



Equiparación estadística de puntuaciones en 4to y 7mo de Educación General Básica del Ecuador utilizando software R.

Statistical equating of scores in 4th and 7th grade of General Basic Education in Ecuador using R software.

Equação estatística das pontuações no 4º e 7º ano do Ensino Básico Geral no Equador utilizando software R.

Héctor Salomón Mullo-Guaminga^I

hmullo@esPOCH.edu.ec

<https://orcid.org/0000-0001-8448-4652>

Jessica Alexandra Marcatoma-Tixi^{II}

jessica.marcatoma@unach.edu.ec

<https://orcid.org/0000-0001-9531-3234>

Danny Gabriel Guaño-Reyes^{III}

danny.guano@evaluacion.gob.ec

<https://orcid.org/0000-0001-9426-5285>

Pavél Wladimir García-Díaz^{IV}

pavel.garcia@evaluacion.gob.ec

<https://orcid.org/0000-0003-2756-0174>

Oswaldo Villacrés-Cáceres^V

ovillacres@esPOCH.edu.ec

<https://orcid.org/0000-0002-5894-5248>

Correspondencia: hmullo@esPOCH.edu.ec

Ciencias de la Educación
Artículo de Investigación

* **Recibido:** 25 de abril de 2022 * **Aceptado:** 16 de mayo de 2022 * **Publicado:** 16 de Junio de 2022

- I. Ingeniero en Estadística Informática, Máster Universitario en Estadística Aplicada, Doctor en Estadística Matemática y Aplicada, Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador.
- II. Ingeniero en Estadística Informática, Máster Universitario en Estadística Aplicada, Universidad Nacional de Chimborazo (UNACH), Riobamba, Ecuador.
- III. Ingeniero en Administración de Empresas, Instituto Nacional de Evaluación Educativa (INEVAL), Quito, Ecuador.
- IV. Ingeniero Estadístico, Instituto Nacional de Evaluación Educativa (INEVAL), Quito, Ecuador.
- V. Ingeniero en Sistemas Informáticos, Magíster en Gestión de Bases de Datos, Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador.

Resumen

La investigación trata sobre la evaluación de los aprendizajes educativos de 4to y 7mo nivel de Educación General Básica del Ecuador, donde se desarrolla una comparativa temporal de los resultados de las evaluaciones en los ciclos 2014, 2015, 2016, 2017, 2018 y 2019, mediante una equiparación con bloques rotados siguiendo la metodología de valores plausibles, considerando un diseño de grupos no equivalentes con covariables (edad y sexo) e ítems ancla. Dentro de los principales resultados, se destaca el uso del estadístico de la mediana, para 4to EGB, la habilidad mediana de los sustentantes desde el ciclo 2014 a 2017 se incrementa. Sin embargo, a partir del ciclo 2017 a 2019 la habilidad disminuye. Luego, para 7mo EGB desde el ciclo 2014 se observa una disminución de la habilidad hasta el ciclo 2016. Seguido, a partir de 2016 se tiene un aumento de la habilidad hasta el ciclo 2019. Finalmente, estos resultados sugieren que los logros de aprendizaje de los estudiantes fluctúan y no muestran un patrón de tendencia concordante para 4to y 7mo en conjunto. Por lo tanto, es necesario una evaluación minuciosa de las políticas a favor del aumento de los logros de aprendizaje del Ministerio de Educación del Ecuador en los periodos de aumento y disminución de los logros de aprendizaje, para detectar fortalezas y áreas de mejora de la calidad educativa en el Ecuador.

Palabras Claves: Evaluación Educativa; Educación General Básica; Equiparación; Psicometría.

Abstract

The research deals with the evaluation of educational learning of the 4th and 7th level of General Basic Education in Ecuador, where a temporary comparison of the results of the evaluations in the 2014, 2015, 2016, 2017, 2018 and 2019 cycles is developed, through a comparison with rotated blocks following the methodology of plausible values, considering a design of non-equivalent groups with covariates and anchor items. Among the main results, the use of the median statistic stands out, for the 4th EGB, the median ability of the supporters from the 2014 to 2017 cycle increases. However, from the 2017 to 2019 cycle the ability decreases. Then, for the 7th EGB from the 2014 cycle, a decrease in ability is observed until the 2016 cycle. Next, from 2016 there is an increase in ability until the 2019 cycle. Finally, these results suggest that the learning achievements of students fluctuate and do not show a concordant trend pattern for 4th and 7th grades combined. Therefore, a thorough evaluation of the policies in favor of the increase in learning achievements of the Ministry of Education of Ecuador

is necessary in the periods of increase and decrease in learning achievements, to detect strengths and areas for quality improvement. education in Ecuador.

Keywords: Educational Evaluation; Basic General Education; Equating; Psychometry.

