point out that determining objective criteria to measure the quality of higher education institutions and designing the basis on which they should be evaluated is a very difficult, if not impossible task. However, rankings are still widely used in various grounds and they

continue to have an impact that goes beyond what their arbitrary design would deserve (Harvey, 2008).

Most university rankings around the world have been built and published by private institutions and the media (Buela-Casal *et al.*, 2006). According to these authors, there are three important aspects that should be considered in the analysis of a higher education ranking system: who ranks, why rank, and the audience for rankings. Harvey (2008) highlights that rankings serve a number of purposes. First, they provide consumers with easily interpretable information on the standing of higher education institutions. According to Harvey (2008), many people see this as the primary purpose of published rankings. Second, rankings can help stimulate competition among institutions; and third, they can be useful tools for assessment and accountability in the national higher education system (Harvey, 2008).

The literature has claimed that rankings can contribute to a better understanding of quality in higher education. However, the connection between rankings and quality is naïve. Harvey (2008) remarks how the operationalization of the concept of quality is at best superficial, rankings being inadequate attempts to operationalize aspects of excellence. In fact, a crucial part of the learning process, namely transformation, is barely considered by rankings (Harvey, 2008). Furthermore, according to Carey (2006), rankings might lead institutions to a loss of independence and freedom when it comes to controlling their academic priorities and programs.

Coming from various sources, criticism on rankings is based on methodological, pragmatic, moral and philosophical concerns (Harvey, 2008). A recurrent critique notes how rankings offer a simplistic evaluation perspective. In fact, Ashworth and Harvey (cited by Tambi *et al*., 2008) claim that performance indicators cannot be used to evaluate the quality of the processes that take place in a higher education institution because of their over-simplistic approach to quality. Rankings are also highly criticized because of the arbitrary selection of the indicators they consider. Stella and Woodhouse (2006, cited by Harvey, 2008) mention that rankings are often based on what can be measured instead of what is relevant and important. Also in this sense, convenience data have frequently taken the place of consistent theoretical backgrounds (Harvey, 2008). These observations constitute a very strong criticism on the validity of rankings. In fact, Harvey (2008) mentions that there is little evidence that the selection of the indicators used in the construction of rankings is part of a rigorous process of theoretical reflection. As highlighted by this author, the arbitrary way publishers frequently choose indicators is truly problematic.

Harvey (2008) has mentioned how the political, social and cultural contexts in which higher education institutions operate affect the way they perform and what they can do. In fact, according to Tambi *et al*. (2008), the indicators used in the construction of a ranking should relate to the mission statement of the higher education institution. As a result, it is necessary to consider these contexts when comparisons are made as part of the process of constructing a ranking. However, as mentioned by Harvey (2008), this is rarely done.

Nevertheless, it is important not just to point out the weaknesses of these rankings. It is essential to highlight alternatives that could be applied in order to benchmark the performance of higher education institutions. The objective of this paper is precisely to contribute in this matter by providing an alternative approach to the available rankings in order to quantify and analyze the performance of students in national exit exams. The specific case of the students enrolled at a Business Administration Program in a Colombian university is addressed, but the proposed methodology could be easily applied to other programs and universities. Nevertheless, the existence of other innovative approaches to university rankings must be noted as well. Tofallis (2012) proposed a multiplicative approach intended to facilitate the aggregation of indicators, a methodology that assists in the process of comparing equivalent students and institutions in a ranking system.

1.2 Ranking Educational Quality Through Exit Exams

Aghion *et al*. (2005) have shown that the provision of qualified tertiary education is an important determinant of economic growth and development. This finding has provoked a widespread concern amongst policy makers over the quality of higher education. Its assessment has led governments from all over the world to design standardized exit exams for evaluating their students. An important question arises from this strategy: can we really measure the quality of instruction us-

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

205

ing the results of standardized exams? The literature on economics has shown that this might not be the case. According to Card and Krueger (1994), economists are skeptical about standardized testing because the tests are not only arbitrarily scaled, but they can be manipulated by teachers and test writers as well. Additionally, Becker (1997) highlights that education is a multi-product output which cannot be reflected in a single multiple-choice test score. It has been shown, indeed, that highly knowledgeable students in subjects like economics already had a high aptitude on the topic (Becker, 1997).

A student might have acquired an enormous amount of knowledge and skills that are not covered or cannot be tested in a standardized test. Additionally, as mentioned by Popham (1999), there might be a confounded causation problem, which arises when the performance of students in a standardized exam is influenced by more than one factor. This author identifies three main factors: what is being taught in college, a student's native intellectual ability, and a student's out of college learning. It is noteworthy that only the first factor is related to educational quality. Since it is impossible to calculate the importance of each factor in the performance of a student, it can be argued that the results of a standardized exam are not a good measure of instruction quality. We simply don't know which factor we are actually measuring.

According to Card and Krueger (1994), a product or skill is only worth what the consumer is willing to pay for it, and there is no guarantee that standardized exams measure skills of economic value. According to this argument, the quality of higher education is given by its expected return to society. Becker (1997) mentions that the beliefs of an instructor, a test design committee or an entire faculty about the importance of certain forms of knowledge and intellectual skills are not always consistent with what students desire and what employers expect and pay for.

However, the literature has shown that standardized test results are a good source of information to compare student specific skills nationally (Popham, 1999). It could be argued, then, that the results of standardized exams are a good measure of the level of knowledge of basic skills that are necessary for graduate school, but are probably not a good measure of educational quality. While the literature presents arguments in favor and against this notion, it is precisely the way standard exam results are used in the current paper. Still, other variables such as earnings can also have a good performance in measuring educational quality (Card and Krueger, 1994). In fact, Saavedra (2007 and 2009) has used labor market variables to measure educational quality in Colombia.

2. Flawed Rankings and the Media

The Colombian media have used the results of the ECAES exams to establish rankings of what they consider to be the best academic programs. Apart from the contrasting uses mentioned above for this type of results, inadequacy in the methodology employed for the analysis might also take a toll on the validity of this ranking approach. By using descriptive tools such as simple averages and standard deviation, the big picture is missed.

The performance of the students from a given university should be compared to that of other students that are similar in observable characteristics. When the comparison is set between students that are very different in terms of academic, socioeconomic and family background variables, the results might actually capture such differences instead of the effect of the educational program. Hence, as the performance of students in a standardized exam is influenced by more than one factor (Popham, 1999), it is important to highlight that the rankings of the media can be misleading because they do not indicate the improvement in student performance caused by the evaluated program.

The perception of individuals about the Business Administration programs offered in different universities can not only be tainted by these rankings, but also incorporated in the decision making process of these people. Ultimately, these biased assessments can negatively affect the reputation of a given institution.

The goal of this paper is to build on this literature and develop an alternative empirical methodology to analyze the results of standardized exams. Although we have focused on *Universidad Javeriana* as a case study, this analytical framework can be easily applied to other universities. We propose the hypothesis that *Universidad Javeriana* students obtain higher ECAES scores than equivalent students from other universities; which would support the notion that the rankings developed by the media can be misleading.

3. Setting the Scene

3.1. Higher Education in Colombia

Sixty nine percent of the 177 colleges and universities in Colombia are privately owned and operated (Saavedra, 2009). In 2006, 30% of 18 to 24 year old Colombians were enrolled in college, and 47% of them were in private institutions. Seventy three percent of universities are located in the largest cities, and 50% are found in the three largest cities (Saavedra, 2009). Additionally, 31% of all colleges and universities operate in Bogota, the capital city. The most selective institutions have a higher fraction of full time and PhD faculty, greater expenditures per student and higher admission standards (Saavedra, 2009).

As of 2004, there were a total of 574 academic programs in 118 higher education institutions which would award a business administration major upon completion of the academic requirements. Out of those 574 programs, only 13 have a high quality certification from the Colombian Ministry of Education.

