

Estimación De Los Resultados En Matemáticas Y Ciencias De Las Pruebas TIMSS 2015: Un Nuevo Enfoque Desde La Metodología De Áreas Pequeñas

Estimation Of The Mathematics And Science Results Of The 2015 TIMSS Test: A New Approach From The Small Area Estimation Methodology

Cristian Fernando Tellez Piñerez^a
cristiantellez@usantotomas.edu.co

Leonardo Trujillo-Oyola^b
ltrujilloo@unal.edu.co

Andrés Felipe Pedraza Triviño^c
andrespedrazat@usantotomas.edu.co

Resumen

Las pruebas TIMSS son un instrumento que se aplica a nivel internacional con el fin de medir las tendencias en los logros de Matemáticas y Ciencias en los grados cuarto de primaria y octavo de bachillerato. Dado el diseño de estas pruebas, es posible obtener resultados trazables y comparables, que permiten a gobiernos de distintas naciones generar, mantener o eliminar políticas públicas con el objetivo de mejorar su sistema educativo. Naturalmente, estas pruebas han adquirido gran relevancia a nivel mundial. Teniendo en cuenta lo anterior, el objetivo de este trabajo académico es aplicar un estimador de áreas pequeñas que permita mejorar la calidad de las estimaciones que se obtienen actualmente con el diseño de muestreo empleado en TIMSS 2015. Adicional a lo anterior, esta metodología permite estimar el resultado que hubiese obtenido un país que no presentó esta prueba en el año 2015, pero que cuenta con información auxiliar relacionada con el logro académico. El estimador que se aplicará para estimar el logro académico utiliza el modelo de Fay-Herriot. Este modelo emplea variables auxiliares como la información demográfica, socioeconómica, entre otros. Finalmente se mostrará con este trabajo que la utilización de la estimación en áreas pequeñas (SAE, por sus siglas en inglés), en pruebas internacionales, mejora en gran parte la eficiencia de los resultados obtenidos y, por ende, permite una mejor toma de decisiones en política educativa.

Palabras Clave: TIMSS, SAE, Educación, Fay-Herriot

Abstract

The TIMSS tests are an international instrument applied in order to measure trend achievement in mathematics and science in the fourth and eighth grades of high school. Given the test's design, it is possible to obtain traceable and comparable results. Which allows governments across the globe to generate, maintain or eliminate public policies that aim to improve their educational system. Naturally, these tests have become very important and known worldwide. Therefore, this academic work focuses on the application of a smaller areas estimator that will allow an improved quality in the given estimates

^aFacultad de Estadística, Universidad Santo Tomás

^bDepartamento de Estadística, Universidad Nacional de Colombia

^cFacultad de Estadística, Universidad Santo Tomás

with the TIMSS 2015 design. In addition, this methodology generates a speculated result that would have been obtained by a country that did not present the test in 2015 but does count with auxiliary information related to academic achievement. The estimator that will be applied to estimate academic achievement, uses the Fay-Herriot model. This model uses auxiliary variables such as demographic and socioeconomic information, among others. Finally, this document exhibits how the use of Small Area Estimation (SAE), in international tests, greatly improves efficiency.

Keywords: TIMSS, SAE, Education, Fay-Herriot

1. Introducción

La educación es un proceso de formación permanente fundamentado en la integridad del ser humano y considerada motor de desarrollo por las naciones del mundo. Por tal razón, se hace importante para un país contar con un sistema de control que permita regularla. Para tal fin es pertinente utilizar metodologías evaluativas que se encuentren a la vanguardia, que permitan medir con certeza y fiabilidad la calidad y la eficiencia del proceso educativo. Este, al ser un proceso prolongado, requiere evaluaciones en puntos intermedios del ciclo con el fin de realizar ajustes y correcciones al aprendizaje y así garantizar que los jóvenes cumplirán con las expectativas y necesidades de la sociedad.

En Colombia, existe un organismo oficial encargado de administrar la educación en el país. Este organismo es el ministerio de educación. La gestión de esta entidad está enfocada en asegurar la calidad de la educación mediante la creación, mantenimiento y eliminación de políticas públicas; al mismo tiempo que mantienen la sinergia entre las diferentes organizaciones que participan en las etapas de evaluación de las áreas en que se desenvuelve la dinámica educativa.

El sistema educativo colombiano lo conforman: la educación preescolar, la educación básica (primaria y cuatro grados de secundaria), la educación media (últimos dos grados de secundaria) y la educación superior (Mineducación, 2010). Cada etapa del ciclo educativo es medida por unos estándares básicos de competencias que definen la calidad que se aspira a alcanzar. Bajo estos mismos estándares son diseñados o adoptados métodos e instrumentos que permiten capturar información que muestre si una persona, institución educativa o región geográfica no alcanza, alcanza o supera las expectativas de la comunidad.

Para llevar a cabo la tarea de evaluación, se designó al Icfes (Instituto Colombiano para la Evaluación de la Educación, antes: Instituto Colombiano para el Fomento de la Educación Superior). Este organismo aplica diferentes pruebas en todos los niveles educativos y recopila la información con el fin de analizar factores incidentes en la calidad educativa. Dichas pruebas son test estandarizados que permiten tener resultados válidos, fiables, exactos y trazables.

Actualmente, el Icfes mide la calidad de la educación con las siguientes pruebas: Saber 3, 5 y 9, Saber 11, entre otras (ICFES, 2015). Ahora bien, existe la necesidad de comparar los resultados de calidad en educación a nivel internacional. Las comparaciones internacionales brindan la oportunidad de identificar fortalezas y debilidades y, mirando a otros, pensar en la manera de optimizar el sistema educativo actual y que ello se refleje en la transformación de los modelos (Schleicher, 2008); transformación que es importante para que un país pueda ser competitivo en un mundo donde los modelos educativos cambian con rapidez. Para lograr este objetivo, el Icfes administra la aplicación de las pruebas internacionales como TIMSS (Trends in International Mathematics and Science Study) (Mullis et al., 2015). Esta prueba es diseñada por el IEA (International Association for the Evaluation of Educational Achievement), y es aplicada a grupos de estudiantes de 4° y 8° previamente seleccionados utilizando un diseño muestral multietápico. TIMSS busca evaluar los conocimientos adquiridos por los estudiantes en Matemáticas y Ciencias.

Conociendo la importancia y la naturaleza de estas pruebas, es indispensable aplicarlas periódicamente; sin embargo, el costo económico que acarrea la planificación, aplicación, levantamiento, análisis y divulgación de la información es elevado. Esto puede derivar en que escuelas, países, u otras unidades de interés no tengan información suficiente para que las estimaciones directas sean eficientes y preci-

sas. Para este escenario se requiere alguna técnica que mejore dichas estimaciones. Es aquí donde se hace necesario recurrir a las metodologías de estimación en áreas pequeñas. Sin embargo, la inclusión de estimación en áreas pequeñas en la estimación del puntaje medio de la habilidad representa un gran desafío en cuanto a la construcción del estimador mismo, además de su error cuadrático medio (MSE, por sus siglas en inglés), debido a que el estimador debe combinar la teoría de respuesta al ítem, la imputación múltiple y un modelo en áreas pequeñas. Este tema es abordado por Tellez (2021), quién en su trabajo doctoral construye este estimador, incorporando como modelo en áreas pequeñas un modelo de Fay-Herriot. En este trabajo se pone a prueba el estimador propuesto bajo un ejercicio de simulación, estimando el puntaje promedio en las pruebas PISA (Martinez, 2006) y en las pruebas Saber 3°, 5° y 9° y demostró de manera consistente que el estimador es insesgado y, que ante diferentes escenarios de estrés (escenarios en los cuales se disminuía el tamaño de los dominios) presentaba un menor error estándar relativo promedio (*EERP*) comparado con otros estimadores. Es por esto que, en el presente trabajo, se plantea como objetivo principal aplicar este estimador para medir la habilidad promedio a nivel país en las pruebas TIMSS 2015. Como segundo objetivo, se pretende demostrar que el modelo propuesto es capaz de predecir de manera confiable los resultados de algún país que por algún motivo no haya podido presentar las pruebas.

