Universidad del Atlántico

MATUA

Revista MATUA Páginas: 68–80

Facultad de Ciencias Básicas
©Programa de Matemáticas
Vol. VII , Nº 1, (2020)

Received 16 Sept 2020; Approved 21 Nov 2020

Un Diseño Factorial Aplicado a las Pruebas Saber Pro de Estudiantes de Matemáticas

A Factorial Design Applied to Saber Pro Test of Math Students

Greisy Paola Morillo

Universidad Nacional Evangélica, Recinto Santiago, Santiago de los Caballeros, República Dominicana

gmorillo@unev.edu.do

Resumen

El Valor Agregado o Efecto Escuela, ver [3], es útil para determinar que tanto puede aportar una institución en el buen desempeño de un estudiante. Se hizo el estudio utilizando la base de datos de la prueba Saber Pro realizada por los estudiantes de Matemáticas y áreas afines en el año 2013, ver [5, 6] y la técnica estadística Diseño de Experimentos, ver [2] y ver [4]. El estudio arroja conclusiones interesantes que permitirán a los directivos de las escuelas y colegios potenciar a los futuros matemáticos.

Palabras claves: Análisis factorial, efecto escuela, programas de matemáticas, Saber 11, Saber Pro

Abstract

The school effect, see [3], is useful to determine how an institution can help to reach the good development of students. I is realised an study using the data base of Saber Pro test made by the students of math and closer areas during the year 2013, see [5, 6], as well the statistical techniques know as experimental design, see [2] and also [4]. Interesting results and conclusiones are obtained that will allow to principals of school and colleges to a good performance to future mathematicians.The school effect, see [3], is useful to determine how an institution can help to reach the good development of students. I is realised an study using the data base of Saber Pro test made by the students of math and closer areas during the year 2013, see [5, 6], as well the statistical techniques know as experimental design, see [2] and also [4]. Interesting results and conclusiones are obtained that will allow to principals of school and colleges to a good performance to future mathematicians.

Keywords: School effect, Saber 11 test, Saber Pro test, bachelor in mathematics, factorial analysis.

Introduction

This quasi-experimental research article corresponds to a statistical study using design of experiments, specifically the technique known as Factor Analysis, see[1].

Statement and justification of the problem

According to the databases of the Saber 11 and Saber Pro exams, studying Mathematics (or related) is not very popular among students who are finishing high school in Colombia. In these databases it is evident that studying mathematics, except in exceptional cases, is a second option.

Little or almost nothing has been done to study the population of Students of Mathematics in Colombia, which is very small compared to other careers such as engineering a, medicine, administration, etc.

Studying the school effect, also known as added value (the school of origin affects the future of students) in this population through the use of design of experiments will allow to advise future students .

Delimitation of the Problem

It is ambitious to study all the factors of the school that can affect the students of Mathematics of Colombia. In this investigation *Saber Pro* contribute to the response variables and the factors are variables of*Saber 11*.

In this article, the score obtained in the Saber Pro tests is considered as a dependent variable (response). As factors to analyze if there is a school effect, the city of presentation and the score obtained in Exam Language are taken Saber 11.

Research question

The problem (or research question) is to determine if there is evidence of a school effect or added value in the students of Matematicas de Barranquilla, Bogota, Cali, Cartagena and Medellin who presented the Saber Pro tests in 2013 and Saber 11 tests from 2006 to 2009.

1. Methodology and Factorial Design

ICFES provided the databases of the results of the Saber Pro tests from 2013 and the Saber 11 tests from 2006 to 2009.

The database was manually built that related the students who presented Saber Pro with their respective results in Saber 11.

The database was refined so that the study was carried out on a group of 144 math students from Barranquilla, Bogota, Cali, Cartagena and Medellin. , who presented the Saber Pro tests in 2013 and the Saber 11 tests from 2006 to 2009 in those same cities.

A two-factor factorial design with a single response was applied. The factors correspond to city and language score in the Saber 11 tests, while the response variable is the score of the Saber Pro tests.

Language Score	City				
	Barranquilla	Bogota	Cali	Cartagena	Medellin
<50	10,45	10,82	10,88	9,75	10,67
50 – 60	10,75	11,16	10,58	10,33	10,8
>60	11,04	11,85	11,59	10,04	11,18

Figura 1. Table of cities and scores in language

As levels of the city factor for the presentation of the Saber 11 tests, the cities Barranquilla, Bogota, Cali, Cartagena and Medellin were taken.