3.2 Business Administration at Universidad Javeriana

Universidad Javeriana is one of the major research and teaching centers in Colombia. The Business Administration Department has 130 professors, out of whom 27 work full time. Nine of them 27 have a PhD in Business, while 10 have a MSc in Business. The Department offers both undergraduate

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

207

and graduate degrees. At the graduate level, there are 5 different one year specialization programs (*Universidad Javeriana*, 2011).

The undergraduate program requires students to complete 160 academic credits. One hundred and four of these credits correspond to core courses; 24 to concentration courses; 16 to electives, and another 16 to complementary subjects. The core courses include Introduction to Management, Basic Finance, Corporate Finance, Marketing, Marketing Management, Organizational Behavior, Human Resources, and Micro/Macro Economics. In 2008, the Colombian National Ministry of Education certified the high quality of this undergraduate program for a period of six years. Students are expected to take any time from 4 to 5 years to complete the graduation requirements (*Universidad Javeriana*).

3.3. Admission Process at Universidad Javeriana

The criteria used at *Universidad Javeriana* to determine acceptance to the Business Administration program is based on observables. The main component to determine the acceptance of a new student are the Saber11 exam results. High School Seniors take this exam as a requirement for college admission[1]. The Saber11 exam tests specific subjects like Mathematics, Social Sciences, Spanish, Physics, Chemistry, Biology, and a Foreign Language. The admissions committee uses the average score of all the core subjects of the exam to measure overall pre-

paredness for higher education. Additionally, due to the importance of Mathematics in the Business Administration major (*Universidad Javeriana*), this score is considered separately. *Universidad Javeriana* being a Jesuit University, having attended a Jesuit High School is another factor that is taken into consideration by the admissions committee to determine acceptances.

There are other factors that are not necessarily taken into consideration by the admissions committee to make a decision, but ultimately affect whether the student ends up enrolling at *Universidad Javeriana* or not. Some of these factors are: whether the student works, the level of education of the parents, and his/her socioeconomic stratum.

3.4. The ECAES Exam

The ECAES is a State managed exit exam that seeks to evaluate formal undergraduate education at higher education institutions. The evaluation of Business Administration students started in 2004. According to decree 1781 of 2003, the objectives of the ECAES exam are to:

a) Make sure that students have the sufficient skills/competences when they graduate from university.

b) Build indicators of undergraduate education value.

c) Provide information that enables comparisons amongst academic programs, institutions and learning methodologies, and follow their progress over time.

---

[1]　This exam is similar to the SAT exam that is administered in the United States.

d) Provide information for the construction of quality indicators for academic programs and higher education institutions. This information is meant to support policy design and aid in the decision making process on educational matters.

This same decree establishes that this exam is mandatory for all college seniors.

The Colombian Association of Business Administration Departments (ASCOLFA) was made responsible for the design of the questions for the exam. ASCOLFA has put together a committee with participants from various universities to design the evaluation, which includes a total of 200 questions, out of which 60 deal with basic subjects (mathematics, statistics and economics), 25 with management and organizations, 25 with finance, 25 with production and operations, 25 with marketing, 20 with human resources, and 20 with ethics, corporate social responsibility and law. The questions that are asked in the exam are at the basic to intermediate level. The scale used to report individual results has been normalized, with a theoretical mean of 100 and a standard deviation of 10. This means that a student who gets a score of 104.5 is above the national average by 0.45 standard deviations (ASCOLFA, 2006).

4. Data

Databases containing the nation-wide results of the ECAES and Saber11 exams, as well as information about the socioeconomic features of the students who took the exam, have been made public by ICFES[2] for research purposes. Those datasets were the information source for this empirical paper. The database of the ECAES exam covers the years 2007 to 2009, while that of the Saber11 exam covers the years 2000 to 2010.

The sample used in this empirical exercise has nation-wide observations from Business Administration students about whom ICFES has both ECAES and Saber11 results information. Although ICFES does not have a dataset where the results of both exams are matched for each student, this agency provides a codification strategy to do this. After matching and dropping all the incomplete observations (those lacking information from either ECAES or Saber 11 results), 13,595 observations were left. The missing data does not render systematic differences between complete treated cases and complete control cases. Hence, it can be assumed that the sample is representative of the population.

Table 1 presents the statistics of the scores obtained by students in the Saber11 and ECAES exams. Out of a total of 13,595 observations, 426 correspond to *Universidad Javeriana*'s students. This means that the students of this institution roughly represent 3.13 per cent of the total sample. Specifically developed for this empirical paper, the variable *average score in the main components of the Saber11 exam* captures the average score obtained in the core subjects of this exam. It is calculated as the arithmetic mean of the scores obtained in the biology, philosophy, physics, history, and geography

---

2 The governmental agency in charge of these exams.

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

209

Table 1. Scores in the Saber11 and Ecaes Exams

Variable	Universidad Javeriana		Other Universities		t-test (p-value)
	Mean	Std. Dev.	Mean	Std. Dev.	
Math score in the Saber11 exam	34.34	20.69	30.33	20.70	0.001
Average score in the main components of the Saber11 exam (other than math)	37.08	21.5	32.77	21.95	0.001
Score in the Ecaes exam	107.597	7.34	99.02	11.08	0.0000
Observations	426		13169		

Source: Author's calculations using the ICFES dataset

sections of the exam. The average ECAES score of *Universidad Javeriana's* students is higher (107.6) than those of students from other universities (99.02). The average score across treatment and control groups is 99.29 points, with a standard deviation of 11.09. T-tests of means are also reported for the two categories in order to provide initial evidence of differences among the two groups.

Table 2 presents information about the educational background of the parents of the students in the sample through the variables *level of education of the student's father* and *level of education of the student's mother*, which can go from no formal studies to completion of graduate school. On average, 76.48 per cent of the mothers of *Universidad Javeriana*'s Business Administration students had completed at least High School, while 63.83 percent of the mothers of students from other institutions had reached that educational level. Contrastingly, 75.94 per cent of the fathers of *Universidad Javeriana*'s Business students had completed at least High School, while the same figure for the fathers of students from other institutions is 64.34 percent. In other words, the average level of education

of the student's mother is higher than that of the father in the case of Javeriana's students; whereas in the case of students from other institutions the father was more educated than the mother. T-tests of means are also reported in order to provide initial evidence of differences between the two groups.

Table 3 presents information about the socioeconomic strata of the students in the sample, which in Colombia go from 1 to 6, respectively indicating the lowest to the highest income levels. Most of the students that attend *Universidad Javeriana* come from levels 3 (28.87%) and 4 (28.17%), both of which make up the middle class and sum to 57.04%. Likewise, other institutions receive most of their students (53.66%) from strata 3 or 4. T-tests of means are also reported in table 3 for both categories.

Information about the students that attended a Jesuit High School is presented in table 4, corresponding to 2.1% of all the students in the sample, and 12.7% of the students that attended *Universidad Javeriana*. Given the Jesuit affiliation of the institution, this is a unique characteristic of their students.