La información recopilada de años anteriores evidencia que el desempeño de un establecimiento tiene una alta correlación con el desempeño medido en años anteriores, lo cual sugiere que la información de años pasados puede ser utilizada como predictor del rendimiento futuro bajo la construcción de un modelo estadístico (Tellez, 2021). Del mismo modo variables que expliquen el contexto macroeconómico de un país pueden utilizarse al momento de intentar explicar el logro académico del mismo (Treviño et al., 2016). Si esto es aprovechado e incluido en la estrategia de estimación utilizando SAE, la calidad de las estimaciones puede mejorar significativamente; a su vez, si un país por motivos externos puede llegar a no presentar estas pruebas, los resultados que hubiese obtenido pueden predecirse con gran precisión utilizando SAE.

El presente artículo se estructura de la siguiente manera. En el capítulo 2 se introducen los conceptos teóricos de las pruebas TIMSS, describiendo a grandes rasgos la metodología que se utiliza para estimar la habilidad media en las pruebas de matemáticas y ciencias. El capítulo 3 presenta el modelo que será aplicado, y exhibe los fundamentos matemáticos y la justificación teórica del modelo escogido. En el Capítulo 4 se realiza un ejercicio comparativo entre el estimador actual y el estimador propuesto, logrando evidenciar que este último es mejor en términos de error cuadrático medio relativo. En esa misma dirección, en el mismo capítulo se realiza un ejercicio de predicción, eliminando un país de la base disponible para ajustar los modelos y, haciendo uso de las variables auxiliares, se estimó su puntaje. Este ejercicio arrojó resultados satisfactorios, mostrando que las predicciones diferían de los puntajes observados en menos del 5 %. En el Capítulo 5 se concluye que la metodología propuesta por Tellez (2021) se convierte en una alternativa útil para medir la habilidad media por país de manera eficaz y confiable, capaz de mejorar la calidad de las estimaciones que se obtienen de los diseños de muestreo actuales y con el beneficio de predecir resultados de aquellos países que no presentan la prueba. Sin embargo, cabe resaltar que el objetivo de este trabajo no es que disminuya la participación de países en las pruebas, sino mejorar la eficiencia de las estimaciones obtenidas por la metodología actual y poder llegar a predecir los resultados de algunos países que por algún motivo no presentaron la prueba pero que lo venían haciendo de manera continua.

2. Conceptos teóricos de las pruebas TIMSS

Las pruebas TIMSS son un instrumento que permite medir las tendencias en los logros de Matemáticas y Ciencias en los grados cuarto y octavo. Estas pruebas fueron diseñadas por el IEA y han sido aplicadas desde 1995 hasta la fecha. Desde entonces, han sido actualizadas y aplicadas cada 4 años. Estas pruebas tienen un diseño cuasi-longitudinal con un cohorte de estudiantes de cuarto grado evaluado cuatro años después en octavo grado (Mullis et al., 2015). Las pruebas TIMSS utilizan el currículo como el principal concepto de organización para investigar cómo los países participantes ofrecen oportunidades de educa-

ción en Matemáticas y Ciencias a los estudiantes, y los factores relacionados a cómo los estudiantes están aprovechando estas oportunidades (Ferrera, 2014).

TIMSS permite a los países participantes, que para el caso de este artículo se tomarán como los dominios $d = 1, 2, \dots, D$, tomar decisiones basadas en la evidencia para mejorar las políticas educativas. Esto gracias a que las pruebas permiten medir la eficacia de los sistemas educacionales en un contexto global; adicional a esto, permite identificar brechas en las oportunidades de aprendizaje. A continuación, se describen los conceptos básicos de esta prueba.

2.1. Población y diseño de muestreo

TIMSS define su población objetivo con base en el número de años de educación formal que los estudiantes han recibido. Así, la población objetivo se define en el grado inferior como todos los estudiantes en su cuarto año de educación formal y en el grado superior como todos los estudiantes en su octavo grado de educación formal. La UNESCO provee un esquema de clasificación que describe los niveles de escolaridad (UNESCO, 2012). En este orden de ideas, la población objetivo de TIMSS se define como:

1. **Cuarto grado:** Todos los estudiantes matriculados en el grado que representa cuatro años de escolaridad contando desde el primer año del nivel 1 de la clasificación internacional normalizada de la educación (ISCED, por sus siglas en inglés), siempre que la edad media en el momento del examen sea al menos 9.5 años.
2. **Octavo grado:** Todos los estudiantes matriculados en el grado que representa ocho años de escolaridad contando desde el primer año del nivel 1 del ISCED, siempre que la edad media en el momento del examen sea al menos 13.5 años.

Todas las escuelas que tengan estudiantes matriculados en el grado objetivo son parte de la población objetivo internacional, independientemente el tipo de escuela y si está o no bajo la autoridad del ministerio nacional de educación. Ahora bien, para cumplir con los estándares de precisión, las muestras nacionales de estudiantes deben proporcionar un error estándar no mayor a 0.035 desviaciones para el logro medio del país. Asimismo, cualquier estimación porcentual que se haga a nivel estudiante debe tener un intervalo de confianza de $\pm 3.5\%$. Para la mayoría de los países, los requerimientos de precisión se logran con una muestra de 150 escuelas siendo esto aproximadamente 4000 estudiantes de cada grado objetivo.

Por otro lado, para seleccionar a los estudiantes que presentarán las pruebas en cada país, se utiliza un diseño de muestreo probabilístico multietápico y estratificado. En la primera etapa del diseño, las escuelas son muestreadas con probabilidades proporcionales al número de estudiantes (PPT sin reemplazo). Adicional a esto, las escuelas, en el marco muestral, deben estar estratificadas de acuerdo con variables demográficas tales como, urbano, rural, público privado, entre otros. La selección de las escuelas se hace mediante un mecanismo sistemático, lo anterior conlleva a que, en la primera etapa, el diseño empleado para la selección de las escuelas es un diseño PPT - Sistemático. Para efectos de reemplazos, se asignan por cada escuela seleccionada, dos de reemplazo, esto para el caso en que la escuela originalmente seleccionada rechace participar en la prueba.

La segunda etapa consiste en seleccionar uno o más salones de clases completos del grado objetivo de cada escuela participante. El muestreo de clases es dirigido por el NRC (National Research Coordinator quien es la persona responsable de la implementación de TIMSS en cada país) haciendo uso del software *Within-School Sampling Software* (WinW3S) desarrollado por el IEA y Statistic Canada (Joncas, 2007). Una vez que la escuela acepta participar en la evaluación, el NRC solicita información sobre el número de clases y maestros en la escuela y la ingresa en la base de datos de WinW3S. Las clases más pequeñas que un tamaño mínimo especificado, se agrupan en pseudo-clases antes del muestreo. El software selecciona

clases con probabilidades iguales dentro de las escuelas (utilizando un muestreo aleatorio simple). Todos los estudiantes en cada clase muestreada participan en la evaluación. Las clases muestreadas que se niegan a participar no pueden ser reemplazadas. Ahora bien, para los países que participan en TIMSS, a los estudiantes dentro de una clase muestreada, se les asigna al azar un cuadernillo de TIMSS a través de un sistema de rotación de cuadernillos. Esto se hace para garantizar que TIMSS se administra a muestras probabilísticamente equivalentes.

