

**ESTUDIO SOBRE BILINGÜISMO EN LOS NUEVOS EGRESADOS  
COLOMBIANOS**

**JUAN DAVID MARTIN OCAMPO**

**UNIVERSIDAD ICESI  
FACULTAD DE CIENCIAS ADMINISTRATIVAS Y ECONÓMICAS  
PROGRAMA DE ECONOMÍA Y NEGOCIOS INTERNACIONALES  
SANTIAGO DE CALI  
NOVIEMBRE DE 2013**

**ESTUDIO SOBRE BILINGÜISMO EN LOS NUEVOS EGRESADOS  
COLOMBIANOS**

**JUAN DAVID MARTIN OCAMPO**

**PROYECTO DE GRADO II**

**DIRECTOR DE PROYECTO:  
JULIO CÉSAR ALONSO CIFUENTES**

**UNIVERSIDAD ICESI  
FACULTAD DE CIENCIAS ADMINISTRATIVAS Y ECONÓMICAS  
PROGRAMA DE ECONOMÍA Y NEGOCIOS INTERNACIONALES  
SANTIAGO DE CALI  
NOVIEMBRE DE 2013**

# Índice

<b>1. Resumen</b>	<b>4</b>
<b>2. Introducción</b>	<b>5</b>
<b>3. Metodología: La Distribución Relativa</b>	<b>6</b>
3.1. La función de densidad relativa . . . . .	6
3.2. Índices de polarización mediano . . . . .	10
<b>4. Datos y nuestro ejercicio</b>	<b>13</b>
<b>5. Resultados</b>	<b>15</b>
5.1. Análisis según el carácter de la Institución de Educación Superior . . . . .	16
5.2. Análisis según grupo de referencia . . . . .	19
5.2.1. Grupos de referencia por nivel universitario . . . . .	19
5.2.2. Grupos de referencia por nivel técnico profesional . . . . .	26
5.2.3. Grupos de referencia por nivel tecnológico terminal . . . . .	29
<b>6. Conclusiones</b>	<b>35</b>
<b>7. Referencias bibliográficas</b>	<b>36</b>
<b>8. Anexos</b>	<b>37</b>

## 1. Resumen

En el mundo globalizado actual, el bilingüismo se ha convertido en una competencia clave para el éxito profesional. Este estudio analiza la distribución del nivel de inglés como segundo idioma en la población estudiantil colombiana. Para ello se cuenta con los resultados de las pruebas SABER 11 y SABER PRO obtenidos por estudiantes para el término de su respectivo ciclo educativo, utilizando la calificación estandarizada del Marco Común Europeo de Referencia para Lenguas como variable de análisis. A través del uso de los métodos de Distribución Relativa propuestos por Handcock y Morris (1999), se compararon directamente las distribuciones de las dos poblaciones para identificar cambios luego del paso por la educación superior. Los resultados sugieren que, si bien en general, la distribución del nivel de inglés dista mucho de seguir una forma “ideal”, existe una mejora en la distribución, caracterizada particularmente por un mayor impacto hacia los niveles más altos y un aumento de la polarización.

**Palabras clave:** Bilingüismo, Competencias lingüísticas, Distribución Relativa, Desigualdad, Medidas de polarización.

## 2. Introducción

En el mundo globalizado actual, el bilingüismo se ha convertido en una competencia clave para el éxito profesional. Particularmente, el uso del inglés como segundo idioma se ha convertido en un requisito para lidiar con la creciente competitividad del mercado laboral (Alonso *et al*, 2012). En este orden de ideas, pensar en el conocimiento del inglés como un factor influyente en el nivel de capital humano implica una relación positiva entre la distribución del nivel de competencias en esta área y la distribución de los salarios (Mincer, 1974). En este sentido, se pretende establecer un marco más amplio de análisis dirigido a identificar rasgos de la población que contribuyan en materia de información para la adecuada implementación de políticas públicas dirigidas al sector educativo, siguiendo las pautas propuestas por Alonso *et al* (2012) y Sánchez (2013).

De acuerdo con lo anterior, el objetivo de este estudio es explorar la distribución del nivel de inglés como segundo idioma en la población estudiantil colombiana; en particular, analizar cambios en la forma de la distribución del nivel de esta competencia luego del paso por la educación superior e identificar los grupos de la población con menor desarrollo de esta competencia. Para ello se cuenta con los resultados de las pruebas SABER 11 y SABER PRO obtenidos por estudiantes para el término de su respectivo ciclo educativo, utilizando la calificación estandarizada el Marco Común Europeo de Referencia para Lenguas (CEFR) como variable de análisis.

Comúnmente los investigadores se enfrentan al problema de tener que comparar el comportamiento de una población antes y después de una intervención o comparar el comportamiento de dos poblaciones. Lo común es emplear estadísticas descriptivas como la media para comparar el comportamiento de dos poblaciones, en algunas ocasiones el análisis es complementado empleando otras medidas de tendencia central como la mediana o incluso medidas de dispersión como la varianza. No obstante es bien conocido que el uso de una estadística descriptiva como la media esconde mucho del comportamiento de una población. La metodología empleada en este estudio se basa en los conceptos de Distribución Relativa propuestos por Handcock y Morris (1999). Este enfoque propone comparar directamente las distribuciones de las dos poblaciones y no emplear únicamente medidas de tendencia central. Esta herramienta estadística se emplea para representar de forma plena diferencias entre distribuciones, permitiendo un análisis más completo. Otros estudios como Handcock *et al* (1997) han empleado estos métodos para realizar análisis de desigualdad económica a nivel poblacional; algunos incluso enfocándose directamente en la distribución relativa de los salarios, tal como ocurre en Hermeto & Guimaraes (2009).

La estructura de este estudio se presenta en seis secciones incluyendo la presente introducción. A continuación, se presenta formalmente la metodología empleada. En seguida, se describe la naturaleza y bondades de los datos empleados. Acto seguido, los resultados que se dividen en tres tipos de análisis: a nivel agregado de la muestra total; a nivel de agregación por nivel de institución de educación superior (IES); y a nivel de grupos de referencia de áreas de conocimiento al interior del nivel por IES. Finalmente, las conclusiones indican los principales hallazgos y las recomendaciones en materia de política.

### 3. Metodología: La Distribución Relativa

Como se menciona anteriormente, las medidas de tendencia central que tradicionalmente se emplean para detectar cambios en la población suelen ocultar información relevante para ciertos tipos de análisis. Handcock y Morris (1999) proponen comparar directamente las distribuciones de las dos poblaciones y no emplear únicamente medidas de tendencia central. Es decir, la idea es comparar lo que ocurre en toda la distribución de ambas poblaciones. Este método lo denomina Distribución Relativa (DR) y lo sugieren como una alternativa muy conveniente para el análisis de variables propias de las ciencias sociales.

El método de DR es una herramienta estadística que se emplea para representar de forma plena diferencias entre distribuciones. De manera intuitiva, se compara el comportamiento de ambas distribuciones tanto en los extremos como en el centro. Esta herramienta proporciona un marco integrado para el análisis que incluye, entre otros, un componente gráfico que simplifica el análisis exploratorio de los datos. Adicionalmente, este método permite descomponer las diferencias, de tal manera que se puede analizar características más complejas sobre diferencias dentro y entre grupos de las poblaciones de estudio.

La DR es esencialmente una serie de cuantiles que resulta de asignar a una distribución de probabilidad de referencia el nivel que toma la variable de estudio en la población de análisis (comparación). En otras palabras, la DR representa la proporción de individuos de la población de comparación, asociada a un determinado nivel en la variable de estudio seleccionada, en el hipotético caso que dichos individuos se comportaran como la distribución de referencia. O, de forma similar, es la probabilidad de encontrar un individuo con ciertas características de una población de comparación dentro de una población de referencia.

Para nuestro caso, este método no paramétrico permite hacer un análisis comparativo más completo sobre cambios en la distribución del nivel de inglés en la población estudiantil antes y después de su paso por la educación superior. En lo que resta de esta sección se discute de manera formal el método de DR.

#### 3.1. La función de densidad relativa

Antes de continuar, es importante definir formalmente el concepto de *distribución relativa* y *función de densidad relativa* siguiendo de cerca a Handcock y Morris (1999). Sea  $Y_0$  una variable aleatoria generada por la distribución de referencia. Esta variable cuenta con una función de probabilidad acumulada definida por  $F_0(y)$  y una función de densidad  $f_0(y)$ . A estas funciones de probabilidad acumulada y de densidad las llamaremos funciones de referencia.

Por otro lado, sea  $Y$  la misma variable aleatoria pero generada por otro proceso generador de datos cuyas funciones de probabilidad acumulada y de densidad las denotaremos como  $F(y)$  y  $f(y)$ , respectivamente. A estas funciones las llamaremos función de probabilidad acumulada de comparación y función de densidad de comparación. El objetivo del método de DR es determinar si ambas distribuciones de probabilidad acumulada son iguales o no. Es decir, si  $F_0(y)$  es lo suficientemente parecida en su comportamiento a  $F(y)$ , como para poder afirmar que en ambos casos la variable aleatoria  $Y$  es

generada por la misma distribución y por tanto tienen el mismo comportamiento.

Siguiendo a Handcock y Morris (1999), la *distribución relativa* de  $Y$  con respecto a  $Y_0$  se define como la variable aleatoria:

$$R = F_0(Y). \quad (1)$$

Handcock y Morris (1999) demostraron que  $R$  es continua en el espacio  $[0, 1]$  y cuenta con una distribución de probabilidad acumulada,  $G(r)$ , definida por

$$G(r) = F(F_0^{-1}(r)) = F(Q_0(r)), \quad 0 \leq r \leq 1, \quad (2)$$

donde  $Q_0$  es la función cuantil de  $F_0$  y  $r$  representa la realización de  $R$ . En otras palabras, la probabilidad de encontrar un individuo con ciertas características de una población de comparación ( $Y$ ) dentro de una población de referencia ( $F_0$ ).

Finalmente, la *función de densidad relativa* (FDR), es la función de densidad de  $R$ ,  $g(r)$ , definida como la derivada de  $G(r)$ :

$$g(r) = \frac{f(y_r)}{f_0(y_r)}, \quad y_r \equiv Q_0(r) \geq 0. \quad (3)$$

De acuerdo con Handcock y Morris (1999), la FDR puede interpretarse como una razón de densidad: la razón de probabilidades entre la población de comparación y la población de referencia asociada a un determinado nivel en la variable de estudio. No obstante, los autores recalcan que, a diferencia de una razón de probabilidades común,  $g(r)$  tiene la propiedad básica de una función de densidad, en el sentido que su integral en todo el dominio de  $r$  es 1.

La discusión anterior corresponde al caso cuando la variable aleatoria en estudio es continua. No obstante, estos resultados se pueden extender fácilmente al caso de variables discretas (como es el caso en este estudio). Handcock y Morris (1999), demuestran que si se emplea una transformación de la distribución discreta, los resultados se mantienen. De tal manera que *distribución relativa discreta* es continua a pesar de que las distribuciones que la conforman son discretas. Esta versión conserva el concepto fundamental de la DR y mantiene la maleabilidad y la interpretación de la versión continua.

En este caso, sean  $Y_0$  e  $Y$  variables aleatorias discretas que provienen de la población referencia y la de comparación, respectivamente, con un número finito de realizaciones  $N$  definidas dentro del conjunto  $\{y_i\}_{i=1}^N$ . Luego, sean

$$p_{0i} = P(Y_0 = y_i) \quad y \quad p_i = P(Y = y_i), \quad i = 1, 2, \dots, N$$

las funciones de densidad discretas correspondientes. Por lo que la función de probabilidad acumulada de  $Y_0$  es:

$$F_0(y) = \sum_{i: y_i \leq y} p_{0i}(Y_0 = y_i), \quad y \in \mathbb{R}.$$

Dado que  $F_0(y)$  no es necesariamente continua, para obtener una transformación, como la aplicada en (1), que produzca una escala adecuada para la comparación, Handcock y Morris (1999) modifican la función discreta empleando la distribución uniforme. De esta forma, los “huecos” entre los diferentes posibles valores de la distribución se “llenan” con una la distribución uniforme cuyos valores se integran

a uno en el intervalo que antes estaba vacío. En este sentido, la versión continua de la función acumulada de probabilidad discreta de referencia  $F_0^d(y)$  es:

$$F_0^d(y) = U [F_0(y_{i-1}), F_0(y_i)] \quad \text{con} \quad y_{i-1} < y \leq y_i, \quad i = 1, 2 \dots N$$

donde  $y \in \{y_i\}_{i=1}^N$  y  $U [a, b]$  denota una distribución uniforme en el intervalo  $(a, b]$ . Handcock y Morris (1999) demuestran que  $F_0^d$  es una extensión continua de  $F_0$  y que la transformación realizada no afecta para nada las propiedades de la distribución discreta. De esta forma, la *distribución relativa discreta* se define como:

$$R = F_0^d(Y). \tag{4}$$

Como se mencionó anteriormente, en esta versión  $R$  mantiene las mismas propiedades que cuando se emplean directamente variables continuas;  $R$  es continua y está acotada en el intervalo  $[0, 1]$ . De manera análoga al caso continuo, las funciones de probabilidad acumulada,  $G(r)$  y de densidad,  $g(r)$ , están dadas por:

$$G(r) = (r - F_0(x_{i-1})) \frac{p_i}{p_{0i}} + F_0(x_{i-1}), \tag{5}$$

$$g(r) = g(i) = \frac{p_i}{p_{0i}}. \tag{6}$$

Handcock y Morris (1999) aclaran que  $g(i)$  es discreta, pero conserva la cualidad interpretativa de la FDR definida en (3). Una vez se construye una FDR se puede proceder a realizar la comparación de las distribuciones de referencia y de comparación.