Table 2. Parental Education

	Father's Education					Mother's Education				
	Javeriana		Other Universities		t-test (p-value)	Javeriana		Other Universities		t-test (p-value)
	Number of Students	%	Number of Students	%		Number of Students	%	Number of Students	%	
Parent didn't complete any studies (level 1)	6	1.42	239	1.83	0.8361	1	0.24	125	0.96	0.487
Incomplete Elementary School (level 2)	25	5.9	1,300	9.96	0.9969	17	4.0	1,195	9.14	0.9999
Elementary school was completed (level 3)	27	6.37	1,474	11.29	0.9992	33	7.76	1,477	11.29	0.9875
Incomplete High School (level 4)	44	10.38	1,641	12.57	0.9057	49	11.53	1,937	14.81	0.9674
High school was completed (level 5)	84	19.81	2,904	22.25	0.8738	105	24.71	3,076	23.52	0.2680
Technical program was not completed (level 6)	22	5.19	1,061	8.13	0.9850	27	6.35	1,004	7.68	0.8381
Technical program was completed (level 7)	48	11.32	1,604	12.29	0.7147	52	12.24	1,904	14.56	0.9037
University Degree was not completed (level 8)	37	8.73	730	5.59	0.0028	31	7.29	689	5.27	0.0000
University was completed (level 9)	79	18.63	1,441	11.04	0.0000	69	16.24	1,143	8.74	0.0002
Graduate school was completed (level 10)	52	12.26	658	5.04	0.0000	41	9.65	531	4.06	0.0000

Source: Author's calculations using the ICFES dataset

Table 3. Socieconomic Strata

	Other Universities		Javeriana		t-test (p-value)
	Number of Students	Percentage	Number of Students	Percentage	
Stratum 1	840	6.38	9	2.11	0.9998
Stratum 2	3,386	25.71	59	13.85	0.9898
Stratum 3	5,036	38.24	123	28.87	0.9999
Stratum 4	2,031	15.42	120	28.17	0.0021
Stratum 5	1,113	8.45	71	16.67	0.1002
Stratum 6	763	5.79	44	10.33	0.0000

Source: Author's calculations using the ICFES dataset

Table 4. Jesuit High School Graduates

		Student attended a jesuit school		
		Freq.	Percent	t-test (p-value)
Total Sample	0	12,898	97.94	0.000
	1	271	2.06	
Javeriana	0	372	87.32	
	1	54	12.68	

Source: Author's calculations using the ICFES dataset

The descriptive statistics presented in this section suggest the existence of a reasonable counterfactual for the analysis. The common support and balancing properties are tested in section 6 to further demonstrate the existence of a reasonable counterfactual and therefore justify the use of the Propensity Score Matching technique.

5. Method

As previously mentioned, the goal of this paper is to use an alternative approach to as-

sess the ECAES exam performance of the students enrolled at *Universidad Javeriana*'s Business Administration Program. To do so, a program evaluation technique known as Propensity Score Matching is used to find out whether an education at *Universidad Javeriana* (the intervention) is effective in achieving a higher score in the ECAES exam (the objective). Propensity Score Matching was first used for a similar purpose by Allcott and Ortega (2009) in order to estimate the effects of graduating from the *Fe y Alegría* private school system in Venezuela on standardized test scores (2007).

However, it is also important to assess alternative methodologies to make sure that the results are consistent. Propensity score matching is based only on observables, and therefore it is important to consider other non-experimental methods to test the robustness and consistency of the results. In this paper, ordinary least squares and fixed effects regressions are considered for that purpose.

212

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

5.1 Propensity Score Matching

5.1.1 Standard Framework for Evaluation

According to Heinrich *et al.* (2010), the main challenge of a program evaluation is the construction of the counterfactual outcome, i.e., what would have happened to participants in the absence of treatment. The standard framework to formalize this problem is the potential outcome approach or Roy–Rubin model (Caliendo and Kopeinig, 2008).

The treatment indicator D_i equals one if individual i receives treatment, and zero otherwise. The potential outcomes are defined as $Y_i(D_i)$ for each individual i, where $i = 1, \ldots, N$ and N denotes the total population (Caliendo and Kopeinig, 2008). The treatment effect for an individual i can be represented as:

$$\tau_i = Y_i(1) - Y_i(0)$$

It is not possible to directly estimate this effect because we cannot observe both $Y_i(1)$ and $Y_i(0)$ for the same individual i; the counterfactual outcome cannot be observed (Caliendo and Kopeinig, 2008). Hence, the counterfactual has to be estimated using statistical methods like Propensity Score Matching (PSM). PSM uses information from a pool of units that do not participate in the intervention, in order to identify what would have happened to the participating units in the absence of the intervention (Heinrich *et al.*, 2010).

5.1.2 Average Treatment Effect on the Treated and Selection Bias

Caliendo and Kopeinig (2008) mention that, according to Heckman (1997), the most relevant evaluation parameter is the average treatment effect on the treated (ATT). This parameter focuses on the effects on those for whom the program is intended, and is given by:

$$\tau ATT = E(\tau | D = 1) = E[Y(1)|D = 1] - E[Y(0)|D = 1]$$

The outcomes of individuals from the treatment and comparison groups would differ even in the absence of treatment, leading to a selection bias. According to Caliendo and Kopeinig (2008), this can be expressed as:

$$E[Y(1)|D = 1] - E[Y(0)|D = 0] = \tau ATT + E[Y(0)|D = 1] - E[Y(0)|D = 0]$$

Therefore, the parameter τATT is only identified if $E[Y(0)|D = 1] - E[Y(0)|D = 0] = 0$. When assignment to treatment is random, this condition is always met. In this empirical study, assignment to treatment is nonrandom; therefore, units receiving treatment and those excluded from treatment may differ in characteristics that affect both participation and the outcome of interest. This problem is known as selection bias.

As discussed by Heinrich *et al.* (2010), to avoid biased results the PSM technique finds an untreated unit that is similar to a participating unit. The impact of the intervention can be estimated as the difference between a participant and the matched comparison case (Heinrich *et al.*, 2010).

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

213

5.1.3 Eliminating Selection Bias

To eliminate potential bias, the matching process has to be done considering a full range of variables across which the treatment and comparison units might differ. This can be problematic in terms of dimensionality. PSM allows reducing the problem to a single dimension by defining a propensity score, which is the probability that a unit in the combined sample of treated and untreated units receives the treatment. Instead of trying to match all the variables, individuals can be compared on the basis of propensity scores (Heinrich *et al.* 2010)

According to Heinrich *et al.* (2010), two conditions must be satisfied to implement PSM. First, the variables in which the treated and untreated groups differ must be observable to the researcher. The rich database available from ICFES allows for that condition to be met. This assumption is known as the *conditional independence* or *unconfoundedness* assumption. Second, in order to calculate the difference in mean outcomes, there must be a positive probability of finding either a treated or an untreated unit to ensure that any of the former can be matched with one of the latter. This assumption is known as the *common support* or *overlap condition* (Heinrich *et al.* 2010). The common support condition is tested and discussed in more detail in section 6.

5.1.4 Implementing Propensity Score Matching

The first step to apply PSM is to estimate the propensity score, which is the probability of attending the Business Administration pro-gram at *Universidad Javeriana*. Given that the treatment status is dichotomous, a logit or probit function can be used. According to Heinrich *et al.* (2010), there are no strong differential advantages in using either logit or probit models with binary treatment variables. The variables that are included in the probit and logit models are:

Dependent Variable: dummy variable taking value 1 if the student is from *Universidad Javeriana*, and 0 otherwise.

Independent Variables: whether the student works, the year when they took the exam (a dummy variable indicating each year), the average score in the core subjects of the Saber11 exam, the score in the mathematics section of the Saber11 exam, whether they attended a Jesuit High School, the level of education of the parents (a dummy variable indicating each level of education), and the family's socioeconomic stratum (a dummy variable indicating each stratum). Those variables are included following the admission criteria discussed in section 3.3.

The function used to estimate the propensity score is then:

JAVERIANA=f(STUDENTWORKS, YEARICFES, SUBCORE, SUBMATH, JESUITSCHOOL, EDUCFATHER, EDUCMOTHER, STRATA)

Once the propensity score has been calculated, a matching algorithm is chosen to contrast the outcome of a treated individual with those of the comparison group mem-

bers (Caliendo and Kopeinig, 2008)[3]. In this case, caliper and kernel matchings, as well the nearest neighbor algorithm, were used. These algorithms were used to test the robustness of the results.