Scores less than 50 points, scores between 50 and 60 points, and scores greater than 60 points were taken as levels of the score factor in Language of the Saber 11 tests. The averages by level of each factor were calculated.

The analysis of the factorial design, including the non-additivity test, was carried out first in the Excel program and then, without considering interactions, we worked in SPSS.

2. Results and discussions

This section presents the results of the article and also in the discussions concerning the results.

2.1. Data table and model

The model that we present in this article is the following

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ij}, \tag{1}$$

where y_{ij} is the Saber Pro score on the level i Language score, level j from the city; μ is the global mean; τ_i is the effect i Language score; β_j is the effect j from the city; $(\tau\beta)_{ij}$ is the effect of the interaction at the level i of the Language score and level j from the city.

2.2. Assumptions

The assumptions that are assumed are the following:

$$\sum_{i=1}^{3} \tau_i = 0 \tag{2}$$

$$\sum_{j=1}^{5} \beta_j = 0 \tag{3}$$

$$\sum_{i=1}^{3} (\tau\beta)_{ij} = 0 \tag{4}$$

$$\sum_{j=1}^{5}(\tau\beta)_{ij} = 0 \tag{5}$$

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \quad \text{independent} \tag{6}$$

2.3. *Null hypothesis and alternative hypothesis*

The following notation will be used:

- H_0 is the null hypothesis
- H_1 is the alternate hypothesis

In this way, the following is proposed:

Hypothesis for the first factor:

$H_0 : \tau_i = 0, \forall i \in \{1, 2, 3\}$. Language score levels have the same effect.

$H_1 : \exists k \in \{1, 2, 3\}$ such that $\tau_k \neq 0$. At least one Language score level has a different effect.

Hypothesis for the second factor

$H_0 : \beta_j = 0, \forall j \in \{1, 2, 3, 4, 5\}$. City levels have the same effect.

$H_1 : \exists k \in \{1, 2, 3, 4, 5\}$ such that $\beta_k \neq 0$. At least one city level has a different effect.

Hypothesis for the inter-action of factors

$H_0 : (\tau\beta)_{ij} = 0, \forall ij \in \{1, 2, 3\} \times \{1, 2, 3, 4, 5\}$. Language interactions $*$ City have the same effect.

$H_1 : \exists k\ell \in \{1, 2, 3\} \times \{1, 2, 3, 4, 5\}$ such that $(\tau\beta)_{k\ell} \neq 0$. At least one of the Language interactions $*$ City has a different effect.

2.4. *Analysis with Excel*

Using Excel y [2, §5] we have:

$a = 3$	$b = 5$	$N = 15$	$y_{1.} = 52{,}57$
$y_{2.} = 53{,}62$	$y_{3.} = 55{,}70$	$y_{.1} = 32{,}24$	$y_{.2} = 33{,}83$
$y_{.3} = 33{,}05$	$y_{.4} = 30{,}12$	$y_{.5} = 32{,}75$	$y_{..} = 161{,}89$
$\bar{y}_{..} = 10{,}7927$	$SS_T = 4{,}1555$	$SS_A = 1{,}0151$	$SS_B = 2{,}5838$
$SS_R = 0{,}5566$	$SS_N = 0{,}1908$	$SS_E = 0{,}3658$	$MS_A = 0{,}5075$
$MS_B = 0{,}6460$	$MS_R = 0{,}0696$	$MS_N = 0{,}1908$	$MS_E = 0{,}0523$
$F_{0A} = 9{,}7113$	$F_{0B} = 12{,}3601$	$F_{0R} = 1{,}3313$	$F_{0N} = 3{,}6506$
$PV_A = 0{,}0096$	$PV_B = 0{,}0027$	$PV_R = 0{,}3595$	$PV_N = 0{,}0977$

According to the above results it is observed:

The language score level in Saber 11 ($PValue < \alpha$) and the city's presentation of Saber 11 ($PValue < \alpha$) affect scores earned in Saber Pro by Math students.

The Language*City relationship ($PValue > \alpha$) does not affect the score in Saber Pro math students. Therefore we can take the intersection of the model and repeat the experiment in SPSS.

The non-additive test ($PValue > \alpha$) tells us that there is insufficient evidence to assert that there is an interaction between the data, which is corroborated in the previous item. Therefore, the following formula is proposed as a new model.