De acuerdo con (6) la FDR es esencialmente un ratio entre la probabilidad en la población de comparación y la probabilidad en la población de referencia para un determinado cuantil. Entonces, un determinado valor de  $g(i)$  implicará que la probabilidad de encontrar individuos en la población de comparación con un determinado nivel de la variable de estudio es  $g(i)$  veces la probabilidad correspondiente en la población de referencia. Por lo que si  $g(i) = 1$ , esto indicará que las dos distribuciones coinciden en el cuantil correspondiente;  $g(i) > 1$  implicará que la distribución de probabilidad de la población de comparación es mayor que la de referencia para ese cuantil y si  $g(i) < 1$ , ocurre lo opuesto, es decir, que la distribución de probabilidad de la población de comparación es menor que la de referencia.

Para nuestro caso, la población de comparación corresponde a los resultados de las pruebas SABER PRO y la población de referencia los resultados de SABER 11. En la Figura 1 se ilustra un ejemplo en el que no se presenta diferencia alguna entre las distribuciones, para el cual la FDR toma una forma aplanada, igual a 1. Por otro lado, en la Figura 2 se presenta el caso cuando existe un cambio “favorable”, donde la FDR toma una forma creciente, indicando proporción de individuos menor en niveles bajos y mayor en niveles altos. Mientras que en la Figura 3 ocurre el caso contrario, la distribución “empeora” y la FDR tiene una forma decreciente.



Figura 1: Ejemplo sin cambio

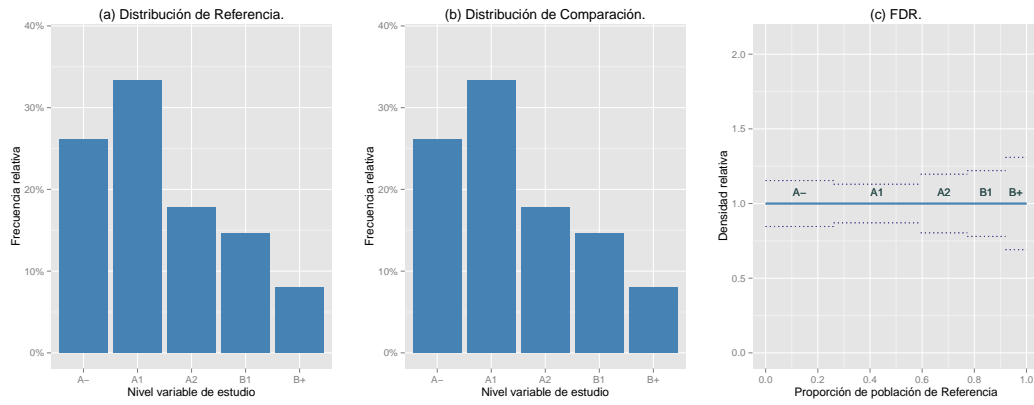


Figura 2: Ejemplo con cambio favorable

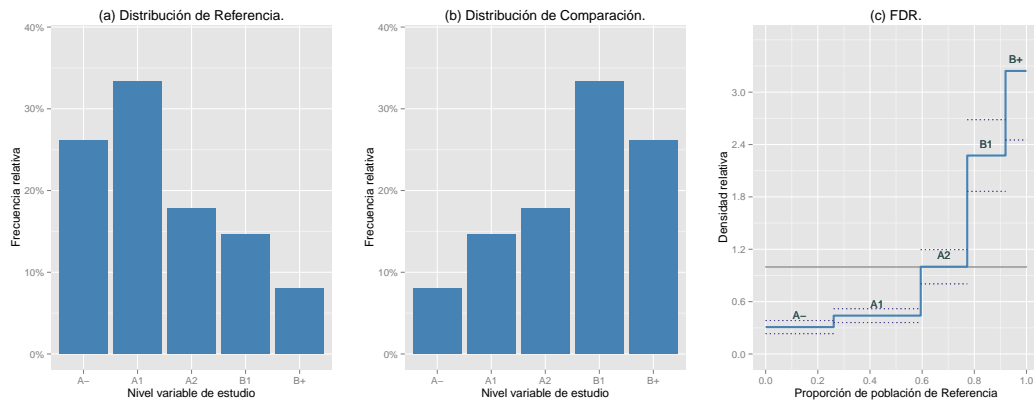
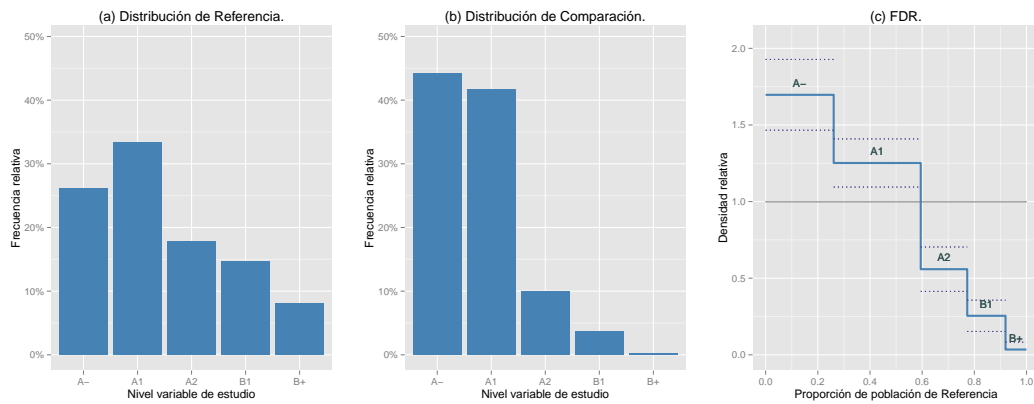


Figura 3: Ejemplo con cambio desfavorable



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Antes de continuar, es importante reconocer que en la práctica es usual que se desconozcan ambas distribuciones, por lo que será necesario estimarlas de alguna manera. Dadas unas muestras de tamaño  $n$  para la población de referencia y otra de tamaño  $m$  para la de comparación, Handcock y Morris (1999) demuestran que un estimador,  $\hat{g}(i)$ , insesgado para  $g(i)$  se define como:

$$\hat{g}_{m,n}(i) = \frac{\hat{p}_i}{\hat{p}_{0i}} \quad i = 1, 2, \dots, N, \quad (7)$$

donde

$$\hat{p}_{0i} = \frac{1}{n} \sum_{l:y_l=y_i}^n y_l, \quad \forall y_l \in Y_0,$$

$$\hat{p}_i = \frac{1}{m} \sum_{l:y_l=y_i}^m y_l, \quad \forall y_l \in Y.$$

En particular, los autores demuestran que  $\hat{g}(i)$  sigue una distribución asintóticamente normal

$$\hat{g}_{m,n}(i) \sim^a N(g(i), \text{VAR}(\hat{g}(i))) \quad (8)$$

Para nuestro caso, (7) permite estimar de manera sencilla la FDR. Función que indicará la diferencia proporcional entre la probabilidad de la distribución de comparación (SABER PRO) y la probabilidad de la distribución de referencia (SABER 11) asociada al  $i$ -ésimo cuantil de la distribución de referencia. Además (8) permite construir intervalos de confianza para los valores estimados de  $g(i)$ . De esta forma se contará con una herramienta gráfica para determinar si las dos distribuciones se parecen o no. Y en caso de no parecerse, dónde se ha realizado el cambio, si en la parte superior o inferior de la distribución.

### 3.2. Índices de polarización mediano

La FDR permite responder la pregunta ¿qué tan iguales son dos distribuciones? No obstante, en algunos casos será importante responder preguntas más relacionadas con la forma de la distribución, como por ejemplo si una distribución es más igualitaria que otra. En nuestro caso, si, después de la educación superior, la distribución de los niveles de inglés se sigue polarizando o por el contrario, ésta es más igualitaria.

Las medidas polarización típicamente permiten establecer niveles de desigualdad entre poblaciones. Estas medidas están diseñadas para captar discrepancias a lo largo de toda la distribución, no solamente respecto a una característica de ésta. Handcock y Morris (1999) desarrollaron un indicador de polarización que se ajusta al concepto de DR, que permite examinar las similitudes o diferencias entre las distribuciones de referencia y de comparación. En especial, el índice de polarización que sugieren los autores permite distinguir polarizaciones en la distribución (cambios en ambas colas de la distribución, aumentos de los individuos en la cola superior de la distribución o en la cola inferior.

Como lo discuten Handcock y Morris (1999) una medida de polarización idealmente debería medir las desviaciones de la forma de la distribución de una población con respecto a una distribución uniforme; y en nuestro caso una distribución relativa uniforme. Consideremos la distribución relativa de  $Y_L$  con

respecto a  $Y_0$  (denotada por  $R_{0L}$ ), donde  $Y_L$  corresponde a  $Y$  ajustada a la mediana de  $Y_0$ . En otras palabras

$$R_{0L} = F_0(Y - \rho)$$

con

$$\rho = Q\left(\frac{1}{2}\right) - Q_0\left(\frac{1}{2}\right).$$

Entonces, Handcock y Morris (1999) definen el índice de polarización mediano, o MRP (por la sigla en inglés del término Median Relative Polarization), de  $Y$  como

$$\text{MRP}(F, F_0) = 4E \left[ \left| R_{0L} - \frac{1}{2} \right| \right] - 1. \quad (9)$$

La justificación para emplear esta expresión que emplea  $R_{0L}$  en lugar de  $R$ , es que ésta elimina el efecto de diferencias en ubicación, lo que permite analizar únicamente diferencias entre distribuciones en cuanto a forma. Un MRP positivo indica un incremento en el grado de polarización; es decir, mayor estiramiento de las colas de la distribución de comparación en relación a la de referencia. Valores negativos del MRP indican menor polarización; en otras palabras, mayor convergencia hacia el centro de la distribución de comparación que aquella de referencia. Por otro lado, si no existen diferencias entre las distribuciones, al menos en cuanto a forma, el MRP es cero.

Adicionalmente, el MRP está contenido en  $[-1, 1]$ , lo que permite una interpretación en términos de proporción de cambio. En especial, un  $\text{MRP} > 0$  ( $\text{MRP} < 0$ ) sugiere un cambio de  $\text{MRP} \times 100\%$  mayor masa poblacional hacia los extremos (centro) de la distribución de comparación en relación a la distribución de referencia.

Otra propiedad útil del MRP es la simetría. Esto es, si MRP es el índice para la DR de  $Y$  en  $Y_0$ , y  $\text{MRP}_0$  es el índice para la DR de  $Y_0$  en  $Y$ , se cumple que:

$$\text{MRP} = -\text{MRP}_0.$$

El MRP puede descomponerse en relación la cola inferior (baja) y superior (alta) de la distribución de  $R_{0L}$ . Entonces, tendremos un índices de polarización baja (LRP por su sigla del inglés Lower Polarization Index) y de polarización alta (URP por la sigla del inglés Upper Polarization Index) que permiten medir el grado de polarización en la cola inferior y superior, respectivamente. Estos índices se definen como:

$$\text{LRP}(F, F_0) = 4E \left[ \left| R_{0L} - \frac{1}{2} \right| \middle| R_{0L} \leq \frac{1}{2} \right] - 1, \quad (10)$$

$$\text{URP}(F, F_0) = 4E \left[ \left| R_{0L} - \frac{1}{2} \right| \middle| R_{0L} > \frac{1}{2} \right] - 1. \quad (11)$$

Estos dos índices permiten descomponer el MRP de la siguiente manera:

$$\text{MRP}(F, F_0) = \frac{1}{2}\text{LRP}(F, F_0) + \frac{1}{2}\text{URP}(F, F_0). \quad (12)$$

Esta descomposición permite identificar la contribución de diferencias en las colas bajas y altas a la polarización mediana. Por tanto, vale la pena resaltar que, un  $\text{MRP} = 0$  no implica necesariamente que

las distribuciones son idénticas en forma, ya que esto puede originarse a partir de una composición entre un LRP y un URP de signos opuestos pero muy similares en magnitud.

Handcock y Morris (1999) sugieren los siguientes estimadores para los índices de polarización:

$$\widehat{\text{LRP}}(F, F_0) = \frac{8}{m} \sum_{j=1}^m \left| \hat{Q}_j - \frac{1}{2} \right| - 1, \quad \forall \hat{Q}_j \leq \frac{1}{2}, \quad (13)$$

$$\widehat{\text{URP}}(F, F_0) = \frac{8}{m} \sum_{j=1}^m \left| \hat{Q}_j - \frac{1}{2} \right| - 1, \quad \forall \hat{Q}_j > \frac{1}{2}, \quad (14)$$

donde  $\hat{Q}_j = F_0^d(y_j - \hat{\rho})$ .