Resumo

A investigação trata da avaliação da aprendizagem educacional nos 4º e 7º níveis do Ensino Básico Geral no Equador, onde é desenvolvida uma comparação temporal dos resultados das avaliações nos ciclos de 2014, 2015, 2016, 2017, 2018 e 2019, através de uma equação com blocos rotativos seguindo a metodologia de valores plausíveis, considerando uma concepção de grupos não equivalentes com covariatos (idade e sexo) e itens de âncora. Entre os principais resultados, destacamos a utilização da estatística mediana, para o 4º EGB, a capacidade mediana dos tomadores de teste do ciclo de 2014 a 2017 aumenta. Contudo, a partir do ciclo de 2017 a 2019 a capacidade diminui. Depois, para o 7º GBS a partir do ciclo de 2014, observa-se uma diminuição da capacidade até ao ciclo de 2016. Depois, a partir de 2016 há um aumento das capacidades até ao ciclo de 2019. Finalmente, estes resultados sugerem que os resultados de aprendizagem dos estudantes flutuam e não mostram um padrão de tendência concordante para o 4º e 7º ano como um todo. Por conseguinte, é necessária uma avaliação minuciosa das políticas pró-realizações do Ministério da Educação equatoriano durante períodos de aumento e diminuição dos resultados de aprendizagem para detectar pontos fortes e áreas para melhorar a qualidade da educação no Equador.

Palavras-chave: Avaliação Educacional; Educação Básica Geral; Equipose; Psicometria.

Introducción

La evaluación educativa es el ejercicio permanente mediante el cual se pretende estimar y formular juicios sobre el proceso educativo del estudiante, así como sobre sus resultados de aprendizaje con el objetivo de aumentar y/o mantener la calidad de la educación (Ministerio de Educación Nacional, 1997). En este sentido el Instituto Nacional de Evaluación Educativa (Ineval) desarrolla periódicamente el proyecto de evaluación educativa Ser Estudiante (SEST) el cual evalúa los aprendizajes educativos del nivel de Educación General Básica (EGB) y Bachillerato General Unificado (BGU) del Ecuador (Ineval, 2022). Los resultados de la evaluación SEST permiten

conocer los logros de aprendizaje de los estudiantes del Sistema Nacional de Educación, y con base a estos se puede detectar fortalezas y áreas de mejora, para elevar la calidad educativa del Ecuador (Ineval, 2021 a). Si bien es cierto el Ineval desarrolla el proyecto de evaluación educativa SEST periódicamente y presenta resultados sobre este proceso (ver por ejemplo: Ineval, 2021 b; Ineval, 2021 c). Sin embargo, no se tiene una comparativa temporal entre los logros de aprendizaje de los estudiantes, esto no permite tener un panorama claro sobre la evolución del aprendizaje en el Sistema Nacional de Educación del Ecuador y reduce la capacidad de tomar decisiones bien informadas en favor de la calidad de la educación. Por lo tanto, en el marco del proyecto Ser Estudiante se pretende en este trabajo realizar una equiparación de calificaciones de los distintos periodos de evaluación 2013-2019 para **4to y 7mo EGB**, para colocar en una misma escala las puntuaciones de los diferentes periodos de evaluación.

Para lograr el objetivo planteado en el acápite anterior, se describe a continuación la teoría de la equiparación siguiendo a González y Wiberg (2017), junto con el procedimiento de recolección de datos en equiparación. Antes de esto, es importante tener en cuenta que en el proceso de equiparación de puntuaciones en evaluaciones educativas, las diferencias de puntuación no se deben exclusivamente a diferencias en la dificultad de las formas de las pruebas¹. Puede suceder que un grupo de examinados sea simplemente más hábil que el otro, en cuyo caso, una comparación entre grupos mediante el uso de las puntuaciones de las pruebas, confunde la habilidad de los individuos de los dos grupos. En este sentido, una definición de equiparación debe considerar los puntajes en dos formas de prueba como equivalentes en un grupo de sustentantes, si representan la misma posición relativa en el grupo.

Formulación estadística del problema de equiparación

Aunque en muchos casos, se deben equiparar más de dos formas de prueba, para facilitar la exposición, se consideran solo dos formas de prueba y las denotaremos como X e Y . Se supondrá que estas dos formas se administran a n_x y n_y sustantivos seleccionados al azar, respectivamente. Las puntuaciones obtenidas de la forma de prueba X se indican con X_i ($i = 1, \dots, n_x$), y las obtenidas con Y como Y_i ($i = 1, \dots, n_y$). Se supone que las puntuaciones X e Y son variables aleatorias que se definen en consecuencia en los espacios muestrales X e Y , y que siguen las distribuciones F_X y F_Y , respectivamente.

¹ Una forma de prueba es una acomodación específica de las preguntas en una evaluación.

El problema estadístico de la equiparación consiste en modelar la relación entre un puntaje en una forma de prueba y su puntaje correspondiente en otra forma. Matemáticamente, esto significa que se debe definir una función que tome valores en X y dé como resultado un valor en Y . La siguiente definición es general para cualquier función de transformación de equiparación.

Definición 1: Sean X e Y dos espacios muestrales. Una función $\varphi : X \mapsto Y$ se denomina transformación de equiparación.

La definición 1 establece que la transformación de equiparación mapea los puntajes de una forma de prueba en la escala de otra. Una forma matemática explícita para φ , dependerá de la definición de equiparación o, en otras palabras, de la forma en que se operacionalice la idea de tratar las puntuaciones como si procediera de la misma prueba.

Definición 2 (Definición de equiparación de Braun y Holland (1982)): Sean X e Y dos formas de prueba que generan ambos datos de puntuación X e Y , respectivamente. Se dice que las formas X e Y se equiparan en la población T por $\varphi(x)$ si $F_Y(y) = F_{\varphi(x)}(y)$.

