

# Modelo Multidimensional de Teoría de Respuesta al Item Jerárquico

DIANA STELLA ANTONIO ROJAS<sup>a</sup>

Diana Stella Antonio Rojas  
Código: 832411

Trabajo de grado presentado para optar al título de Maestría en Ciencias- Estadística

DIRIGIDO POR:  
Alvaro Mauricio Montenegro Díaz

UNIVERSIDAD NACIONAL DE COLOMBIA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADÍSTICA

Bogotá, 2013

---

<sup>a</sup>Estudiante de maestría en estadística. E-mail: dsantonior@unal.edu.co

Modelo Multidimensional de Teoría de Respuesta al Item Jerárquico

Diana Stella Antonio Rojas  
Código: 832411

Maestría en Ciencias- Estadística

Director  
Alvaro Mauricio Montenegro Díaz

UNIVERSIDAD NACIONAL DE COLOMBIA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADÍSTICA

Bogotá, 2013

## Modelo Multidimensional de Teoría de Respuesta al Ítem Jerárquico

### Resumen

Se propone un modelo para analizar datos resultantes de la aplicación de una prueba multidimensional a varias poblaciones, en las que se presentan diferencias en los parámetros de los ítems entre ellas. Se presentan resultados de la aplicación del modelo propuesto a datos simulados y a una muestra de datos reales provenientes del *Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado (PERCE)*, que fue aplicada por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. Los datos fueron suministrados por el Instituto Colombiano para la Evaluación de la Educación (ICFES). Se usa una muestra de datos de tres países: Bolivia, Colombia y Cuba.

**Palabras clave:** Teoría de respuesta al ítem, modelos jerárquicos, funcionamiento diferencial del ítem, dimensiones de los trazos latentes.

## Multidimensional Hierarchical Item Response Theory Model

### Abstract

Proposes a model for analyzing the data resulting from application of a multidimensional test several stocks, which are differences in the parameters of the items between them. We present results of applying the proposed model to simulated data and a sample of real data from the *First International Comparative Study of Language, Mathematics and Associated Factors in Third and Fourth Grade (PERCE)*, which was applied by the Laboratory for Assessment of the Quality of Education. The data were supplied by the *Colombian Institute for the Evaluation of Education (ICFES)*. We use a data sample of three countries: Bolivia, Colombia and Cuba.

**Key words:** Item response theory, hierarchical models, differential item functioning, dimensions of the latent traits.

# Índice

<b>1. Introducción</b>	<b>5</b>
<b>2. Metodología</b>	<b>5</b>
<b>3. Modelos clásicos de TRI</b>	<b>6</b>
3.1. Modelo clásico unidimensional de TRI dicotómico . . . . .	6
3.2. Modelo TRI unidimensional con efectos aleatorios . . . . .	6
3.3. Modelo TRI Multidimensional . . . . .	8
3.3.1. Análisis de las dimensiones . . . . .	9
<b>4. Modelo multidimensional jerárquico TRIMJ</b>	<b>9</b>
4.1. Función de verosimilitud . . . . .	10
4.2. Estimación de los parámetros del modelo TRIMJ . . . . .	11
<b>5. Simulación</b>	<b>11</b>
<b>6. Aplicación a una muestra de datos de la prueba PERCE</b>	<b>15</b>
6.1. Análisis para detectar la dimensionalidad del espacio de los trazos latentes . . . . .	15
6.2. Estimación de los parámetros de los ítems . . . . .	16
6.3. Estimación de los trazos latentes básicos . . . . .	17
6.4. Estimación de los trazos latentes principales . . . . .	18
6.5. Evaluación del ajuste del modelo TRIMJ . . . . .	23
<b>7. Discusión</b>	<b>24</b>
<b>8. Conclusiones</b>	<b>28</b>
<b>9. Propuestas para trabajos posteriores</b>	<b>29</b>
<b>10. Agradecimientos</b>	<b>30</b>
<b>Apéndice A.</b>	<b>31</b>
<b>Apéndice B.</b>	<b>33</b>
<b>Apéndice C.</b>	<b>37</b>
<b>Apéndice D.</b>	<b>38</b>
<b>Apéndice E.</b>	<b>39</b>
<b>Apéndice F.</b>	<b>40</b>
<b>Apéndice G.</b>	<b>42</b>

## 1. Introducción

El modelo desarrollado fue diseñado para el caso en el que existen clusters de respondientes. Se asume que los trazos latentes de todos los respondientes provienen de una única distribución y las diferencias entre los clusters de respondientes son modeladas relajando la hipótesis de invarianza de la medición. Se usa un modelo jerárquico para modelar las diferencias en los parámetros de los items entre los clusters de personas, en lugar de asumir la invarianza de los parámetros de los items a lo largo de cada uno de los clusters de respondientes. Un enfoque similar fue aplicado por Fox (2010), en el diseño del modelo unidimensional de teoría de respuesta al item (TRI) con efectos aleatorios, para el análisis de pruebas unidimensionales aplicadas a personas que pertenecen a diferentes poblaciones.

En este trabajo, se extiende el modelo de (Fox 2010) para el caso multidimensional, usando el enfoque jerárquico en lugar del enfoque con efectos aleatorios. Es decir, se incluye la estrategia jerárquica en el modelo multidimensional de TRI (TRIM) (Reckase 2009, pp. 86), teniendo en cuenta que un item puede ser sensible a diferencias en las personas en múltiples dimensiones. Cuando las dimensiones son objetivo de la evaluación, los resultados sirven como indicadores válidos de las dimensiones de interés. Cuando subgrupos de la población tienen diferencias en dimensiones que no son objetivo de la evaluación, pueden incluir sesgo en el item. Los análisis con modelos TRIM pueden ayudar a aclarar tanto las diferencias entre las poblaciones, como las razones del funcionamiento diferencial del item FDI (Reckase 2009, pp. 74-75). Las diferencias en los parámetros de los items son modeladas jerárquicamente, por lo anterior, el modelo no identifica el FDI.

El modelo desarrollado en este trabajo puede aplicarse a pruebas en las cuales existen varias poblaciones o clusters de respondientes, en cuyo caso, es usual que los clusters tengan diferencias entre ellos, por razones culturales, de idioma, sistemas educativos, entre otras. Un ejemplo de ello es la prueba PERCE, que se aplicó en 1997 a estudiantes de 13 países latinoamericanos, con el objetivo de percibir los logros educativos de los países a través de enfoques comparados, que permitieran generar indicadores comunes (Casassus & Cusato 1998b, pp. 7).

En el enfoque de efectos aleatorios, se calcula un parámetro global para todos los items, independiente de la población, y un parámetro local en cada población. La diferencia entre el parámetro general y el local, es considerada un efecto aleatorio. El valor del parámetro local corresponde al parámetro global más su efecto aleatorio en esa población. En las pruebas internacionales, el parámetro global es llamado internacional, y el de cada país es llamado nacional.

En el modelo diseñado en este trabajo, se estiman parámetros de los items específicos en cada cluster. Al mismo tiempo, se conserva una escala de medición común, porque se estima un parámetro global, que corresponde a la media de toda la población, que sirve como parámetro de comparación de las diferencias en el funcionamiento del item entre cada cluster y la media global. Con el cálculo de parámetros de los items en cada población, no es necesario que los items sean invariantes, ni es necesario clasificarlos como invariantes o no invariantes (Fox 2010, pp. 193).

## 2. Metodología

En este trabajo se generaliza el modelo unidimensional de TRI con efectos aleatorios (Fox 2010), en el sentido de incluir dos conjuntos de parámetros de los items en el modelo TRIM (Reckase 2009, pp. 86). Para estimar los parámetros del modelo diseñado, se desarrolló un algoritmo en base a la metodología descrita en Bolt (2003), se usaron técnicas Bayesianas, como lo han hecho autores como Fox (2010) y Montenegro (2011), entre otros. El modelo propuesto y el procedimiento de estimación se encuentran en la sección 4.

Se simuló datos a través de un procedimiento desarrollado en R, a los que se aplicó el modelo TRIM (Reckase 2009) y el modelo propuesto para recuperar los parámetros simulados. Se compararon los resultados y se evaluó la calidad de la recuperación de los parámetros usando el algoritmo con el que se implementó el modelo. Los resultados se encuentran en la sección 5.

Se aplicó el modelo propuesto a un conjunto de datos reales, tomados del Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado (Casassus & Cusato 1998a), suministrados por el ICFES. Se usó un diseño de muestreo aleatorio simple, para tomar

la muestra de tres países: Bolivia, Colombia y Cuba, correspondiente al seis por ciento del total de datos en cada país.

Para determinar el número de dimensiones de la muestra de datos reales se desarrolló un análisis paralelo (Horn 1965), (Reckase 2009, pp. 215). Para estimar los parámetros del modelo propuesto se implementó un algoritmo en Winbugs . Winbugs fue desarrollado conjuntamente por la MRC Biostatistics Unit, Cambridge y el Imperial College School of Medicine at St Mary's, Londres, se encuentra disponible para descargar en la página <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>.

Se estimaron los parámetros de los items usando el modelo TRIM clásico (Reckase 2009, pp. 86), con el objetivo de comparar el ajuste de los dos modelos. Se compararon los resultados de la aplicación de los modelos con el *Deviance Information Criterion* (DIC) (Carlin 2009, pp. 71). Se evaluó el desempeño del modelo propuesto con los métodos Conditional Predictive Ordinate CPO y CPO inverso (Ntzoufras 2009, pp. 344).

### 3. Modelos clásicos de TRI

En esta sección se describen los modelos de TRI usados en el desarrollo de este trabajo. Se usan modelos dicotómicos, donde las respuestas de las personas a los items se califican como 1 correcto, 0 incorrecto. En este trabajo se usan modelos TRI de dos parámetros. En estos modelos se supone que cada item tiene una discriminación y una dificultad específicas. La discriminación se refiere a la capacidad del item para detectar las diferencias de los trazos latentes entre las personas evaluadas, y se asume que la dificultad de cada item es diferente.

El parámetro  $\theta$  de la persona corresponde al trazo latente que el evaluador busca estimar a través de la aplicación de la prueba. Un modelo es unidimensional cuando el trazo latente es unidimensional. En el caso multidimensional,  $\theta$  corresponde a un vector de trazos latentes. La respuesta de un individuo a los items de una prueba, resulta de la interacción entre los parámetros  $\theta$  y los parámetros de los items (Reckase 2009, pp. 2).

#### 3.1. Modelo clásico unidimensional de TRI dicotómico

En los modelos unidimensionales de TRI (TRIU), se asume que la respuesta de un individuo a un item resulta de la interacción de los parámetros del item y un solo parámetro  $\theta$ . En estos modelos,  $\theta$  representa el parámetro que describe el trazo latente de la persona,  $\xi_k$  representa el vector de parámetros que describe las características de los items de la prueba,  $Y$  corresponde a las respuestas de una persona a los items de la prueba, y  $y$  representa un posible valor del promedio de respuestas correctas.  $f$  representa la función que describe la relación entre los parámetros del modelo y la probabilidad de respuesta del individuo. Así, un modelo TRIU de dos parámetros se define de forma general (Reckase 2009, pp. 12):

$$P(Y = y|\theta) = f(\theta, \xi_k, y)$$

El modelo UIRT de dos parámetros con enlace logístico, se define como:

$$P(Y_{ik} = 1|\theta_i, a_k, b_k) = \frac{e^{a_k(\theta_i - b_k)}}{1 + e^{a_k(\theta_i - b_k)}}$$

donde  $Y_{ik}$  corresponde a la variable aleatoria que representa la respuesta del individuo  $i$  al item  $k$ ,  $\theta_i$  denota el trazo latente del individuo  $i$  y  $a_k$  y  $b_k$  denotan los parámetros del item  $k$  (Fox 2010, pp. 9).