En cuanto a la estratificación del diseño, esta consiste en agrupar las escuelas de acuerdo con características comunes entre ellas, por ejemplo, tipo de escuela, región geográfica, desempeño en exámenes nacionales, etc.). La estratificación puede tomar dos formas: explícita e implícita; en la primera, se construye un marco de muestreo para cada estrato y una muestra de escuelas se extrae de cada estrato. En TIMSS, la mayor razón para considerar la estratificación explícita es cuando la distribución de escuelas entre los estratos es desproporcionada. Ahora bien, la estratificación implícita consiste en ordenar las escuelas de acuerdo con una o más variables de estratificación entre cada estrato explícito o en todo el marco muestral. La combinación entre la estratificación implícita y el muestreo sistemático es una manera efectiva para asegurar una asignación de estudiantes a la muestra de forma proporcional al tamaño de los estratos.

El marco muestral de escuelas es organizado de acuerdo con las variables de estratificación de forma tal, que las escuelas adyacentes son tan similares como es posible. Sin importar que variables se escojan para la estratificación, el tamaño de las escuelas siempre se incluye como una variable de estratificación implícita.

2.2. Modelo de calificación

Para la calificación de las pruebas y obtención de la habilidad estimada θ , para cada individuo j en el dominio d , TIMSS utiliza tres modelos distintos de la Teoría de respuesta al ítem (TRI, por sus siglas en inglés), dependiendo del tipo del ítem i que se califique. Cada uno de estos modelos describe la probabilidad de que un estudiante j responda de cierta manera a un ítem i dependiendo de su habilidad θ_j , la cual es inobservable o latente. Para los ítems de selección múltiple con única respuesta, se utiliza un modelo logístico de tres parámetros (3PL) desarrollado por Birnbaum (1958), y para los ítems que solo cuentan con dos opciones de respuestas se utiliza los modelos de dos parámetros logísticos (2PL) (de Andrade et al., 2000) los cuales se definen como:

$$P_i(\mathbf{U}_{ij} = 1 \mid \theta_j, a_i, b_i, c_i) = c_i + \frac{(1 - c_i)}{1 + \exp [(-a_i(\theta_j - b_i))]} \quad (1)$$

donde \mathbf{U}_{ij} es una variable dicotómica y toma el valor de uno cuando el individuo j responde correctamente al ítem i y cero si no responde de forma correcta con $j = 1, 2, \dots, n_d$, donde n_d es el número de individuos en el colegio o dominio d , b_i es la dificultad del ítem, a_i es la discriminación y c_i se define como el pseudo azar del ítem. A su vez, θ_j y b_i están en la misma escala y toman valores en el intervalo $(-\infty, \infty)$. Para el caso del modelo de 2PL, se asume $c_i = 0$.

Por último, se utiliza el modelo de crédito parcial para calificar aquellos ítems que tiene 0, 1 y 2 como los niveles de calificación posibles. Estos ítems políticos se escalaron utilizando un modelo generalizado de crédito parcial (Muraki, 1992). El modelo de calificación de un estudiante con habilidad θ_j en la escala k , para el i -ésimo ítem, una respuesta ϕ_{ij} que se califica en la l -ésima de m_i categorías ordenadas de puntuación se define como,

$$P(\phi_{ij} = l \mid \theta, a_i, b_i, d_{i,1}, \dots, d_{i,m_i-1}) = \frac{\exp \left(\sum_{v=0}^l 1.7 \cdot a_i \cdot (\theta_j - b_i + d_{i,v}) \right)}{\sum_{g=0}^{m_i-1} \exp \left(\sum_{v=0}^g 1.7 \cdot a_i \cdot (\theta_j - b_i + d_{i,v}) \right)} \quad (2)$$

donde, $d_{i,l}$ es el parámetro de umbral de categoría l , m_i es el número de categorías de respuesta para el ítem i , usualmente 3.

3. Estimador empleado

La estimación en áreas pequeñas hace referencia a la estimación para dominios que tienen un tamaño de muestra efectivo relativamente pequeño, el cual hace que las estimaciones directas calculadas tengan errores de muestreo altos y por lo tanto inhabilitan su utilización. Los métodos para llevar a cabo esta clase de estimaciones están basados en el ajuste de modelos mixtos, que toman en cuenta las variaciones existentes dentro de cada dominio (países). El ajuste de este tipo de modelos permite aumentar la información disponible en el área pequeña, integrando la información provista por variables auxiliares o por la misma variable analizada en un periodo de tiempo distinto, este aumento de información viene acompañado de una disminución en el error de muestreo y una mejora sustancial en la calidad de las estimaciones (Rao and Molina, 2015).

Puesto que las estimaciones en áreas pequeñas, en su mayoría, son modelo-dependiente, la calidad de las estimaciones están sujetas a la calidad de la información auxiliar que se le suministre. Es por esto que, basados en el trabajo de Treviño et al. (2016) se propone utilizar, para la construcción del estimador propuesto en Tellez (2021), variables macroeconómicas tales como, el Producto Interno Bruto de cada país o dominio (PIB), Porcentaje del PIB destinado a investigación y desarrollo en el d -ésimo país, Porcentaje del PIB destinado a la educación en el d -ésimo país, entre otros, para modelar el logro académico de los estudiantes en el país o dominio d . En la sección 4 se presentan todas las variables que se utilizarán en este artículo. Cabe resaltar que estas no son las únicas variables que se pueden considerar en el modelo, toda aquella variable que esté correlacionada con el logro académico es susceptible a ser utilizada. El estimador propuesto en Tellez (2021) el cuál se basó en el modelo de Fay and Herriot (1979) supone lo siguiente:

1. Sean $\mathbf{x}_d = (\mathbf{x}_{d1}, \dots, \mathbf{x}_{dp})$ vectores conocidos de variables explicativas relacionadas con $\mu_d = \bar{Y}_d$, $d = 1, \dots, D$. μ_d está relacionado con la habilidad promedio real de los estudiantes, ya sea en Matemáticas o Ciencias en un país o dominio d .
2. Las μ_d son independientes y con distribución $N(\mathbf{x}_d\boldsymbol{\beta}, \sigma_u^2)$, donde $\boldsymbol{\beta}$ es un vector de dimensión p de parámetros desconocidos, de modo que $\boldsymbol{\mu} = (\mu_1, \dots, \mu_D)' \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_\mu)$ con $\boldsymbol{\Sigma}_\mu = \sigma_u^2 \mathbf{I}_D$.
3. Existe un vector de estimadores muestrales (Para este caso son las estimaciones directas que reporta TIMSS en el 2015 en cada país) $\bar{\mathbf{y}} = (\bar{y}_1, \dots, \bar{y}_D)'$ de $\boldsymbol{\mu}$ con distribución $\sim N(\boldsymbol{\mu}, \mathbf{V}_e)$, donde $\mathbf{V}_e = diag(\sigma_1^2, \dots, \sigma_D^2)$ y los elementos σ_d^2 son las varianzas estimadas por TIMSS en cada país.