Finalmente, el estimador para el MRP es el promedio simple entre los índices bajo y alto estimados:

$$\widehat{\text{MRP}}(F, F_0) = \frac{1}{2} \widehat{\text{LRP}}(F, F_0) + \frac{1}{2} \widehat{\text{URP}}(F, F_0). \quad (15)$$

## 4. Datos y nuestro ejercicio

Los datos empleados en este estudio corresponden a los resultados de la prueba SABER 11 y SABER PRO obtenidos por estudiantes quienes presentaron ambas pruebas para el término de su respectivo ciclo educativo. Los datos provienen directamente de Instituto Colombiano para la Evaluación de la Educación (IFCES). Para analizar cambios en la distribución poblacional del nivel de inglés como segundo idioma, tomamos como variable la calificación estandarizada del Marco Común Europeo de Referencia para las Lenguas (CEFR, por las siglas en inglés para *Common European Framework of Reference*) obtenida en cada prueba.

Para efectos del análisis de la distribución relativa, la escala ordinal del CEFR se ha ajustado como se ilustra en el Cuadro 1

Cuadro 1: Escala numérica para las categorías CEFR

Categoría	Escala numérica
A-	1
A1	2
A2	3
B1	4
B+	5

**Fuente:** Marco Común Europeo de Referencia para las Lenguas

La muestra se describe en mayor detalle en el Anexo 6. En total se cuentan con 28,422 estudiantes que tomaron la prueba SABER 11 y SABER PRO.

Nuestra aproximación para lograr el objetivo de determinar si existe un cambio o no en la distribución del dominio del inglés antes y después de pasar por el sistema de educación superior, es la siguiente. Primero se estimará la MPR, empleando como población referencia la clasificación de los estudiantes en la prueba de inglés de SABER 11 y como población de comparación la clasificación a partir de la prueba SABER PRO. En caso de identificar un cambio en la distribución de SABER 11 a SABER PRO, emplearemos los índices MRP, LRP y ULRP evaluar donde ocurrió el cambio en la distribución. En general, un patrón de cambio “ideal” debería estar representado por una reducción de masa poblacional en niveles de inglés muy bajos (A-), acompañada por un incremento hacia los niveles más altos de competencia (B1 o B+). La combinación de índices de polarización que describe un patrón de cambio de este tipo estaría compuesta por un LRP negativo y un URP positivo (preferiblemente de mayor magnitud al inferior).

No obstante, dado que los índices de polarización se estiman con base en la distribución relativa ajustada por medianas,  $\hat{Q}_j$ , es posible encontrar valores estimados imprecisos. En nuestro estudio, esto se presenta cuando la mediana de la distribución de referencia (SABER 11) es igual al límite inferior en que se define la variable de estudio; es decir, 1. En consecuencia el estimador del LRP resulta inadecuado y computa valores inferiores a -1, lo cual carece de sentido dada la definición teórica. Luego, como el MRP se compone del promedio entre LRP y URP, su valor estimado tenderá a tomar valores muy por debajo de su valor real.

En la mayoría de los casos, este problema se soluciona gracias a la propiedad de simetría del MRP.

Si la muestra proveniente de SABER PRO tiene mediana mayor que 1, podemos calcular los índices respectivos usando muestra de SABER PRO como la distribución de referencia y SABER 11 como la distribución de comparación. Esto hace que el problema de subestimación del LRP desaparezca conservando las propiedades estadísticas del MRP se mantienen. En caso de que la mediana, tanto de SABER 11 como de SABER PRO, sea igual a 1, el análisis de polarización relativa no aplica.

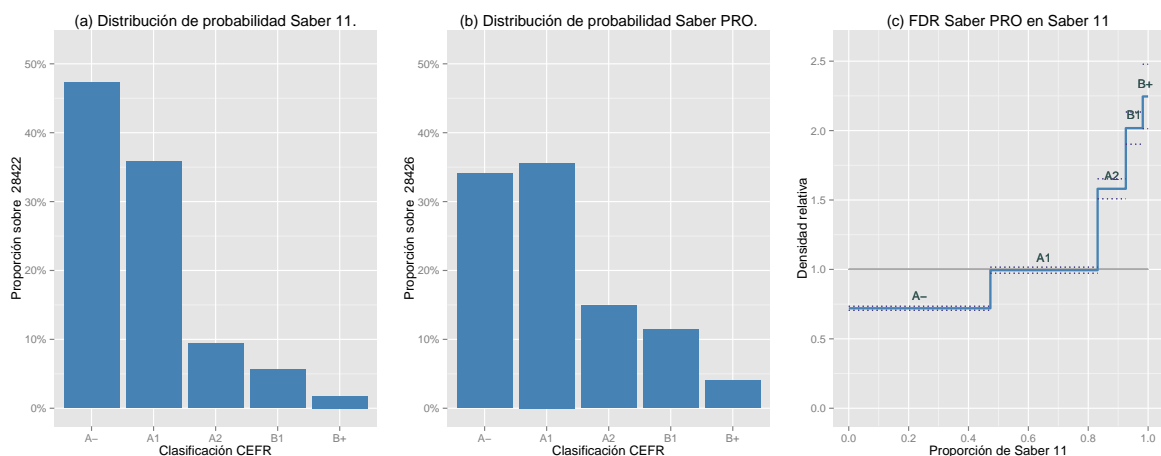
El análisis descrito anterior se realizará para los siguientes subconjuntos de nuestra muestra:

1. Total de la población evaluada
2. A nivel de agregación por nivel de IES
3. A nivel de grupos de referencia por áreas de conocimiento al interior de cada nivel de IES

## 5. Resultados

La Figura 4 presenta los histogramas de probabilidad de la muestra SABER 11 y SABER PRO (paneles a y b). La moda para la prueba SABER 11 es A- (el nivel más bajo) y la moda para la prueba SABER PRO es A1. Este análisis gráfico sugiere que en efecto sí existe un cambio en la distribución. Al observar los paneles a y b, encontramos que más de la mitad de la población parece concentrarse en niveles por debajo de A1 tanto en SABER 11 como en SABER PRO. Mientras tanto, en los cuantiles más altos, se puede notar un aumento en la proporción de individuos con niveles A2 o superior.

Figura 4: Total Agregado



**Fuente:** Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Por otro lado en el panel c de la Figura 4 se presenta la FDR correspondiente y los respectivos intervalos de confianza. En el eje horizontal se ordenan los cuantiles ( $i$ ) de corte de la distribución de referencia (SABER 11), cada uno de estos asociado a una clasificación CEFR. en el eje vertical se ordena el valor de la FDR para cada cuantil ( $\hat{g}(i)$ ). Las líneas punteadas alrededor de la curva indican el respectivo intervalo de confianza del 95 %. Para cuantiles por debajo de la mediana de SABER 11, correspondientes a niveles A-, la FDR está cerca de 0.70 (significativamente diferente de 1 con un 95 % de confianza). En otras palabras, las probabilidades de encontrar individuos con calificación A- en SABER PRO es del 30 % (100 %–70 %) más bajas que en SABER 11. Para cuantiles entre 0.4 y 0.8 el intervalo de confianza incluye la unidad, señalando que para niveles A1 las distribuciones de probabilidad entre SABER 11 y SABER PRO son estadísticamente iguales. Por otra parte, para menos del 20 % de la población con niveles A2, B1 y B+, la FDR es aproximadamente 1.5, 2.0 y 2.3, respectivamente, lo que sugiere que en SABER PRO existe cerca de 50 % más probabilidades de encontrar individuos con A2, y que para niveles B1 o B+ las probabilidades son más del doble frente a la prueba SABER 11.

De forma complementaria, los índices de polarización confirman que existe un cambio en la forma de la distribución. El MRP estimado para la FDR de la Figura 4 es 0.07, significativo con un 99 % de confianza. Por su parte, el LRP es no significativo y el URP es 0.15, significativo al 1 %. Estos resultados describen un patrón de cambio “deseable”. Un aumento del 7 % en la masa de la distribución desde niveles de inglés medios a niveles más altos.

Hasta ahora, se pueden resaltar tres importantes hallazgos. En primer lugar, a nivel agregado, existe un cambio en la dirección “deseable” en la forma de la distribución del nivel de inglés luego del paso por la educación superior. En segundo lugar, este cambio se caracteriza por un aumento en la desigualdad, expresado en mayor proporción de la población hacia las colas altas de la distribución. Finalmente, de acuerdo con la FDR, la magnitud del cambio es mayor conforme el nivel de competencias es mayor.

En los siguientes apartados se realizan las comparaciones de las distribuciones de las dos pruebas desagregando la población total en diferentes subconjuntos.

## 5.1. Análisis según el carácter de la Institución de Educación Superior

Luego de encontrar evidencias que respaldan cambios favorables en la forma de la distribución del nivel de inglés para el agregado poblacional, resulta pertinente verificar si existe una relación entre la magnitud del cambio y el nivel tipo de Institución de Educación Superior (IES). En esta sección nos dedicamos a explorar esta posibilidad. Para efectos del análisis desagregamos la muestra de acuerdo con la clasificación de las instituciones de educación superior en Colombia: institución universitaria (UNI), institución de técnica profesional (TNAL) e institución tecnológica terminal (TNOL).

Figura 5: Universidades



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Con el análisis de la distribución relativa identificamos cambios más sobresalientes en los tres niveles de agregación. Las Figuras 5, 6 y 7 muestran las distribuciones de probabilidad SABER 11 y SABER PRO y la FDR estimada para UNI, TNAL y TNOL. Al observar los histogramas, vemos que los tres niveles de agregación presentan características muy similares: la proporción de individuos con calificaciones A- y A1 es mayor en SABER 11, mientras que la proporción de individuos con calificaciones A2, B1 y B+ es mayor en SABER PRO. Lo anterior se confirma al observar la FDR correspondiente a cada caso, la cual toma valores cercanos a 0.3 para A- y cercanos a 1 para niveles A1.

No obstante, existen diferencias que vale la pena mencionar en lo que respecta a los niveles B1 y B+. En primer lugar, el cambio es relativamente más pronunciado en las IES de nivel más básico. Para Instituciones Universitarias (UNI) la FDR para B1 y B+ es aproximadamente 1.8 y 2.2, respectivamente. Para instituciones de técnica profesional (TNAL) es 2.6 y 3.0 y para instituciones de tecnológica termi-



Figura 6: Técnica Profesional

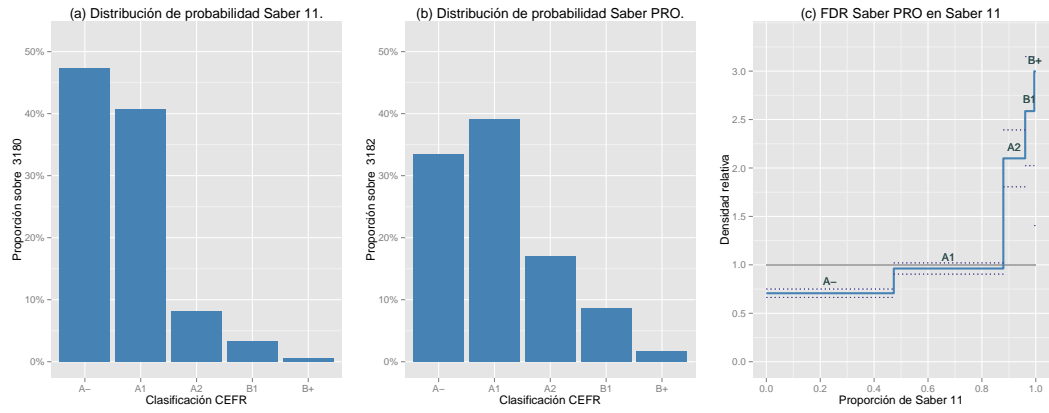
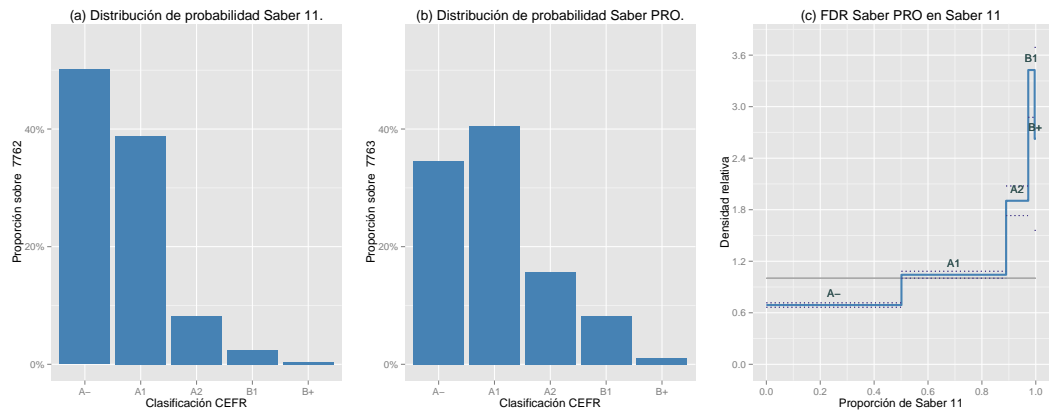


Figura 7: Tecnológica Terminal



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

nal (TNOL) es 3.4 y 2.6. En segundo lugar, cuanto mayor es el nivel educativo más grande es el cambio. Para las instituciones universitarias el cambio se presenta en más del 10% de la población (1–0.89), mientras que en las instituciones técnicas y tecnológicas la proporción de la población beneficiada con el cambio no supera el 4% y 3%, respectivamente.

