

# Análisis secuencial de los resultados de la Evaluación Censal de Estudiantes

Roger Asencios Núñez

Encuentro de Economistas del BCRP, 2014

# Esquema

## Motivación

Rendimientos en la prueba censal de estudiantes.  
Trabajos Previos

## Nuestro Trabajo

Los Datos  
Metodología  
Principales Resultados

# Esquema

## Motivación

Rendimientos en la prueba censal de estudiantes.  
Trabajos Previos

## Nuestro Trabajo

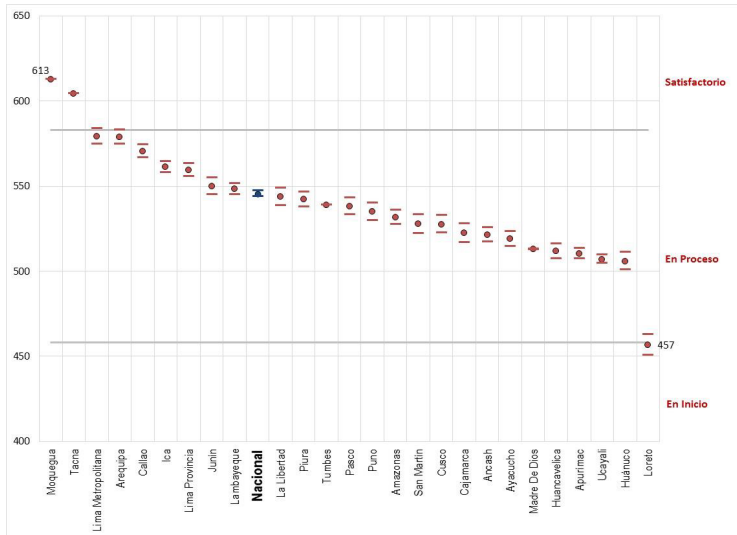
Los Datos  
Metodología  
Principales Resultados

## Una gran proporción de los alumnos no aprueban los exámenes.

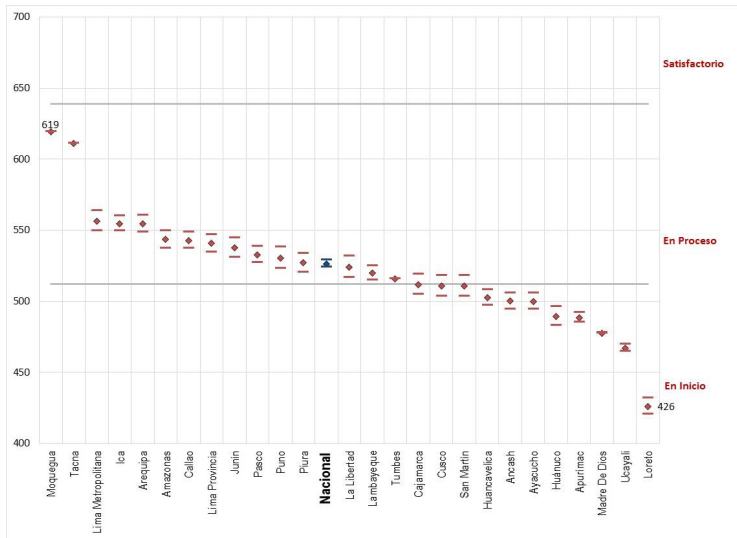
### RESULTADOS EN LA ECE (En porcentajes)

	Matemáticas		C. Lectora	
	2012	2013	2012	2013
Satisfactorio	12.8	16.8	30.9	33.0
En proceso	38.2	32.8	49.3	51.3
En inicio	49.0	50.9	19.8	15.8