Teniendo en cuenta lo anterior, el modelo se puede plantear como un modelo lineal,

$$\bar{y}_d = \mu_d + e_d \quad ; \quad \mu_d = \mathbf{x}_d\boldsymbol{\beta} + u_d, \quad d = 1, \dots, D \quad (3)$$

donde $\mathbf{e} = (e_1, \dots, e_D)$ y $\mathbf{u} = (u_1, \dots, u_D)$ son independientes con distribución $N(\mathbf{0}, \mathbf{V}_e)$ y $N(\mathbf{0}, \boldsymbol{\Sigma}_u)$ respectivamente. Si escribimos el modelo lineal en la forma matricial $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, se obtiene:

$$\begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_D \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{D1} & \cdots & x_{Dp} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_D \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_D \end{pmatrix} \quad (4)$$

Donde, $\mathbf{Z} = \mathbf{I}_D$, $tr(\mathbf{Z}'\mathbf{Z}) = \mathbf{D}$, $rango(\mathbf{X}, \mathbf{Z}) = \mathbf{D}$, $cov[\mathbf{y}, \mathbf{u}] = \mathbf{Z}\boldsymbol{\Sigma}_u$,

$$\mathbf{V} = \text{var}(\mathbf{y}) = \mathbf{Z}\boldsymbol{\Sigma}_u\mathbf{Z}' + \mathbf{V}_e = \boldsymbol{\Sigma}_u + \mathbf{V}_e = \text{diag}(\sigma_u^2 + \sigma_1^2, \dots, \sigma_u^2 + \sigma_D^2) \quad (5)$$

y

$$\mathbf{V}^{-1} = \text{diag}[(\sigma_u^2 + \sigma_1^2)^{-1}, \dots, (\sigma_u^2 + \sigma_D^2)^{-1}] \quad (6)$$

Por su parte, los estimadores y predictores insesgados BLUP, según Prasad and Rao (1990) de β y \mathbf{u} son:

$$\begin{aligned}\tilde{\beta} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\bar{\mathbf{y}} \\ \tilde{\mathbf{u}} &= \boldsymbol{\Sigma}_u\mathbf{Z}'\mathbf{V}^{-1}(\bar{\mathbf{y}} - \mathbf{X}\tilde{\beta})\end{aligned}$$

Ahora bien, según Ghosh and Rao (1994) las componentes de $\tilde{\mathbf{u}}$ son:

$$\tilde{u}_d = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_d^2}(\bar{y}_d - \mathbf{x}_d\tilde{\beta}), d = 1, \dots, D \quad (7)$$

donde \mathbf{x}_d es la fila d de \mathbf{X} .

El estimador BLUP de $\mu_d = \mathbf{x}_d\beta + u_d$ es:

$$\hat{y}_d^{BLUP} = \mathbf{x}_d\tilde{\beta} + \tilde{u}_d = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_d^2}\bar{y}_d + \frac{\sigma_d^2}{\sigma_u^2 + \sigma_d^2}\mathbf{x}_d\tilde{\beta} \quad (8)$$

Por último, el estimador EBLUP para estimar la habilidad promedio de los estudiantes en un país o dominio d está dado por,

$$\hat{Y}_d^{FH} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \sigma_d^2}\bar{y}_d + \frac{\sigma_d^2}{\hat{\sigma}_u^2 + \sigma_d^2}\mathbf{x}_d\hat{\beta} \quad (9)$$

donde, se obtiene sustituyendo σ_u^2 por un estimador $\hat{\sigma}_u^2$ en la ecuación (8). Particularmente, utilizaremos este estimador para estimar los promedios de las habilidades en Matemática y Ciencias en los grados cuarto y octavo de la prueba TIMSS 2015. Como medida de calidad para el estimador propuesto en este artículo, se utilizará el error cuadrático medio (MSE) mostrado en Tellez (2021),

$$MSE\left(\hat{Y}_d^{FH}\right) = MSE\left(\hat{y}_d^{BLUP}\right) + E\left(\hat{Y}_d^{FH} - \hat{y}_d^{BLUP}\right)^2 + 2E\left[\left(\hat{Y}_d^{FH} - \hat{y}_d^{BLUP}\right)\left(\hat{y}_d^{BLUP} - \bar{y}_d\right)\right]$$

4. Aplicación del estimador con los datos de TIMSS 2015

El objetivo de esta sección es mostrar, mediante la utilización de los resultados de las pruebas TIMSS 2015, las bondades que tiene el estimador mostrado en la sección 3 del presente artículo. Para tal fin, se ajustará un modelo de modelo Fay-Herriot en el cual se utilizarán como variables explicativas del logro académico, variables relacionadas con el contexto de aprendizaje (entiéndase el contexto de aprendizaje como el entorno en el cual el estudiantado aprende; por ejemplo la infraestructura física de la escuela, situación socio-económica, entre otras) que según el estudio realizado por el laboratorio latinoamericano

de evaluación de la calidad de la educación, tiene una relación directa con el logro académico (Treviño et al., 2016). De la misma manera, variables que den muestra de la situación socioeconómica de un país, resultan importantes a la hora de analizar los resultados en materia de educación, la producción interna, la tasa de pobreza, el nivel de oportunidades y el bienestar humano están fuertemente asociados con la educación. Es tan así que, Treviño et al. (2016), refieren “Países con economías fuertes y estables, altos niveles de equidad y mejores oportunidades de vida para sus habitantes, proveen condiciones más favorables para el desarrollo de la educación y el logro de aprendizajes. Del mismo modo, las mejoras al sistema educativo repercuten positivamente en las condiciones de desarrollo de las naciones”.

Otros autores (Thieme, 2017) también evidencian que existe relación entre los resultados académicos y el nivel socioeconómico de su población. Tomando como ejemplo Colombia, se evidencia que la desigualdad en los resultados de las pruebas Saber se debe en gran medida, a la inequidad de recursos en las diferentes regiones del país; reflejado en las condiciones físicas de las escuelas y su acceso a servicios públicos, la jornada completa y el grado de satisfacción docente Duarte et al. (2011).

Dado esto, se consultan fuentes de datos abiertos como las del World Bank¹, la UNESCO² o la ONU³ para extraer información correspondiente a indicadores macroeconómicos para el año 2015. Es importante aclarar que para algunos países no se encontraba la información disponible para el año 2015; es por esto que, se decide completar los datos vacíos con la información del año más cercano. Por otro lado, TIMSS cuenta con dos bases de datos para cada ciclo de aplicación de las pruebas; una, con los resultados de los estudiantes de cuarto grado, y otra con los resultados de los estudiantes de octavo grado. Tanto la base de cuarto grado como la de octavo contienen información sobre la habilidad de los estudiantes en Matemáticas y Ciencias, información del contexto de aprendizaje, los padres, los maestros, la escuela y el currículo (63 países participantes).