In the nearest neighbor algorithm an individual from the comparison group is chosen as a match for a treated individual through the closest propensity score (Heinrich *et al.*, 2010)*.* In the case with replacement, an untreated individual can be used more than once as a match, whereas in the case without replacement the individual is considered only once (Caliendo and Kopeinig, 2008). According to Caliendo and Kopeinig (2008), it might be problematic to use the nearest neighbor algorithm because it is possible that the closest neighbor is far away. For that reason, it is important to consider additional algorithms like caliper matching and kernel matching. In caliper matching, a tolerance level for the maximum propensity score distance is established, so as to avoid bad matches. However, there is a possibility that fewer matches can be performed with this algorithm, and this is problematic because it increases the variance of the estimates. Another concern with caliper matching is that reasonable levels of tolerance are difficult to determine (Caliendo and Kopeinig, 2008). On the other hand, kernel matching resorts to non-parametric matching estimators that employ the weighted averages of all individuals in the control group to construct

the counterfactual outcome. In this case, the variance is lower because more information is used (Caliendo and Kopeinig, 2008).

5.2 Ordinary Least Squares and Fixed Effects Regressions

Ordinary least squares regressions can also be estimated to determine the impact of attending *Universidad Javeriana* on the results in the ECAES exit exam. OLS regressions are often estimated because of their ease of use and interpretation, but it is important to recognize that they are potentially affected by endogeneity and do not control for fixed effects. As mentioned by Kennedy (2003), the dominant role of the ordinary least squares estimator is that of a standard against which other estimators are compared. In the current paper, PSM was preferred over OLS regression because the former effectively compares only equivalent individuals, whereas OLS does not. In addition, potential problems associated to misspecification can be overcome by the non-parametric approach to PSM, which is not the case of OLS regression.

Another methodology that is considered in this analysis is fixed effects regression. A fixed effects model assumes differences in intercepts across groups or time periods. According to Yaffee (2003), in a fixed effects model the slope is constant but intercepts differ according to the cross-section, which in this case corresponds to the high school institution. Despite the lack of significant temporal effects, there are significant differences among high schools in this type of model. While the intercept differs among high school institutions, it may or may not

---

[3]   To do the matching and calculate the impact of the program, it is necessary to use a program called pscore which is available in Stata. PSMatch2 can also be used in Stata to perform the matching. The PSMatch results are also available upon request.

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

215

differ over time. Torres (2007) highlights that when using a fixed effects model, it is assumed that something within the individual may impact or bias the predictive or outcome variables and is therefore necessary to control it. This model removes the effect of time-invariant characteristics from the predictive variables so that it is possible to assess the net effect of the predictor. According to Kennedy (2003), fixed effect models allow controlling for individual heterogeneity, thus reducing aggregation bias and improving efficiency through more variable data and reduced collinearity. However, it is important to highlight that the advantage of PSM over fixed effects regression is that PSM does not rely on the assumption of a specific functional form, whereas fixed effects regression does.

6. Results

Table 5 reports probit estimates for the propensity score of attending *Universidad Javeriana*. The model is specified using the variables outlined in section 6.4. The logit estimates are also available upon request. As mentioned earlier, Heinrich *et al*. 2010 suggest that for a binary treatment variable there is no strong advantage in using either a logit or a probit model. Three variables are significantly and positively related to the probability of attending *Universidad Javeriana*: the dummy variable expressing whether the student graduated from a Jesuit High School and high levels of education for both the mother and the father. The fact that those variables are significant and the coefficients are positive (as expected) is a good sign of result suitability. As mentioned earlier, and according to *Universidad Javeriana*'s Business

Administration Program, the average score obtained by the student in the core subjects of the exam and whether they attended a Jesuit high school are taken by the admissions committee as a measure of overall preparedness for higher education, thus constituting the main aspects considered to determine acceptances. The level of education of the student's parents is not directly considered by the admissions committee to take a decision, but according to the results presented here, high levels of education amongst the parents of a student ultimately affect whether he/she ends up enrolling at *Universidad Javeriana* or not. However, it is worth noting that a good number of the dummy variables for parental education are not significant. The results of the joint significance tests (F-statistics) between parental education and socioeconomic strata (appendix 1) justify the inclusion of the latter in the model.

As mentioned by Heinrich *et al*. (2010), an important step in investigating the validity of the PSM estimation is the verification of the common support condition. The odds of participating in an intervention, as conditioned by observed characteristics, lie between 0 and 1. Heinrich *et al*. (2010) highlight that the common support condition ensures that units with the same X_i values have a positive probability of being either participants or nonparticipants. In this case, the condition of common support between treatment and comparison groups is checked through visual inspection of their propensity score distributions. Graph 1 shows the kernel distribution and the common support area across the treated [Javeriana students] and untreated [those from other universities] in

Table 5. Propensity Score for Attending Universidad Javeriana

	Probit	
	Coefficient	P>|z|
Student took the Saber11 Exam 98	0.0656449	0.557
Student took the Saber11 Exam 99	0.1052732	0.413
Student took the Saber11 Exam 00	0.0548776	0.532
Student took the Saber11 Exam 01	0.1335094	0.113
Student took the Saber11 Exam 02	-0.0604439	0.567
Student took the Saber11 Exam 03	0.104237	0.327
Student took the Saber11 Exam 04	-0.1331099	0.464
Student took the Saber11 Exam 05	0.062265	0.587
Student Socieconomic Strata (1)	-0.5151172	0.001***
Student Socieconomic Strata (2)	-0.3314912	0.001***
Student Socieconomic Strata (3)	-0.2286372	0.009**
Student Socieconomic Strata (4)	0.0772175	0.382
Student Socieconomic Strata (5)	0.0727177	0.447
Graduated from a Jesuit High School	0.8798918	0.000***
Math score in the Saber11 exam	-0.0012904	0.768
Average score in main components of Saber11 exam (other than math)	0.0050321	0.224
Level of education of the student's father - Level 1	0.075954	0.667
Level of education of the student's father - Level 2	-0.0576831	0.740
Level of education of the student's father - Level 3	0.0346157	0.834
Level of education of the student's father - Level 4	-0.0199692	0.900
Level of education of the student's father - Level 5	-0.0976609	0.585
Level of education of the student's father - Level 6	0.0496695	0.763
Level of education of the student's father - Level 7	0.1035503	0.550
Level of education of the student's father - Level 8	0.1678998	0.104*
Level of education of the student's mother - Level 1	-0.2670015	0.049**
Level of education of the student's mother - Level 2	-.0029415	0.979
Level of education of the student's mother - Level 3	-0.1147975	0.235

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

217

	Probit	
	Coefficient	P>|z|
Level of education of the student's mother - Level 4	-0.0338425	0.680
Level of education of the student's mother - Level 5	-.0372692	0.750
Level of education of the student's mother - Level 6	-0.1403504	0.132
Level of education of the student's mother - Level 7	-0.068484	0.531
Level of education of the student's father - Level 8	0.1085609	0.302
	# obser. = 13595 LR chi2(33) = 91.43 Prob > chi2 = 0.0000 Log likelihood = -263.789 Pseudo R2 = 0.0715	

*** p<0.01, ** p<0.05, * p<0.1

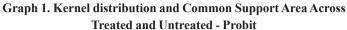Source: Author's calculations using the ICFES dataset

the sample. There is a great degree of over-lap in the propensity scores of the treatment and comparison units[4]. By setting the common support condition, it is ensured that any combination of characteristics observed in the treatment group can also be observed in the control group (Caliendo and Kopeinig, 2008). In other words, we can make sure that each treatment unit has a corresponding matching unit in the comparison group.

To test the Balancing Property Hypothesis, the sample is divided in equally spaced groups defined over the estimated propensity score, and it is checked whether the average propensity score of the treatment and control units differ within each group (Dehejia and Wahba, 2002). If there are significant differences in the average propensity score between treatment and control units in at least one group, the latter is split in half and the test is repeated again. This process continues until the average propensity score of treatment and control units do not differ in any group. In other words, when the balancing property is met, individuals with the same propensity score have the same distribution of observable characteristics independently of treatment status (Dehejia and Wahba, 2002). Table 6 shows the distribution of students attending *Universidad Javeriana* across blocks within the common support. The final number of blocks is 6, which ensures that the mean propensity score is not different for the treated and the controls in each block.