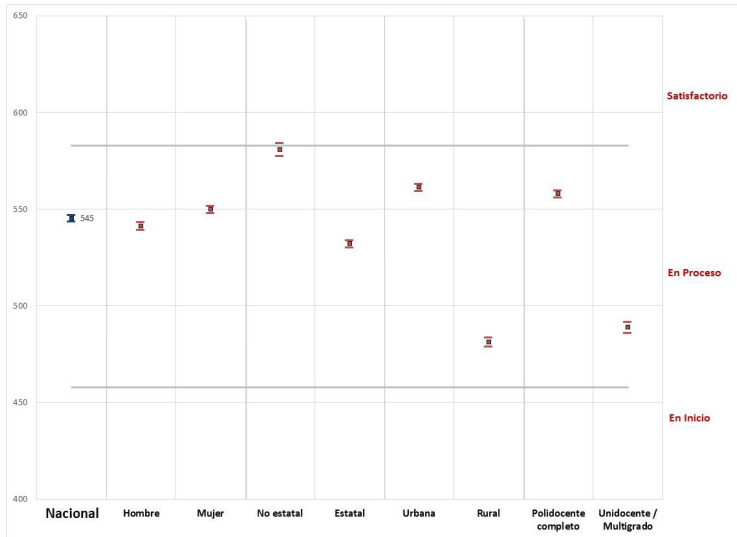
## Los resultados varían según departamentos.



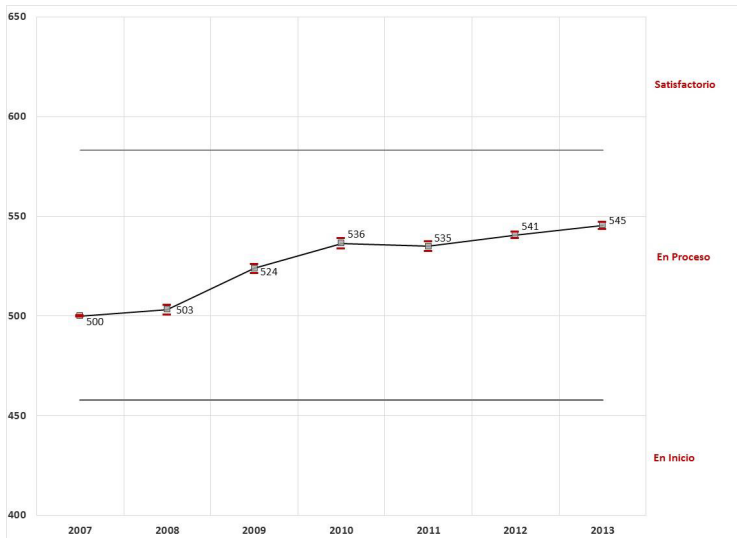
## Los resultados varían según departamentos.



## También son diferentes según características.

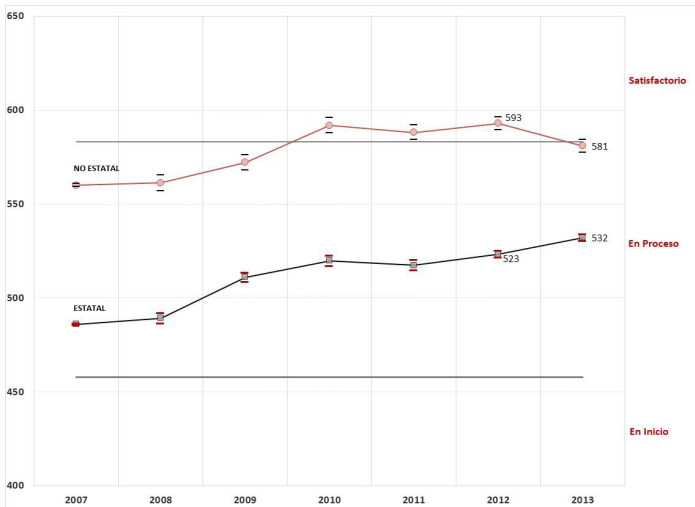


## Los resultados han ido mejorando.





## La mejora ha permitido reducir la diferencia.





# Esquema

## Motivación

Rendimientos en la prueba censal de estudiantes.  
Trabajos Previos

## Nuestro Trabajo

Los Datos  
Metodología  
Principales Resultados

## Factores asociados.

Uribe (2004) y Beltrán y Seinfeld (2012) han resaltado la existencia de factores asociados al rendimiento escolar que se pueden agrupar en:

- ▶ Factores asociados al estudiante, como la lengua materna, peso, talla, sexo, asistencia a la educación inicial, edad y otros.
- ▶ Factores asociados a la familia: nivel socioeconómico, características del hogar (luz, agua, desagüe en casa, número de miembros del hogar y el capital cultural (nivel educativo de los padres, tener libros en casa)

## Factores asociados.

- ▶ Factores asociados al centro educativo, como características de los docentes (título y años de experiencia) y las características del centro educativo (polidocente, computadoras, conexión a internet, estado de las aulas, electricidad, y otras)

Coleman (1966) encuentra que los factores socioeconómicos tienen mayor preponderancia que los factores escolares. No obstante, varios estudios han encontrado luego que los factores de la escuela tienen una mayor importancia, en especial en países en vías de desarrollo (Banco Mundial 1999).

## Factores asociados.

Jacocy (1999) resalta que existe evidencia que los factores asociados al rendimiento en lengua estarían más influenciados por el ambiente familiar, mientras que el rendimiento en matemática podría ser más influenciado por la escuela.

## Factores asociados.

Harbison y Hanushek (1992) subdividen los factores del lado de la oferta en tres categorías:

- ▶ Hardware: asociado a la infraestructura del centro escolar
- ▶ Software, asociado con los insumos que intervienen en el proceso de enseñanza, como material educativo, currículo, entre otros.
- ▶ Docentes, asociado al capital humano que está relacionado con el proceso de aprendizaje.

# Esquema

## Motivación

Rendimientos en la prueba censal de estudiantes.  
Trabajos Previos

## Nuestro Trabajo

Los Datos

Metodología

Principales Resultados



## Base de Datos.

Obtener los datos necesarios para el trabajo implicó enlazar tres diferentes bases de datos.

- ▶ Evaluación Censal de Estudiantes 2010 (ECE 2010)
- ▶ Encuesta Nacional de Hogares ENAHO 2011 -2012
- ▶ Censo Escolar 2011

## Base de Datos.

- ▶ ENAHO, se consideró la información proveniente del año 2011 (Segundo Trimestre al Cuatro Trimestre) y 2012 (Primer Trimestre), puesto que la recolección de información de la pregunta 318 “¿Ha participado en la evaluación censal de estudiantes realizada por el Ministerio de Educación en diciembre del 2010?”, se refiere al ECE 2010.
- ▶ El procedimiento para juntar las bases de datos constó en enlazar en base a: Apellido Paterno, Apellido Materno, Nombres y Distrito Idénticos. En cuanto se agotaron las coincidencias, se pasó a buscar Apellido Paterno, Apellido Materno y Nombres tomando en cuenta la presencia de diferencias en algún carácter, dentro del mismo distrito.

## Base de Datos.

- ▶ Finalmente para esta base de 1 396 individuos, se agregó información del Censo Escolar 2011 realizado por la Unidad de Estadística del Ministerio de Educación. El procedimiento consistió en usar la variable código modular, que identifica el colegio al cual pertenecen los individuos de la muestra, para adjuntar las variables de características de los docentes e infraestructura del colegio.

## Muestra en la ECE y la ENAHO.

<b>TIPO DE COINCIDENCIAS</b>	<b>Personas</b>	<b>%</b>
<b>Total personas en la Encuesta</b>	<b>2503</b>	<b>100</b>
<b>Total coincidencias</b>	<b>1396</b>	<b>55.8</b>
Coincidencias a nivel de Distrito	1188	47.5
Coincidencias a nivel de Departamento	156	6.2
Coincidencias Inter Departamentos	52	2.1
<b>No coinciden</b>	<b>1107</b>	<b>44.2</b>

Table : Descripción de la muestra.