La definición 2 indica que dos puntuaciones x e $y = \varphi(x)$ se equiparan si la distribución de Y y la de las puntuaciones transformadas (equiparadas) $\varphi(X)$ son iguales. Bajo esta definición, podemos obtener una forma explícita para φ que iguale X a Y en T . De hecho, si x e y son los cuantiles en las distribuciones de las puntuaciones de las pruebas X e Y para una proporción acumulativa común arbitraria p de la población, es decir, $x(p) = F_X^{-1}(p)$ y $y(p) = F_Y^{-1}(p)$, se sigue que, para cada $p \in [0,1]$, con el requisito de que $F_Y(y) = F_{\varphi(x)}(y)$, se puede obtener una puntuación equivalente y en la prueba Y para una puntuación x en X como

$$y = \varphi(x) = F_Y^{-1}(F_X(x)).$$

Aunque la ecuación anterior se puede utilizar en general para la comparación de dos muestras o distribuciones de variables aleatorias, en la literatura de igualación la función φ se conoce como la transformación equipercentil. La heurística detrás de esta definición es simple: definir puntajes equiparados como aquellos que tienen el mismo rango percentil en un grupo dado de examinados.

Recopilación de datos en equiparación

En la práctica, se necesitan datos de puntuación para la estimación de φ . Debido a que las diferencias en la distribución de puntajes pueden atribuirse a diferencias en la dificultad de las formas de prueba administradas o debido a diferencias en la habilidad de los sustentantes, es necesario desentrañar la confusión antes de que se produzca cualquier vínculo entre X e Y . La

equiparación corrige las diferencias en la dificultad de las formas de prueba, por lo que primero necesitamos un método de corregir las diferencias en la habilidad. El modo en que se realice dicha corrección se guiará por el método en que se recopilen los datos de puntuación. Dependiendo de la situación, se usarán personas o ítems comunes para desentrañar la confusión entre estos dos factores.

Aunque en principio es posible obtener datos de cualquier número de grupos diferentes de examinados dentro de cualquier número de poblaciones diferentes, la exposición que sigue considerará sólo dos poblaciones y dos grupos dentro de cada población. Del mismo modo, sólo se considerarán dos formas de prueba y una prueba de anclaje², aunque en teoría cualquier número de ellos podría usarse para igualar.

Al definir los diseños de equiparación, la población de interés o la población objetivo de la que se obtendrán los datos de puntuación se indicará como T . Si los examinados no pueden considerarse equivalentes, sino que se considera que provienen de diferentes poblaciones, entonces las dos diferentes poblaciones se denotan como P y Q . En tal caso, T será una especie de mezcla de P y Q . La Tabla siguiente muestra un resumen de los diseños de equiparación para recopilar datos de puntuación que se utilizan comúnmente en la literatura de equiparación.

Descripción de los diseños de equiparación con respecto a poblaciones, grupos y qué información se recopila

Tabla 1. Descripción de diseños de equiparación según poblaciones, grupos e información recopilada

Diseño	Grupo	T				P				Q			
		X	Y	A	C	X	Y	A	C	X	Y	A	C
SG	G	x	x										
EG	G_1	x											
	G_2		x										
CB	G_1	x	x										
	G_2	x	x										
NEAT	G_1					x		x					
	G_2										x	x	
NEC	G_1					x			x				
	G_2										x		X

Fuente: (González y Wiberg, 2017)

Elaborado: Autores

² Una prueba de anclaje es un subconjunto de preguntas de una prueba que se repiten en diferentes aplicaciones de la evaluación. Estas preguntas se conocen como ítems ancla.

Donde en la Tabla 1 se representa como: *C*: Covariables, *A*: Prueba de anclaje, *T*: Población objetivo, *X*: Forma de prueba *X*, *Y*: Forma de prueba *Y*, *P*: Población *P*, *Q*: Población *Q*, *SG*: Solo un grupo, *EG*: Grupos equivalentes, *CB*: Contrapesada, *NEAT*: Grupos no equivalentes con prueba de anclaje, *NEC*: Grupos no equivalentes con covariables.

En las diferentes evaluaciones que desarrolla el Ineval, como Ser Estudiante, Ser Maestro, Quiero Ser Maestro Intercultural Bilingüe, etc. Se plantea el uso de varias formas de prueba con ítems anclados a diferentes grupos de la población. Además, se incluye una Encuesta de Factores Asociados que responde a circunstancias que pueden incidir en el desempeño de las y los estudiantes evaluados. Por lo tanto, el diseño de equiparación implícito que ocupa el Ineval es de Grupos no equivalentes con prueba de anclaje (*NEAT*) y/o Grupos no equivalentes con covariables (*NEC*).

Por otro lado, el Ineval en los últimos años tiene una apuesta decidida a la utilización de la Teoría de Respuesta al Ítem (*TRI*) para analizar pruebas, en este sentido, se utiliza para la equiparación la *TRI* en el contexto de una combinación de *NEAT* y *NEC*. Luego, como el objetivo es realizar la equiparación de puntuaciones de los distintos periodos de evaluación para 4to y 7mo EGB del proyecto SEST, se realiza la equiparación con bloques rotados (Kolen y Brennan, 2004) siguiendo la metodología de valores plausibles (Fuentealba, 2020).