En el Cuadro 2 se presentan los índices de polarización correspondientes a este análisis. Para UNI el comportamiento de los índices estimados es muy similar al observado en el agregado total, con un MRP estimado de 0.06 (significativo con un 99% de confianza) y mayor polarización hacia niveles altos. En el caso de TNAL, el MRP estimado es mayor que el nivel agregado, indicando un cambio del 9% en la masa de la distribución hacia los niveles más altos. Mientras tanto, el subgrupo de TNOL presenta la combinación de índices más “deseable”; aún con un MRP igual a similar al de UNI, presenta un LRP negativo y un URP positivo más grandes (-0.08 y 0.20, ambos significativos al 95% y 99% de confianza respectivamente).

Cuadro 2: Índices de polarización relativa por nivel de IES

Nivel de agregación	MRP (p-value)	LRP (p-value)	URP (p-value)
Total Agregado	0.07 *** (0.00)	-0.02 (0.31)	0.15 *** (0.00)
UNI	0.06 *** (0.00)	-0.02 (0.34)	0.15 *** (0.00)
TNAL	0.09 *** (0.00)	-0.02 (0.75)	0.19 *** (0.00)
TNOL	0.06 *** (0.00)	-0.08 ** (0.03)	0.20 *** (0.00)

(\*) Nivel de significancia: 10%

(\*\*) Nivel de significancia: 5%

(\*\*\*) Nivel de significancia: 1%

**Fuente:** Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Se tiene entonces que el cambio entre la distribución del nivel de inglés SABER 11 y SABER PRO que observamos en el agregado total, se manifiesta de manera muy similar entre los tres agregados por nivel de IES. En otras palabras, la relación entre el paso por la educación superior y el desarrollo en las competencias de inglés se presenta de manera casi uniforme a lo largo de los tres niveles de educación superior. Sin embargo, existe una relación inversa entre la magnitud del cambio y la proporción de la población para las calificaciones más altas: a menor nivel de IES mayor es el impacto sobre el nivel de competencias, pero menor es la proporción de la población beneficiada.

## 5.2. Análisis según grupo de referencia

Hemos identificado diferencias significativas entre las distribuciones SABER 11 y SABER PRO, las cuales se manifiestan de forma muy similar al descomponer la muestra por nivel de IES. Ahora concentramos nuestra atención en identificar las áreas de conocimiento que más se ven afectadas dentro

de cada nivel de educación superior. Para ello desagregamos cada agregado, UNI, TNAL y TNOL, de acuerdo a los grupos de referencia por área de conocimiento.

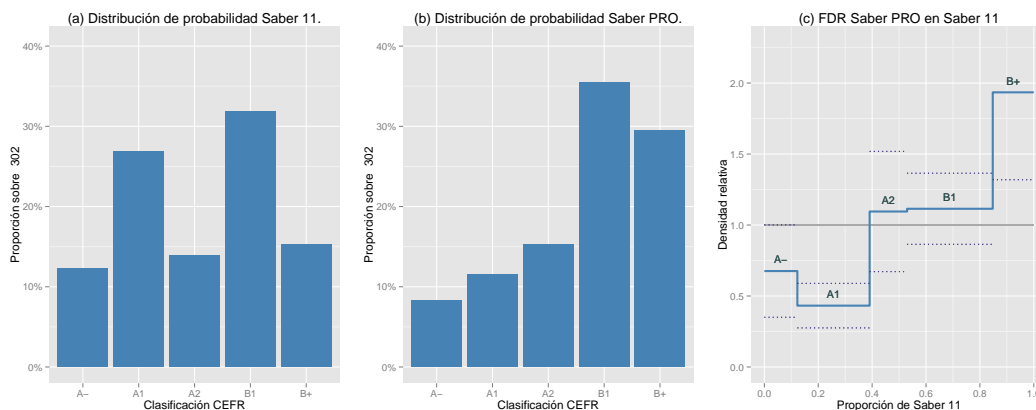
### 5.2.1. Grupos de referencia por nivel universitario

En esta sección se analizan los cambios en la distribución del nivel de inglés dentro los grupos de referencia universitarios. Aquí se encuentra mayor heterogeneidad entre las características de cambio. Pocos grupos, conformados por la mayoría de la población, presentan las mejoras relativamente más importantes, mientras que muchos grupos, menos representativos, no dan evidencia estadística de cambios en su distribución.

Aquellos grupos que presentan mejoras significativas siguen un patrón de cambio muy similar al del agregado UNI, aunque mucho más pronunciado. El análisis de la FDR sugiere que el aumento en la proporción de individuos con calificaciones altas es mayor que la disminución en la proporción de individuos con calificaciones bajas. El grupo de Administración, contaduría y afines (109) el cual presenta las condiciones de cambio más favorables en su distribución, la FDR sugiere que en SABER PRO la proporción de individuos con calificaciones B1 y B+ es respectivamente 1.7 y 2.3 veces la de SABER 11. Este cambio representa cerca del 25 % de la población (Figura 16). El grupo de Comunicación, periodismo y publicidad (106) le sigue de cerca con características de cambio muy similares (Figura 13). Algunos como Ingeniería (112) y Economía (131) siguen un patrón de cambio similar aunque un poco más modesto (Figuras 17 y 21). Para otros grupos, como Bellas artes y diseño (101), Derecho (105) y Salud (113), los cambios de mayor representatividad ocurren para calificaciones A- y A1, mientras que los de mayor magnitud se presentan en un porcentaje muy bajo de la población (Figuras 8, 12 y 18).

Por otra parte, en los grupos de Ciencias naturales y exactas (102), Ciencias sociales (103), Humanidades (104), Ciencias militares y navales (107), Ciencias agropecuarias (108), Medicina (114) y Recreación y deportes (130), no se encuentra evidencia sobre cambios importantes en cuanto a la distribución. En la mayoría de los casos, la FDR sólo es estadísticamente diferente de 1 para menos del 35 % de la población. Adicionalmente, cuando esto ocurre, el cambio sólo se presenta en términos de disminuciones en la proporción de individuos con calificaciones medias y bajas (Figuras 9, 10, 11, 14, 15, 19 y 20).

Figura 8: Bellas artes y diseño (101)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Figura 9: Ciencias naturales y exactas (102)

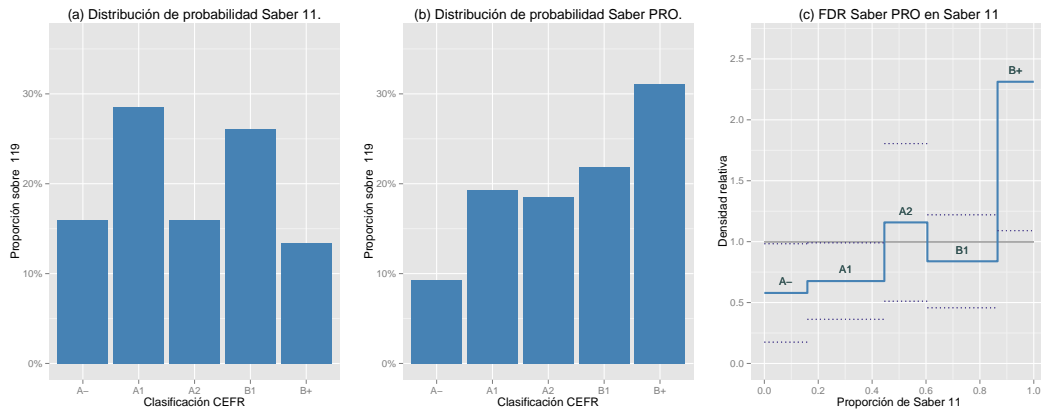


Figura 10: Ciencias sociales (103)

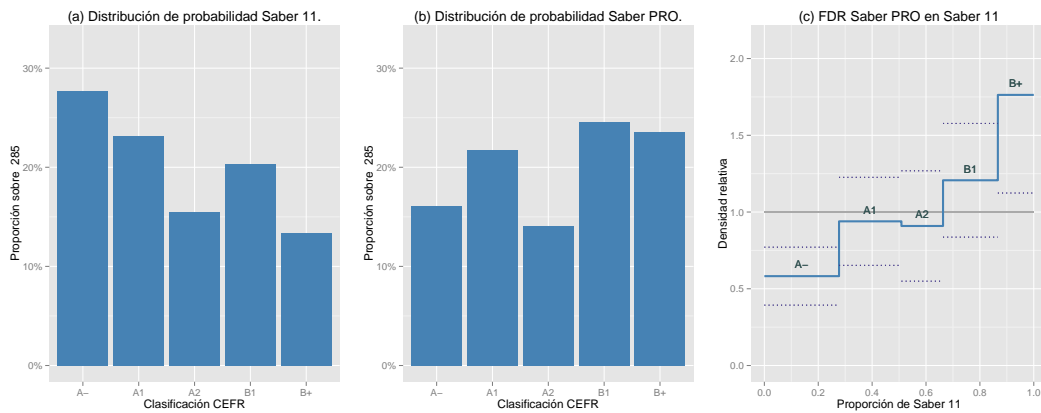
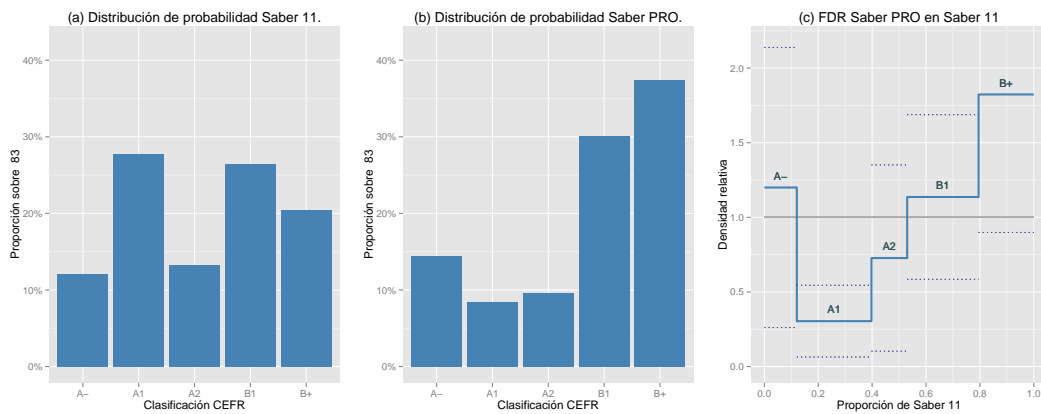


Figura 11: Humanidades (104)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Figura 12: Derecho (105)

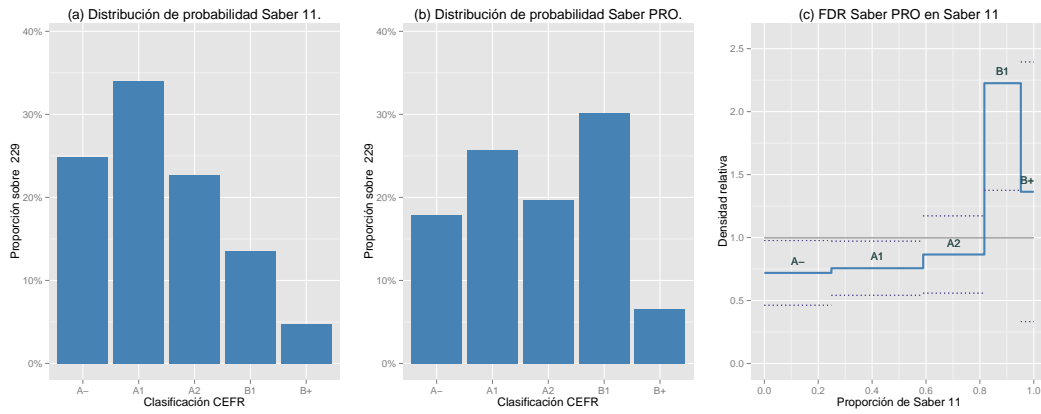


Figura 13: Comunicación, periodismo y publicidad (106)

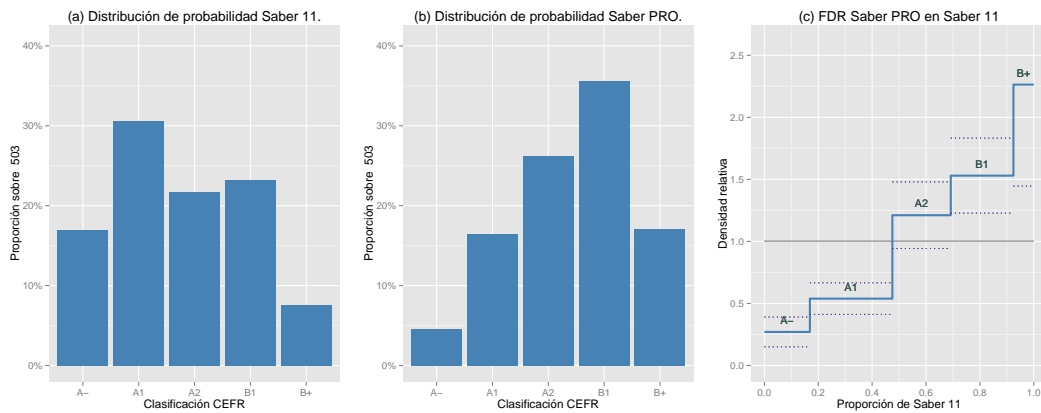
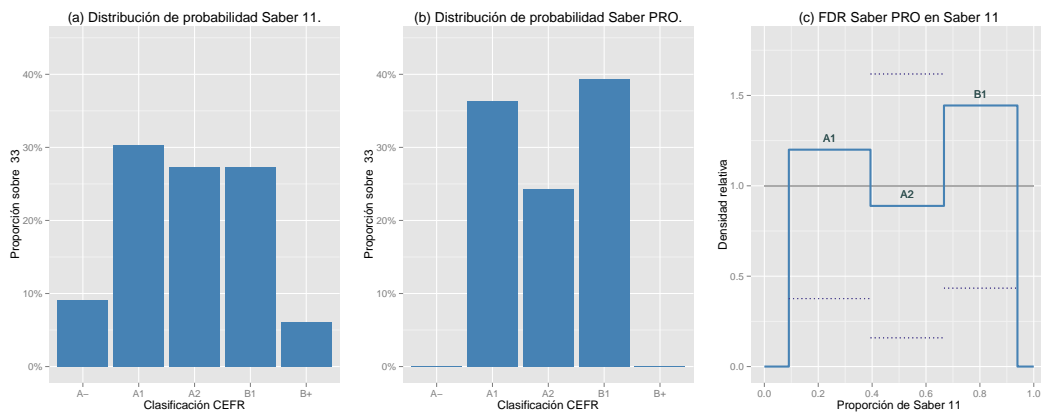