Una vez definida la información auxiliar que será utilizada, se procede con la etapa de modelamiento. En esta etapa, el objetivo se centra en vincular la información auxiliar x_d y las estimaciones generadas por TIMSS 2015 (\hat{Y}_d). La idea es entonces, encontrar un modelo cuyas variables sean significativas y que dichas variables, en el contexto de la educación, expliquen el logro académico. En este sentido los modelos resultantes fueron los siguientes:

Modelo Matemáticas 4°:

$$\hat{Y}_d^{Mat4} = \beta_0 + \beta_1 X_{1d} + \beta_2 X_{2d} + \beta_3 X_{3d} + \beta_4 X_{5d} + \beta_5 X_{6d} + \beta_6 X_{8d} + \beta_7 X_{9d} + \beta_8 X_{11d} + v_d + e_d$$

Modelo Ciencias 4°:

$$\hat{Y}_d^{Ci4} = \beta_0 + \beta_2 X_{2d} + \beta_3 X_{3d} + \beta_5 X_{5d} + \beta_8 X_{8d} + \beta_9 X_{9d} + \beta_{11} X_{11d} + v_d + e_d$$

Modelo Matemáticas 8°:

$$\hat{Y}_d^{Mat8} = \beta_0 + \beta_4 X_{4d} + \beta_6 X_{6d} + \beta_7 X_{7d} + \beta_8 X_{8d} + \beta_{10} X_{10d} + \beta_{15} X_{15d} + v_d + e_d$$

Modelo Ciencias 8°:

$$\hat{Y}_d^{Ci8} = \beta_0 + \beta_4 X_{4d} + \beta_5 X_{5d} + \beta_6 X_{6d} + \beta_7 X_{7d} + \beta_8 X_{8d} + \beta_{15} X_{15d} + v_d + e_d$$

donde, $\hat{Y}_d^{Mat(i)}$ denota el resultado en Matemáticas para el grado $i = 4^\circ$ y 8° en el d -ésimo país y $\hat{Y}_d^{Ci(i)}$ denota el resultado en Ciencias para el grado $i = 4^\circ$ y 8° en el d -ésimo país. Las variables auxiliares X_{id} se describen en la tabla 1, y en la tabla 2, se presentan los valores que asumen los coeficientes regresores estimados $\hat{\beta}_i$ para cada uno de los modelos presentados.

¹Accesible desde <https://datos.bancomundial.org>

²Accesible desde <http://data UIS.unesco.org/>

³Accesible desde <https://www.un.org/en/databases/>

Variable	Descripción	Año
X_{1d}	PIB per cápita (US\$ a precios actuales) del d -ésimo país	2015
X_{2d}	Porcentaje del PIB destinado a investigación y desarrollo en el d -ésimo país	2015
X_{3d}	Porcentaje del PIB destinado a la educación en el d -ésimo país	2015
X_{4d}	Porcentaje del PIB destinado al gasto público en el d -ésimo país	2015
X_{5d}	Coeficiente de Gini en el d -ésimo país	2010-2017
X_{6d}	Estimación de desempleo de la OIT en el d -ésimo país	2015
X_{7d}	Índice de desarrollo humano en el d -ésimo país	2015
X_{8d}	Artículos en publicaciones científicas en el d -ésimo país	2015
X_{9d}	Porcentaje de escuelas primarias con acceso a electricidad en el d -ésimo país	2012-2018
X_{10d}	Porcentaje de escuelas secundarias con acceso a electricidad en el d -ésimo país	2012-2018
X_{11d}	Porcentaje de escuelas primarias con acceso a agua potable en el d -ésimo país	2012-2018
X_{12d}	Porcentaje de escuelas secundarias con acceso a agua potable en el d -ésimo país	2012-2018
X_{13d}	Horas de clase al año en grado 8° del d -ésimo país	2011
X_{14d}	Horas de clase al año en grado 4° del d -ésimo país	2011
X_{15d}	Recursos del hogar para el aprendizaje en grado 8° ¹	2011

[1] La variable *Recursos del Hogar para el Aprendizaje* es tomada del capítulo 4 del libro *TIMSS 2011 International Results in Mathematics* en el Exhibit 4.3. Esta variable es un puntaje construido a partir de las respuestas que estudiantes dieron con respecto a la disponibilidad de tres recursos en el hogar para el aprendizaje.

Tabla 1: Lista de variables auxiliares.

Coeficiente	Modelo			
	Matemáticas 4	Ciencias 4	Matemáticas 8	Ciencias 8
β_0	$2.02 \times 10^{+2}$	$1.61 \times 10^{+2}$	$1.22 \times 10^{+2}$	$1.73 \times 10^{+2}$
β_1	1.07×10^{-5}	—	—	—
β_2	$2.00 \times 10^{+1}$	$1.98 \times 10^{+1}$	—	—
β_3	3.55×10^{-2}	1.43	—	—
β_4	-2.20	—	$1.46 \times 10^{+1}$	9.21
β_5	-3.47×10^{-2}	-2.17	—	-0.73
β_6	1.81×10^{-4}	—	-1.43	-0.99
β_7	2.69	—	$1.26 \times 10^{+2}$	$2.33 \times 10^{+2}$
β_8	—	1.32×10^{-4}	1.57×10^{-4}	0.15×10^{-3}
β_9	—	3.15	—	—
β_{10}	—	—	7.39×10^{-3}	—
β_{11}	8.25×10^{-1}	6.85×10^{-1}	—	—
β_{12}	—	—	—	—
β_{13}	—	—	—	—
β_{14}	—	—	—	—
β_{15}	—	—	$2.36 \times 10^{+1}$	$1.32 \times 10^{+1}$

Tabla 2: Valor de los coeficientes de regresión de los modelos ajustados.

Algo para tener en cuenta es que, en el año 2015, la distribución de países que presentaron las pruebas es la siguiente: para grado 4°, 47 países presentaron Ciencias y 49 países presentaron Matemáticas; para grado 8°, 39 países presentaron pruebas de Ciencias y 39 países presentaron Matemáticas. Haciendo uso del software estadístico R y su librería *intsvy*, se obtienen los estimadores directos junto con sus desviaciones estándar para los países que presentaron las pruebas. Ahora bien, dado que el objetivo de este trabajo es mostrar que la metodología propuesta en Tellez (2021) mejora las estimaciones en términos de eficiencia, se realiza la comparación entre el error estándar de la estimación publicada por TIMSS y la raíz cuadrada del error cuadrático medio (*RMSE*) de las estimaciones utilizando la metodología. En

las tablas 3, 4, 5 y 6 se consignan el puntaje $\hat{\theta}$ reportado por TIMSS en 2015, el puntaje estimado con la metodología propuesta \hat{Y}_d^{FH} , y el error relativo porcentual ER entre el error estándar reportado por TIMSS (SE) y la raíz del error cuadrático medio $RMSE$, para los 3 casos con mejor ER . Se debe tener en cuenta que entre mayor es el ER mejor es la estimación hecha por la metodología utilizada en este artículo. En el Anexo A se presentan las tablas con la comparación de los resultados de todos los países que presentaron las pruebas en 2015.

País	$\hat{\theta}_{Mat4}$	SE	\hat{Y}_d^{FH}	$RMSE$	$ER(%)$
Abu Dhabi, UAE	419.29	4.62	419.67	4.60	0.37
Bulgaria	524.28	5.33	523.89	5.30	0.48
Canada, Quebec	535.83	3.93	535.75	3.92	0.29

Tabla 3: Resultados publicados y resultados estimados de las pruebas de Matemáticas para 4º grado.

País	$\hat{\theta}_{Ci4}$	SE	\hat{Y}_d^{FH}	$RMSE$	$ER(%)$
Abu Dhabi, UAE	414.71	5.53	415.38	5.49	0.67
Bulgaria	535.70	5.80	534.91	5.76	0.72
Saudi Arabia	390.33	4.80	391.56	4.78	0.52

Tabla 4: Resultados publicados y resultados estimados de las pruebas de Ciencias para 4º grado.

País	$\hat{\theta}_{Mat8}$	SE	\hat{Y}_d^{FH}	$RMSE$	$ER(%)$
Abu Dhabi, UAE	441.67	4.71	442.17	4.69	0.47
Kazakhstan	527.81	5.33	526.77	5.30	0.60
Russian Federation	538.00	4.56	537.57	4.54	0.43

Tabla 5: Resultados publicados y resultados estimados de las pruebas de Matemáticas para 8º grado.