T-statistics for the equality of covariate means within estimated propensity score blocks show that students that attended *Universidad Javeriana* and those that attended other institutions have very similar observable characteristics within each block in all cases. In the full sample, there were no significant differences in the means of the

---

[4]   The reported results correspond to the Probit regression, but the results of the Logit regression are also available upon request.

Graph 1. Kernel distribution and Common Support Area Across Treated and Untreated - Probit

Source: Author's calculations using the ICFES dataset

Table 6. Lower Bound - Numbers of Treated and Untreated Units in Each Block

pscore lower block	Javeriana		Total
	0	1	
0.0032876	7,591	125	7,716
0.025	3,706	130	3,836
0.05	1,582	116	1,698
0.1	217	20	237
0.2	63	26	89
0.3	3	9	12
Total	13,162	426	13,588

The Balancing Property is Satisfied

Source: Author's calculations using the ICFES dataset

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

219

covariates between students attending Javeriana and those who did not.

The balancing property is also satisfied, and therefore, the estimated propensity score can be used to calculate the average treatment effect through the nearest neighbor, caliper and kernel matching algorithms. Table 7 presents the results of the matching process, revealing a strong treatment effect (which is consistent across algorithms) of attending *Universidad Javeriana* on the performance in the ECAES exam. The average treatment effect varies from 5.863 to 9.051 points. To put these results in context, the average score across treatment and control groups is 99.29 points, with a standard deviation of 11.09. Thus, the treatment effect corresponds to more than half a standard deviation, which is a very considerable improvement in raw ECAES results for the students from *Universidad Javeriana*. This suggests that, as hypothesized, students from this institution have higher ECAES scores than others with similar observable characteristics.

It could be argued, then, that the rankings developed by the media are misleading because students from *Universidad Javeriana* are actually performing better than equivalent students. This shows that simple comparisons of mean test scores between students from different universities may be biased measures of the true impact of the treatment, because the two groups may be very different in their observable and unobservable characteristics. The performance of students should be compared to that of equivalent students. As Stella and Woodhouse (2006, cited by Harvey, 2008) point out, rankings are often based on what can be measured rather than what is relevant and important.

Nevertheless, as mentioned previously, it is important not just to point out the weaknesses of rankings, but to highlight alternatives for benchmarking the performance of higher education institutions. The objective of this paper is precisely to contribute in this matter by providing an alternative approach to the available rankings, in order to quantify and analyze the performance of students in national exit exams. Following literature, and in contrast to previous studies, the results of standardized exams should not be analyzed as a measure of educational quality in higher education institutions. According to Becker (1997), education is a multi-product

Table 7. ATT Estimation Test Scores

	Nearest Neighbor Matching	Kernel Matching (bootstrapped)	Caliper Matching
Treated	426	426	407
Control	971	13162	12890
ATT	5.863	7.929	9.051
Standard Error	0.796	0.429	0.378
t - value	7.370	18.493	23.951

Source: Author's calculations using the ICFES dataset

output which cannot be reflected in a single multiple-choice test score. But there are alternative ways to analyze the results. Popham (1999) showed that standardized test results are a good source of information to make comparisons of students' specific skills at the national level. Therefore, these results should be seen as a measure of the general basic skills knowledge that is necessary for graduate school. By supporting that approach, it follows that students who attended *Universidad Javeriana* are better prepared for graduate school than equivalent students who attended other institutions. In contrast to what previous studies have shown using rankings derived from the simple comparison of mean test scores, students from *Universidad Javeriana* have been found to perform better than equivalent students from other institutions.

Although a first check for consistency on the PSM methodology results had already been undertaken using alternative matching approaches, an additional set of checks for robustness was also employed, namely a straightforward ordinary least squares (OLS) regression and a fixed effects regression. The

output of those regressions is included in appendix 2.

Table 8 presents the results of all the different approaches used to assess the impact of studying Business Administration at *Universidad Javeriana* on the results of the ECAES exam. Consistent with the PSM results, the average impact of studying at Javeriana University is 6.312 points for the least squares regression and 5.42 points for the fixed effects regression.

The fact that the results in all the different approaches explored in this paper are similar is a good indicator of result reliability. The contrasts observed between the results of the kernel and caliper matching algorithms are related to differences in the way the pairing is done in each of them.

Given the estimation problems that might come up in each of these approaches, double-checking the robustness of the results is very important, as far as it allows claiming more confidently that the effect of the intervention is significant, and that these results are suitable for policy decisions.

Table 8. ATT Estimation Test Scores

	Nearest Neighbor Matching	Kernel Matching (bootstrapped)	Caliper Matching	Least Squares Regression	Fixed Effects
Treated	426	426	407	424	424
Control	971	13162	12890	13,094	13,094
Average Impact	5.863	7.929	9.051	6.312	5.42
Standard Error	0.796	0.429	0.378	0.396	0.668
t - value	7.370	18.493	23.951	15.94	8.11

Source: Author's calculations using the ICFES dataset

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

221

7. Heterogeneous Program Impacts

Due to its important policy implications, the specific impact of an education at *Universidad Javeriana* on variables such as gender and socioeconomic stratum is particularly interesting, all the more in light of the particular nature of the data analyzed in this paper. In effect, an education at this institution seems to reduce gender and income inequality amongst the treated individuals. According

to Khandker *et al.* (2010), there are several ways to present the distributional impacts of a program, depending on the interests of policy makers. In this case, by running fixed effects regressions and interpreting the coefficients of interacting variables, it was possible to assess impact heterogeneity (see Appendix 3 for detailed results and Table 9 for a summarized version). The default category is Socioeconomic Stratum 6. *Specification (1)* (presented in column 1) estimates heteroge-

Table 9. Heterogeneity of the Impact

VARIABLES	(1)	(2)
	Score in the ECAES Exam	Score in the ECAES Exam
Student studied at Universidad Javeriana (treatment)	6.179***	5.886***
	(0.916)	(0.774)
Interaction Stratum 1 and Javeriana		12.07***
		(4.567)
Interaction Stratum 2 and Javeriana		10.57***
		(2.665)
Interaction Stratum 3 and Javeriana		9.528***
		(2.233)
Interaction Stratum 4 and Javeriana		4.989**
		(2.214)
Interaction Stratum 5 and Javeriana		3.957*
		(2.398)
Interaction Males and Javeriana	-1.728	
	(1.263)	
Observations	11,691	11,691
R-squared	0.097	0.1
Number of Schools	4,018	4,018

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations using the ICFES dataset

neous impact by gender; and *specification (2)* (presented in column 2) estimates heterogeneous impact by socioeconomic stratum. Ranging from 6.179 to 5.886 points in both specifications, the impact of attending *Universidad Javeriana* on the performance in the ECAES exam is very significant.

The sign and magnitude of the heterogeneous impact coefficient by gender can be further interpreted. According to these results, attending *Universidad Javeriana* has a positive effect of around 1.728 points on the performance of female students in the ECAES exam. Although the results are not significant, they could have important policy implications in terms of the impact of attending *Universidad Javeriana* on gender inequality reduction in education. This finding should be explored in more detail in future research.

Similarly, the heterogeneous impact results by socioeconomic stratum are quite interesting for policy makers, inasmuch as they indicate that the students from the lowest socioeconomic strata are the ones that benefit the most from an education at *Universidad Javeriana* in terms of the results in the ECAES exam. On average, a student from the lowest socioeconomic stratum (1) has an increase of 12.07 points in the score of the exam by attending this institution. The results show that as the household income level increases, the effect of attending *Universidad Javeriana* on the ECAES exam scores gets smaller, but it is still positive. This graded benefit suggests that the educational gap with the students from higher strata might be closing because of this intervention. The specialists in charge of social policy should further explore this

finding because it opens the debate about the need to implement policies serving the educational needs of students from different socioeconomic strata.