Metodología

Participantes: Los participantes de la evaluación de resultados de aprendizaje para 4to EGB de Ecuador fueron 10680, 7589, 8850, 7340, 10715 y 13769 para los ciclos 2013-2014 (2014), 2014-2015 (2015), 2015-2016 (2016), 2016-2017 (2017), 2017-2018 (2018) y 2018-2019 (2019). Mientras que el número de participantes para 7mo EGB fueron 11453, 112368, 10423, 7853, 10930 y 13892 para los ciclos 2014, 2015, 2016, 2017, 2018 y 2019 respectivamente.

Evaluación: El instrumento de evaluación constó de un número determinado de ítems de los campos disciplinares de Matemática, Lengua y Literatura, Ciencias Naturales y Ciencias Sociales. El formato de evaluación fue presencial, esta evaluación se aplica en 4 días, donde se ocupa un día por campo disciplinar y el tiempo de evaluación fue de 120 minutos.

Proceso: La obtención de los datos que se muestran a continuación, se realizó a través de una solicitud ciudadana de datos codificados y anonimizados al Ineval. Esta información responde a

las políticas de datos abiertos del Ineval para salvaguardar los lineamientos de confidencialidad que exige la ley.

1. Mapas técnicos³ de 4to y 7mo de Educación General Básica de los ciclos 2014, 2015, 2016, 2017, 2018 y 2019.
2. Matriz de 0/1⁴ de 4to y 7mo de Educación General Básica de los ciclos 2014, 2015, 2016, 2017, 2018 y 2019.
3. Matriz de factores asociados⁵ (FA) con información de variables primarias como sexo y edad.

Posteriormente se definió el proceso de equiparación del siguiente modo:

1. Se verifica la consistencia entre la información de los mapas técnicos y las matrices de 0/1 correspondientes, verificado el código único de los ítems de ambas matrices.
2. Se verifica la consistencia entre la información de las matrices de 0/1 y la matriz de factores asociados, verificado los códigos de identificación de los sustentantes, según las variables clip, amie y cédula según sea el caso.
3. Se realiza la equiparación con bloques rotados (Kolen y Brennan, 2004) siguiendo la metodología de valores plausibles (Fuentealba, 2020) de 4to y 7mo y 7mo de Educación General Básica de los ciclos 2014, 2015, 2016, 2017, 2018 y 2019. Esto mediante la información de las matrices de 0/1 y de factores asociados. Se considera como escala base al ciclo 2014. La equiparación se desarrolla considerando un diseño de grupos no equivalentes con covariables⁶ e ítems ancla, siguiendo los pasos:
 - a) Preparación de la matriz 0/1 y covariables de los ciclos iniciales y final.
 - b) Calibración de los ciclos inicial y final con covariables mediante regresión latente (Adams y Wu, 2007; Adams, Wilson y Wu, 1997).
 - c) Calibración concurrente de los ciclos inicial y final (Wingersky y Lord, 1984).
 - d) Calibración del ciclo inicial con parámetros de los ítems fijos obtenidos del paso dos (Kim, 2006). Los ítems fijos son aquellos ítems considerados como anclas.

³ Un mapa técnico es un documento que describe la distribución de los ítems en las diferentes formas de la prueba.

⁴ Una matriz de 0/1 presenta información sobre la calificación de los sustentantes de la evaluación, es decir, es un arreglo de filas (sustentantes) y columnas (preguntas) donde se coloca 0 o 1, para incorrecto o correcto respectivamente.

⁵ Una matriz de factores asociados son las respuestas a una encuesta que es llenada por los sustentantes previos a la evaluación.

⁶ Las covariables utilizadas fueron edad y sexo debido a que estas covariables presentaron información completa en la encuesta de factores asociados.

- e) Obtención de constantes de transformación lineal, utilizando el método de momentos y el enfoque de valores plausibles. Es importante, guardar las constantes de transformación lineal en todas las equiparaciones entre ciclos.
 - f) Escalamiento del ciclo final a la escala del ciclo inicial utilizando los resultados del paso anterior. Se guardan, las equiparaciones obtenidas entre los ciclos.
4. Mediante un proceso de encadenamiento (es decir, transformado secuencialmente), las constantes de transformación lineal obtenidas en el paso anterior se utilizan para colocar las equiparaciones de los diferentes ciclos en la escala base 2014. Después de este proceso de encadenamiento, todas las distribuciones de habilidad estimadas están en la escala base 2014. Además, la media y la desviación estándar de las distribuciones de habilidad estimadas transformadas a la escala base 2014 se pueden utilizar para comparar la diferencia en la habilidad media y la variabilidad en los diferentes ciclos de evaluación (Kolen y Brennan, 2004).

Es importante indicar que para el procedimiento de equiparación se utilizó básicamente el paquete “TAM” de los autores Kiefer, Robitzsch y Wu (2016) del software estadístico R.

Resultados

Consistencia entre la información de las matrices de 0/1 y la matriz de factores asociados

En general para 4to y 7mo EGB, existe una buena consistencia entre la información de las matrices de 0/1 y la matriz de factores asociados, esto permite seguir el procedimiento planteado para la equiparación mediante covariables.