Figura 14: Ciencias militares y navales (107)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Figura 15: Ciencias agropecuarias (108)

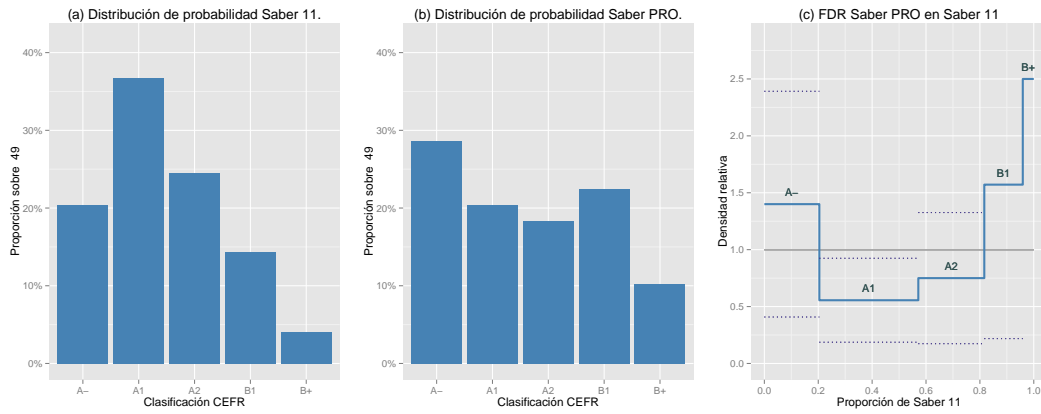


Figura 16: Administración, contaduría y afines (109)

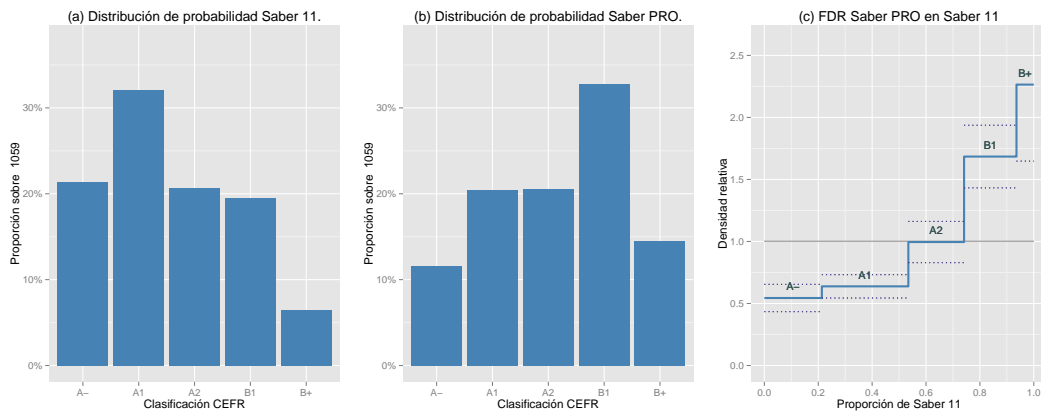
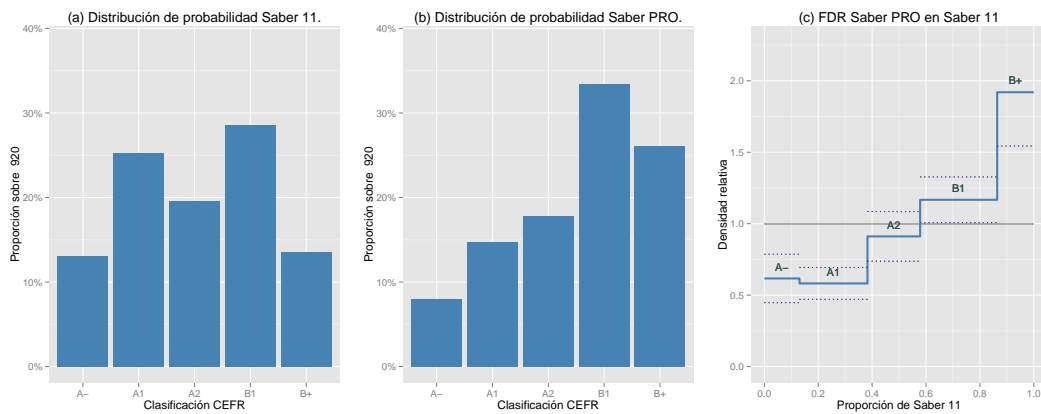


Figura 17: Ingeniería (112)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Figura 18: Salud (113)

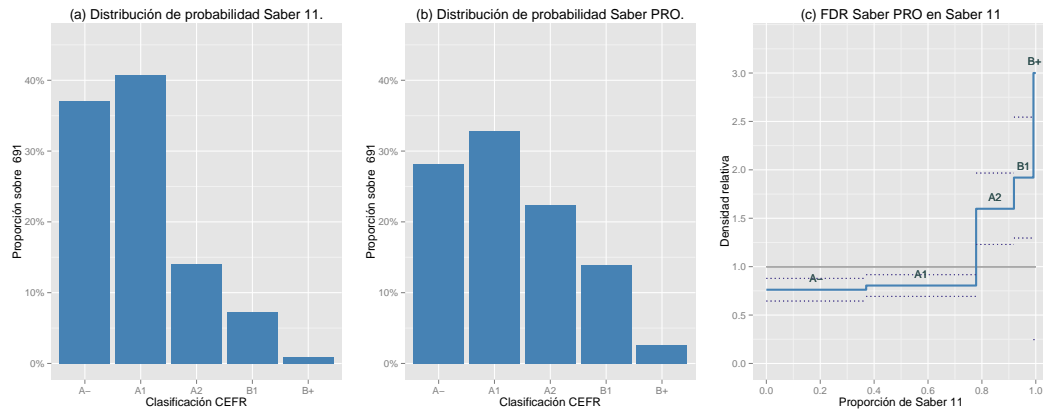


Figura 19: Medicina (114)

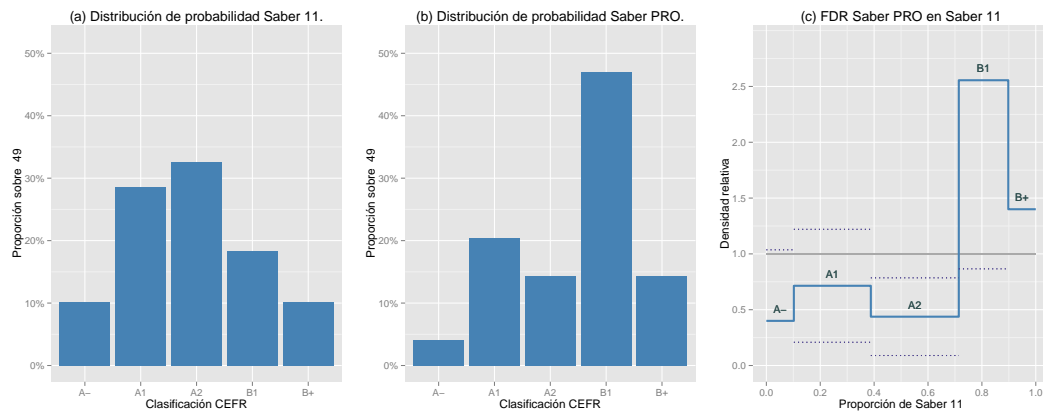
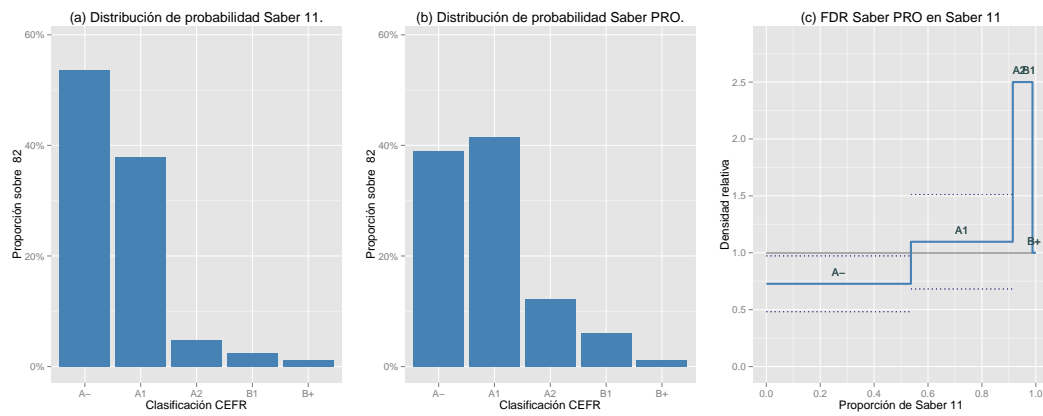
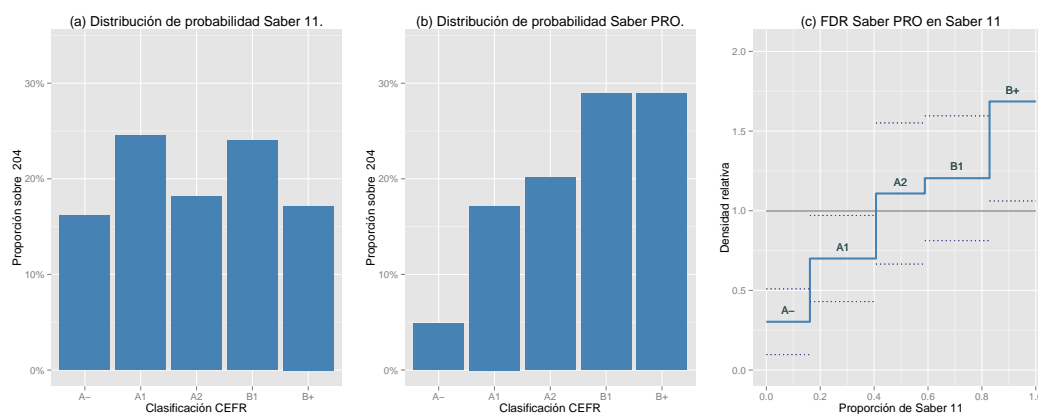


Figura 20: Recreación y deportes (130)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Figura 21: Economía (131)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

El Cuadro 3 resume el análisis de índices de polarización relativa correspondiente a cada caso. El grupo 109 presenta un patrón de cambio muy similar al “ideal”, con un MRP estimado de 0.07 (significativo al 99 % de confianza). Para 113 el MRP (significativo al 95 % de confianza) sugiere que la masa de la distribución es 10 % mayor hacia las calificaciones más altas. Por otro lado, el grupo 105 presenta un grado de polarización más alto (0.10) aunque significativo sólo al 90 % de confianza. El caso del grupo de Ciencias agropecuarias (108) es diferente. Éste presenta el grado de polarización más alto, indicando un cambio del 24 % en la masa de la distribución hacia niveles tanto más altos como más bajos.

### 5.2.2. Grupos de referencia por nivel técnico profesional

En esta sección se analizan los cambios en la distribución al interior del nivel de IES técnico profesional. En general, este caso presenta características muy similares a lo observado en el nivel UNI respecto a la heterogeneidad en los patrones de cambio. No obstante, aquí, los cambios favorables de mayor relevancia ocurren en la mayoría de los grupos de referencia, los cuales, además, representan la mayor parte de la población.

Los cambios más pronunciados en la distribución del nivel de inglés ocurren en Ingeniería, industria y minas (121), Administración y turismo (122) y TICs (123). Estos tres grupos siguen un patrón muy similar: mientras la proporción de individuos con niveles A- se reduce, para niveles A2 o mayores, ésta se incrementa. Sin embargo, como se ha podido observar en casos anteriores, los cambios más sobresalientes ocurren en los niveles más altos de la distribución, donde la proporción de la población beneficiada es relativamente baja. De hecho, para todos estos es común que el cambio en B+ no sea significativo. Por ejemplo, para 122 el análisis de la FDR sugiere que en SABER PRO hay cerca de 32 % menos probabilidades de encontrar individuos con nivel A-, mientras que las probabilidades asociadas a A2 y B1 son respectivamente 2.1 y 3.8 veces las de SABER 11. Estos cambios se manifiestan para cerca del 60 % de la población, del cual sólo un 2.2 % está asociado a B1 (Figura 23).