8. Conclusions and Policy Implications

Higher education rankings have been criticized on many fronts, but their publication keeps attracting a lot of attention and generating both academic and non-academic discussion. Determining objective criteria to measure the quality of higher education institutions and designing the basis on which they should be evaluated is a very difficult task, if not impossible (Tambi *et al.* 2008). The literature has shown that standardized tests (*e.g.*, exit exams) are a good source of information to compare students' specific skills nationally, but they are not a good measure of quality (Popham, 1999 and Card and Krueger, 1994). The objective of this paper was to build on literature and develop an alternative empirical and analytical approach to analyze the results of the ECAES exam. PSM is employed to quantify and compare the performance of *Universidad Javeriana*'s Business Administration students to that of equivalent students in the ECAES exam. Since the construction of a strong counterfactual relies on observable characteristics, the existence of a rich database, which has been made available by ICFES, is crucial for the implementation of this econometric strategy.

The results show a strong treatment effect (ranging from 5.863 to 9.051 points) of attending *Universidad Javeriana* on the performance in the ECAES exam. The students

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

223

of this institution attain higher scores in the ECAES exam than those exhibiting similar observable characteristics. Hence, the naïve approach, which ranked Javeriana ninth, is likely to be misrepresenting results instead of being a good source of information for the design of educational policy. In this respect, ranking results should be dismissed as tools of analysis.

The robustness of the average treatment effect estimated in this paper was checked by comparing the PSM results to those obtained using other methodologies, namely least squares and fixed effects regressions, among other techniques and algorithms, all of which rendered very robust results. In fact, the outcome obtained through the fixed effects regression was very similar to the one attained by the nearest neighbor algorithm (5.42 and 5.863, respectively). Similarly, the least squares regression shows that the effect of attending *Universidad Javeriana* on the ECAES result is of around 6.312 points.

Given its important policy implications, the specific impact of an education at *Universidad Javeriana* on variables like gender and socioeconomic stratum is particularly interesting, especially in face of the nature of the data analyzed in this paper. Significant differences in the impact of the treatment across different socioeconomic strata suggest that students from lower levels benefit the most from an education at *Universidad Javeriana*. Attending this institution also has a positive effect of around 1.728 points on the performance of female students in the ECAES exam. All this indicates that an education at

Universidad Javeriana can have a positive social impact in terms of reducing income and gender inequality.

It is important to have a better understanding of where a university stands in terms of preparing students for further education. The approach used in this paper can be easily extended to the results of other subject tests. Given the high level of competitiveness currently featuring graduate school admission processes, examining such results is extremely relevant.

The results of this paper have important research and policy implications. The simple comparison of mean test scores between cohorts of students of different universities may be a biased measure of the true impact of the education they have received. This is so because the two groups may be very different in their observable and unobservable characteristics. Future research efforts should not only take this into account, but they should resort to appropriate econometric approaches too. It should also be noted that the results of standardized exams are not necessarily a good measure of educational quality. Standardized tests seem to obviate enormous amounts of student knowledge and skills that are finally uncovered and untested. In fact, it has been shown in the literature that those students that are highly knowledgeable in subjects like economics had already acquired such high aptitude (Becker, 1997). The current educational quality ranking analyses, which are built from the simple comparison of mean test scores, may be misleading because of their erroneous interpretation, and might therefore be misinforming policy makers. It

is suggested then to dismiss those rankings for policy-making purposes.

A possible weakness of this research is that information for both the Saber11 and ECAES exams only covers a limited time span. A balanced panel data set would have made the analysis much stronger, but unfortunately such information is not available. It is advisable to explore this research idea later on, when a new panel data set becomes available.

For future research purposes, it would be interesting to use the methodology applied in this paper to analyze the results of other universities in the ECAES exam (currently SaberPro). Doing so would help university officials and policy makers get a better understanding of where other universities stand. Taking that into consideration, it would be relevant to carry out new research in order to develop an alternative ranking system that is capable of exploiting the strengths of the methodology used in this paper.

References

Aghion, P., Boustan, L., Hoxby, C. and Vandenbussche, J. (2005). *Exploiting states' mistakes to identify the causal impact of higher education on growth*. Cambridge, MA: Harvard University.

Allcott, H. and Ortega, D. (2009). *The performance of decentralized school systems: Evidence from Fe y Alegría in Venezuela*. World Bank Policy Research Working Paper Series.

Asociación Colombiana de Facultades de Administración – ASCOLFA (2006). *Examen de estado para la calidad de la educación – ECAES en administración*. Retrieved on April 10, 2012, from: http://www.ascolfa.edu.co/archivos/ECAES_2006_-_ANALISIS_DE_RESULTADOS_-_ASCOLFA.pdf

Becker, W. E. (1997). Teaching economics to undergraduates. *Journal of Economic Literature*, 35 (3), 1347-1373.

Buela-Casal, G., Gutiérrez-Martínez, O., Bermúdez-Sánchez, M. P. and Vadillo-Muñoz, O. (2007). Comparative study of international academic rankings of universities. *Scientometrics*, 71 (3), 349-365.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22 (1), 31-72.

Card, D., and Krueger, A. (1994). *The economic return to school quality: A partial survey*. Industrial Relations Section Working Paper 334: Princeton University.

Carey, K. (2006). *College rankings reformed: The case for a new order in higher education*. Education Sector. Retrieved from http://www.educationsector.org/publications/college-rankings-reformed

Dehejia, R. H. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84 (1), 151-161.

Harvey, L. (2008). Rankings of higher education institutions: A critical review. *Quality in Higher Education*, 14 (3), 187-207.

Heinrich, C., Maffioli, A. and Vázquez, G. (2010). *A primer for applying propensity-score matching*. Working paper 1005. Inter-American Development Bank, Office of Strategic Planning and Development Effectiveness (SPD).

Instituto Colombiano para la Evaluación de la Educación (ICFES) (2011). *ECAES and Saber11 Databases*. Programa de Fomento a la Investigación.

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

225

Kennedy, P. (2003). *A guide to econometrics*. The MIT press. 5th Edition.

Popham, W. J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56, 8-16.

Revista Dinero (2007). *Universidades ¿dan la talla?* Retrieved on March 15, 2012, from: http://www.dinero.com/caratula/edicion-impresa/articulo/universidades-dan-talla/32928

Saavedra, J. E. (2007). *Selective universities and skill acquisition: Evidence from Colombia*. Working Paper. Harvard University.

Saavedra, J. E. (2009). *The learning and early labor market effects of college quality: A regression discontinuity analysis*. Universidad de los Andes, School of Government.

Tambi, A. M. B. A., Ghazali, M. C. and Yahya, N. B. (2008). The ranking of higher education institutions: A deduction or delusion?. *Total Quality Management and Business Excellence*, 19 (10), 997-1011

Tofallis, C. (2012). A different approach to university rankings. *Higher Education*, 63 (1), 1-18.

Torres, O. (2007). *Panel data analysis: Fixed and random effects*. Data and Statistical Services, Princeton University. Retrieved on April 10, 2012, from: http://www.princeton.edu/~otorres/Panel101.pdf

Universidad Javeriana (2011). *Programa de administración de empresas*. Retrieved on March 15, 2012, from: http://www.javeriana.edu.co/fcea/

Yaffee, R. (2003). *A primer for panel data analysis*. Connect - Information Technology at New York University. Retrieved on April 10, 2012, from: http://www.nyu.edu/its/pubs/connect/fall03/pdfs/yaffee_primer.pdf