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}. \tag{7}$$

EVIDENCE OF INTER-SUBJECT EFFECTS					
DEPENDENT VARIABLE: SCORE					
Origen	**SUM OF SQUARES**	**gl**	**Quadratic Mean**	**F0**	**P_VALUE**
Language	1.0151	2	0.5075	9.7113	**0.0096**
City	2.5838	4	0.6460	12.3601	**0.0027**
Language*City	0.5566	8	0.0696	1.3313	0.3595
Non-Addition	0.1908	1	0.1908	3.6506	0.0977
Error	0.3658	7	0.0523		
Total	4.1555				

Figura 2. Table of the effects in question

Descriptive statistics

Depent variable

Language		**Mean**	**typical deviation**	N
<50	Barranquilla	10.4500		1
	Bogota	10.8200		1
	Cali	10.8800		1
	Cartagena	9.7500		1
	Medellin	10.6700		1
	Total	10.5140	.45818	5
50-60	Barranquilla	10.7500		1
	Bogota	11.1600		1
	Cali	10.5800		1
	Cartagena	10.3300		1
	Medellin	10.8000		1
	Total	10.7240	.30517	5
>60	Barranquilla	11.0400		1
	Bogota	11.8500		1
	Cali	11.5900		1
	Cartagena	10.0400		1
	Medellin	11.1800		1
	Total	11.1400	.69430	5
Total	Barranquilla	10.7467	.29501	3
	Bogota	11.2767	.52482	3
	Cali	11.0167	.51868	3
	Cartagena	10.0400	.29000	3
	Medellin	10.8833	.26502	3
	Total	10.7927	.54481	15

Figura 3. Descriptive Statistics Table of Language vs City

Dependent variable score — **Test inter subjects effects**

Source	Square sum type III	gl	Quadratic Mean	F	Sig.	Partial Square Ela
Modelo	1750,824[a]	7	250.118	3594.850	.000	1.000
Lenguaje	1.015	2	.508	7.294	.016	.646
Ciudad	2.584	4	.646	9.284	.004	.823
Error	.557	8	.070			
Total	1751.380	15				

a. R cuadrado = 1.000 (R cuadrado corregida = .999)

Figura 4. Inter-subject effects table

Inter subject factors

		Label value	N
Language	1	<50	5
	2	50-60	5
	3	>60	5
City	1	Barranquilla	3
	2	Bogota	3
	3	Cali	3
	4	Cartagena	3
	5	Medellin	3

Figura 5. Table of inter-subject factors

Dependente variable	Scores			
Language	**Mean**	**Tip. Error**	**Conf. interval 95%**	
			Inf. limit	**Sup. limit**
<50	10.514	.118	10.242	10.786
50-60	10.724	.118	10.452	10.996
>60	11.140	.118	10.868	11.412
City	**Mean**	**Tip. Error**	**Conf. interval 95%**	
			Inf. limit	**Sup. limit**
Barranquilla	10.747	.152	10.395	11.098
Bogota	11.277	.152	10.925	11.628
Cali	11.017	.152	10.665	11.368
Cartagena	10.040	.152	9.689	10.391
Medellin	10.883	.152	10.532	11.235

Figura 6. Table of descriptive statistics of the score

Scores				
Language			**Subset**	
		N	1	2
Student-Newman-Keuls[a,b]	<50	5	10.5140	
	50-60	5	10.7240	
	>60	5		11.1400
	Sig.		.244	1.000
Scheffe[a,b]	<50	5	10.5140	
	50-60	5	10.7240	10.7240
	>60	5		11.1400
	Sig.		.485	.100

Homogeneous subsets
alpha = .05

Figura 7. Scorecard with different tests

Scores

City		N	Subset 1	Subset 2
Student-Newman-Keuls[a,b]	Cartagena	3	10.0400	
	Barranquilla	3		10.7467
	Medellin	3		10.8833
	Cali	3		11.0167
	Bogota	3		11.2767
	Sig.		1.000	.142
Scheffe[a,b]	Cartagena	3	10.0400	
	Barranquilla	3	10.7467	10.7467
	Medellin	3	10.8833	10.8833
	Cali	3		11.0167
	Bogota	3		11.2767
	Sig.		.050	.286