País	$\hat{\theta}_{Ci8}$	SE	\hat{Y}_d^{FH}	$RMSE$	$ER(%)$
Abu Dhabi, UAE	454.25	5.67	455.16	5.61	1.02
Kuwait	410.74	5.25	411.62	5.21	0.83
Lebanon	398.16	5.21	399.22	5.17	0.83

Tabla 6: Resultados publicados y resultados estimados de las pruebas de Ciencias para 8º grado.

En las tablas 10, 11, 12 y 13 se puede observar que en todos los países que presentaron las pruebas, se logró mejorar la estimación en términos de variabilidad utilizando la metodología aquí descrita. Si bien, las mejoras no son tan grandes para algunos países, esto va ligado a la calidad de la información auxiliar que se utilizó y por ende se puede mejorar refinando la información auxiliar.

Por otro lado, como segundo objetivo del presente artículo, se pretende mostrar que utilizando la metodología propuesta por Tellez (2021) se logran predicciones de buena calidad. Para esto, se elimina aleatoriamente 1 país de cada modelo ajustado. Para el caso de los modelos de Matemáticas del grado 4º, se elimina Holanda y del grado 8º Lebanon, y para los modelos de Ciencias, del grado 4º se elimina Portugal y del grado 8º Georgia. Luego, se procede a ajustar nuevamente los modelos. Una vez obtenidos los nuevos modelos, se hace uso de las variables auxiliares disponibles para los países eliminados y se estima su puntaje para las pruebas de Matemáticas y Ciencias (\hat{Y}_d^{FH}). En las tablas 7 y 8 se presentan dichas predicciones junto con los puntajes reportados por TIMSS ($\hat{\theta}$) y el sesgo relativo porcentual del estimador ($BR\hat{Y}_d^{FH}$).

Prueba	País	$\hat{\theta}_4$	\widehat{Y}_d^{FH}	BRY_d^{FH} (%)
Matemáticas	Holanda	529.82	536.07	1.18
Ciencias	Portugal	508.06	501.65	1.26

Tabla 7: Sesgo relativo de la predicción en Matemáticas y Ciencias para 4° grado.

Prueba	País	$\hat{\theta}_8$	\widehat{Y}_d^{FH}	BRY_d^{FH} (%)
Matemáticas	Lebanon	442.43	440.66	0.40
Ciencias	Georgia	443.17	459.02	3.58

Tabla 8: Sesgo relativo de la predicción en Matemáticas y Ciencias para 8° grado.

De las tablas 7 y 8 se puede evidenciar que en ningún caso el sesgo relativo supera el 5 %, sin embargo, existen algunos países cuya predicción presenta un sesgo más amplio (superiores al 10 %), como es el caso de Singapur, y esto es debido a que la información auxiliar que se consiguió para este país no estaba actualizada en los repositorios que se descargó la información. En la tabla 9 se presenta la información de la estimación para este país.

Prueba	País	$\hat{\theta}_4$	\widehat{Y}_d^{FH}	BRY_d^{FH} (%)
Ciencias	Singapur	590.48	520.38	11.87

Tabla 9: Sesgo relativo de la predicción de Singapur en Ciencias para 4° grado.

5. Discusión y Conclusiones

El estimador propuesto en Tellez (2021) utilizado en este artículo para estimar las habilidades promedio de los estudiantes por países, basado en el enfoque de estimación en áreas pequeñas para la prueba TIMSS, genera resultados positivos dado que al comprar las medidas de calidad de la estimación resultante y las reportadas por TIMSS en 2015 (tablas del Anexo A), se observa que en todos los países hay una disminución en la variabilidad de la estimación, lo cual era uno de los objetivos de esta metodología. Si bien, la disminución en todos los países no es muy grande, el objetivo que se quiso plantear en este artículo es proponer una nueva forma para estimar, con mayor precisión, los resultados de los países en la prueba TIMSS 2015. La precisión se puede mejorar refinando las variables utilizadas como explicativas o incorporando en el modelo otras variables altamente correlacionadas con el logro académico.

Por otro lado, otra ventaja que tiene utilizar este estimador es que, como se menciona anteriormente, si hay países que no puedan presentar esta prueba en un año particular, se pueden estimar sus resultados con alta eficiencia y bajo sesgo y así poder tomar acciones a nivel general de política educativa.

Se pudo observar también, que después de haber probado diferentes modelos combinando las diferentes variables disponibles, es posible mejorar las estimaciones haciendo uso de factores asociados a la educación. Esto va en línea con los hallazgos reportados en el informe de resultados del TERCE (Treviño et al., 2016) que extiende el alcance del entorno socioeconómico dentro del proceso de aprendizaje. Es así como la metodología propuesta, se perfila como una buena opción para mejorar la calidad de la estimación del puntaje que obtendrá un país a partir de la observación de indicadores y estadísticas relacionadas al contexto educativo.

Algo para tener en cuenta es que, a pesar de contar con un número reducido de países para la construcción de los modelos, se encontró que el sesgo entre el valor observado y el valor predicho del puntaje en Matemáticas y Ciencias de 4° y 8° grado no supera el 5 % en la mayoría de los casos. Esto permite concluir

que esta metodología puede extenderse a otras pruebas internacionales que presenten más países y esto permitirá un mejor ajuste del modelo y por tanto unas mejores estimaciones de la habilidad promedio.

6. Agradecimientos

Este artículo es un resultado derivado de la tesis de doctorado de Cristian Fernando Tellez Piñerez la cual fue dirigida por el profesor Leonardo Trujillo Oyala.

Referencias

- Estudio Internacional de Tendencias en Matemática y Ciencias (TIMSS).* Agencia de calidad de la educación, Santiago, 2018.
- A. Birnbaum. On the estimation of mental ability. In *Series report n°15*, pages 7755–7723. USAF School of Aviation Medicine, Randolph Air Force Base, Texas, 1958.
- D. F. de Andrade, H. R. Tavares, and R. da Cunha Valle. Teoria da Resposta ao Item: Conceitos e Aplicações. *ABE, Sao Paulo, Brasil*, 2000.
- J. Duarte, C. Gargiulo, and M. Moreno. *Infraestructura Escolar y Aprendizajes en la Educación Básica Latinoamericana: Un análisis a partir del SERCE*. Banco Interamericano de Desarrollo, Santiago de Chile, 2011.
- R. Fay and R. Herriot. Estimation of income from small places: An application of jamesstein procedures to census data. *Journal of the American Statistical Association*, 74:269–277, 1979.
- J. M. C. Ferrera. Factores explicativos del rendimiento en educación primaria: un análisis a partir de timss 2011. *Estudios sobre educación*, 27:9–35, 2014.
- M. Ghosh and J. Rao. Small area estimation: An appraisal. *Statistical Science*, 9(1):55–83, 1994.
- ICFES. Manual de calificación. prueba cognitiva saber 3, 5, 7 y 9. Technical report, ICFES, 2015.
- M. Joncas. Pirls 2006 sampling weights and participation rates. *PIRLS*, page 105, 2007.
- S. LaRoche, M. Joncas, and P. Foy. Sample design in timss 2015. In M. Martin, I. Mullis, and M. Hooper, editors, *Methods and procedures in TIMSS 2015*. TIMSS & PIRLS International Study Center, Boston, 2015.
- M. R. Martinez. La metodología de los estudios PISA. *Revista de Educación*, pages 1,111–129, 2006.
- Mineducación. *Sistema educativo colombiano*. Ministerio de educación, Bogotá, 2010.
- I. V. Mullis, M. O. Martin, A. Kennedy, K. Trong, and M. P. Sainsbury. Assessment frameworks. *TIMMS and Pirls International Study Center, Boston College*, 2015.
- E. Muraki. A generalized partial credit model: Application of an em algorithm. *Applied Psy-chological Measurement*, 16:159–176, 1992.
- N. Prasad and J. Rao. The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85(409):163–171, 1990.
- J. N. Rao and I. Molina. *Small Area Estimation*. John Wiley & Sons, 2 edition, 2015.
- A. Schleicher. La importancia de compararse globalmente. In *Altablero*, volume 44. Ministerio de educación, Bogotá, 2008.