Appendix 1. Joint Test of Significance-Household Strata and Parental Education

Joint Test of Significance - Strata and Father's Education		
Level of Education	Stratum	p-score
Level of education of the student's father - Level 1	Stratum 1	0.0026
Level of education of the student's father - Level 1	Stratum 2	0.003
Level of education of the student's father - Level 1	Stratum 3	0.0311
Level of education of the student's father - Level 1	Stratum 4	0.6221
Level of education of the student's father - Level 1	Stratum 5	0.6841
Level of education of the student's father - Level 2	Stratum 1	0.003
Level of education of the student's father - Level 2	Stratum 2	0.0031
Level of education of the student's father - Level 2	Stratum 3	0.0325
Level of education of the student's father - Level 2	Stratum 4	0.6427
Level of education of the student's father - Level 2	Stratum 5	0.7066
Level of education of the student's father - Level 3	Stratum 1	0.0028
Level of education of the student's father - Level 3	Stratum 2	0.0031
Level of education of the student's father - Level 3	Stratum 3	0.0332
Level of education of the student's father - Level 3	Stratum 4	0.6679
Level of education of the student's father - Level 3	Stratum 5	0.7333
Level of education of the student's father - Level 4	Stratum 1	0.003
Level of education of the student's father - Level 4	Stratum 2	0.0032
Level of education of the student's father - Level 4	Stratum 3	0.034
Level of education of the student's father - Level 4	Stratum 4	0.6753
Level of education of the student's father - Level 4	Stratum 5	0.7421
Level of education of the student's father - Level 5	Stratum 1	0.0029
Level of education of the student's father - Level 5	Stratum 2	0.0029
Level of education of the student's father - Level 5	Stratum 3	0.0297
Level of education of the student's father - Level 5	Stratum 4	0.5864
Level of education of the student's father - Level 5	Stratum 5	0.6443
Level of education of the student's father - Level 6	Stratum 1	0.0026
Level of education of the student's father - Level 6	Stratum 2	0.003
Level of education of the student's father - Level 6	Stratum 3	0.0324

Joint Test of Significance - Strata and Father's Education		
Level of Education	Stratum	p-score
Level of education of the student's father - Level 6	Stratum 4	0.6519
Level of education of the student's father - Level 6	Stratum 5	0.7157
Level of education of the student's father - Level 7	Stratum 1	0.002
Level of education of the student's father - Level 7	Stratum 2	0.0022
Level of education of the student's father - Level 7	Stratum 3	0.0258
Level of education of the student's father - Level 7	Stratum 4	0.5765
Level of education of the student's father - Level 7	Stratum 5	0.6306
Level of education of the student's father - Level 8	Stratum 1	0.0004
Level of education of the student's father - Level 8	Stratum 2	0.0004
Level of education of the student's father - Level 8	Stratum 3	0.0056
Level of education of the student's father - Level 8	Stratum 4	0.1971
Level of education of the student's father - Level 8	Stratum 5	0.2081

Joint Test of Significance - Strata and Mother's Education		
Level of Education	Stratum	p-score
Level of education of the student's mother - Level 1	Stratum 1	0.0003
Level of education of the student's mother - Level 1	Stratum 2	0.0003
Level of education of the student's mother - Level 1	Stratum 3	0.0038
Level of education of the student's mother - Level 1	Stratum 4	0.1007
Level of education of the student's mother - Level 1	Stratum 5	0.1076
Level of education of the student's mother - Level 2	Stratum 1	0.003
Level of education of the student's mother - Level 2	Stratum 2	0.003
Level of education of the student's mother - Level 2	Stratum 3	0.033
Level of education of the student's mother - Level 2	Stratum 4	0.6818
Level of education of the student's mother - Level 2	Stratum 5	0.7489
Level of education of the student's mother - Level 3	Stratum 1	0.0013
Level of education of the student's mother - Level 3	Stratum 2	0.0011
Level of education of the student's mother - Level 3	Stratum 3	0.0125
Level of education of the student's mother - Level 3	Stratum 4	0.3549
Level of education of the student's mother - Level 3	Stratum 5	0.3775
Level of education of the student's mother - Level 4	Stratum 1	0.0027

Joint Test of Significance - Strata and Mother's Education		
Level of Education	Stratum	p-score
Level of education of the student's mother - Level 4	Stratum 2	0.0028
Level of education of the student's mother - Level 4	Stratum 3	0.0289
Level of education of the student's mother - Level 4	Stratum 4	0.6333
Level of education of the student's mother - Level 4	Stratum 5	0.6917
Level of education of the student's mother - Level 5	Stratum 1	0.0026
Level of education of the student's mother - Level 5	Stratum 2	0.0027
Level of education of the student's mother - Level 5	Stratum 3	0.0293
Level of education of the student's mother - Level 5	Stratum 4	0.657
Level of education of the student's mother - Level 5	Stratum 5	0.7142
Level of education of the student's mother - Level 6	Stratum 1	0.0008
Level of education of the student's mother - Level 6	Stratum 2	0.0007
Level of education of the student's mother - Level 6	Stratum 3	0.0078
Level of education of the student's mother - Level 6	Stratum 4	0.2338
Level of education of the student's mother - Level 6	Stratum 5	0.2463
Level of education of the student's mother - Level 7	Stratum 1	0.0025
Level of education of the student's mother - Level 7	Stratum 2	0.0027
Level of education of the student's mother - Level 7	Stratum 3	0.027
Level of education of the student's mother - Level 7	Stratum 4	0.5665
Level of education of the student's mother - Level 7	Stratum 5	0.62
Level of education of the student's father - Level 8	Stratum 1	0.002
Level of education of the student's father - Level 8	Stratum 2	0.0022
Level of education of the student's father - Level 8	Stratum 3	0.0234
Level of education of the student's father - Level 8	Stratum 4	0.3781
Level of education of the student's father - Level 8	Stratum 5	0.4231

Source: Author's calculations using the ICFES dataset

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

229

Appendix 2. OLS and Fixed Effects Regression Output

Least Squares Regression - ECAES Score	
Variables	Model
stratum 1	-10.39***
	(0.568)
stratum 2	-7.27***
	(0.475)
stratum 3	-6.34***
	(0.459)
stratum 4	-4.32***
	(0.485)
stratum 5	-2.29***
	(0.528)
saber11_95	0.09
	(1.142)
saber11_96	-0.35
	(1.023)
saber11_97	-1.07
	(0.98)
saber11_98	-0.58
	(0.951)
saber11_99	-1.88
	(1)
saber11_00	-2.12*
	(0.93)
saber11_01	-1.1
	(0.926)
saber11_02	-0.01
	(0.954)
saber11_03	1.29
	(0.97)

Least Squares Regression - ECAES Score	
Variables	Model
saber11_04	-3.07**
	(1.17)
saber11_05	-0.79
	(1.022)
saber11_06	-2.60*
	(1.061)
saber11_07	-2.19*
	(1.086)
saber11_09	1.75
	(3.855)
mathscore	-0.12***
	(0.02)
coresubjects	0.15***
	(0.019)
Level of education of the student's father - Level 1	0.57
	(0.641)
Level of education of the student's father - Level 2	-0.32
	(0.632)
Level of education of the student's father - Level 3	0.5
	(0.625)
Level of education of the student's father - Level 4	-0.3
	(0.608)
Level of education of the student's father - Level 5	0.37
	(0.668)
Level of education of the student's father - Level 6	0.56
	(0.632)
Level of education of the student's father - Level 7	2.64***
	(0.711)
Level of education of the student's father - Level 8	1.78**
	(0.663)

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

231

Least Squares Regression - ECAES Score	
Variables	Model
Level of education of the student's father - Level 9	3.79***
	(0.765)
Level of education of the student's mother - Level 1	0.93
	(0.807)
Level of education of the student's mother - Level 2	1.22
	(0.8)
Level of education of the student's mother - Level 3	2.61***
	(0.79)
Level of education of the student's mother - Level 4	1.23
	(0.779)
Level of education of the student's mother - Level 5	1.96*
	(0.833)
Level of education of the student's mother - Level 6	2.36***
	(0.794)
Level of education of the student's mother - Level 7	3.02***
	(0.867)
Level of education of the student's mother - Level 8	3.24***
	(0.849)
Level of education of the student's mother - Level 9	
Jesuit	1.85**
	(0.609)
Javeriana	6.32***
	(0.396)
Observations	13,518
R-squared	0.12
Adj. R-squared	0.12

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Source: Author's calculations using the ICFES dataset