Means for homogeneous subsets **Harmonic Mean = 3.0**
Alpha = .05

Scores **Multiple comparison**

(I) Language			Difference means i-j	Tip Error	Sig.	Conf. interval 95% Inf. lim	Sup. lim
Scheffe	<50	50-60	-.2100	.16683	.485	-.7082	.2882
		>60	-,6260*	.16683	.017	-1.1242	-.1278
	50-60	<50	.2100	.16683	.485	-.2882	.7082
		>60	-.4160	.16683	.100	-.9142	.0822
	>60	<50	,6260*	.16683	.017	.1278	1.1242
		50-60	.4160	.16683	.100	-.0822	.9142

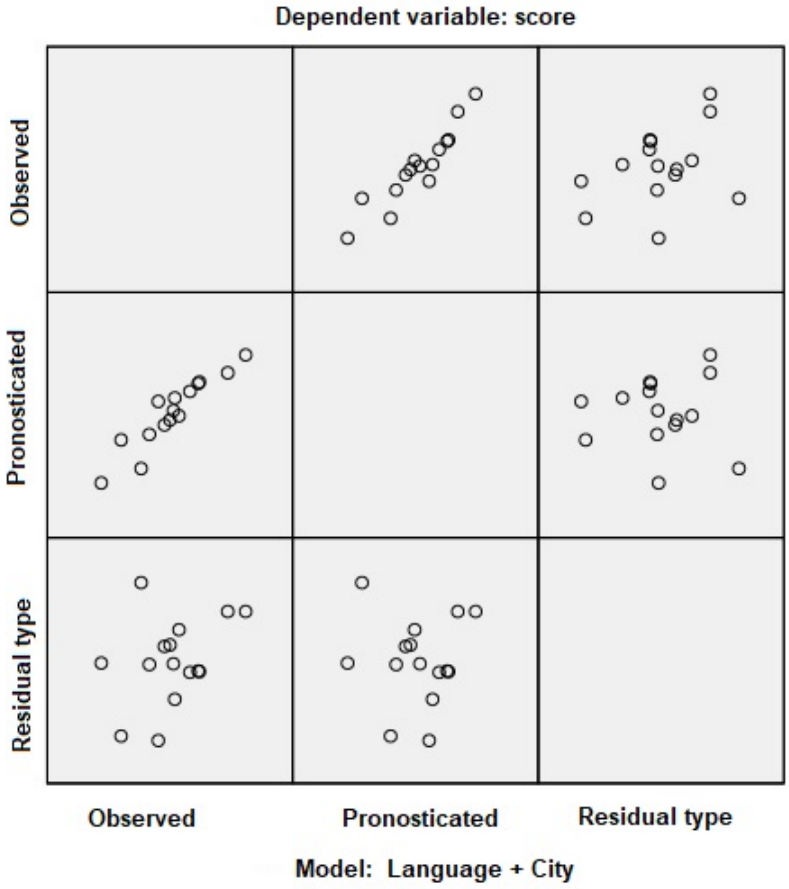
Observed means **Differences of means is significant** **alpha = .05**