Tabla 2. 4to EGB: Consistencia entre la información de las matrices de 0/1 y la matriz de factores asociados, según ciclos

Ciclo	Sustentantes, matriz 0/1	Ítems matriz 0/1	Sustentantes matriz 0/1 y FA	cruce	NAs variable sexo	NAs variable edad
2014	10680	120	10659		0	0; 3*
2015	7589	150	7589		0	0
2016	8850	142	8850		0	0; 1155*
2017	7340	149	7276		0	0
2018	10715	148	10715		0	2
2019	13769	123	13686		0	0

*Valores iguales o inferiores a 0, que fueron imputados mediante la mediana.

Fuente: (Ineval, 2022)

Elaborado: Autores

Tabla 3. 7mo EGB: Consistencia entre la información de las matrices de 0/1 y la matriz de factores asociados, según ciclos

Ciclo	Sustentantes, matriz 0/1	Ítems matriz 0/1	Sustentantes matriz 0/1 y FA	cruce	NAs variable sexo	NAs variable edad
2014	11453	166	11432		0	0; 4*
2015	112368	219; 186**	112350		0	0
2016	10423	175	10422		0	0; 1119*
2017	7853	169	7846		0	0
2018	10930	168	10930		0	1
2019	13892	136	13803		0	0

*Valores iguales o inferiores a 0, que fueron imputados mediante la mediana.

**Número de ítems luego de quitar aquellos con más del 96% de valores faltantes. Esto, debido a que el proceso de calibración no converge sin este cambio.

Fuente: (Ineval, 2022)

Elaborado: Autores

Cantidad de ítems ancla entre ciclos

Tanto para 4to y 7mo EGB, se tiene una buena cantidad de ítems ancla entre ciclos contiguos que permiten desarrollar el procedimiento planteado para la equiparación mediante covariables e ítems ancla.

Tabla 4. Cantidad de ítems ancla según ciclo y nivel educación general básica

Ciclo Inicial	Ciclo Final	4to EGB: Ítems ancla	7mo EGB: Ítems ancla
2014	2015	75	77
2015	2016	75	109
2016	2017	69	81
2017	2018	84	102
2018	2019	88	104

Fuente: (Ineval, 2022)

Elaborado: Autores

Escalamiento y constantes de transformación lineal

Para realizar la equiparación se utiliza la transformación $\varphi : \theta_{ciclo\ final} \mapsto \theta_{ciclo\ final\ equiparado\ a\ ciclo\ inicial}$, esto en forma de ecuación lineal es: $\theta_{ciclo\ final\ equiparado\ a\ ciclo\ inicial} = B * \theta_{ciclo\ final} - A$. Las constantes de transformación lineal se obtienen, utilizando el método de momentos y el enfoque de valores plausibles como se indica a continuación:

$$B = \frac{\sigma_{VP \text{ del ciclo inicial con covariables}}}{\sigma_{VP \text{ del ciclo inicial con parámetros fijos}}}$$

$$A = \mu_{VP \text{ del ciclo inicial con covariables}} - B \mu_{VP \text{ del ciclo inicial con parámetros fijos}}$$

Los resultados del cálculo de las constantes de transformación lineal se presentan en la Tabla 5 y 6.

Tabla 5. Constantes de transformación lineal de 4to EGB, según ciclo inicial y final de la equiparación

Ciclo Inicial	Ciclo Final	A	B
2014	2015	0.1225	0.9352
2015	2016	0.2187	0.8228
2016	2017	-0.0179	0.9324
2017	2018	0.2417	0.9702
2018	2019	-0.4270	0.9645

Fuente: (Ineval, 2022)

Elaborado: Autores

Tabla 6. Constantes de transformación lineal de 7mo EGB, según ciclo inicial y final de la equiparación

Ciclo Inicial	Ciclo Final	A	B
2014	2015	0.1222	1.0162
2015	2016	-0.2860	0.9873
2016	2017	0.3528	0.9577
2017	2018	0.2747	0.9822
2018	2019	0.2089	0.9598

Fuente: (Ineval, 2022)

Elaborado: Autores

Equiparación con bloques rotados siguiendo la metodología de valores plausibles y proceso de encadenamiento de 4to y 7mo EGB

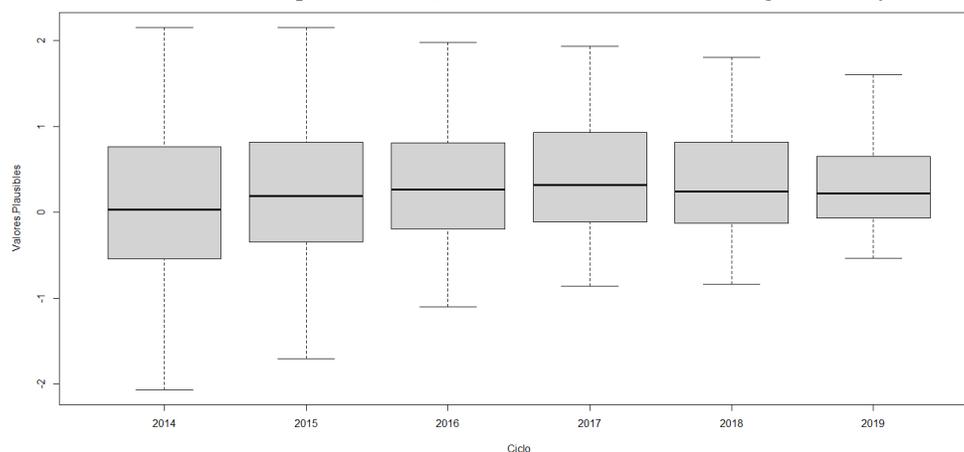
Los resultados del proceso de equiparación de 4to EGB se pueden observar en la Figura 1 y Tabla 7. El estadístico de la mediana muestra que la habilidad de los sustentantes desde el ciclo 2014 a 2017 se incrementó. Sin embargo, a partir del ciclo 2017 a 2019 la habilidad disminuyó. La variabilidad de las distribuciones de habilidad disminuye conforme pasan los ciclos.