#### 3.2. Modelo TRI unidimensional con efectos aleatorios

El modelo con efectos aleatorios para el caso unidimensional (Fox 2010, pp. 193-209), fue desarrollado para analizar datos provenientes de pruebas aplicadas a poblaciones divididas en clusters, suponiendo que los individuos requieren de un trazo latente para responder a los items de la prueba. La estimación de ese trazo latente es el objetivo principal de la teoría.

En este modelo el autor asumió que debido a la clasificación de los evaluados en clusters, existen diferencias en los parámetros de los items entre clusters, más no dentro de ellos. Debido a que los items no son invariantes, se calcula un conjunto de parámetros de los items en cada cluster, y un conjunto de parámetros globales, que corresponde a la media de los parámetros de los items en toda la población. Así, se incluyeron las diferencias de los parámetros en cada cluster con el parámetro global, como efectos aleatorios. Se asumió que los items no son invariantes, pero las observaciones son condicionalmente independientes, es decir, las respuestas a los items son independientes entre las personas evaluadas.

Los parámetros globales de los items son invariantes en toda la población y se denotan  $\xi_k = (a_k, b_k)'$ . Se admite que las variables aleatorias que generan los parámetros de los items están correlacionadas. Fox (2010) propuso la siguiente distribución a priori para los parámetros de los items:

$$(a_k, b_k)' \sim NM(\mu_\xi, \Sigma_\xi) I_{A_k}(a_k),$$

donde  $(a_k, b_k)'$  provienen de una distribución normal bivariada truncada  $NM$ ,  $\mu_\xi$  y  $\Sigma_\xi$  denotan los hiperparámetros de los items,  $\Sigma_\xi$  denota la matriz de covarianza de los parámetros de los items,  $I$  denota la función indicadora y  $A_k = \{a_k \in R, a_k > 0\}$ , para admitir solo valores positivos en el parámetro  $a_k$ . Para modelar los hiperparámetros de los items, Fox (2010, pp. 36) propuso una distribución inversa de Wishart denotada como  $IW$ , como distribución a priori conjugada para la matriz de covarianza  $\Sigma_\xi$ , con matriz de escala  $\Sigma_0$  y grados de libertad  $v \geq 2$ ; y una distribución normal como a priori para  $\mu_\xi | \Sigma_\xi$ . La distribución conjunta propuesta a priori para los hiperparámetros  $(\mu_\xi, \Sigma_\xi)$ , se define:

$$\Sigma_\xi \sim IW(v, \Sigma_0),$$

$$\mu_\xi | \Sigma_\xi \sim N(\mu_0, \Sigma_\xi / K_0).$$

donde  $K_0$  denota el número de mediciones a priori,  $\mu_0 = (1, 0)'$ ,  $v=2$  y  $\Sigma_0$  es una matriz de escala que es mínimamente informativa de  $\Sigma_\xi$ .

Hasta ahora, no se han tenido en cuenta las variaciones de los items entre poblaciones. Para permitir estas variaciones, el autor incluyó efectos aleatorios en los parámetros de los items y propuso una distribución a priori para los efectos aleatorios que se define como:

$$\xi_{kg} = (a_{kg}, b_{kg})' \sim NM((a_k, b_k)', \Sigma_{\bar{\xi}}),$$

donde  $\xi_{kg}$  denota los parámetros del item  $k$  en la población  $g$  y  $NM$  denota la distribución normal bivariada. El autor propuso el modelo incluyendo los dos grupos de parámetros de los items, globales y locales, y usó enlace normal  $\Phi$  para facilitar la estimación de los parámetros del modelo (Fox 2010, pp. 198).

Teniendo en cuenta los efectos aleatorios de los items, el primer nivel del modelo de respuesta al item para datos binarios fue definido por Fox (2010) como:

$$P(Y_{igk} = 1 | \theta_i, \xi_{kg}) = \Phi(a_{kg}\theta_i - b_{kg})$$

donde  $Y_{igk}$  denota la variable aleatoria que representa la respuesta del individuo  $i$  de la subpoblación  $g$  al item  $k$ . En esta primera parte del modelo no se tiene en cuenta la agrupación de la población en clusters, se describe la probabilidad de respuesta correcta dado el parámetro  $\theta$  y los parámetros de los items en cada cluster.

El modelo TRIU con efectos aleatorios, es un modelo que integra los parámetros globales y sus efectos aleatorios, y se define de la siguiente manera:

$$P(Y_{igk} = 1 | \theta_i, \xi_k) = \Phi(a_k\theta_i - b_k - \epsilon_{bkg} + \epsilon_{akg}\theta_i)$$

donde  $Y_{igk}$  denota la respuesta del individuo  $i$  de la subpoblación  $g$  al item  $k$ ,  $\epsilon_{bkg}$  y  $\epsilon_{akg}$  denotan los efectos aleatorios de los parámetros de los items. Los efectos aleatorios de los parámetros de los items se distribuyen normalmente con media cero y varianza  $\sigma_{bk}^2$  y  $\sigma_{ak}^2$ , respectivamente (Fox 2010, pp. 198).

### 3.3. Modelo TRI Multidimensional

En el modelo TRI multidimensional, se asume que las respuestas de los individuos a una prueba, resultan de la interacción entre el conjunto de parámetros de los items de la prueba y un conjunto de trazos latentes de los individuos. Por lo anterior, en los modelos TRIM se tiene para modelar un vector de  $L$  parámetros  $\theta$  para cada persona evaluada, donde  $L$  corresponde a la dimensión del espacio de los trazos latentes, es decir, un vector de trazos latentes cuya estimación es el objetivo general de los modelos. Los parámetros de discriminación de los modelos TRIM conforman un vector  $\mathbf{a}_k$  en cada dimensión de la prueba.

Es importante tener en cuenta que se debe distinguir entre las dimensiones definidas por los modelos TRIM, que representan abstracciones estadísticas de los datos observados, y los trazos latentes que representan dimensiones de variación cognitivas o afectivas en una población (Reckase 2009, pp. 79).

Las respuestas de los items se pueden considerar como señales parciales de los trazos latentes de los examinados. Si una prueba está compuesta por  $K$  items, el patrón de respuesta de un individuo es un vector en el espacio  $\{0, 1\}^K$ . En la práctica, el número de dimensiones es inferior al número de items. Por lo anterior, el vector de trazos latentes de un individuo es una representación de su patrón de respuesta en un espacio euclidiano de dimensiones reducidas (Montenegro 2011, pp. 38).

En el modelo TRIM se calcula un parámetro de intercepto  $d_k$  para cada item, que es un escalar. El parámetro de dificultad en el caso multidimensional, se define como:

$$B_k = \frac{-d_k}{\sqrt{\sum_{l=1}^L a_{kl}^2}} \quad (1)$$

donde  $d_k$  denota el parámetro de intercepto del item  $k$ ,  $a_{kl}$  denota el parámetro de discriminación del item  $k$  en la dimensión  $l$  y  $L$  denota el número de dimensiones del espacio de los trazos latentes. El valor de  $B_k$  tiene la misma interpretación que en los modelos unidimensionales, también llamado MDIFF. Esta interpretación de  $B$  es válida solo en la dirección especificada por el  $\alpha$ -vector, que corresponde al vector de ángulos entre el eje de coordenadas y la línea desde el origen al punto que representa la ubicación de la persona  $i$  en el espacio solución (Reckase 2009, pp. 116-117).

El índice de discriminación multidimensional, se define como:

$$A_k = \sqrt{\sum_{l=1}^L a_{kl}^2} \quad (2)$$

donde  $a_{kl}$  denota el parámetro de discriminación del item  $k$  en la dimensión  $l$ .  $A_k$  es la discriminación multidimensional para el item  $k$ , también llamada  $MDISC_k$ . Dado que  $A_k$  tiene la misma forma matemática que el denominador de  $B_k$ , otra forma de expresar la dificultad multidimensional es  $B_k = -d_k/A_k$  (Reckase 2009, pp. 118).

La dirección de máxima discriminación desde el origen del espacio  $\theta$  hasta el punto del espacio  $\theta$  que tiene la mayor inclinación considerando todas las posibles direcciones, se obtiene de la siguiente expresión:

$$\cos\alpha_{kl} = \frac{a_{kl}}{\sqrt{\sum_{l=1}^L a_{kl}^2}}$$

donde  $\alpha$  corresponde al vector de ángulos que especifica la dirección donde la superficie tiene mayor pendiente. Los ángulos se obtienen a través del arcoseno del coseno de  $\alpha$ . Los ángulos y cosenos son características del item (Reckase 2009, pp. 116-117).

La extensión multidimensional del modelo TRI con enlace logístico fue presentada por Reckase (2009, pp. 86) y se define:

$$P(Y_{ki} = 1 | \theta_i, \mathbf{a}_k, d_k) = \frac{e^{\mathbf{a}'_k \theta_i + d_k}}{1 + e^{\mathbf{a}'_k \theta_i + d_k}}$$



donde  $Y_{ki}$  denota la variable aleatoria que representa la respuesta del individuo  $i$  al ítem  $k$ . A continuación se expande el exponente de  $e$  en el modelo, para mostrar la manera en que interactúan los parámetros de los ítems y de las personas:

$$a'_k \theta_i + d_k = a_{k1} \theta_{i1} + a_{k2} \theta_{i2} + \cdots + a_{kL} \theta_{iL} + d_k = \sum_{l=1}^L a_{kl} \theta_{il} + d_k$$

El exponente es una función lineal de los elementos de  $\theta$  con el parámetro  $d$  como intercepto y los elementos del vector  $a$  como los parámetros de la pendiente. La expresión en el exponente define una línea en un espacio  $L$ -dimensional.

### 3.3.1. Análisis de las dimensiones

Esta investigación se limita a modelos multidimensionales, que requieren la previa especificación de la dimensión del espacio de los trazos latentes para su uso. Cuando las pruebas están divididas en subpruebas, cada subprueba se diseña para medir un trazo latente, y obtener una estimación de él (Montenegro 2011, pp. 9), por lo que se puede definir la dimensión de la prueba como el número de subpruebas, aunque la dimensión de la prueba puede no coincidir con la dimensión de los datos.

En el espacio euclidiano en el cual pueden ser representados por vectores los trazos latentes de los individuos, los trazos latentes individuales no tienen una interpretación directa. Sin embargo, el objetivo de la prueba es medir algunos trazos latentes de los examinados, que se pueden interpretar a través de una teoría subyacente (Montenegro 2011, pp. 41).

Aunque el espacio de los trazos latentes es multidimensional, el conjunto de ítems usados en una prueba es sensible principalmente a diferencias a lo largo de una dirección en el espacio de los trazos latentes, según el concepto de unidimensionalidad esencial de Stout (1990). Este concepto se puede generalizar a dimensionalidad esencial, donde los ítems pueden ser agrupados en clases, sensibles principalmente a lo largo de una dirección en el espacio de los trazos latentes. En este caso, las dimensiones esenciales medidas por las clases no necesariamente son ortogonales. Estos trazos latentes no ortogonales se convierten en ortogonales a través de transformaciones lineales que no cambian los patrones de probabilidad, pero cambian la correlación de los trazos latentes (Montenegro 2011, pp. 3).