Fixed Effects Regression - ECAES Score	
School Fixed Effects	
Variables	Model
stratum 1	-8.05***
	(0.746)
stratum 2	-5.85***
	(0.585)
stratum 3	-5.42***
	(0.549)
stratum 4	-3.40***
	(0.564)
stratum 5	-1.62**
	(0.596)
saber11_95	-0.89
	(1.601)
saber11_96	-0.06
	(1.424)
saber11_97	-0.77
	(1.374)
saber11_98	-0.71
	(1.309)
saber11_99	-1.01
	(1.366)
saber11_00	-1.56
	(1.264)
saber11_01	-0.68
	(1.261)
saber11_02	0.2
	(1.297)
saber11_03	0.81
	(1.318)
saber11_04	-2.45
	(1.479)

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

233

Fixed Effects Regression - ECAES Score	
School Fixed Effects	
Variables	Model
saber11_05	-0.82
	(1.344)
saber11_06	-1.92
	(1.173)
saber11_07	-1.03
	(1.477)
saber11_09	4.12
	(5.452)
mathscore	-0.11***
	(0.023)
coresubjects	0.16***
	(0.021)
Level of education of the student's father - Level 1	-0.17
	(0.933)
Level of education of the student's father - Level 2	-0.88
	(0.917)
Level of education of the student's father - Level 3	0.46
	(0.899)
Level of education of the student's father - Level 4	-0.29
	(0.876)
Level of education of the student's father - Level 5	0.02
	(0.953)
Level of education of the student's father - Level 6	0.51
	(0.899)
Level of education of the student's father - Level 7	2.64**
	(0.967)
Level of education of the student's father - Level 8	1.48
	(0.928)
Level of education of the student's father - Level 9	4.16***
	(1.024)

Fixed Effects Regression - ECAES Score	
School Fixed Effects	
Variables	Model
Level of education of the student's mother - Level 1	1.47
	(1.122)
Level of education of the student's mother - Level 2	1.37
	(1.1)
Level of education of the student's mother - Level 3	2.13*
	(1.075)
Level of education of the student's mother - Level 4	1.04
	(1.063)
Level of education of the student's mother - Level 5	1.86
	(1.133)
Level of education of the student's mother - Level 6	2.04
	(1.075)
Level of education of the student's mother - Level 7	2.40*
	(1.149)
Level of education of the student's mother - Level 8	2.79*
	(1.125)
Level of education of the student's mother - Level 9	2.34
	(1.218)
Jesuit	-1.19
	(2.171)
Javeriana	5.42***
	(0.668)
Observations	11,623
Number of cole_codigo_colegio	3,997
R-squared	0.09

Robust standard errors in parentheses

*** $p<0.001$, ** $p<0.01$, * $p<0.05$

Source: Author's calculations using the ICFES dataset

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

235

Appendix 3. Heterogeneous Effects Regression Output

Variables	(1)	(2)
	Score in the ECAES Exam	Score in the ECAES Exam
Student studied at Universidad Javeriana (treatment)	6.179***	5.886***
	(0.916)	(0.774)
Graduated from a Jesuit High School	-1.326	0.328
	(2.136)	(2.159)
Level of education of the student's mother - Level 1	1.378	1.353
	(1.114)	(1.112)
Level of education of the student's mother - Level 2	1.435	1.387
	(1.095)	(1.094)
Level of education of the student's mother - Level 3	2.151**	2.100**
	(1.072)	(1.07)
Level of education of the student's mother - Level 4	0.995	0.945
	(1.059)	(1.058)
Level of education of the student's mother - Level 5	1.974*	1.894*
	(1.128)	(1.127)
Level of education of the student's mother - Level 6	2.013*	1.964*
	(1.071)	(1.07)
Level of education of the student's mother - Level 7	2.360**	2.348**
	(1.145)	(1.143)
Level of education of the student's mother - Level 8	2.245*	2.058*
	(1.214)	(1.213)
Level of education of the student's father - Level 1	-0.00937	-0.0429
	(0.925)	(0.924)
Level of education of the student's father - Level 2	-0.92	-0.926
	(0.909)	(0.908)
Level of education of the student's father - Level 3	0.373	0.375
	(0.892)	(0.891)

Variables	(1)	(2)
	Score in the ECAES Exam	Score in the ECAES Exam
Level of education of the student's father - Level 4	-0.382	-0.379
	(0.87)	(0.869)
Level of education of the student's father - Level 5	-0.026	-0.0000362
	(0.946)	(0.944)
Level of education of the student's father - Level 6	0.485	0.444
	(0.893)	(0.892)
Level of education of the student's father - Level 7	2.542***	2.517***
	(0.961)	(0.96)
Level of education of the student's father - Level 8	4.097***	4.230***
	(1.017)	(1.017)
Average score in main components of Saber11 exam (other than math)	0.166***	0.165***
	(0.0213)	(0.0213)
Math score in the Saber11 exam	-0.112***	-0.111***
	(0.0225)	(0.0224)
Student took the Saber11 Exam 95	-0.348	-0.322
	(1.203)	(1.201)
Student took the Saber11 Exam 96	0.453	0.408
	(0.973)	(0.972)
Student took the Saber11 Exam 97	-0.234	-0.243
	(0.901)	(0.9)
Student took the Saber11 Exam 98	-0.21	-0.2
	(0.803)	(0.802)
Student took the Saber11 Exam 99	-0.549	-0.56
	(0.886)	(0.885)
Student took the Saber11 Exam 00	-1.076	-1.079
	(0.694)	(0.693)
Student took the Saber11 Exam 01	-0.186	-0.182
	(0.686)	(0.685)

Variables	(1)	(2)
	Score in the ECAES Exam	Score in the ECAES Exam
Student took the Saber11 Exam 02	0.701	0.712
	(0.774)	(0.773)
Student took the Saber11 Exam 03	1.359*	1.393*
	(0.815)	(0.814)
Student took the Saber11 Exam 04	-2.346**	-2.357**
	(1.045)	(1.044)
Student took the Saber11 Exam 05	-0.642	-0.563
	(0.809)	(0.808)
Student Socieconomic Stratum (1)	-7.899***	-8.337***
	(0.735)	(0.744)
Student Socieconomic Stratum (2)	-5.649***	-6.073***
	(0.584)	(0.594)
Student Socieconomic Stratum (3)	-5.253***	-5.682***
	(0.548)	(0.559)
Student Socieconomic Stratum (4)	-3.319***	-3.575***
	(0.562)	(0.577)
Student Socieconomic Stratum (5)	-1.549***	-1.768***
	(0.594)	(0.613)
Interaction Stratum 1 and Javeriana		12.07***
		(4.567)
Interaction Stratum 2 and Javeriana		10.57***
		(2.665)
Interaction Stratum 3 and Javeriana		9.528***
		(2.233)
Interaction Stratum 4 and Javeriana		4.989**
		(2.214)
Interaction Stratum 5 and Javeriana		3.957*
		(2.398)

Variables	(1)	(2)
	Score in the ECAES Exam	Score in the ECAES Exam
Gender		1.697***
		(0.261)
Interaction Males and Javeriana	-1.728	
	(1.263)	
Males	1.766***	
	(0.265)	
Observations	11691	11691
R-squared	0.097	0.1
Number of Schools	4018	4018

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations using the ICFES dataset

Joint Test of Significance Javeriana (Treatment) and Socioeconomic Strata		
Stratum	F (2, 7629)	p-score
Stratum 1	63.18	0.000
Stratum 2	52.87	0.000
Stratum 3	52.52	0.000
Stratum 4	19.31	0.000
Stratum 5	4.16	0.0156
Joint Test of Significance Javeriana (Treatment) and Gender (Male)		
	F (2, 7633)	p-score
Male	41.05	0.000

Source: Author's calculations using the ICFES dataset

Cuad. admon.ser.organ. Bogotá (Colombia), 26 (46): 201-239, enero-junio de 2013

239