	Comprensión Lectora			Total
	En inicio	En proceso	Satisfactorio	
Pobre Multidimensional	73.8	48.6	29.3	50.1
Pobre Monetario	61.3	40.9	24.3	41.9
Padre con solo primaria	65.4	43.2	28.2	45.0
Castellano	77.0	91.2	97.1	89.1
Biblioteca	40.5	61.8	69.9	58.6
Laboratorio	16.4	31.4	41.0	30.1
Colegio privado	4.7	13.2	28.2	14.8
Colegio Polidocente	53.5	79.4	90.3	75.7
Area rural	72.7	40.9	22.0	44.1
N observaciones	344	704	341	1389

Table : Descripción de la muestra.

	Lógico Matemático			
	En inicio	En proceso	Satisfactorio	Total
Pobre Multidimensional	57.7	43.7	36.0	50.1
Pobre Monetario	47.8	36.6	32.0	41.9
Padre con solo primaria	49.8	43.3	29.8	45.0
Castellano	86.3	91.2	95.5	89.1
Biblioteca	53.9	61.2	70.5	58.6
Laboratorio	26.5	31.3	41.6	30.1
Colegio privado	10.9	17.0	24.7	14.8
Colegio Polidocente	70.3	80.0	86.0	75.7
Area rural	52.4	37.0	29.2	44.1
N observaciones	344	704	341	1389

# Esquema

## Motivación

Rendimientos en la prueba censal de estudiantes.  
Trabajos Previos

## Nuestro Trabajo

Los Datos  
**Metodología**  
Principales Resultados

## Metodología.

- ▶ Por lo general se estiman funciones de producción de logro de aprendizaje en CL y LM para cada niño. El logro de aprendizaje se aproxima con el resultado en la ECE. Los insumos son los factores de oferta y demanda.
- ▶ En este trabajo se usa el enfoque de logit secuencial, donde la variable a explicar depende de un evento anterior. Por ejemplo, en un proceso de tres resultados la primera transición consiste en la probabilidad de pasar al nivel A o al B. La segunda transición, nos da la probabilidad de pasar a B o C para aquellos que pasaron el nivel A.



## Función de producción.

$$ECE_{n,2010} = F \left( X_n^d, X_n^o, \mu \right)$$

Donde:

- ▶  $ECE_{n,2010}$ : Puntaje obtenido por el estudiante n en la ECE
- ▶  $X_n^d$ : Factores de demanda asociados al estudiante "n", que incluye variables como lengua materna, nivel de educación del jefe de hogar.
- ▶  $X_n^o$ : Factores de oferta asociados al colegio al que asiste el estudiante "n", como la calidad de la infraestructura, acceso a internet, biblioteca, docente calificados, y otros.
- ▶  $\mu$ : Componente no observable.

## Logit secuencial.

Para cada nivel de decisión se asume que existe una variable latente  $y^*$  definida por la ecuación:

$$y_i^* = \beta' x_i + \mu_i$$

Donde:

- ▶  $y_i^*$  es no observable y depende del índice lineal  $\beta' x_i = E(y_i^* | x_i)$
- ▶  $x_i$  es el vector de características observables del individuo  $i$ .
- ▶  $\mu_i$  es el vector de características no observables o término de error del individuo  $i$ .

Lo que se observa es una variable dicotómica y que se define como:

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases}$$

Entonces, la variable latente  $y_i^*$  es la propensión a que el evento, representado por la variable  $y_i$ , ocurra.

Si  $y_i = 1$ , en términos de probabilidades se tiene:

$$Pr(y_i = 1/x_i) = Pr(y_i^* > 0/x_i)$$

$$Pr(y_i = 1/x_i) = Pr(\mu_i > -\beta'x_i/x_i)$$

$$Pr(y_i = 1/x_i) = F(\beta'x_i)$$

El modelo secuencial se construye a partir de este mismo modelo, repetido tantas veces como secuencias, eventos o niveles de decisión se quieran incluir. El análisis de cada nivel de decisión es condicional al anterior y sólo se practica para observaciones que, al menos, hayan cumplido satisfactoriamente el nivel de decisión que le antecede.