En Artes, diseño y comunicación (125), por otra parte, los cambios asociados a niveles B1 y B+, aunque



Cuadro 3: Índices de polarización relativa por grupo de referencia universitario

Código	Nombre	MRP (p-value)	LRP (p-value)	URP (p-value)
101	Bellas artes y diseño	0.01 (0.82)	-0.11 (0.56)	0.13 (0.48)
102	Ciencias naturales y exactas	0.04 (0.57)	-0.10 (0.74)	0.18 (0.53)
103	Ciencias sociales	0.01 (0.84)	-0.12 (0.53)	0.14 (0.46)
104	Humanidades	0.15 (0.10)	0.16 (0.65)	0.14 (0.70)
105	Derecho	0.10 * (0.08)	-0.01 (0.95)	0.20 (0.33)
106	Comunicación, periodismo y publicidad	-0.04 (0.28)	-0.19 (0.18)	0.11 (0.43)
107	Ciencias militares y navales	-0.083 (0.56)	-0.14 (0.80)	-0.03 (0.96)
108	Ciencias agropecuarias	0.24 ** (0.04)	0.23 (0.61)	0.26 (0.57)
109	Administración, contaduría y afines	0.07 *** (0.00)	-0.04 (0.69)	0.19 * (0.07)
112	Ingeniería	0.03 (0.28)	-0.08 (0.45)	0.14 (0.19)
113	Salud	0.10 *** (0.00)	-0.01 (0.93)	0.20 * (0.08)
114	Medicina	0.11 (0.33)	-0.01 (0.99)	0.23 (0.61)
130	Recreación y deportes	0.04 (0.64)	-0.06 (0.86)	0.15 (0.67)
131	Economía	-0.07 (0.22)	-0.23 (0.30)	0.09 (0.68)

(\*) Nivel de significancia: 10 %

(\*\*) Nivel de significancia: 5 %

(\*\*\*) Nivel de significancia: 1 %

**Fuente:** Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

relativamente más modestos, son todos significativos al 95 % de confianza y más representativos, con más del 10 % de la población beneficiada (Figura 26).

Por último, las Figuras 25 y 27 ilustran el caso de los grupos de Salud (124) y Ciencias agropecuarias (126). Mientras, para 124 sólo se identifican cambios significativos asociados a calificaciones A- y A2, para 126 en los histogramas notamos que, a pesar de la inclusión de cierto porcentaje de individuos en B1, el intervalo de confianza de la FDR sugiere que la distribución no ha cambiado significativamente entre SABER 11 y SABER PRO.

Figura 22: Ingeniería, industria y minas (121)

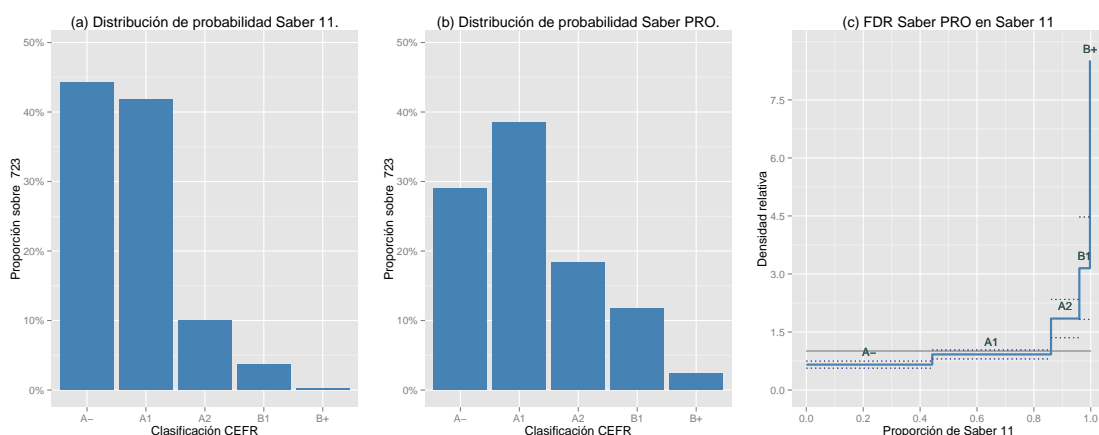
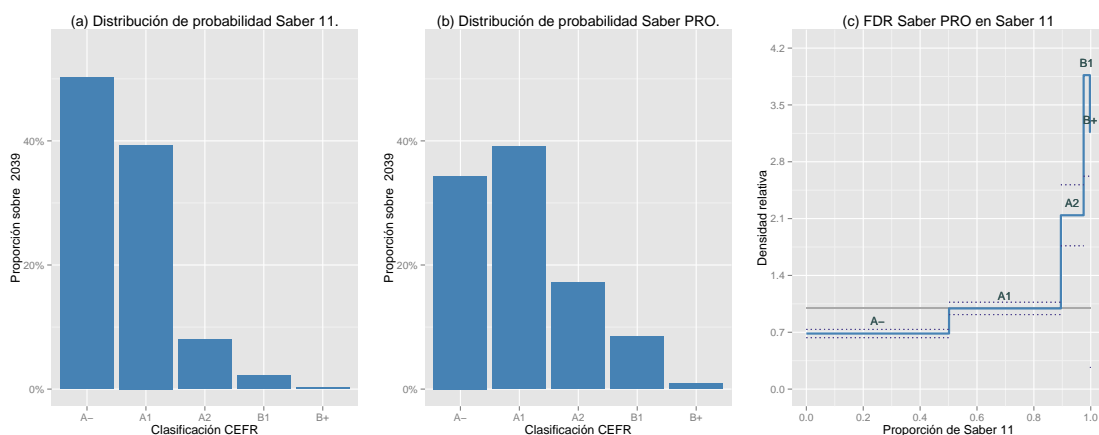


Figura 23: Administración y turismo (122)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Sin embargo, *a priori*, sólo asegurar cambios en la forma de la distribución para el grupo 122. Los índices polarización relativa sugieren un aumento en la desigualdad del 8 %, compuesta principalmente por un crecimiento en la cola superior de la distribución. Encontramos polarización significativa en Ingeniería, industria y minas (121) y Artes, diseño y comunicación (125). El MRP estimado de 121 indica un aumento en la desigualdad del 9.92 %, mientras que para 125 el aumento es de 6.53 %. Ambos índices son significativamente diferentes de 0 con un nivel de confianza del 99 %. La composición de estos sugiere que el cambio en la forma de la distribución es favorable y sigue de cerca patrón “ideal”.

Figura 24: TICs (123)

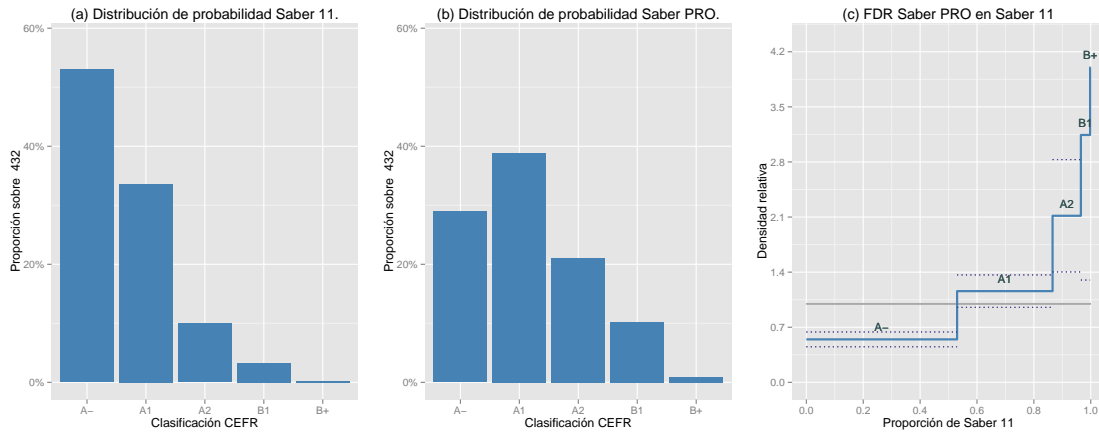


Figura 25: Salud (124)

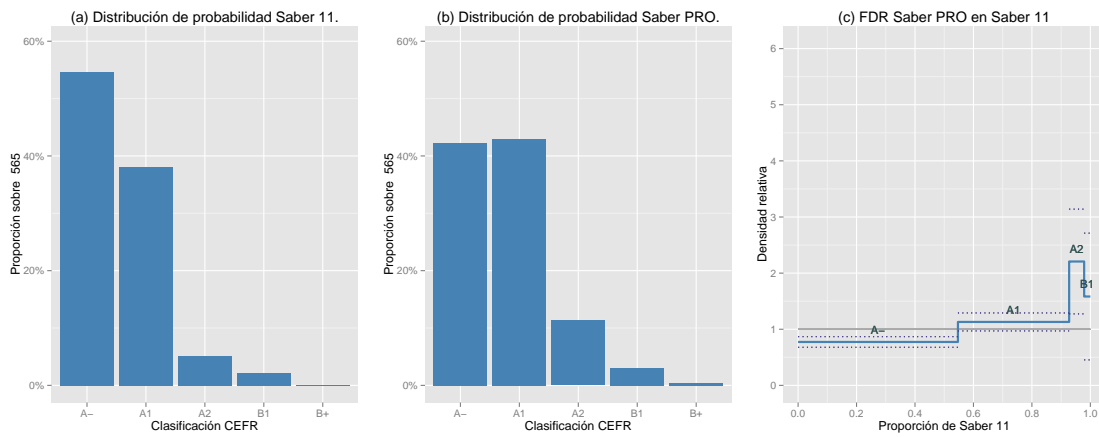
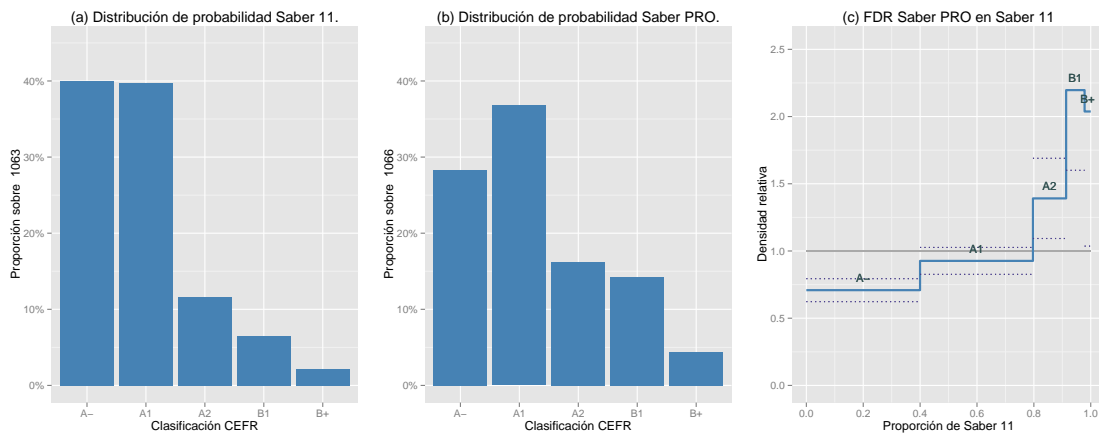
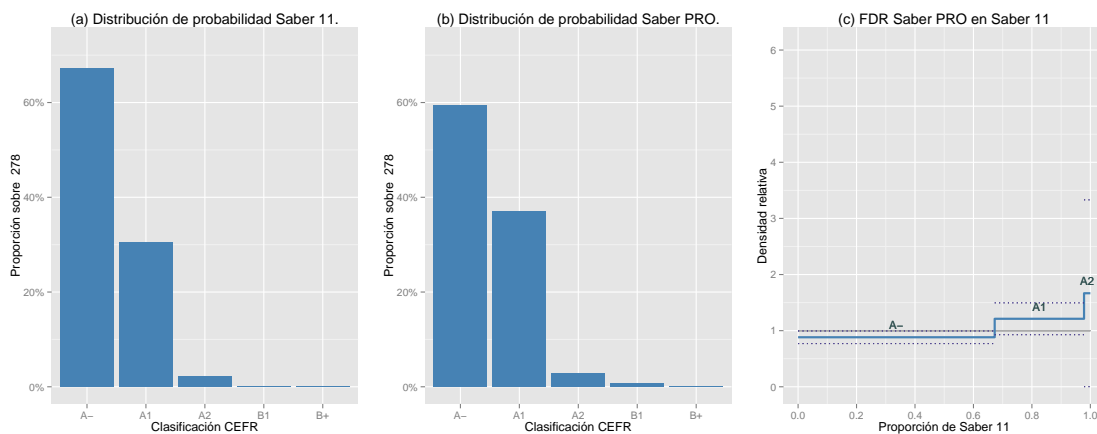


Figura 26: Artes, diseño y comunicación (125)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Figura 27: Ciencias agropecuarias (126)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Cuadro 4: Índices de polarización relativa por grupo de referencia nivel técnico profesional