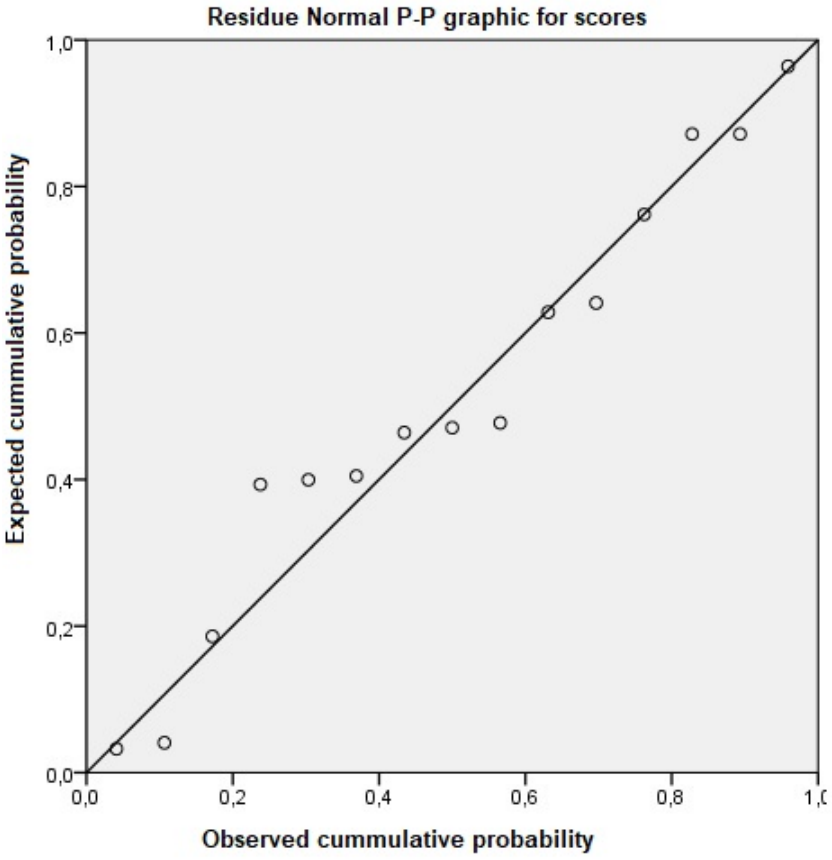
Multiple comparisons

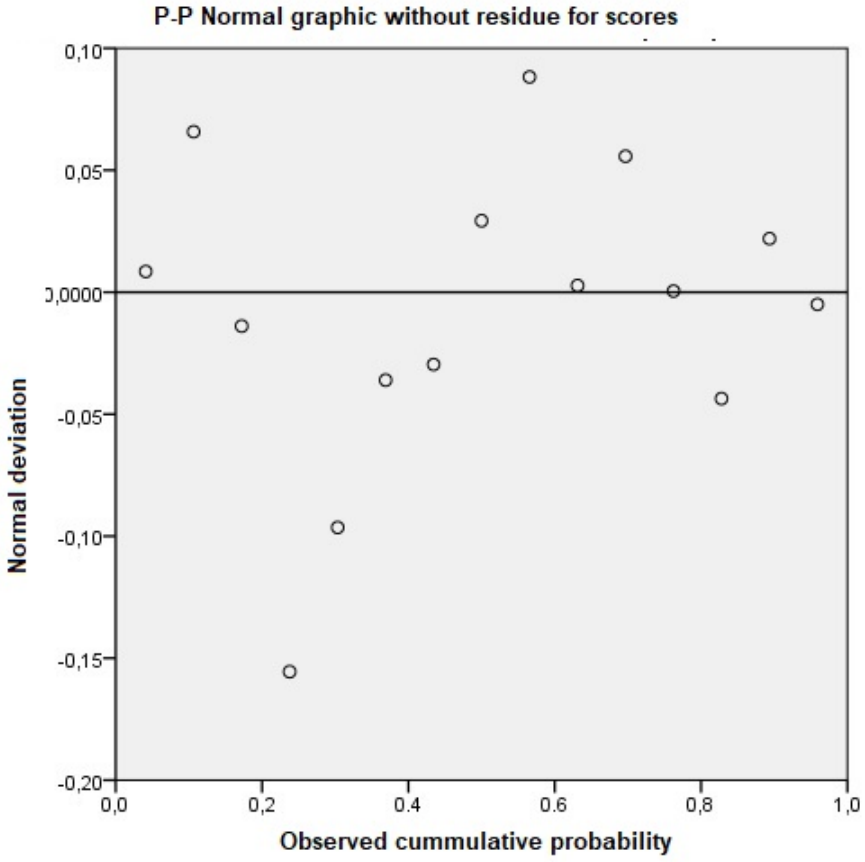
Dependent variable score

(I) City			Difference of means i-j	Tip. error	Sig.	Conf. interval 95% Inf. lim	Sup. lim
Scheffe	Barranquilla	Bogota	-.5300	.21537	.286	-1.3738	.3138
		Cali	-.2700	.21537	.809	-1.1138	.5738
		Cartagena	.7067	.21537	.109	-.1372	1.5505
		Medellin	-.1367	.21537	.979	-.9805	.7072
	Bogota	Barranquilla	.5300	.21537	.286	-.3138	1.3738
		Cali	.2600	.21537	.828	-.5838	1.1038
		Cartagena	1,2367*	.21537	.006	.3928	2.0805
		Medellin	.3933	.21537	.540	-.4505	1.2372
	Cali	Barranquilla	.2700	.21537	.809	-.5738	1.1138
		Bogota	-.2600	.21537	.828	-1.1038	.5838
		Cartagena	,9767*	.21537	.024	.1328	1.8205
		Medellin	.1333	.21537	.981	-.7105	.9772
	Cartagena	Barranquilla	-.7067	.21537	.109	-1.5505	.1372
		Bogota	-1,2367*	.21537	.006	-2.0805	-.3928
		Cali	-,9767*	.21537	.024	-1.8205	-.1328
		Medellin	-.8433	.21537	.050	-1.6872	.0005
	Medellin	Barranquilla	.1367	.21537	.979	-.7072	.9805
		Bogota	-.3933	.21537	.540	-1.2372	.4505
		Cali	-.1333	.21537	.981	-.9772	.7105
		Cartagena	.8433	.21537	.050	-.0005	1.6872