Un índice sintético unidimensional del vector de trazos latentes en un modelo TRIM es usualmente llamado una *composición*. Se define la composición  $\Theta_\beta$  del vector completo de trazos latentes  $\theta$ , como la combinación lineal de  $\theta$  que es  $\Theta_\beta = \beta' \theta = \sum_{k=1}^L \beta_k \theta_k$ , donde  $\beta = (\beta_1, \beta_2, \dots, \beta_L)$  es un vector constante llamado la dirección de la composición  $\Theta_\beta$ . Si  $Var(\Theta_\beta) = 1$ ,  $\Theta_\beta$  será una composición normalizada (Montenegro & Cepeda 2010).

Montenegro (2011, pp. 41) propuso los nombres de trazos latentes básicos a los componentes del vector de trazos latentes y trazos latentes principales, a los trazos latentes que la prueba intenta medir. Así, los principales trazos latentes se definen de acuerdo al diseño de la prueba. En pruebas con varias subpruebas, se calcula una dirección de referencia para cada uno, a partir de las direcciones de los ítems que lo conforman.

## 4. Modelo multidimensional jerárquico TRIMJ

En esta sección se introduce el modelo propuesto en este documento, que se denominará "*Modelo Multidimensional Jerárquico de dos Parámetros TRIMJ*", que es una generalización del modelo de (Fox 2010) en el caso multidimensional. Se estiman dos conjuntos de parámetros de los ítems: globales y locales; y los parámetros de las personas están compuestos por un vector de trazos latentes para cada individuo.

El conjunto de parámetros globales se denota  $\xi_k = (\mathbf{a}_k, d_k)'$ , donde  $\mathbf{a}_k$  es el vector de parámetros globales de discriminación del ítem  $k$ , y  $d_k$  es el parámetro de intercepto global del ítem  $k$ .  $\mathbf{a}_k$  corresponde a un vector de  $L$  parámetros, donde  $L$  denota el número de dimensiones del espacio de los trazos latentes, hay un parámetro de discriminación en cada dimensión.

El conjunto de parámetros locales se denota  $\xi_{kg} = (\mathbf{a}_{kg}, d_{kg})'$ , donde  $\mathbf{a}_{kg}$  es el vector de  $L$  parámetros de discriminación del item  $k$  en la población  $g$ , y  $d_{kg}$  es el parámetro de intercepto del item  $k$  en la población  $g$ .

Los parámetros  $a_{kg}$  conforman un vector de  $K$  parámetros en cada dimensión  $l$ , y estos vectores conforman una matriz de  $L$  vectores en cada población  $g$ . El modelo incluye  $K*G*L$  parámetros de discriminación en total. Los parámetros  $d_{kg}$  conforman un vector de  $K$  parámetros en cada subpoblación, donde  $K$  es el número total de items. El modelo incluye en total una matriz de  $K*G$  parámetros de intercepto, donde  $G$  es el número total de subpoblaciones.

El vector de trazos latentes se denota  $\theta_{ig}$ , donde  $i$  corresponde al individuo y  $g$  a la población a la cual pertenece. El vector  $\theta_{ig}$  está compuesto por  $L$  parámetros para cada individuo  $i$ . Se denota  $n_g$  al número de personas que conforman la subpoblación  $g$ , y el número de personas de la población total se denota  $N$ . Es decir,  $N = (n_1 + n_2 + n_3 + \dots + n_G)$ . En total, se estiman  $N*L$  parámetros  $\theta$ .

A partir de la anterior descripción de los parámetros, el modelo TRIMJ se define:

$$P(Y_{igk} = 1 | \theta_{ig}, \xi_{kg}) = \frac{e^{\mathbf{a}'_{kg} \theta_{ig} + d_{kg}}}{1 + e^{\mathbf{a}'_{kg} \theta_{ig} + d_{kg}}}, \quad (3)$$

donde  $Y_{igk}$  denota la variable aleatoria que representa la respuesta del individuo  $i$  de la subpoblación  $g$  al item  $k$ , para  $i=1, \dots, n_g$ ,  $g = 1, \dots, G$  y  $k=1, \dots, K$ . Se usa enlace logístico en el modelo TRIMJ porque facilita el proceso de estimación, y las diferencias en los resultados con el enlace normal son mínimas, como se detalla en la sección 6.

El exponente expandido de  $e$  se escribe como:

$$a'_{kg} \theta_{ig} + d_{kg} = a_{kg1} \theta_{ig1} + a_{kg2} \theta_{ig2} + a_{kg3} \theta_{ig3} + \dots + a_{kgL} \theta_{igL} + d_k = \sum_{l=1}^L a_{kg l} \theta_{ig l} + d_k$$

el exponente es el predictor lineal latente del modelo TRIMJ.

Para completar el modelo, se propone que:

$$(d_{kg} | d_k) \sim N(d'_k, \sigma_{d_k}^2)$$

$$(\mathbf{a}_{kg} | \mathbf{a}_k) \sim NM(a'_k, \sigma_{a_{kg}}^2) I_{A_{kg}}(a_k),$$

para  $k=1, 2, \dots, K$ , y  $g=1, 2, \dots, G$ .  $N$  denota la distribución normal,  $NM$  denota la distribución normal multivariada,  $I$  denota la función indicadora y  $A_{kg} = \{\mathbf{a}_{kg} \in R^L, \mathbf{a}_{kg} > 0\}$ , donde  $\mathbf{a}_{kg} > 0$  se interpreta  $a_{kg1} > 0$  y  $a_{kg2} > 0$ , hasta  $a_{kgL} > 0$ , para admitir solo valores positivos en el parámetro  $a_{kg}$ . En adelante, la pareja de varianzas  $\sigma_{d_k}^2, \sigma_{a_k}^2$ , se denotará  $\sigma_{\xi_k}^2$ .

Finalmente, para los trazos latentes, se propone que el trazo latente  $\theta_{ig}$  es la realización de una variable aleatoria  $\Theta_{I_g}$  con distribución normal multivariada, de acuerdo con Bolt (2003):

$$\Theta_{I_g} \sim NM(0, I_L),$$

donde  $I_L$  es la matriz identidad de orden  $L$ . Nótese que si  $\Sigma_\theta$  no fuera  $I$ , siempre es posible hacer una reparametrización para tener una distribución de los trazos latentes con la matriz  $\Sigma_\theta = I$ , lo que implica que la correlación entre las direcciones es muy cercana a cero.

#### 4.1. Función de verosimilitud

Bajo la hipótesis de independencia entre las personas evaluadas e independencia entre las respuestas de la misma persona (independencia condicional), la función de verosimilitud del modelo TRIMJ está dada por:

$$f(y | \xi, \theta) = \prod_{g=1}^G \prod_{i=1}^{n_g} \prod_{k=1}^K P_{igk}^{y_{igk}} (1 - P_{igk})^{1 - y_{igk}},$$

con  $P_{igk} = P(Y_{igk} = y_{igk} | \boldsymbol{\xi}_{kg}, \boldsymbol{\theta}_{ig})$ .  $y$  representa las respuestas de todos los individuos,  $\boldsymbol{\xi}$  representa todos los parámetros de los items y  $\boldsymbol{\theta}$  representa los trazos latentes de todos los individuos. Nótese que:

$$\begin{aligned} P(Y_{igk} = 0 | \boldsymbol{\theta}_{ig}, \boldsymbol{\xi}_{kg}) &= 1 - P(Y_{igk} = 1 | \boldsymbol{\theta}_{ig}, \boldsymbol{\xi}_{kg}) \\ &= \frac{1}{1 + e^{\mathbf{a}'_{kg} \boldsymbol{\theta}_{ig} + d_{kg}}}. \end{aligned}$$

## 4.2. Estimación de los parámetros del modelo TRIMJ

Los parámetros del modelo propuesto se estimaron con una estrategia Bayesiana. Para estimar los parámetros de discriminación del modelo, es común usar a priori una distribución lognormal (Fox 2010, pp. 35) o una distribución normal truncada (Glas 2001, pp. 545). Para el modelo TRIMJ, se propone una distribución normal multivariada truncada:

$$\mathbf{a}_k \sim NM(0, I_L) I_{A_k(a_k)}$$

donde  $I$  denota la función indicadora y  $A_k = \{\mathbf{a}_k \in R^L, \mathbf{a}_k > 0\}$ , para admitir solo valores positivos en el parámetro  $a_k$ . Para los parámetros de intercepto, se propone a priori una distribución normal:

$$d_k \sim N(0, 1000).$$

Siguiendo la estrategia usada en Fox (2010, pp. 35) y McKay (2010, pp. 13), se propone para la varianza de los parámetros locales  $\sigma_{\xi}^2$  una distribución gamma inversa:

$$\sigma_{\xi}^2 \sim \text{gammainv}(0.1, 1000).$$

El modelo TRIMJ no es identificable. Para hacerlo identificable, se definieron tres vectores de discriminación como (1,0,0), (0,1,0) y (0,0,1), respectivamente y un vector de parámetros  $d_k$  se fijó en cero, como es usual en estos casos.

El algoritmo desarrollado para estimar los parámetros del modelo TRIMJ a partir de las estrategias descritas, es el siguiente:

1. Definir los valores iniciales para los parámetros de los items:  $a=1$ ,  $d=0$ .
2. Muestrear los parámetros locales  $\mathbf{a}_{kg}$  y  $d_{kg}$  para todo  $kg$ .
3. Muestrear los parámetros globales  $\mathbf{a}_k$  y  $d_k$  para todo  $k$ .
4. Muestrear los trazos latentes  $\boldsymbol{\theta}_{ig}$  para todo  $ig$ .
5. Muestrear la varianza de los parámetros locales  $\sigma_{a_k}^2$  para todo  $k$ .
6. Muestrear la varianza de los parámetros locales  $\sigma_{d_k}^2$  para todo  $k$ .
7. Calcular los estadísticos de ajuste del modelo.

El algoritmo de estimación fue implementado en Winbugs para estimar los parámetros de los items y los trazos latentes. La implementación se desarrolló en base al procedimiento descrito por Bolt (2003). Winbugs fue desarrollado conjuntamente por la MRC Biostatistics Unit, Cambridge y el Imperial College School of Medicine at St Mary's, Londres, se encuentra disponible para descargar en la página <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>.

## 5. Simulación

Con el objetivo de evaluar la calidad de la recuperación de los parámetros con el algoritmo desarrollado, este se usó para recuperar los parámetros de los items y de los trazos latentes haciendo uso del modelo TRIMJ con varios conjuntos de datos simulados; que varían en número de items, dimensiones y subpoblaciones. Se usó el modelo TRIM clásico con algunos conjuntos de datos, para observar el comportamiento de este modelo con datos con subpoblaciones y para comparar la calidad de la recuperación de parámetros de los dos modelos. Al generar los datos simulados y luego recuperados, se siguió estrictamente

el modelo teórico y las restricciones propuestas en la sección 4.2 para hacer identificable el modelo. Se calculó la correlación, el error absoluto y el error cuadrático medio (ECM) entre los parámetros simulados y recuperados.