Como una forma de contrastar los resultados de equiparación se muestra los resultados de la estimación de las distribuciones de habilidad sin considerar la equiparación. En el caso de 4to EGB se muestra esta información en la Figura 2 y Tabla 8. Al comparar los resultados de manera rústica

y utilizando la mediana, se tiene que en generar la habilidad aumenta y luego disminuye en el siguiente ciclo. Claramente, estos resultados son erróneos y se presentan sólo para analizar, qué pasaría si no se considera la equiparación en la comparación de resultados de evaluación estandarizada.

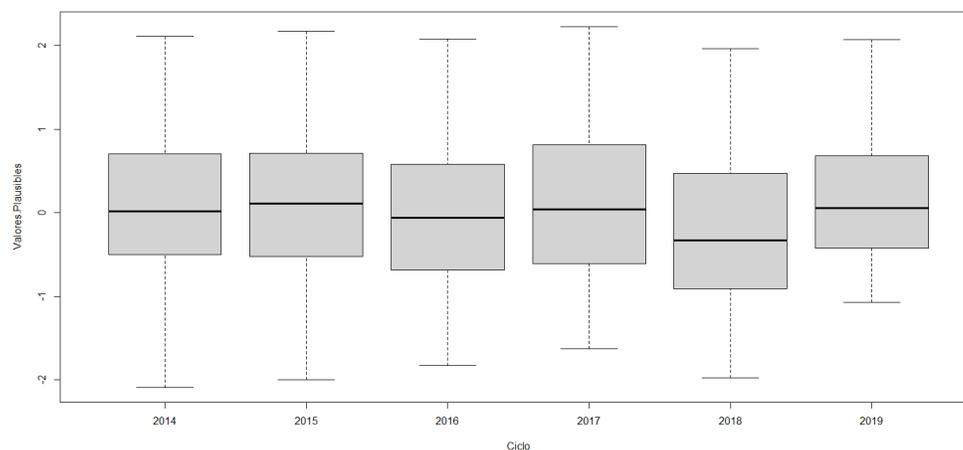
Por otro lado, utilizando la mediana para 7mo EGB desde el ciclo 2014 se observa una disminución de la habilidad hasta el ciclo 2016. Luego, a partir de 2016 se tiene un aumento de la habilidad hasta el ciclo 2019. La desviación estándar se mantiene estable para todos los ciclos de evaluación (ver Figura 3 y Tabla 9). Estos resultados contrastan con los obtenidos en la comparación sin considerar la equiparación (ver Figura 4 y Tabla 10).

Figura 1. Distribución de valores plausibles EQUIPARADOS de 4to EGB, según ciclos y sin datos atípicos



Elaborado: Autores

Figura 2. Distribución de valores plausibles NO EQUIPARADOS de 4to EGB, según ciclos y sin datos atípicos



Elaborado: Autores

Tabla 7. Estadísticas descriptivas de los valores plausibles EQUIPARADOS de 4to EGB, según ciclos

Estadístico	2014	2015	2016	2017	2018	2019
Mínimo	-2.0679	-1.7072	-1.1011	-0.8582	-0.8350	-0.5340
1er cuartil	-0.5416	-0.3442	-0.1931	-0.1093	-0.1277	-0.0691
Mediana	0.0302	0.1922	0.2687	0.3167	0.2471	0.2239
Media	0.1035	0.2356	0.3186	0.4155	0.3498	0.3151
3er cuartil	0.7645	0.8150	0.8135	0.9300	0.8189	0.6514
Máximo	2.1568	2.1571	1.9790	1.9351	1.8061	1.6022
Desviación Estándar	0.8819	0.7990	0.6669	0.6502	0.6085	0.4984

Fuente: (Ineval, 2022)