Based on observed means **Difference of means is significative with alpha=.05**


Dependent variable: score
Observed
Pronosticated
Residual type
Observed
Pronosticated
Residual type
Model: Language + City


Residue Normal P-P graphic for scores
Expected cummulative probability
Observed cummulative probability
0,0
0,2
0,4
0,6
0,8
1,0


P-P Normal graphic without residue for scores
0,10
0,05
0,0000
-0,05
-0,10
-0,15
-0,20
Normal deviation
0,0
0,2
0,4
0,6
0,8
1,0
Observed cummulative probability

2.5. Analysis with SPSS

According to the tables and graphs it is observed:

- Cartagena students scored the lowest score in Saber Pro at each level of the Language score in Saber 11.
- Medellin students and Barranquilla students had very similar scores in Saber Pro at each level of the Language score in Saber 11, but it wasn't the highest scores.
- Bogota students and Cali students had very similar scores in Saber Pro at each level of the Language score in Saber 11, and their scores were the highest.
- Both the levels of the score in Language and in the city of presentation of Saber 11 affect the score in Saber Pro.

It is seen in the table of multiple comparisons of the language factor, applying Scheffe test, there are significant differences in test scores saber pro among students who scored under 50 and over 60 in the language test saber 11. In the multiple comparison table the city factor applying Scheffe test, there are significant differences in test scores saber pro between students from Cartagena and students from Bogota. In the multiple comparison table the city factor applying Scheffe test, there are significant differences in test scores Saber Pro between students from Cartagena and students from Cali. In the predicted graph Vs Observed we see that the data observed with the forecasts are similar forming a trend on the right identity. In the PP Normal graph of residue for score we see atypical data but also a tendency towards the correct identity so we could say that the assumption of normality is met. In the Normal PP chart with no residue trend for scoring there are points above and below zero which are scattered.

3. Conclusions

This investigation proved the following facts:

1. The score of the tests Saber Pro Math students are affected by the test language score Saber 11.
2. The score of the tests Saber Pro matematics students is affected by the city of presentation of the tests Saber 11.

This indicates that there is evidence of school effect or added value in the math students who took the tests Saber Pro in the year 2013 and the evidence Saber 11 between the year 2006 and the year 2009.

Acknowledgements

To God for all the blessings He has given me; to my husband for helping me and advising me on the writing of this article; professors Jesus Alonso and Martin Diaz for their teachings during my master's studies in Estaística Aplicada at the Universidad del Norte and finally the editors of Matua Magazine for considering this article for publication.

Referencias

[1] Ruiz Escorcia, R. R., Arévalo Medrano, J. B., Morillo, G. P., Acosta-Humánez, P. B. (2018). Análisis de componentes principales aplicado a la prueba estatal Colombiana Saber 11. Revista Espacios, **39**, No. 10, 1–12

[2] Montgomery, D.C., *Diseño y Análisis de Experimentos*, Segunda Edición, Limusa-Wiley, México D.F., 2004

[3] R. Cervini, *Los efectos de la escuela y del aula sobre el logro en matemáticas y en lengua de la educación secundaria. Un modelo multinivel*, Perfiles educativos **28** (2006) # 112

[4] E. Gutiérrez Moya, *Una herramienta didáctica para el aprendizaje de las técnicas de diseño estadístico de experimentos*, Revista de Enseñanza Universitaria, **27** (2006), 23–31

[5] ICFES, *Resultados Saber 11*, 2006 – 2009. www.icfes.gov.co

[6] ICFES, *Resultados Saber Pro*, 2013. www.icfes.gov.co