Para generar los datos, se siguió la estrategia propuesta por Kromrey et al. (1999), Montenegro (2011), y se tomó como base el procedimiento desarrollado en *R* (Team 2008) por Montenegro (2011), que puede ser consultado en el Apéndice B. *R* se encuentra disponible para descargar en <http://cran.r-project.org/>. En el procedimiento de simulación de los conjuntos de datos se asignaron valores diferentes a los parámetros de los ítems de cada subpoblación. Los parámetros de intercepto se simularon de tal manera que tuvieran distribución normal, con media cero y varianza uno. Los parámetros de discriminación fueron simulados de tal manera que tuvieran distribución lognormal, media cero y varianza 1/4. Los trazos latentes  $\theta$ , fueron simulados de tal manera que tuvieran distribución normal, media cero y matriz de varianza la identidad.

En el proceso de recuperación de los parámetros se usó el código de Winbugs que se encuentra en el Apéndice A. Se generaron para todos los casos 10000 muestras de las distribuciones posteriores más 1000 iniciales como calentamiento de la cadena. En todos los casos se observó que eran suficientes para la convergencia de las cadenas. El tiempo de estimación osciló entre 1.5 y 12 horas, en un computador con procesador Pentium (R) T4500 2.3 GHz, con velocidad de memoria 2 GB.

En la tabla 1 se encuentra la descripción de las simulaciones con sus características. Se puede observar que se usaron conjuntos de datos con diferentes características, y se recuperaron los parámetros de los ítems con el modelo TRIMJ excepto en la simulación 1, y en algunos casos con el modelo TRIM clásico. En los casos donde se describen varias simulaciones con las mismas características, estas se diferencian en que se generó un conjunto de datos diferente para cada simulación, aunque con las mismas características. En la tabla 1 se detalla que en algunos conjuntos de datos los parámetros de discriminación se generaron con distribución normal truncada.

TABLA 1: Descripción de las simulaciones

Simulación	Número de subpoblaciones	Número de ítems	Número de personas	Número de dimensiones $\theta$	Modelos usados
1	2	20	1000	2	TRIM clásico
2-4	2	20	1000	2	TRIMJ sin recuperación y con recuperación de los parámetros globales de los ítems
5-9	2	20	1000	2	TRIMJ sin recuperación de los parámetros globales de los ítems
10	2	100	1000	2	TRIMJ sin recuperación y con recuperación de los parámetros globales de los ítems
11	3	50	1000	3	TRIMJ sin recuperación de los parámetros globales de los ítems
12	2	30	1000	2	TRIMJ sin recuperación de los parámetros globales de los ítems
13-15	2	30	1000	3	TRIMJ recuperación de los parámetros globales de los ítems
16	3	20	1500	2	TRIMJ sin recuperación y con recuperación de los parámetros globales de los ítems
17	3	30	1500	3	TRIMJ sin recuperación y con recuperación de los parámetros globales de los ítems
18-20	2	20	1000	2	TRIMJ, distribución normal truncada para los parámetros de discriminación
21	2	20	5050	2	TRIMJ, distribución normal truncada para los parámetros de discriminación, tamaños de las poblaciones: 5000 y 50 respectivamente
22	2	20	1000	2	TRIMJ, y TRIM clásico. Distribución normal truncada para los parámetros de discriminación

Los resultados de la simulación 1 de la tabla 1 mostraron que cuando se usó el modelo TRIM clásico (Reckase 2009, pp. 86), para observar la calidad de la recuperación de los parámetros cuando no se tienen en cuenta las subpoblaciones, esta fue de baja calidad, como se espera. La correlación entre los parámetros de los ítems simulados y recuperados fue superior a 0,9 en la mayoría de los casos, lo que se puede considerar adecuado. Sin embargo, el ECM fue superior a 0.7 en general, lo que indica la baja calidad de la recuperación de estos parámetros, a pesar tener una correlación adecuada, de acuerdo con la literatura, ver por ejemplo Patz et al. (2002) y Bolt (2003). En cuanto a los trazos latentes, en la dimensión 1 la correlación entre los trazos simulados y recuperados fue adecuada, pero el ECM fue demasiado alto,

y en la dimensión 2 la correlación fue demasiado baja. En el Apéndice D pueden leerse los resultados de esta simulación.

Los resultados de la simulación 22, descrita en la tabla 1, se pueden observar en las tablas 2 y 3. En esta simulación, se usaron los modelos TRIM clásico y TRIMJ con el mismo conjunto de datos, para comparar la calidad de la recuperación de los parámetros con los dos modelos. Obsérvese en la tabla 2 que los parámetros de los items recuperados con el modelo TRIM clásico tuvieron valores de correlación por encima de 0.9, con valores de ECM menores a 0.1, lo que indica que los parámetros recuperados son similares a los parámetros globales simulados, como se espera. Los parámetros locales recuperados con el modelo TRIMJ, tuvieron una recuperación adecuada, de acuerdo con los altos valores de correlación y los valores bajos de error absoluto y ECM. Entre los parámetros  $\theta$  reales y recuperados, las correlaciones oscilaron entre 0.7 y 0.8 con los dos modelos, pero nótese que los valores de los errores absolutos y el ECM, fueron mucho más grandes con el modelo TRIM clásico que con el modelo TRIMJ, lo que indica que la recuperación de parámetros del modelo TRIMJ fue de mejor calidad respecto al modelo TRIM clásico, como se espera.

TABLA 2: Resultados simulación 22, con el modelo TRIM clásico

Media global	Correlación	Error absoluto	Varianza	Error cuadrático
$a_1$	0,976	0,069	0,147	0,089
$a_2$	0,962	0,075	0,126	0,100
$d$	0,985	0,053	0,198	0,073
Subpoblación 1	Correlación	Error absoluto	Varianza	Error cuadrático
$\theta_1$	0,77	0,507	0,632	0,637
$\theta_2$	0,79	0,485	0,57	0,611
Subpoblación 2	Correlación	Error absoluto	Varianza	Error cuadrático
$\theta_1$	0,78	1,006	0,571	1,268
$\theta_2$	0,82	1,042	0,626	1,299

TABLA 3: Resultados simulación 22, con el modelo TRIMJ

Media global	Correlación	Error absoluto	Varianza	Error cuadrático
$a_1$	0,984	0,075	0,143	0,089
$a_2$	0,96	0,102	0,16	0,123
$d$	0,99	0,061	0,23	0,073
Subpoblación 1	Correlación	Error absoluto	Varianza	Error cuadrático
$a_1$	0,977	0,083	0,179	0,101
$a_2$	0,967	0,083	0,152	0,112
$d$	0,991	0,084	0,576	0,109
$\theta_1$	0,776	0,506	0,65	0,636
$\theta_2$	0,792	0,484	0,606	0,609
Subpoblación 2	Correlación	Error absoluto	Varianza	Error cuadrático
$a_1$	0,95	0,099	0,152	0,13
$a_2$	0,925	0,134	0,208	0,175
$d$	0,976	0,099	0,306	0,118
$\theta_1$	0,783	0,49	0,582	0,623
$\theta_2$	0,823	0,444	0,623	0,567

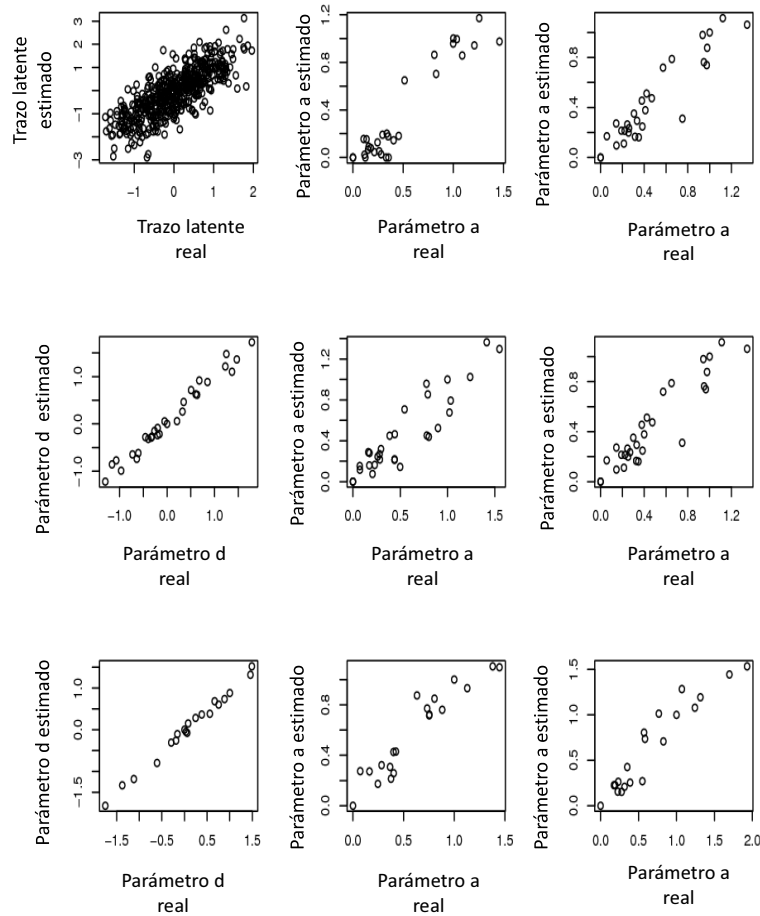


FIGURA 1: Resultados de algunas simulaciones.

Primera fila, simulación 11. Segunda fila, simulación 16. Tercera fila, simulación 15.

Los resultados de las simulaciones 2 a 21, descritas en la tabla 1, mostraron que la recuperación de los parámetros con el modelo TRIMJ fue de buena calidad en todos los casos. En el Apéndice E se encuentran los resultados de las correlaciones, los errores absolutos y el ECM de algunas de las simulaciones descritas en la tabla 1. En general, las correlaciones entre los parámetros de los ítems de los datos simulados y los parámetros recuperados fueron superiores a 0.9, lo que se considera adecuado. Los errores absolutos oscilaron entre 0.07 y 0.15, y el ECM fue menor a 0.22, estos resultados también se consideran adecuados de acuerdo con la literatura. La correlación entre los trazos latentes simulados y recuperados osciló entre 0.7 y 0.8, el error absoluto entre 0.4 y 0.5, y el ECM estuvo alrededor de 0.6. Aunque el error absoluto y el ECM son notablemente más altos que con los parámetros de los ítems, este resultado no es motivo de preocupación porque se simuló más personas que ítems en todos los casos. En la figura 1 se observan los resultados de algunas simulaciones. Nótese en la gráfica que la recuperación de los parámetros de intercepto tienen mejor calidad respecto a los parámetros de discriminación, y la recuperación de los trazos latentes es la de menor calidad. En general, los trazos latentes se recuperan con más error respecto a los parámetros de los ítems, lo que se puede considerar normal porque la cantidad de ítems fue siempre menor a la cantidad de personas en las simulaciones. Se observa también que la recuperación de los

parámetros de intercepto, fue de mejor calidad que la de los parámetros de discriminación, resultado que ya ha sido observado por autores como Fox (2010) y Montenegro (2011, pp. 51).