Elaborado: Autores

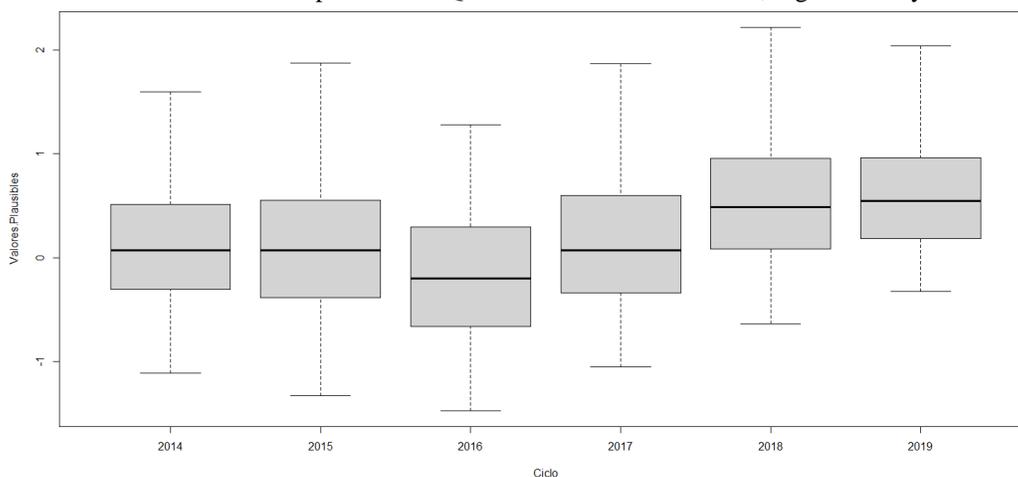
Tabla 8. Estadísticas descriptivas de los valores plausibles NO EQUIPARADOS de 4to EGB, según ciclos

Estadístico	2014	2015	2016	2017	2018	2019
Mínimo	-3.2021	-3.6337	-3.4757	-3.0909	-3.3719	-2.5864
1er cuartil	-0.5229	-0.5319	-0.6959	-0.6231	-0.9220	-0.4388
Mediana	0.0169	0.1073	-0.0586	0.0384	-0.3264	0.0599
Media	0.0989	0.1211	-0.0162	0.1363	-0.1993	0.1925
3er cuartil	0.7472	0.7258	0.5975	0.8310	0.4898	0.7195
Máximo	3.3809	2.9133	3.0128	3.0396	2.9109	2.8442
Desviación Estándar	1.0032	0.9105	0.9142	0.9553	0.9284	0.8186

Fuente: (Ineval, 2022)

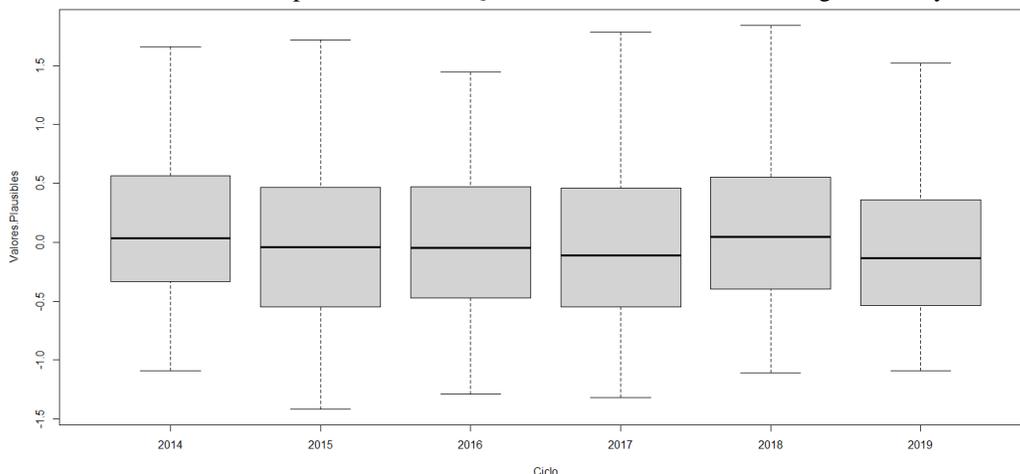
Elaborado: Autores

Figura 3. Distribución de valores plausibles EQUIPARADOS de 7mo EGB, según ciclos y sin datos atípicos



Elaborado: Autores

Figura 4. Distribución de valores plausibles NO EQUIPARADOS de 7mo EGB, según ciclos y sin datos atípicos



Elaborado: Autores

Tabla 9. Estadísticas descriptivas de los valores plausibles EQUIPARADOS de 7mo EGB, según ciclos

Estadístico	2014	2015	2016	2017	2018	2019
Mínimo	-1.1157	-1.3296	-1.4785	-1.0546	-0.6385	-0.3245
1er cuartil	-0.3064	-0.3865	-0.6616	-0.3398	0.0842	0.1860
Mediana	0.0736	0.0690	-0.1994	0.0696	0.4863	0.5499
Media	0.1221	0.1009	-0.1703	0.1556	0.5515	0.6064
3er cuartil	0.5102	0.5560	0.2937	0.5992	0.9560	0.9612
Máximo	1.6012	1.8739	1.2839	1.8726	2.2226	2.0419
Desviación Estándar	0.5927	0.6700	0.6228	0.6390	0.6064	0.5272

Fuente: (Ineval, 2022)

Elaborado: Autores

Tabla 10. Estadísticas descriptivas de los valores plausibles NO EQUIPARADOS de 7mo EGB, según ciclos

Estadístico	2014	2015	2016	2017	2018	2019
Mínimo	-1.0912	-1.4179	-1.2877	-1.3169	-1.1101	-1.0907
1er cuartil	-0.3320	-0.5491	-0.4737	-0.5496	-0.3953	-0.5363
Mediana	0.0364	-0.0417	-0.0454	-0.1094	0.0440	-0.1341
Media	0.1118	-0.0191	0.0023	-0.0304	0.1142	-0.0551
3er cuartil	0.5626	0.4686	0.4710	0.4614	0.5542	0.3603
Máximo	1.6604	1.7166	1.4474	1.7825	1.8436	1.5184
Desviación Estándar	0.5946	0.6719	0.5986	0.6717	0.6381	0.5954