Código	Nombre	MRP (p-value)	LRP (p-value)	URP (p-value)
121	Ingeniería, industria y minas	0.19 *** (0.00)	-0.04 (0.74)	0.24 ** (0.04)
122	Administración y turismo	0.08 *** (0.00)	-0.07 (0.29)	0.24 *** (0.00)
123	TIC	0.02 (0.63)	-0.17 (0.28)	0.20 (0.18)
124	Salud	0.03 (0.39)	-0.04 (0.75)	0.10 (0.44)
125	Artes - diseño - comunicación	0.07 *** (0.01)	-0.05 (0.63)	0.18 * (0.07)
126	Ciencias agropecuarias	NA	NA	NA

(\*) Nivel de significancia: 10 %

(\*\*) Nivel de significancia: 5 %

(\*\*\*) Nivel de significancia: 1 %

Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

### 5.2.3. Grupos de referencia por nivel tecnológico terminal

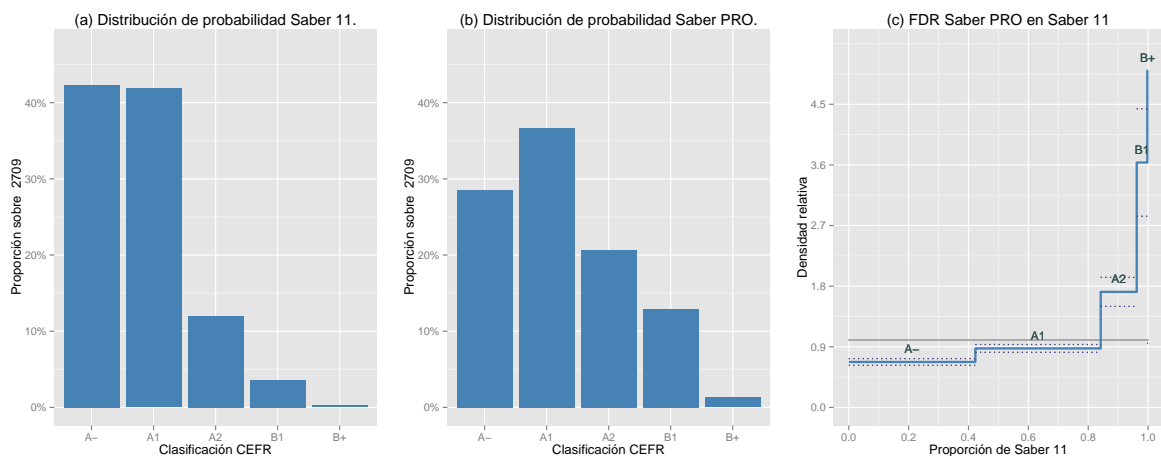
En este nivel de educación superior, se presenta un menor grado de heterogeneidad en los patrones de cambio en comparación con los agregados anteriores. Si bien para la gran mayoría el cambio es relativamente favorable, entre SABER 11 y SABER PRO la distribución de probabilidad mantiene una forma con mayor concentración de masa hacia los niveles más bajos y muy poca hacia los niveles medios y altos.

Los grupos de Ingeniería, industria y minas (115), Administración y turismo (116), TICs (117), Salud (118), Ciencias agropecuarias (120) y Militar y policial (129) se caracterizan principalmente por esta cualidad. El análisis de la FDR indica que mientras la probabilidad de encontrar individuos con calificación A- se reduce entre 20 % y 40 %, la probabilidad para A2 y B1 alcanza a representar más del doble y triple respectivamente en SABER PRO (Figuras 28, 29, 30, 31, 33 y 35). Para B+ el incremento de la probabilidad es incluso más grande, aunque no es significativamente diferente de 1. No obstante lo anterior, los casos de 115, 116, 117 y 118 resultan un tanto más “deseables” que los de 120 y 129.

A parte de los anteriores, el grupo de Artes, diseño y comunicación (119) sí evidencia un cambio en la forma de su distribución. En la Figura 32 los histogramas indican que la distribución del nivel de inglés pasa a ser un poco más “acampanada”. Esto se refleja en la representatividad del cambio: los valores de FDR más alejados de la unidad ocurren tanto para el 30 % de la población con calificaciones más bajas (A-), como para el 15 % con calificaciones más altas (B1 y B+), es decir, el 45 %.

El caso contrario es el correspondiente a Judicial (128). Para este grupo la FDR es significativamente diferente de 1 sólo la parte de la población con nivel A-, la cual se reduce en casi 40 % (Figura 34). Para el resto de niveles, no se encuentra evidencia suficiente de que exista un cambio significativo en la distribución, más allá de un posible incremento de su mediana.

Figura 28: Ingeniería, industria y minas (115)



**Fuente:** Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

El Cuadro 5 presenta los índices de polarización para este nivel de educación superior. Los grupos 115 y 116 registra el MRP estimado más alto (significativo al 99 % de confianza), sugiriendo un cambio en

Figura 29: Administración y turismo (116)

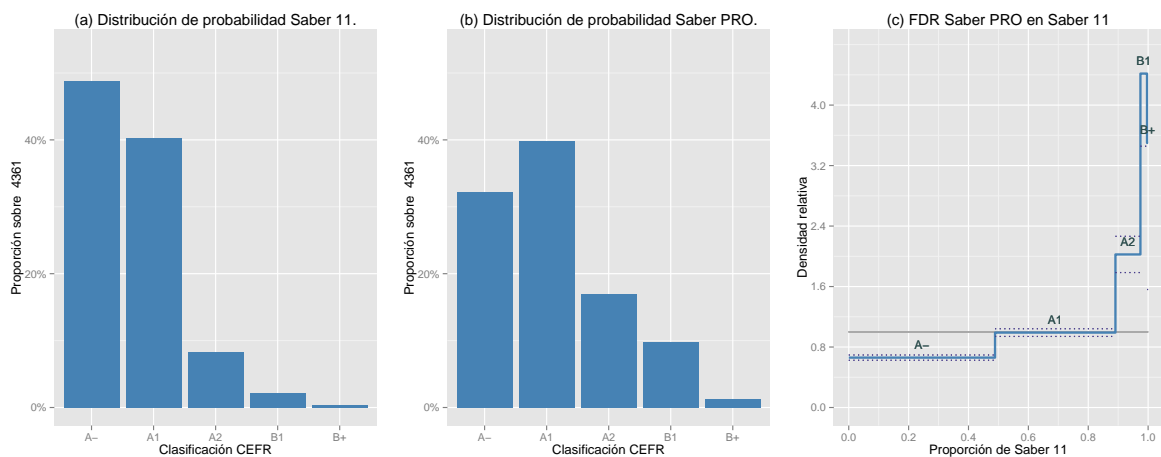
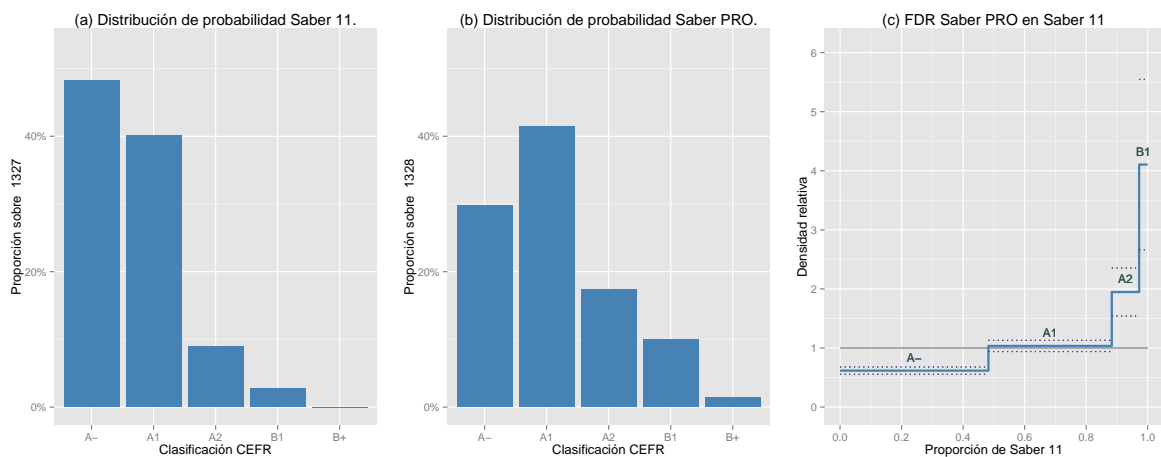


Figura 30: TICs (117)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Figura 31: Salud (118)

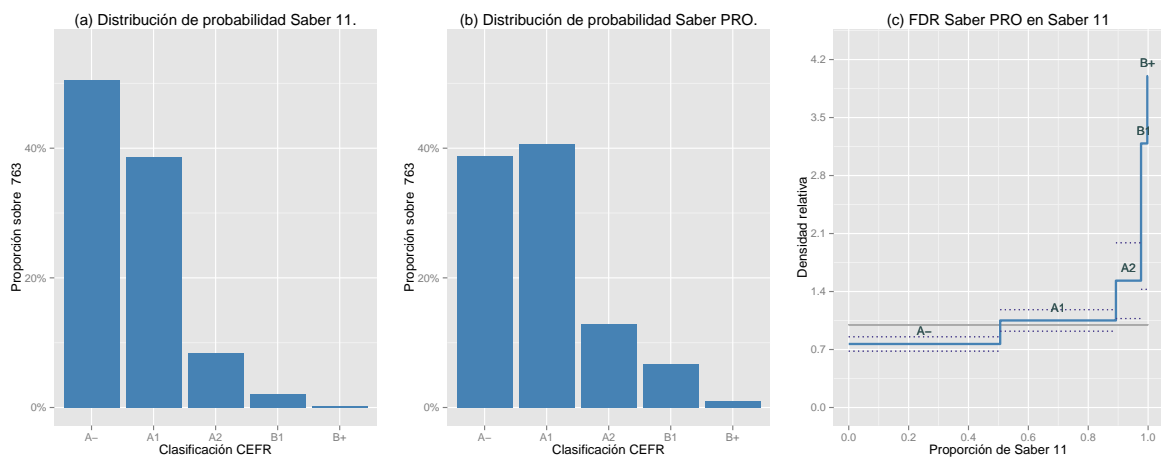
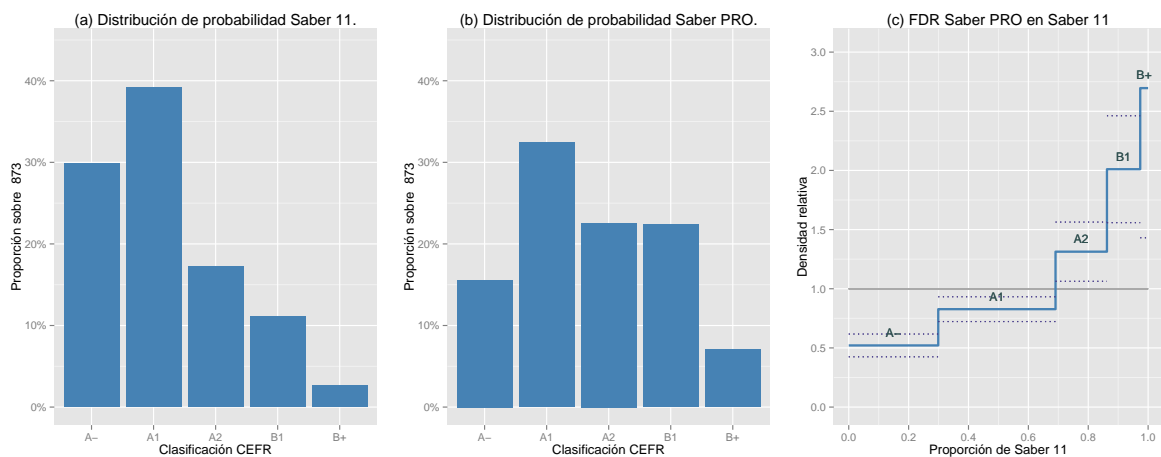


Figura 32: Artes, diseño y comunicación (119)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

Figura 33: Ciencias agropecuarias (120)