## 6. Aplicación a una muestra de datos de la prueba PERCE

Para este trabajo se tomó una muestra de datos del *Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado (PERCE)* (Casassus & Cusato 1998b). Los datos fueron suministrados por el ICFES. La prueba PERCE fue aplicada por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), con el objetivo de comprender el estado de la educación en los países a través de enfoques comparados, generar indicadores comunes y diseñar recomendaciones políticas (Casassus & Cusato 1998b, pp. 7). Uno de los aspectos de mayor interés por parte del LLECE, fue considerar los contenidos de los currículos de los países participantes, por lo que el contenido de la prueba fue aprobado por consenso entre los coordinadores de educación de los países participantes.

La prueba fue aplicada a niños de tercer y cuarto grado entre los meses de junio y noviembre de 1997 en trece países de América Latina. La prueba tiene dos componentes: matemática que consta de 32 ítems, y lenguaje que consta de 19 ítems. Se diseñaron dos cuestionarios: A y B, que están compuestos por diferentes ítems, pero el número de ítems es el mismo. A cada estudiante le correspondió responder uno de los dos cuestionarios al azar, posiblemente con el fin de evitar fraude en las respuestas.

Para este trabajo se tomó una muestra aleatoria de respuestas del cuestionario B, de tres de los países evaluados. Se eligió Colombia para observar su desempeño, Cuba por ser un país con desempeño atípico en la prueba, y Bolivia porque tuvo un desempeño diferente en la prueba (Casassus & Cusato 1998b, pp. 31). El tamaño de la muestra corresponde al seis por ciento de los estudiantes de cada país, se obtuvieron para Bolivia 252 estudiantes, Colombia 194 estudiantes y Cuba 156 estudiantes.

En los datos de la prueba PERCE recibidos del ICFES, las no respuestas fueron calificadas como cero. Las respuestas de las dos subpruebas se encuentran separadas, sin identificación de los estudiantes. Para la aplicación de un modelo MIRT, se requiere el uso de las respuestas de cada estudiante en las dos subpruebas, para obtenerlas, se realizó por parte del ICFES y de la autora la identificación de los estudiantes en cada subprueba, comparando la información de los cuestionarios adicionales que fueron aplicados.

### 6.1. Análisis para detectar la dimensionalidad del espacio de los trazos latentes

Para detectar la dimensión de los trazos latentes se usó el análisis paralelo propuesto por Horn (1965). De acuerdo con Reckase (2009, pp. 216), este procedimiento se ha sugerido por Ledesma (2007) y otros y tiene una larga historia en la literatura sobre análisis factorial, y es tan bueno como los métodos más elaborados.

El análisis paralelo consta de dos pasos: un análisis dimensional que proporciona los primeros  $n$  valores propios de la matriz de correlaciones entre ítems, y la generación de un conjunto de datos que no tienen relación entre los ítems, pero tiene la misma proporción de respuestas correctas de los datos reales para cada ítem. Se analizan los datos generados con el mismo procedimiento de los datos reales, para obtener valores propios. Posteriormente, se dibujan los valores propios de los datos generados y los valores propios de los datos reales que son más grandes que los de los generados. Este es el número de dimensiones sugerido para el análisis (Reckase 2009, pp- 216). El análisis paralelo se encuentra implementado en R, en paquetes como `pcaPA` desarrollado por el ICFES y `paran` de Dinno (2012).

El análisis paralelo aplicado a la matriz de la muestra de datos de las dos subpruebas con el paquete `paran` de R (Dinno 2012), mostró que puede ser adecuado el uso de 3 dimensiones para los trazos latentes de la muestra. En la figura 2 izquierda se observa que hay 3 valores propios de los datos más grandes que los valores propios de los datos generados, por lo que el análisis sugiere el uso de 3 dimensiones de los trazos latentes. Este resultado se contrastó con un análisis en componentes principales, que sugirió igualmente la retención de 3 dimensiones, como se observa en el histograma de valores propios de la figura 2 derecha.

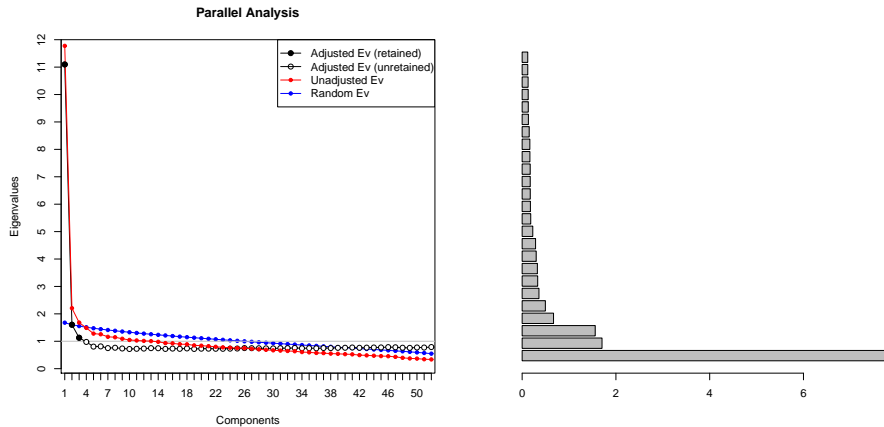


FIGURA 2: Análisis de dimensiones de los trazos latentes de la muestra de datos de la prueba PERCE  
Izquierda: análisis paralelo. Derecha: histograma de valores propios del análisis en componentes principales.

Para estimar los parámetros haciendo uso del modelo TRIMJ, se usó el código de Winbugs que se encuentra en el apéndice Apéndice A. Se desarrollaron 1000 iteraciones para calentamiento, 10000 iteraciones para el muestreo y 2000 adicionales para calcular los estadísticos de bondad de ajuste. Se fijaron los vectores de parámetros de discriminación de los items 1, 15 y 51 para definir las direcciones, con los siguientes valores: (0,0,1), (1,0,0) y (0,1,0), respectivamente. Se fijó el parámetro de intercepto del item 1 en 0. Se asumió que la matriz de covarianza de los trazos latentes es igual a la identidad. En la figura 3 se observan algunas cadenas.

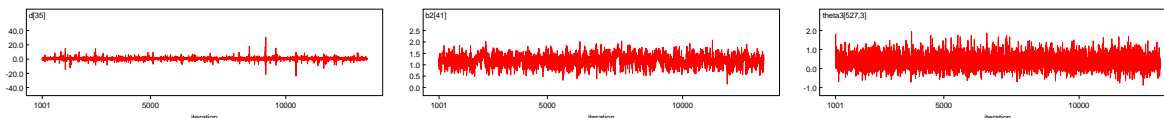


FIGURA 3: Muestra de las cadenas de Markov de la aplicación del modelo TRIMJ  
Izquierda: parámetro d internacional, centro: parámetro  $a_2$  Colombia, derecha: parámetro  $\theta_3$  de uno de los estudiantes de Cuba.

## 6.2. Estimación de los parámetros de los items

Se estimaron dos conjuntos de parámetros de los items con el modelo TRIMJ: los parámetros globales, que en adelante se nombran parámetros internacionales, y los parámetros locales, que corresponden a un conjunto de parámetros en cada país, que en adelante se nombran parámetros nacionales. En el Apéndice G pueden leerse los resultados de las estimaciones con el modelo TRIMJ.

Los resultados de la estimación de los parámetros de los items fueron coherentes con lo que se esperaba observar. En la subprueba de matemática, se presentaron diferencias entre países especialmente en los parámetros de discriminación nacionales. En la subprueba de lenguaje, se presentaron diferencias entre países especialmente en los parámetros de intercepto, y las diferencias entre los parámetros internacionales y nacionales fueron también más grandes en los items de esta subprueba.

En los resultados de la subprueba de matemática, los valores de los parámetros de intercepto oscilaron entre -1.926 y 3.2. Los parámetros de intercepto nacionales tuvieron valores diferentes en Cuba, en los items 2, 10, 12, 15, y 16. En Colombia, los items 6, 7 y 17 tuvieron diferencias en los valores de los parámetros, y en Bolivia el item 18 tuvo un valor diferente. Los resultados pueden leerse en la tabla 18 del Apéndice G.

A continuación, se describen los parámetros de discriminación de la subprueba de matemática por dimensión. En la dimensión 1, los parámetros de discriminación tuvieron valores que oscilaron entre 0.19 y



1.9. En esta dimensión, el ítem 28 tuvo un valor más grande en Cuba. Los parámetros de discriminación de la dimensión 2, tuvieron un valor promedio de 1. En la dimensión 3, los valores de los parámetros de discriminación oscilaron entre 0.15 y 3.82, además de un valor extremo de 7.8 en el ítem 7. En esta dimensión se pueden observar las diferencias más grandes entre países en los valores de los parámetros de discriminación. Los ítems 5, 7, 8, 12, 13, 14 y 28, tuvieron valores más grandes en Cuba en esta dimensión. Estos resultados pueden leerse en las tablas 15, 16 y 17 del Apéndice G.

En cuanto a la subprueba de lenguaje, los valores de los parámetros de intercepto oscilaron entre -1.4 y 3.3. Se observa que 12 de los 19 ítems de lenguaje fueron más fáciles en Cuba. Los resultados pueden leerse en la tabla 18 del Apéndice G. Las diferencias entre los parámetros de intercepto nacionales e internacionales, son más grandes en los parámetros de intercepto, en las dos subpruebas. De acuerdo con Fox (2010, pp. 218), esta variación se presenta porque las dificultades de los ítems incluyen la variación en las medias de los trazos latentes entre países, que no es explícitamente modelada.

Los parámetros de discriminación por dimensión de la subprueba de lenguaje, se describen a continuación. En la dimensión 1, los valores de los parámetros oscilaron entre 0.12 y 0.84. El ítem 34 tuvo un valor más grande en Cuba, el ítem 37 en Colombia, y los ítems 36, 42, 47 y 48 en Bolivia. En la dimensión 2, los valores de los parámetros oscilaron entre 0.47 y 1.82. Los ítems 34, 35 y 38 tuvieron valores más grandes en Cuba, el ítem 44 en Colombia y el ítem 36 en Bolivia. En la dimensión 3, los valores de los parámetros oscilaron entre 0.28 y 1, el ítem 38 tuvo un valor más grande en Cuba y el ítem 36 en Colombia.

Para una mejor interpretación de los parámetros de discriminación e intercepto estimados con el modelo TRIMJ, se calcularon los parámetros MDISC y MDIFF, con el objetivo de hacer una interpretación análoga al caso unidimensional. La definición de estos parámetros puede leerse en la sección 3, y las estimaciones respectivas aparecen en las tablas 19 y 20 del Apéndice G.

El parámetro de discriminación multidimensional MDISC, es la pendiente de la superficie de respuesta del ítem en el punto de máxima pendiente y en dirección del origen del espacio  $\theta$ . Éste parámetro es análogo al parámetro de discriminación en el modelo TRI unidimensional. El parámetro de dificultad multidimensional MDIFF, se interpreta como el parámetro  $b$  del modelo TRI unidimensional, es decir, un ítem es más difícil entre mayor sea su valor de MDIFF, y más fácil entre menor sea su valor de MDIFF. Esta interpretación solo aplica en la dirección especificada por el vector  $\alpha$  (Reckase 2009, pp. 117-118). Los resultados de los parámetros MDISC y MDIFF mostraron diferencias entre países, y con el parámetro internacional. A continuación se describen los resultados de los ítems por subpruebas.