- C. Tellez. *Estimación de áreas pequeñas utilizando imputación múltiple en modelos logísticos de tres parámetros*. PhD thesis, Universidad Nacional, Bogotá, Colombia, 2021.
- C. Thieme. *Los recursos y capacidades como factores explicativos de la calidad, desempeño e impacto de los establecimientos de educación parvularia en Chile*. FONDECYT, Santiago de Chile, 2017.
- E. Treviño, P. Fraser, A. Meyer, L. Morawietz, P. Inostroza, and E. Naranjo. *Informe de resultados TERCE, Factores Asociados*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, París, 2016.
- UNESCO. *International standard classification of education: ISCED 2011*. UNESCO Institute for Statistics Montreal, 2012.

A. Tablas de resultados

A continuación se presentan las tablas que permiten comparar los resultados de las pruebas TIMSS aplicadas en grado 4° y 8° en el año 2015 ($\hat{\theta}$), y el puntaje estimado con la metodología empleada (\widehat{Y}_d^{FH}).

País	$\hat{\theta}_{Mat4}$	SE	\hat{Y}_d^{FH}	RMSE	ER (%)
Abu Dhabi, UAE	419.29	4.62	419.67	4.60	0.37
Armenia	481.03	3.39	480.99	3.38	0.16
Australia	517.23	3.16	517.27	3.15	0.18
Bahrain	450.24	1.89	450.31	1.89	0.05
Belgium (Flemish)	545.66	2.11	545.66	2.11	0.08
Bulgaria	524.28	5.33	523.89	5.30	0.48
Canada	510.56	2.37	510.59	2.37	0.11
Canada, Ontario	512.46	2.27	512.48	2.27	0.10
Canada, Quebec	535.83	3.93	535.75	3.92	0.29
Chile	458.58	2.40	458.44	2.40	0.07
Chinese Taipei	596.62	1.82	596.59	1.82	0.03
Croatia	502.34	1.79	502.34	1.79	0.06
Cyprus	523.03	2.64	522.93	2.64	0.10
Czech Republic	528.14	2.24	528.16	2.24	0.08
Denmark	538.65	2.63	538.70	2.63	0.11
Dubai, UAE	510.64	1.52	510.59	1.52	0.04
England	546.19	2.82	546.15	2.82	0.15
Finland	535.29	1.98	535.32	1.98	0.07
France	488.17	2.84	488.35	2.84	0.15
Georgia	463.15	3.59	463.24	3.58	0.22
Germany	521.63	2.02	521.70	2.02	0.07
Hong Kong SAR	614.52	2.91	614.15	2.91	0.14
Hungary	529.19	3.24	529.13	3.23	0.19
Indonesia	389.14	4.23	389.42	4.22	0.20
Iran, Islamic Rep. of	423.57	4.10	423.96	4.09	0.28
Ireland	547.34	2.13	547.27	2.13	0.07
Italy	506.85	2.46	506.88	2.46	0.11
Japan	592.83	1.96	592.67	1.96	0.04
Kazakhstan	544.42	4.50	543.92	4.49	0.26
Korea, Rep. of	608.04	2.24	607.98	2.24	0.06
Kuwait	351.05	5.13	352.07	5.12	0.17
Lithuania	535.34	2.49	535.23	2.49	0.11
Morocco	379.47	3.86	379.87	3.85	0.23
Netherlands	529.82	1.63	529.83	1.63	0.05
New Zealand	490.56	2.41	490.58	2.41	0.10
Northern Ireland	570.26	2.88	570.13	2.88	0.15
Norway	549.08	2.46	549.04	2.46	0.09
Oman	425.48	2.55	425.56	2.55	0.11
Poland	534.77	2.24	534.72	2.24	0.09
Portugal	541.20	2.23	541.12	2.23	0.09
Qatar	439.00	3.50	439.26	3.49	0.17
Russian Federation	563.92	3.49	563.60	3.48	0.21
Saudi Arabia	383.49	4.00	384.26	3.99	0.29
Serbia	518.00	3.60	517.95	3.59	0.20
Singapore	617.67	3.77	617.12	3.76	0.25
Slovak Republic	498.25	2.43	498.30	2.43	0.10
Slovenia	519.87	1.90	519.91	1.90	0.06
Spain	505.09	2.42	505.10	2.42	0.08
Sweden	518.65	2.76	518.78	2.76	0.12
Turkey	483.15	3.07	483.05	3.07	0.16
United Arab Emirates	451.58	2.39	451.60	2.39	0.10
United States	539.16	2.23	539.28	2.23	0.05

Tabla 10: Resultados publicados y resultados estimados de las pruebas de Matemáticas para 4º grado.
Comunicaciones en Estadística, octubre 2020, Vol. 0, No. 0

País	$\hat{\theta}_{Ci4}$	SE	\hat{Y}_d^{FH}	RMSE	ER (%)
Abu Dhabi, UAE	414.71	5.53	415.38	5.49	0.67
Armenia	443.83	3.94	444.03	3.93	0.32
Australia	523.63	2.87	523.63	2.86	0.19
Bahrain	458.81	2.36	458.90	2.36	0.11
Belgium (Flemish)	511.51	2.34	511.61	2.34	0.12
Bulgaria	535.70	5.80	534.91	5.76	0.72
Canada	524.78	2.68	524.76	2.68	0.16
Canada, Ontario	530.37	2.59	530.33	2.59	0.15
Canada, Quebec	524.51	4.01	524.47	4.00	0.37
Chile	477.71	2.74	477.40	2.74	0.11
Chinese Taipei	555.28	1.81	555.28	1.81	0.04
Croatia	533.44	2.15	533.37	2.15	0.10
Cyprus	481.30	2.52	481.33	2.52	0.12
Czech Republic	534.38	2.34	534.39	2.34	0.12
Denmark	527.03	2.21	527.09	2.21	0.10
Dubai, UAE	517.94	1.80	517.84	1.80	0.07
England	535.83	2.44	535.81	2.44	0.13
Finland	553.81	2.27	553.82	2.27	0.11
France	487.40	2.60	487.57	2.60	0.15
Georgia	451.25	3.74	451.44	3.73	0.30
Germany	528.47	2.38	528.55	2.38	0.12
Hong Kong SAR	556.55	2.96	556.32	2.95	0.19
Hungary	541.98	3.29	541.82	3.28	0.25
Indonesia	396.67	4.78	397.00	4.77	0.31
Iran, Islamic Rep. of	421.01	4.05	421.47	4.04	0.33
Ireland	528.88	2.39	528.81	2.39	0.13
Italy	516.47	2.64	516.45	2.64	0.16
Japan	569.01	1.73	568.88	1.73	0.04
Kazakhstan	549.56	4.38	548.92	4.37	0.30
Korea, Rep. of	589.32	1.99	589.29	1.99	0.06
Kuwait	337.21	6.25	339.09	6.23	0.32
Lithuania	527.67	2.55	527.55	2.55	0.14
Morocco	352.21	4.54	353.15	4.52	0.40
Netherlands	517.12	2.60	517.18	2.60	0.15
New Zealand	505.52	2.70	505.50	2.70	0.15
Northern Ireland	519.77	2.19	519.79	2.19	0.11
Norway	537.60	2.55	537.60	2.55	0.12
Oman	430.97	3.16	431.09	3.15	0.21
Poland	547.19	2.39	547.07	2.39	0.13
Portugal	508.06	2.21	508.04	2.21	0.11
Qatar	436.26	4.12	436.70	4.10	0.37
Russian Federation	567.20	3.28	566.82	3.27	0.23
Saudi Arabia	390.33	4.80	391.56	4.78	0.52
Serbia	524.51	3.85	524.38	3.84	0.32
Singapore	590.48	3.64	590.03	3.63	0.30
Slovak Republic	520.49	2.60	520.48	2.60	0.14
Slovenia	542.57	2.33	542.57	2.33	0.12
Spain	518.20	2.58	518.15	2.58	0.15
Sweden	540.19	3.51	540.32	3.50	0.24
Turkey	483.40	3.24	483.24	3.23	0.22
United Arab Emirates	451.24	2.79	451.27	2.79	0.17
United States	545.91	2.26	545.98	2.26	0.06