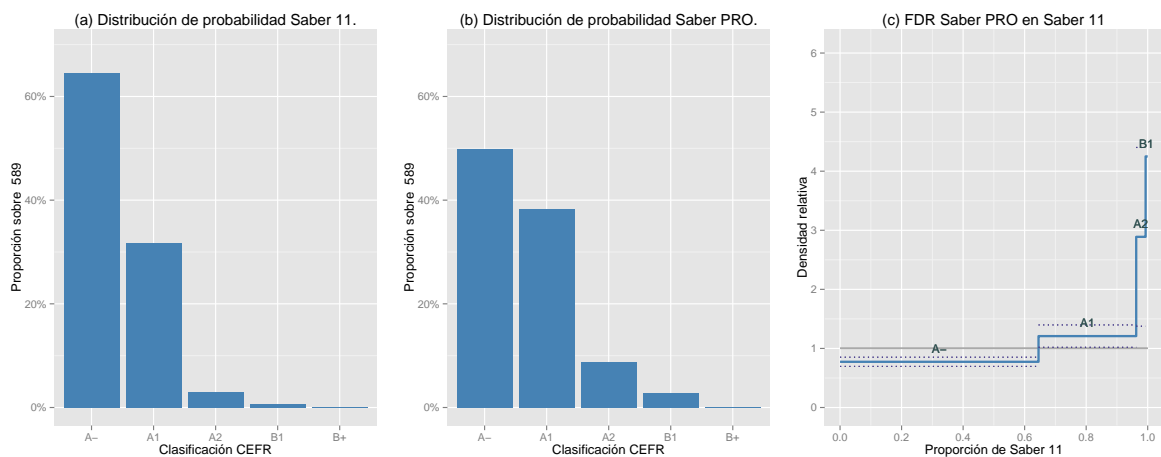
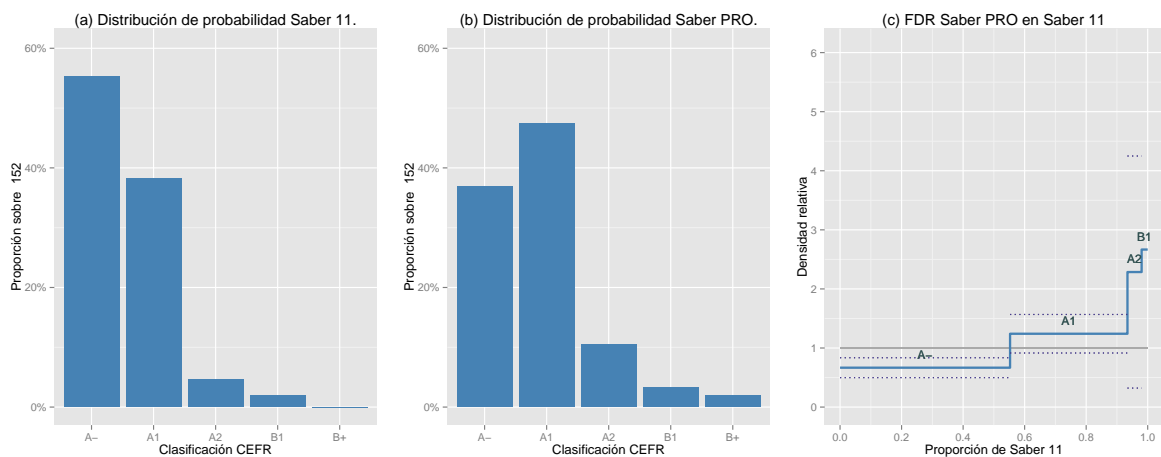


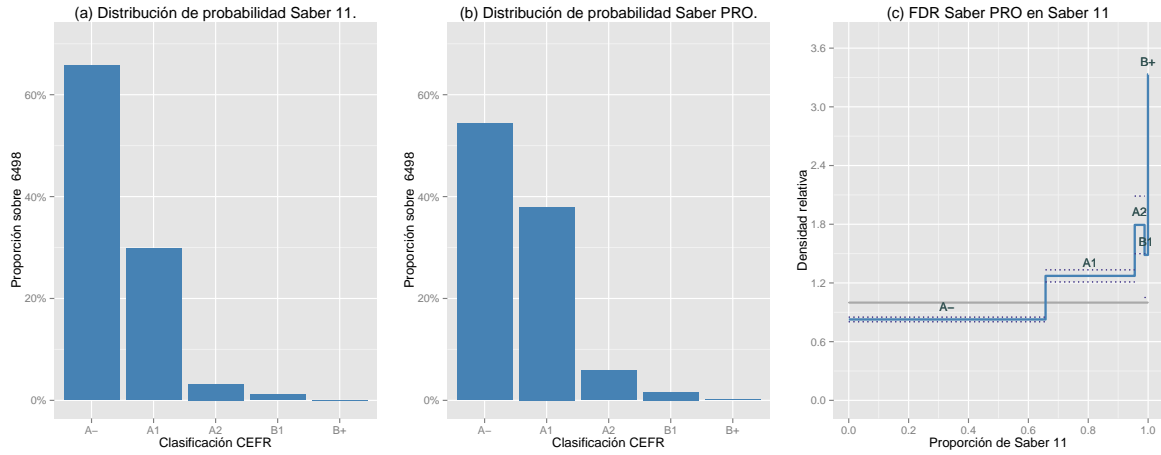
Figura 34: Judicial (128)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.



Figura 35: Militar y policial (129)



Fuente: Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

la forma de la distribución con 10 % más masa hacia los niveles más altos. Les sigue de cerca 117 con un MRP de 0.08 también significativo.

Por otra parte, la combinación de índices de polarización de 119 refleja perfectamente su patrón de cambio. El MRP si bien sólo es significativamente diferente de cero con un 90 % de confianza, éste se compone por unos índices bajo y alto que indican respectivamente un desplazamiento de la población en niveles bajos hacia niveles más centrales y de niveles medios hacia niveles más altos.

Cuadro 5: Índices de polarización relativa por grupo de referencia, nivel tecnológico terminal

Código	Nombre	MRP (p-value)	LRP (p-value)	URP (p-value)
115	Ingeniería, industria y minas	0.11 *** (0.00)	-0.04 (0.54)	0.25 *** (0.00)
116	Administración y turismo	0.10 *** (0.00)	-0.01 (0.82)	0.21 *** (0.00)
117	TIC	0.09 *** (0.00)	-0.02 (0.79)	0.20 ** (0.02)
118	Salud	0.05 * (0.08)	-0.05 (0.66)	0.15 (0.18)
119	Artes - diseño - comunicación	0.05 * (0.08)	-0.11 (0.29)	0.21 ** (0.05)
120	Ciencias agropecuarias	0.06 * (0.05)	0.00 (0.99)	0.13 (0.32)
128	Judicial	0.01 (0.85)	-0.10 (0.69)	0.13 (0.62)
129	Militar y policial	NA	NA	NA

(\*) Nivel de significancia: 10 %

(\*\*) Nivel de significancia: 5 %

(\*\*\*) Nivel de significancia: 1 %

**Fuente:** Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

## 6. Conclusiones

En general, la distribución del nivel de inglés como segundo idioma dista mucho de seguir una forma “ideal”: más del 80 % de los bachilleres se ubican en los niveles más bajos. Existe, sin embargo, cierto nivel de mejora en la distribución luego del paso por la educación superior. Los resultados del análisis de Distribución Relativa sugieren que, en el agregado, existe un cambio en la dirección “deseable” en la forma de la distribución del nivel de inglés luego del paso por la educación superior; esto es una reducción del porcentaje de la población en los niveles más bajos, acompañada por un incremento de éste en los niveles medios y altos. La magnitud del cambio es más grande conforme el nivel de competencias es mayor, donde además se evidencia un incremento en la polarización.

Este patrón de cambio se mantiene, en cierta medida, al descomponer la población por nivel de IES. La diferencia radica en que la magnitud del cambio es mucho mayor en la población egresada de las IES de técnica profesional y tecnología terminal en comparación con la población universitaria y el total agregado. Incluso, el incremento en los niveles de polarización es relativamente más alto en los primeros. En un sentido de eficiencia, esta particularidad probablemente se debe a que, para esta parte de la población estudiantil, la proporción de bachilleres que obtienen las calificaciones más altas es muy baja, lo cual deja un margen mucho más grande para alcanzar mejoras en la distribución.

En un grado de desagregación mayor los resultados sugieren un mayor grado de heterogeneidad entre los grupos de referencia por áreas de conocimiento, al interior de los agregados por IES. El mayor grado de heterogeneidad ocurre entre grupos de IES universitarias, para los cuales los casos de cambio de más favorables son pocos. Los hallazgos separados para los tres agregados coinciden al identificar las áreas de ciencias afines, economía, ingeniería, industria y minas y tecnologías de información, comunicación y periodismo como los grupos referencia con mejor desempeño; mientras que grupos relacionados con medicina, ciencias naturales y exactas, ciencias sociales, ciencias militares, ciencias agropecuarias y judicial no evidencian cambios significativos importantes.

Los hallazgos de este estudio, si bien se limitan a un análisis descriptivo, permiten inferir algunos aspectos a ser considerados: 1) El paso por la educación superior no pasa desapercibido en cuanto a mejoras en el nivel de competencias en bilingüismo, particularmente en el uso del inglés como segundo idioma. Durante este proceso intervienen ciertos factores que una necesidad de mayor capital humano, incluyendo el desarrollo esta competencia. 2) Varias de las características de la distribución del nivel de inglés en la población bachiller está relacionada en cierta medida con el tipo de educación superior en el que se desempeñan luego los estudiantes. Aunque no se puede establecer una relación de causalidad, es importante notar que la distribución de inglés en SABER 11 alberga un porcentaje de la población en los niveles más altos, mucho menor en el agregado de IES tecnológicas que por IES universitarias. 3) A lo anterior, se asocia la particularidad de que en esta parte de la población estudiantil se encuentra el mayor margen de mejora, lo que implica un alto potencial en términos de intervención en materia de política. 4) Las áreas de conocimiento que mejor patrón de cambio presentan están asociadas a los negocios, la comunicación y las tecnologías de información. Es probable que para estas áreas la necesidad capital humano en esta competencia sea mayor.

De acuerdo con lo anterior, las recomendaciones en materia de política coinciden con lo propuesto por Alonso *et al* (2012), en el sentido de que el enfoque debe corresponder a medidas que se ajusten en el largo plazo de forma continua y extensa. Esto, sin embargo, debe responder a los grupos de la población con mayor rezago, con el ánimo de explotar el mayor potencial de mejora que estos presentan.

## 7. Referencias bibliográficas

1. Alonso, J., Gallo, B. & Torres, G. (2012). Elementos para la construcción de una política pública de bilingüismo en el Valle del Cauca: un análisis descriptivo a partir del censo ampliado de 2005. *Estudios Genrenciales*, 28:59-67.
2. Handcock, M. & Morris, M. (1999). *Relative Distribution Methods in the Social Sciences*. New York: Springer-Verlag.
3. Handcock, M., Morris, M. & Bernhardt, A. (1997). *A Distributional Approach to Measuring Changes in Economic Inequality*. Department of Statistics, PennState.
4. Hermeto, A & Guimaraes, R. (2009). Cohort Trends in the Relative Wage Distribution by Gender in Brazil. In: 2009 PAA Annual Meeting, 2009, Detroit. *Proceedings of 2009 Population Association of America Annual Meeting*, 2009.
5. Mincer, J. (1974). *Schooling, Experience and Earnings*. Columbia University Press
6. Sánchez, A. (2013). *Bilingüismo en Colombia*. Centro de Estudios Económicos Regionales, BAN-REP.

## 8. Anexos

En la Tabla 6 se desglosa el tamaño de muestra por examen y por nivel de agregación.

Cuadro 6: Tamaño de muestra por nivel de agregación

Nivel de agregación		N	Faltantes SB11	Faltantes SBPRO	
Código	Nombre				Nivel
<b>Total agregado</b>		<b>28426</b>	<b>4</b>	<b>0</b>	
<i>Clasificación por IES</i>					
UNI	Universitaria	Universitaria	17481	1	0
TNAL	Técnica profesional	Técnica profesional	3182	2	0
TNOL	Tecnológica terminal	Tecnológica terminal	7763	1	0
	Faltantes		0		
<i>Clasificación por grupo de referencia</i>					
101	Bellas artes y diseño	Universitaria	302	0	0
102	Ciencias naturales y exactas	Universitaria	119	0	0
103	Ciencias sociales	Universitaria	285	0	0
104	Humanidades	Universitaria	83	0	0
105	Derecho	Universitaria	229	0	0
106	Comunicación, periodismo y publicidad	Universitaria	503	0	0
107	Ciencias militares y navales	Universitaria	33	0	0
108	Ciencias agropecuarias	Universitaria	49	0	0
109	Administración, contaduría y afines	Universitaria	1059	0	0
110	Educación	No especificado	372	0	0
111	Arquitectura y Urbanismo	Universitaria	6	0	0
112	Ingeniería	Universitaria	920	0	0
113	Salud	Universitaria	691	0	0
114	Medicina	Universitaria	49	0	0
115	Ingeniería, industria y minas	Tecnológica terminal	2709	0	0
116	Administración y turismo	Tecnológica terminal	4361	0	0
117	TIC	Tecnológica terminal	1328	1	0
118	Salud	Tecnológica terminal	763	0	0
119	Artes - diseño - comunicación	Tecnológica terminal	873	0	0
120	Ciencias agropecuarias	Tecnológica terminal	589	0	0
121	Ingeniería, industria y minas	Técnica profesional	723	0	0
122	Administración y turismo	Técnica profesional	2039	0	0
123	TIC	Técnica profesional	432	0	0
124	Salud	Técnica profesional	565	0	0
125	Artes - diseño - comunicación	Técnica profesional	1066	3	0
126	Ciencias agropecuarias	Técnica profesional	278	0	0
127	Normales superiores	Normalista	0	0	0
128	Judicial	Tecnológica terminal	152	0	0
129	Militar y policial	Tecnológica terminal	6498	0	0
130	Recreación y deportes	Universitaria	82	0	0
131	Economía	Universitaria	204	0	0
	Faltantes		1064		

**Fuente:** Instituto Colombiano para la Evaluación de la Educación. Cálculos propios.

El total agregado se compone de 28,426 observaciones que incluyen 4 faltantes. En el nivel de agregación por IES no hay faltantes en términos de clasificación, es decir, todos los datos corresponden, de manera excluyente, a instituciones universitarias, instituciones de técnica profesional o instituciones tecnológicas terminales. En el siguiente nivel de agregación, por grupo de referencia, 1,064 observaciones no se clasifican dentro de ningún grupo y se tratan como datos faltantes. Los grupos de referencia Arquitectura y Urbanismo (111) y Normales superiores (127) se excluyen del análisis a causa de insuficiencia de datos.