En la subprueba de matemática, se presentaron mayores diferencias en el parámetro MDISC respecto al MDIFF. El parámetro MDISC de Cuba tuvo valores marcadamente diferentes con el parámetro MDISC internacional. En los ítems 2, 3, 5, 7 y 8 el valor del parámetro MDISC de Cuba fue más grande respecto al parámetro MDISC internacional, y en los ítems 4, 10 y 28 el valor fue menor. En Colombia, el ítem 16 tuvo un parámetro MDISC con menor valor respecto al internacional e igualmente en Bolivia en el ítem 18. Con relación a los valores del parámetro MDIFF, en 17 de los 32 ítems de matemática se presentaron diferencias notorias entre los parámetros MDIFF internacionales y los nacionales. Se destacan nuevamente las diferencias de los valores del parámetro en Cuba, donde 14 de los 32 ítems tuvieron un valor de MDIFF menor al parámetro MDIFF internacional, y los ítems 14, 16, 19, 21, 28, 29 y 31 tuvieron un valor de MDIFF mayor al del parámetro internacional. En Colombia, el ítem 16 tuvo un valor de MDIFF mayor al internacional. En Bolivia, los ítems 21 y 28 tuvieron un MDIFF mayor al internacional.

En la subprueba de lenguaje, se presentaron mayores diferencias en el parámetro MDIFF respecto al MDISC. El parámetro MDISC tomó valores similares en los tres países, solo se destacó la diferencia del valor del parámetro MDISC del ítem 36 de Bolivia, que fue mayor que el parámetro internacional. En cuanto al parámetro MDIFF, este tuvo valores más pequeños respecto al parámetro internacional en Cuba, en 11 de los 19 ítems de la subprueba. En Colombia, los ítems 44 y 48 tuvieron valores de MDIFF menores al internacional y lo mismo ocurrió en Bolivia, en el ítem 43.

### 6.3. Estimación de los trazos latentes básicos

En el modelo TRIMJ las diferencias son modeladas con la estrategia usada en los modelos jerárquicos. Siguiendo a Fox (2010), se asumió que los trazos latentes provienen de una única distribución. Para estimar los trazos latentes, se fijó la matriz de covarianza como la identidad, por lo cual estos parámetros

no tienen interpretación directa (Montenegro 2011). Los resultados de la estimación de los trazos latentes se encuentran en las tablas 18 a la 23 del Apéndice G.

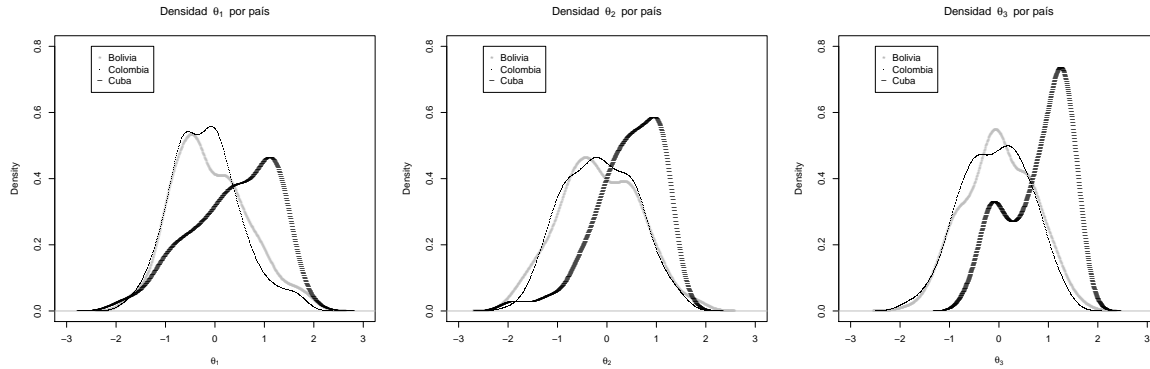


FIGURA 4: Densidad de los trazos latentes estimados  $\theta$  por país

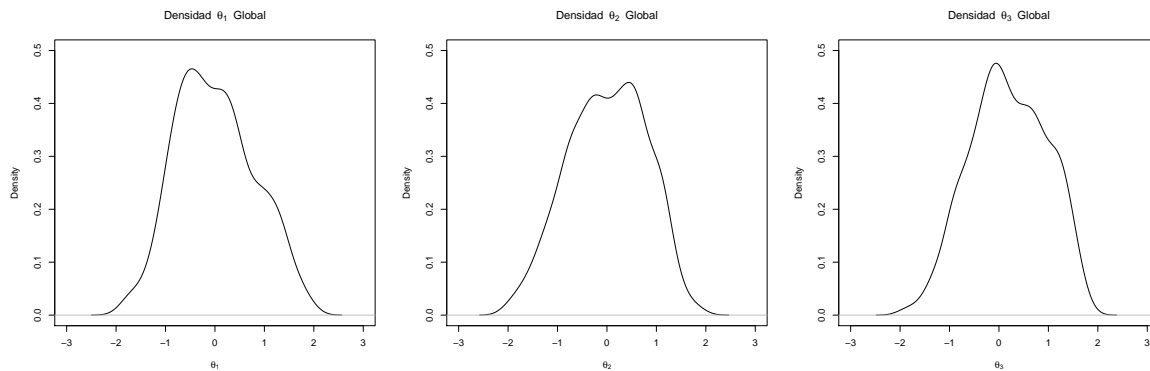


FIGURA 5: Densidad de los trazos latentes estimados  $\theta$  globales

Los trazos latentes básicos estimados con el modelo TRIMJ, mostraron diferencias en las densidades por país. En la figura 4, se observa que la densidad del parámetro  $\theta$  en Cuba no parece coincidir con los supuestos, porque parece ser asimétrica. La densidad de los parámetros  $\theta$  de Bolivia y Colombia es bastante similar.

#### 6.4. Estimación de los trazos latentes principales

Aunque en el modelo TRIMJ los componentes del vector de trazos latentes no tienen una interpretación directa, es posible intentar hacerla. Por ejemplo, se puede calcular el ángulo entre la dirección de cada ítem y cada uno de los ejes de coordenadas, y con base en esta información, tratar de construir una interpretación. Por otro lado, como el interés de la prueba es la estimación de los trazos latentes atribuibles a matemática y lenguaje, Montenegro (2011) propuso calcular la dirección principal de cada subprueba. Otros autores han trabajado en el mismo problema, ver por ejemplo, Wang (1986), Stout (1990) y Yen (1985).

En este trabajo, se usó la transformación de Wang (1986) para calcular el trazo latente unidimensional. Sea  $A$  la matriz cuyas columnas son los vectores de parámetros de discriminación de cada subprueba, y sea  $\theta$  el trazo latente estimado por el modelo TRIMJ, se realizó para cada subprueba la siguiente transformación:

$$G = w * \theta$$

donde  $w$  es el primer vector propio de la matriz  $A * A'$ , que se conoce como la dirección principal de la subprueba (Stout 1990) y (Montenegro 2011).  $G$  es el trazo latente unidimensional en el modelo aproximado, asociado a la dirección principal de la prueba (Montenegro & Cepeda 2010), por lo que en este caso  $G$  corresponde al trazo latente asociado a la subprueba (matemática ó lenguaje). Sea  $A_1$  la matriz cuyas columnas son los vectores de parámetros de discriminación de la subprueba de matemática, para calcular el trazo latente asociado a la componente de matemática  $G_1$ , se calculó el primer vector propio de la matriz  $A_1 * A_1'$ , que corresponde a la dirección principal de la subprueba de matemática, que se denotará  $w_1$ . Sea  $A_2$  la matriz cuyas columnas son los vectores de parámetros de discriminación de la subprueba de lenguaje, para calcular el trazo latente asociado a la componente de lenguaje  $G_2$ , se calculó el primer vector propio de la matriz  $A_2 * A_2'$ , el vector la dirección principal de la subprueba de lenguaje, que se denotará  $w_2$ . Se calculó un trazo latente unidimensional  $G$  para cada país, con los parámetros de discriminación nacionales, y un trazo latente  $G$  internacional, con los parámetros de discriminación internacionales. En las tablas 4 y 5 se encuentran los vectores  $w$  calculados para matemática y lenguaje, respectivamente.

TABLA 4: Vectores de dirección principal  $w$  de Matemática

$w_1$	Dimensión 1	Dimensión 2	Dimensión 3
Internacional	0.65	0.38	0.65
Bolivia	0.65	0.37	0.65
Colombia	0.65	0.38	0.65
Cuba	0.28	0.16	0.94

TABLA 5: Vectores de dirección principal  $w$  de Lenguaje

$w_2$	Dimensión 1	Dimensión 2	Dimensión 3
Internacional	0.26	0.85	0.44
Bolivia	0.29	0.85	0.43
Colombia	0.25	0.84	0.45
Cuba	0.24	0.85	0.44

Las densidades de los trazos latentes  $G$  tienen diferencias entre Cuba y los demás países en las dos subpruebas, como se puede observar en la figura 6. En Bolivia y Colombia, las gráficas de densidad son similares en las dos subpruebas. En Cuba, la media de la distribución es mayor a cero en las dos subpruebas, además, las distribuciones parecen ser asimétricas.

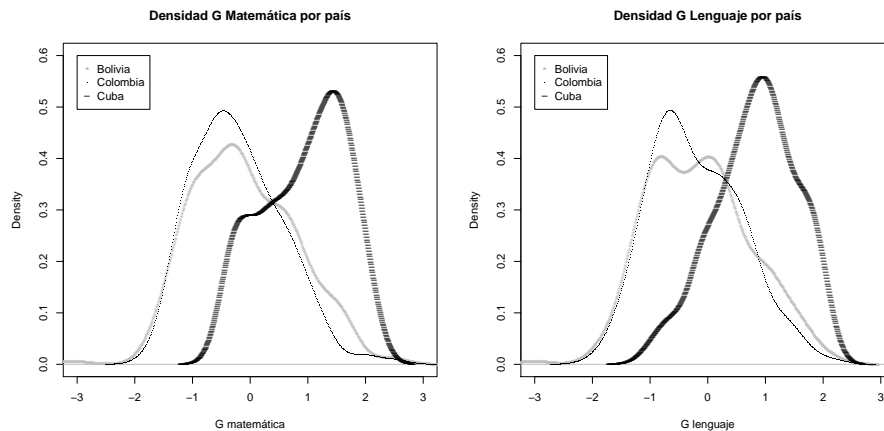


FIGURA 6: Densidad del trazo latente unidimensional  $G_1$  y  $G_2$  por país

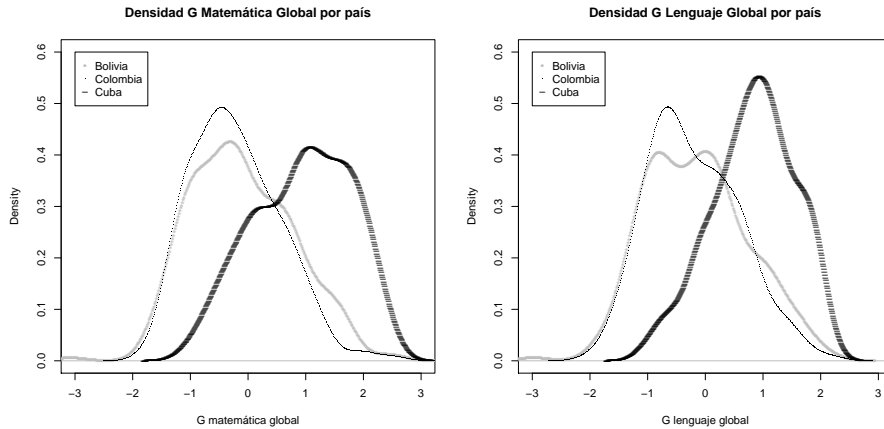


FIGURA 7: Densidad del trazo latente unidimensional  $G_1$  y  $G_2$  Internacional por país