Tabla 11: Resultados publicados y resultados estimados de las pruebas de Ciencias para 4º grado.
Comunicaciones en Estadística, octubre 2020, Vol. 0, No. 0

País	$\hat{\theta}_{Mat8}$	SE	\hat{Y}_d^{FH}	RMSE	ER (%)
Abu Dhabi, UAE	441.67	4.71	442.17	4.69	0.47
Armenia	471.46	3.13	471.35	3.13	0.14
Australia	504.96	3.05	505.10	3.04	0.20
Bahrain	453.95	1.46	453.96	1.46	0.04
Botswana	390.84	2.02	390.85	2.02	0.03
Canada	527.28	2.18	527.29	2.18	0.10
Canada, Ontario	522.30	2.88	522.34	2.87	0.18
Canada, Quebec	543.36	3.80	543.23	3.79	0.31
Chile	427.43	3.16	427.57	3.15	0.19
Chinese Taipei	599.11	2.50	598.96	2.50	0.05
Dubai, UAE	511.85	1.97	511.81	1.97	0.08
Egypt	392.23	4.02	392.43	4.01	0.30
England	518.26	4.13	518.34	4.11	0.37
Georgia	453.20	3.35	453.22	3.34	0.22
Hong Kong SAR	594.25	4.54	593.18	4.52	0.39
Hungary	514.41	3.79	514.27	3.78	0.31
Iran, Islamic Rep. of	436.35	4.62	436.16	4.60	0.34
Ireland	523.49	2.63	523.40	2.63	0.14
Israel	510.90	3.98	511.25	3.97	0.24
Italy	494.39	2.43	494.38	2.43	0.12
Japan	586.47	2.29	586.38	2.29	0.10
Jordan	385.55	3.29	385.78	3.28	0.21
Kazakhstan	527.81	5.33	526.77	5.30	0.60
Korea, Rep. of	605.74	2.65	605.62	2.65	0.11
Kuwait	392.47	4.59	392.93	4.57	0.42
Lebanon	442.43	3.48	442.42	3.47	0.26
Lithuania	511.31	2.68	511.19	2.68	0.16
Malaysia	465.31	3.58	465.25	3.57	0.26
Malta	493.54	0.99	493.53	0.99	0.02
Morocco	384.39	2.21	384.41	2.21	0.09
New Zealand	492.72	3.29	492.80	3.28	0.24
Norway	511.54	2.20	511.61	2.20	0.10
Oman	403.16	2.45	403.27	2.45	0.12
Qatar	437.11	2.92	437.35	2.92	0.16
Russian Federation	538.00	4.56	537.57	4.54	0.43
Saudi Arabia	367.72	4.54	368.67	4.52	0.43
Singapore	620.96	3.15	620.42	3.14	0.20
Slovenia	516.34	2.03	516.33	2.03	0.09
Sweden	500.72	2.74	500.90	2.74	0.15
Thailand	431.42	4.67	431.37	4.65	0.39
Turkey	457.63	4.71	457.25	4.69	0.38
United Arab Emirates	464.78	1.98	464.82	1.98	0.08
United States	518.30	3.08	518.68	3.08	0.10

Tabla 12: Resultados publicados y resultados estimados de las pruebas de Matemáticas para 8° grado.

País	$\hat{\theta}_{Ci8}$	SE	\hat{Y}_d^{FH}	RMSE	ER (%)
Abu Dhabi, UAE	454.25	5.67	455.16	5.61	1.02
Armenia	460.17	3.28	460.10	3.27	0.23
Australia	511.99	2.63	512.11	2.62	0.22
Bahrain	465.85	2.13	465.88	2.13	0.14
Botswana	391.80	2.72	391.81	2.72	0.11
Canada	526.17	2.27	526.19	2.27	0.16
Canada, Ontario	523.87	2.56	523.91	2.55	0.21
Canada, Quebec	529.72	4.58	529.69	4.55	0.67
Chile	453.97	3.06	454.02	3.05	0.25
Chinese Taipei	569.47	2.10	569.34	2.10	0.06
Dubai, UAE	524.72	2.03	524.63	2.03	0.13
Egypt	370.78	4.24	371.52	4.22	0.47
England	536.63	3.89	536.57	3.87	0.49
Georgia	443.17	3.08	443.27	3.07	0.27
Hong Kong SAR	545.76	3.95	545.36	3.93	0.44
Hungary	527.26	3.51	526.96	3.50	0.39
Iran, Islamic Rep. of	456.42	4.03	456.21	4.01	0.46
Ireland	530.10	2.84	529.98	2.83	0.24
Israel	506.73	3.84	507.04	3.83	0.31
Italy	498.93	2.37	498.93	2.37	0.17
Japan	570.90	1.76	570.84	1.76	0.09
Jordan	426.16	3.47	426.25	3.46	0.36
Kazakhstan	532.59	4.40	531.66	4.37	0.57
Korea,Rep.of	555.60	2.22	555.60	2.22	0.12
Kuwait	410.74	5.25	411.62	5.21	0.83
Lebanon	398.16	5.21	399.22	5.17	0.83
Lithuania	519.11	2.61	518.92	2.60	0.22
Malaysia	470.82	4.13	470.64	4.11	0.48
Malta	481.36	1.65	481.38	1.65	0.09
Morocco	393.25	2.54	393.28	2.54	0.17
New Zealand	512.68	3.15	512.70	3.14	0.32
Norway	508.83	2.83	509.04	2.82	0.25
Oman	454.56	2.64	454.58	2.63	0.21
Qatar	456.52	2.96	456.77	2.95	0.25
Russian Federation	544.12	4.26	543.39	4.24	0.54
Saudi Arabia	396.42	4.56	397.66	4.53	0.65
Singapore	596.64	3.17	596.07	3.16	0.31
Slovenia	551.11	2.36	550.97	2.36	0.16
Sweden	522.27	3.54	522.48	3.53	0.36
Thailand	455.84	4.25	455.59	4.23	0.49
Turkey	493.40	4.00	492.75	3.98	0.45
United Arab Emirates	476.65	2.35	476.71	2.35	0.18
United States	530.00	2.88	530.36	2.88	0.13

Tabla 13: Resultados publicados y resultados estimados de las pruebas de Ciencias para 8° grado.