Fuente: (Ineval, 2022)

Elaborado: Autores

Conclusiones

Esta investigación desarrolló con éxito la equiparación con bloques rotados siguiendo la metodología de valores plausibles, para 4to y 7mo de Educación General Básica de los ciclos 2014, 2015, 2016, 2017, 2018 y 2019. Esto utilizando covariables de encuesta de factores asociados. Se utilizó un proceso de encadenamiento, es decir, una transformación secuencia entre ciclos contiguos, para que sean comparables las diferencias en la media y la desviación estándar de las distribuciones de habilidad estimadas de los diferentes ciclos, con respecto a la escala base 2014. El estadístico de la mediana muestra que, para 4to EGB, la habilidad mediana de los sustentantes desde el ciclo 2014 a 2017 se incrementa. Sin embargo, a partir del ciclo 2017 a 2019 la habilidad disminuye. Con respecto a la variabilidad de las distribuciones de habilidad, esta disminuye conforme pasan los ciclos. Por otro lado, utilizando la mediana para 7mo EGB desde el ciclo 2014 se observa una disminución de la habilidad hasta el ciclo 2016. Luego, a partir de 2016 se tiene un aumento de la habilidad hasta el ciclo 2019. La desviación estándar se mantiene estable para todos los ciclos de evaluación.

Los resultados obtenidos del proceso de equiparación para 4to y 7mo EGB en los ciclos 2014, 2015, 2016, 2017, 2018 y 2019 muestran que los logros de aprendizaje de los estudiantes fluctúan y no muestran un patrón de tendencia concordante para 4to y 7mo en conjunto. Por lo tanto, es necesario una evaluación minuciosa de las políticas a favor del aumento de los logros de aprendizaje del ministerio de educación del Ecuador en los periodos de aumento y disminución de los logros de aprendizaje, para detectar fortalezas y áreas de mejora de la calidad educativa en el Ecuador.

Dentro de los limitantes de este trabajo se encuentra la imposibilidad de realizar la equiparación discriminando entre los campos evaluados de Matemática, Lengua y Literatura, Ciencias Naturales y Ciencias Sociales. Esto debido a que no se cuenta con ítems ancla que permitan la implementación de la metodología utilizada en esta investigación.

Finalmente, se detectó que es necesario en el Ineval, establecer un diseño de equiparación estándar para el Proyecto de evaluación Ser Estudiante, entre ciclos. También, sería óptimo instaurar un diseño de equiparación, entre niveles de Educación General Básica, para evaluar el crecimiento del desempeño según los niveles de EGB.

Referencias

1. Adams, R. J., & Wu, M. L. (2007). The mixed-coefficients multinomial logit model. A generalized form of the Rasch model. In M. von Davier & C. H. Carstensen (Eds.): *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 55-76). New York: Springer. doi:10.1007/9780387498393_4
2. Adams, R. J., Wilson, M. R., and Wu, M. L. (1997). Multilevel item response modelling: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, 22, 47-76.
3. Braun, H., & Holland, P. (1982). Observed-score test equating: a mathematical analysis of some ETS equating procedures. In P. Holland & D. Rubin (Eds.), *Test equating* (Vol. 1, pp. 9–49). New York: Academic Press.
4. Fuentealba, F. (2020). Sesión 3 – ERCE y Rosseta Stone, ERCE – Metodología y Aplicación. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.
5. González, J., & Wiberg, M. (2017). *Applying test equating methods*. Cham: Springer International Publishing.
6. Ineval. (2021 a). Ficha técnica, Ser Estudiante, Cuarto de Educación General Básica. <http://evaluaciones.evaluacion.gob.ec/BI/ser-estudiante/>
7. Ineval. (2021 b). Acción 21-3: Capacidad en la resolución de problemas en los sistemas informatizados, de las personas que participaron en la tercera ronda del Programa para la Evaluación Internacional de las Competencias de los Adultos (PIAAC). <http://evaluaciones.evaluacion.gob.ec/BI/ser-estudiante/>
8. INEVAL. (2021 c). Acción Boletines de Investigación y Evaluación: Involucramiento parental y rendimiento académico en escolares de 7mo de Educación General Básica.
9. Ineval. (2022). Ser Estudiante. Consultado el 24 de febrero de 2022. <http://evaluaciones.evaluacion.gob.ec/BI/ser-estudiante/>
10. Kiefer, T., Robitzsch, A., & Wu, M. (2016). TAM: Test analysis modules. R Package Version 1.995-0.
11. Kim, S. (2006). A comparative study of IRT fixed parameter calibration methods. *Journal of Educational Measurement*, 43(4), 355–381.

12. Kolen, M. J., Brennan, R. L., & Kolen, M. J. (2004). Test equating, scaling, and linking: Methods and practices (pp. 177-180). New York: Springer.
13. Ministerio de Educación Nacional. (1997, p. 17). Evaluación en el aula y más allá de ella. Bogotá.
14. Wingersky, M. S., & Lord, F. M. (1984). An investigation of methods for reducing sampling error in certain IRT procedures. *Applied Psychological Measurement*, 8(3), 347–364.

© 2022 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).