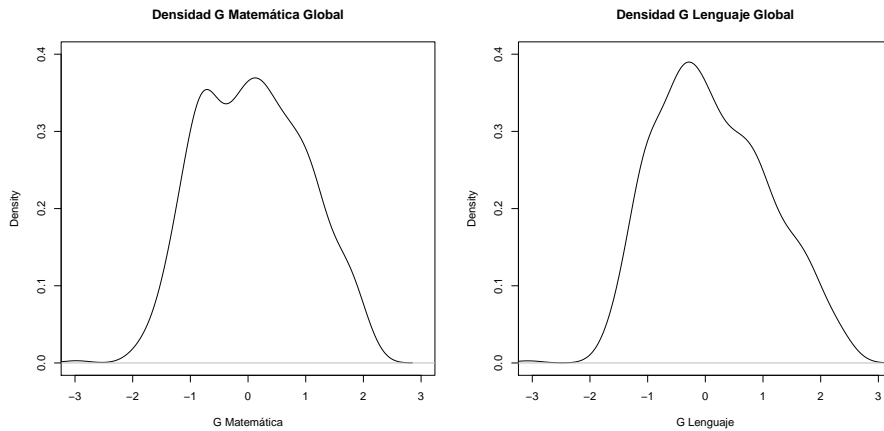


FIGURA 8: Densidad del trazo latente  $G_1$  y  $G_2$  Internacional

En la figura 8 se observa la densidad del trazo latente unidimensional  $G$  internacional de matemática y lenguaje, respectivamente. Esta densidad tiene distribución simétrica, contrario a lo ocurrido en Cuba con la densidad de los trazos latentes  $G$  estimados por país. En la figura 9 se encuentra la densidad conjunta de los trazos latentes  $G$  internacionales estimada por métodos Kernel. Para estimar la gráfica de la izquierda, se usó la función `kde` del paquete `ks` (Duong 2013) de R y para estimar la gráfica en tres dimensiones de la derecha, se usaron las funciones `kde2d` del paquete `MASS` (Ripley 2013) de R y `persp` de R.

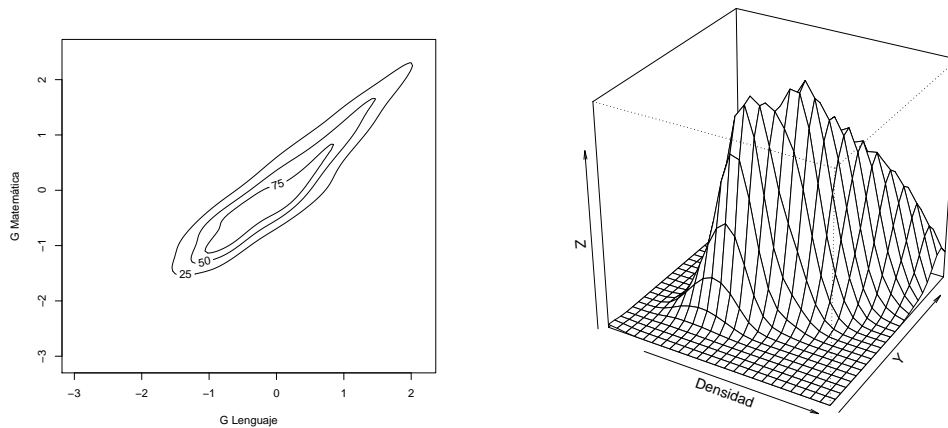


FIGURA 9: Densidad del trazo latente  $G_1$  y  $G_2$  Internacional

En la tabla 6 se pueden leer las correlaciones del puntaje clásico (media de las respuestas correctas por individuo), con el trazo latente unidimensional  $G$  de cada subprueba, el internacional y los parámetros  $\theta$ , por subprueba. La correlación entre el puntaje clásico y el trazo latente  $G$  por país en las dos subpruebas, estuvo en un rango de 0.88 a 0.97. La correlación más baja por país estuvo en Cuba en la subprueba de lenguaje. En cuanto a la correlación del puntaje clásico con los trazos latentes  $\theta$ , el puntaje clásico de la subprueba de matemática tuvo mayor correlación con los parámetros  $\theta_1$  y  $\theta_3$ , y de la subprueba de lenguaje con los parámetros  $\theta_2$ .

Se calculó la correlación del puntaje clásico entre los datos de matemática y lenguaje, el resultado fue 0.716, y la correlación entre los trazos latentes  $G_1$  y  $G_2$  fue de 0.874, por lo que pueden existir dimensiones de los trazos latentes de los individuos presentes en las dos subpruebas. Lo anterior ilustra el hecho de que las componentes del vector de trazos latentes no puede interpretarse directamente, como los constructos teóricos.

TABLA 6: Correlaciones entre el puntaje clásico y los trazos estimados con el modelo TRIMJ

Puntaje clásico/trazos estimados	$G_{Bolivia}$	$G_{Colombia}$	$G_{Cuba}$	$G_{Internacional}$	$\theta_1$	$\theta_2$	$\theta_3$
Matemática	0.97	0.97	0.90	0.97	0.776	0.447	0.791
Lenguaje	0.96	0.954	0.88	0.94	0.347	0.881	0.586

En la figura 10 se puede observar la comparación entre el puntaje clásico de las pruebas de matemática y lenguaje y los trazos latentes  $G$  calculados con los parámetros internacionales de discriminación. La gráfica de lenguaje tiene mayor dispersion que la gráfica de matemática, porque estudiantes con el mismo puntaje clásico, tuvieron valores diferentes del trazo latente unidimensional  $G_2$ .

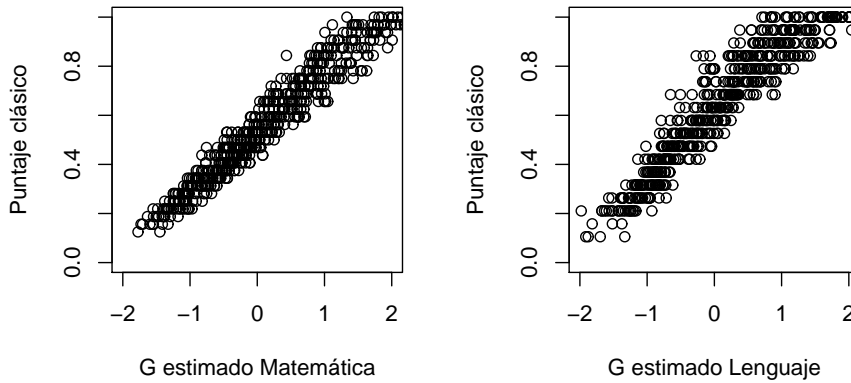


FIGURA 10: Puntaje clásico vs. trazos estimados con los parámetros internacionales modelo TRIMJ  
Izquierda: matemática, derecha: lenguaje. Y: puntaje clásico, X: trazos latentes unidimensionales G

A continuación, se describen los resultados de cada una de las subpruebas. Los resultados de matemática se pueden observar en figura 11, donde se compara el puntaje clásico con el trazo latente  $G_1$  calculado para cada país. La gráfica de Bolivia (izquierda) tiene poca dispersión, lo que muestra que los estudiantes con igual puntaje clásico tuvieron valores similares del trazo latente  $G_1$ . En la gráfica de Colombia (centro), se observa que pocos estudiantes obtuvieron puntajes clásicos altos en la prueba, la mayoría obtuvieron puntajes clásicos bajos y medios, y los estudiantes con igual puntaje clásico tuvieron valores similares de  $G_1$ . En la gráfica de Cuba (derecha), se observa que los estudiantes tuvieron puntajes clásicos más altos que los estudiantes de los otros países, la mayoría de los estudiantes se ubican en los promedios de la mitad hacia arriba del puntaje clásico. Hay mayor dispersión en la gráfica de Cuba, donde estudiantes con el mismo puntaje clásico tuvieron diferente valor de  $G_1$ , incluso cuando este es igual a 1, lo que muestra el impacto de la correlación entre las estimaciones de los trazos latentes principales.

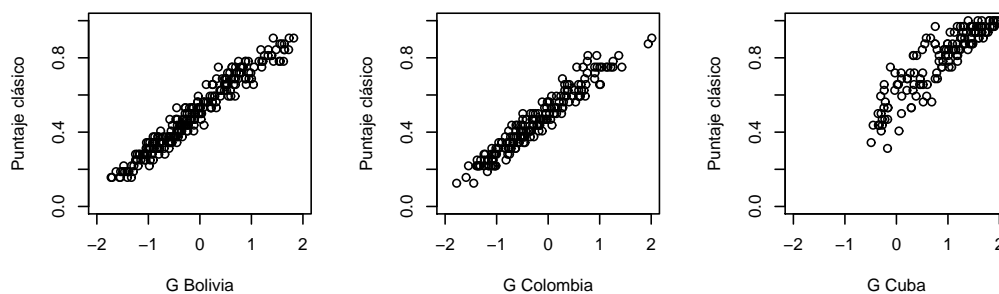


FIGURA 11: Puntaje clásico vs. trazos estimados con los parámetros nacionales de los ítems de matemática por país. Izquierda: Bolivia, centro: Colombia, derecha: Cuba. Y: puntaje clásico, X: trazos latentes unidimensionales  $G_1$

Los resultados de lenguaje se pueden observar en figura 12, donde se compara el puntaje clásico con el trazo  $G_2$  calculado para cada país. Las gráficas de Bolivia (izquierda) y Colombia (centro) no muestran diferencias significativas. En la gráfica de Cuba (derecha) se observa nuevamente que los puntajes clásicos de los estudiantes fueron más altos que en los otros países.

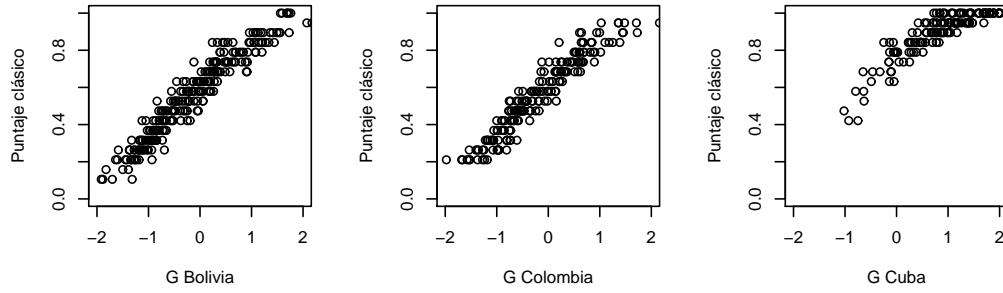


FIGURA 12: Puntaje clásico vs. trazos estimados con los parámetros nacionales de los ítems de lenguaje por país Izquierda: Bolivia, centro: Colombia, derecha: Cuba. Y: puntaje clásico, X: trazos estimados

### 6.5. Evaluación del ajuste del modelo TRIMJ

Para la evaluación de la calidad del modelo TRIMJ, se tuvieron en cuenta varios criterios: el *Deviance Information Criterion* DIC comparado con el modelo TRIM clásico, el *Conditional Predictive Ordinate* CPO y el CPO inverso. Como criterio de bondad de ajuste, se usó la evaluación de las predicciones. El código usado en winbugs para la estimación de los criterios de evaluación y bondad de ajuste del modelo se encuentra en el Apéndice C.