- ▶  $y_1 = 1$  si el individuo alcanzo el nivel 0 en la ECE
- ▶  $y_2 = 1$  si el individuo alcanzo el nivel 1 dado que paso el nivel 0.
- ▶  $y_3 = 1$  si el individuo alcanzo el nivel 2 dado que paso el nivel 1.

En términos de probabilidades se tiene:

$$Pr(y_1 = 1/x_1) = F(\beta'_1 x_1)$$

$$Pr(y_2 = 1/x_2, y_1 = 1) = F(\beta'_2 x_2)$$

$$Pr(y_3 = 1/x_3, y_2 = 1) = F(\beta'_3 x_3)$$

O en forma general:

$$Pr(y_j = 1/x_j, y_{j-1} = 1) = F(\beta'_j x_j)$$

donde  $j = 1, 2$  o  $3$

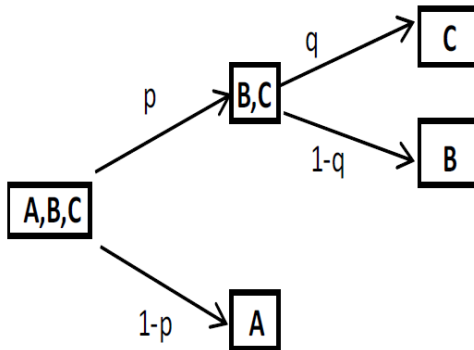
La formulación del índice lineal  $\beta_j'x_j$  tiene en cuenta dos consideraciones:

- ▶ El efecto marginal de un cambio en alguno de las características sobre el índice lineal del nivel de decisión  $j$ , representado por el coeficiente  $\beta_j$ , puede ser distinto al efecto marginal de esas mismas características sobre el índice lineal de otro nivel de decisión distinto de  $j$ .
- ▶ Se permite que en cada secuencia el vector  $x_j$  de atributos del individuo varíe, de acuerdo al nivel de decisión que se trate. Los posibles determinantes de que una persona complete el nivel 1 pueden diferir de los posibles determinantes de que una persona complete el nivel 2.

Habiendo evaluado los distintos niveles de decisión se pueden construir las probabilidades de cada categoría de la siguiente forma:

- ▶ Nivel 0  $[1 - F(\beta'_0 x_0)]$
- ▶ Nivel 1  $F(\beta_0 x_0) [1 - F(\beta'_1 x_1)]$
- ▶ Nivel 2  $F(\beta_0 x_0) F(\beta_1 x_1) [1 - F(\beta'_2 x_2)]$

# Metodología.





## Metodología.

Esta metodología se puede aplicar a los datos de la ECE. En ésta los resultados se presentan en tres niveles (0, 1, 2). Pasar al nivel 1 implica haber completado los mínimos requisitos del nivel anterior, y pasar al nivel 2 implica haber completado tanto los requisitos del nivel 0 como del nivel 1.

Pal (2010) usa un modelo probit/logit secuencial para mostrar que las mismas características individuales y familiares afectan de manera diferente las decisiones de pasar al siguiente nivel de estudios.

# Esquema

## Motivación

Rendimientos en la prueba censal de estudiantes.  
Trabajos Previos

## Nuestro Trabajo

Los Datos  
Metodología  
Principales Resultados

Table : MCGF: Factores Relacionados con el Resultado de la ECE

	LM	CL
Pobre Multidimensional	-18.47*** (-3.322)	-22.53*** (-2.356)
Padre con solo primaria	-5.033 (-3.124)	-5.661* (-2.419)
Lengua	13.59*** (-4.041)	25.50*** (-2.919)
Rural	-18.36*** (-3.296)	-23.57*** (-2.431)
Salón B	12.32*** (-3.63)	9.358*** (-2.806)
Salón C	-5.536 (-4.915)	-3.112 (-3.466)
Salón D	-17.54** (-5.361)	-6.922 (-4.039)
Otros salones	-30.15*** (-8.281)	-16.86* (-7.955)
Biblioteca	12.87*** (-2.824)	15.33*** (-2.185)
Laboratorio	8.075** (-3.046)	5.246* (-2.373)
Aulas	12.46*** (-3.546)	17.24*** (-3.816)
Colegio Privado	29.78*** (-4.787)	34.02*** (-3.327)
Profesores Titulados	24.73** (-7.614)	0.174 (-6.155)
Polidocente Completo	12.76*** (-3.558)	23.83*** (-2.5)
N	1264	1264

Errores estándar en paréntesis  
\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

Table : Resultados Modelo secuencial

	Lógico Matemático		Comprensión Lectora	
	Pasar a 1 o 2	Pasar a 2	Pasar a 1 o 2	Pasar a 2
Pobre Multidimensional	-0.334* (-0.146)	0.116 (-0.233)	-0.514** (-0.186)	-0.405* (-0.18)
Padre con solo primaria	0.0477 (-0.144)	-0.526* (-0.235)	-0.233 (-0.175)	-0.0817 (-0.184)
Salón B	0.142 (-0.166)	0.577* (-0.24)	-0.0333 (-0.231)	0.408* (-0.187)
Salón C	-0.288 (-0.222)	-0.168 (-0.388)	-0.148 (-0.296)	-0.059 (-0.268)
Salón D	-0.39 (-0.288)	-0.349 (-0.502)	-0.248 (-0.408)	-0.288 (-0.346)
Otros salones	-0.513 (-0.32)	0.0289 (-0.528)	-0.736 (-0.399)	0.345 (-0.365)
Lengua	0.209 (-0.21)	0.51 (-0.482)	0.618** (-0.209)	0.744 (-0.426)
Biblioteca	0.197 (-0.135)	0.206 (-0.236)	0.588*** (-0.165)	0.11 (-0.172)
Laboratorio	0.0196 (-0.146)	0.269 (-0.222)	0.00694 (-0.209)	0.0971 (-0.166)
Aulas	0.0323 (-0.225)	0.604 (-0.442)	0.444 (-0.247)	0.117 (-0.334)
Colegio Privado	0.343 (-0.195)	0.452 (-0.277)	0.664* (-0.332)	0.682** (-0.209)
Profesores Titulados	0.25 (-0.37)	1.061 (-0.6)	-0.119 (-0.492)	0.138 (-0.444)
Polidocente Completo	0.112 (-0.172)	-0.00516 (-0.329)	0.415* (-0.186)	0.184 (-0.262)
Rural	-0.474** (-0.159)	0.112 (-0.271)	-0.720*** (-0.199)	-0.337 (-0.203)
N	1264	1264	1264	1264

Errores estándar en paréntesis  
\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

# Resultados

- ▶ Los resultados del MCGF muestran que las variables son significativas, pero tenemos el problema en la modelización de la heterocedasticidad.
- ▶ Se aprecia el efecto distinto según sea el examen tomado.
- ▶ En cuanto al modelo secuencial, parecen existir factores que explican y son significativos para un nivel, pero no para otro. Por ejemplo el nivel educativo del padre, o la condición socioeconómica del hogar.

## Pendientes

- ▶ La muestra puede crecer, INEI cuenta con una base amplia que enlazar.
- ▶ La muestra no es representativa de la población, un mayor  $n$  puede mejorar el resultado de las estimaciones. En este sentido, corregir la estimación por el truncamiento en los datos ayudaría a mejorar los resultados
- ▶ Existen aún bases de datos que enlazar para capturar la heterocedasticidad de la muestra. Por ejemplo, la base de los colegios que son parte del PALE.

## Resumen

- ▶ Las características **afectan de manera diferente** según sea el tipo de examen.
- ▶ Las mismas características que afectan la probabilidad de pasar un nivel **lo hacen de una manera diferente, o no lo hacen** para explicar la probabilidad de pasar otro nivel.
- ▶ Pendiente
  - ▶ Existen unos problemas a corregir con la **heterocedasticidad y posible sesgo**, de la muestra.
  - ▶ Incorporar más observaciones a la muestra, y otros datos que capturen la heterogeneidad de los individuos y colegios.