Para detectar los outliers y valores extremos, se usó el criterio de CPO y CPO inverso. De acuerdo con Ntzoufras (2009, pp. 344), el CPO se define:

$$CPO_i = \frac{f(y)}{f(y_B)} f(y_i|y_B) = \int f(y_i|\theta) f(\theta|y_B) d\theta$$

donde  $y_i$  es la  $i$ -ésima observación de  $y$  y  $y_B$  es  $y$  después de omitir  $y_i$ . Asumiendo normalidad aproximada, los valores de CPO inverso mayores de 40 se pueden considerar como posibles outliers y mayores de 70 como valores extremos (Ntzoufras 2009, pp. 376). Los valores de CPO inverso del modelo resultaron en un 0.09 por ciento de outliers y un 0.04 por ciento de valores extremos. Dado que el porcentaje de outliers es inferior al 5 por ciento, este resultado no es motivo de preocupación (Fox 2010, pp. 111).

En el análisis de CPO inverso en cada país, los resultados fueron los siguientes: en Bolivia, 0.03 por ciento de outliers y 0.01 por ciento de valores extremos; en Colombia, 0.03 por ciento de outliers y cero valores extremos; y en Cuba, 0.2 por ciento de outliers y 0.12 por ciento de valores extremos. En Cuba, el porcentaje de outliers y valores extremos es aceptable de acuerdo con Fox (2010), sin embargo, es marcadamente superior a los otros países.

Para seleccionar el mejor modelo, se usó el DIC, que de acuerdo con Carlin (2009, pp. 71) se define:

$$DIC = \bar{D} + p_D = 2\bar{D} - D(\bar{\theta})$$

donde  $\bar{D} = E_{\theta|y}[D]$  y  $p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta})$ . Valores más pequeños de DIC indican un mejor ajuste del modelo (Carlin 2009, pp. 71). El valor total del DIC para el modelo TRIMJ fue de 31099.5, inferior al DIC estimado para el modelo TRIM (Reckase 2009), 31556.4, lo que indica un mejor ajuste del modelo TRIMJ. Estos resultados pueden leerse en la tabla 7.

TABLA 7: Comparación DIC modelos TRIMJ y TRIM

Parámetro	TRIM	TRIMJ
a1	66.542	66.542
a2	66.542	30.115
a3	66.542	53.012
Response	31356.7	30875.5
Total	31556.4	31099.5

Como criterio de bondad de ajuste del modelo, se usó la evaluación de la predicción (Fox 2010). El método consiste en generar predicciones con el modelo, y comparar los datos replicados con los datos observados.

Las predicciones también son llamadas medidas de discrepancia o estadísticas de la prueba. Estas se eligen para detectar diferencias sistemáticas entre un modelo  $M$  de TRI con parámetros  $(\theta, \xi)$  y los datos (Fox 2010). La hipótesis nula, asume que los datos de respuesta se distribuyen condicionalmente como  $p(y|\xi)$  y los parámetros desconocidos tienen distribución a priori  $p(\theta, \xi)$ , sin tener una hipótesis alterna específica. El interés se enfoca en el estadístico  $T(y)$ , para investigar la compatibilidad del modelo  $M$  con los datos observados  $y_{obs}$ . Un p-valor se puede estimar así:

$$p_0(y_{obs}) = P(T(Y) \geq T(y_{obs})|M).$$

Para calcular los residuales, las predicciones y los p-valores, se usó el modelo TRIMJ con enlace logístico y se añadió en el exponente el valor de 1.702, para obtener estimaciones equivalentes al modelo con enlace normal (Reckase 2009, pp. 95), porque se usó el código de Winbugs de Fox (2010, p. 111), que está diseñado para ser usado con modelos con enlace normal. Se calculó un p-valor para estimar la compatibilidad del modelo con los datos observados, resultando un total de 0.08 por ciento de los p-valores menores a 0.05, y con el criterio de 0.1, se encontraron 0.38 por ciento de p-valores menores. Por lo anterior, se considera que el modelo se ajusta adecuadamente a los datos.

## 7. Discusión

Existen pruebas que se diseñan para ser aplicadas a varias poblaciones, por ejemplo, a varios países. Entre ellas, se encuentran pruebas como PISA (Program for International Student Assessment) y PERCE (Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado), utilizada en esta investigación, entre otras.

A través de la información aportada por la aplicación de pruebas internacionales, es posible evaluar los sistemas de educación de los países participantes, en los temas contenidos en la prueba. De esa manera, es posible percibir los procesos y resultados de los países en el campo educativo (Casassus & Cusato 1998b, pp. 5). El análisis a fondo de los resultados de las evaluaciones, aporta información para diseñar y poner en práctica políticas públicas adecuadas para la atención de los problemas descubiertos y de esta manera mejorar la calidad de los sistemas educativos (OCDE 2006, pp. 29). Teniendo en cuenta el fuerte vínculo entre la calidad de la educación y el progreso de los países, ya reconocido internacionalmente, estas acciones son de gran importancia y pueden tener alto impacto en el desempeño de los países en muchos aspectos.

En el diseño de una prueba internacional se tiene en cuenta que los resultados deben ser comparables entre los países evaluados, para cumplir el objetivo de su aplicación. Para tal fin, se han usado diferentes estrategias. Por ejemplo, el contenido de la prueba PERCE fue aprobado por los coordinadores de educación de cada uno de los países participantes, para garantizar que los contenidos de la prueba estuvieran incluidos en los currículos de todos los países. Por otro lado, en la prueba PISA se evalúan las competencias desarrolladas en el estudiante durante su proceso educativo. Sin embargo, a pesar de los esfuerzos que se realizan en el diseño y aplicación de las pruebas, estas en ocasiones tienen diferencias entre poblaciones, que pueden ser de dos formas: en las medias de las distribuciones de los trazos latentes de las personas o en los valores de los parámetros de los ítems (Fox 2010, pp. 195). Por ejemplo, Fox (2010) estudió datos provenientes de la prueba PISA, y encontró diferencias en los parámetros de los ítems entre poblaciones, es decir, los ítems no son invariantes.

En cuanto al análisis de resultados de pruebas internacionales, es frecuente el uso de modelos de TRI, como en el caso de la prueba PERCE (Casassus & Cusato 1998b), con la que se usó un modelo de Rasch unidimensional para el análisis de los resultados de cada subprueba. Para que la aplicación de este modelo TRI sea significativa, es necesario que se cumplan al menos dos condiciones: que los trazos latentes sean unidimensionales y que los ítems sean invariantes entre subpoblaciones. A pesar de que para el análisis de la prueba PERCE se usaron modelos de Rasch para cada una de las subpruebas, en los informes de resultados no se hace referencia alguna a la razón por la cual se eligió este modelo, ni a la condición de unidimensionalidad de los trazos latentes.



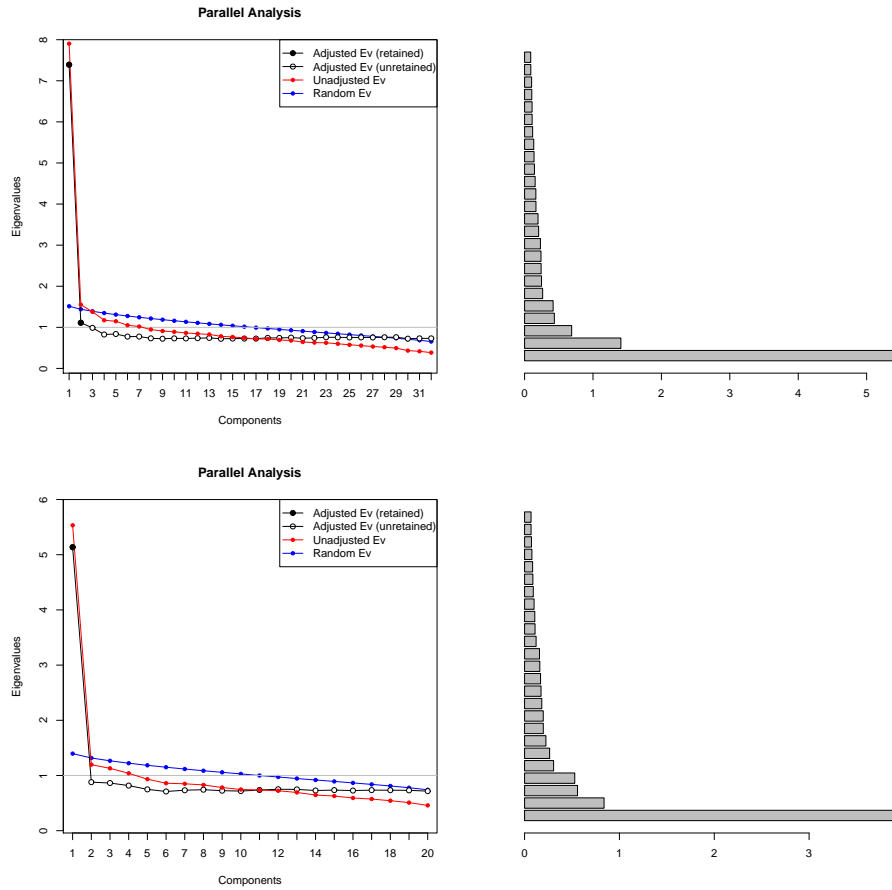


FIGURA 13: Análisis de dimensiones de la muestra de datos de la prueba PERCE por subpruebas Arriba: matemática, abajo: lenguaje.

La primera condición requerida para que la aplicación de un modelo de Rasch sea significativa, se verificó en este estudio para los datos de la prueba PERCE a través de la aplicación de un análisis paralelo y un análisis en componentes principales a los datos de la muestra tomada de cada subprueba, para determinar la dimensión sugerida para los trazos latentes de cada una. El resultado de este análisis sugirió 2 dimensiones para matemática y 2 dimensiones para lenguaje. Los resultados se pueden observar en la figura 13. Adicionalmente, en la sección 6 se mostró que el mismo procedimiento aplicado a la muestra de datos de matemática y lenguaje de manera conjunta, sugirió la retención de tres dimensiones para los trazos latentes (ver figura 2). Los anteriores resultados, indican que el uso de un modelo de Rasch no es la metodología más adecuada para el análisis de los resultados de esta prueba, porque los trazos latentes no son unidimensionales en ninguna de las dos subpruebas.

La segunda condición requerida para ajustar los datos de todos los países para cada subprueba, es la invarianza de los ítems. El informe Casassus & Cusato (2001, pp. 182) de la prueba PERCE, indica que se verificó esta condición a través del desarrollo de un procedimiento denominado por los evaluadores “análisis clásico de ítems”, en los ítems y subpruebas para cada país separadamente. Los estadísticos clásicos de los ítems fueron: índice de dificultad, medido como el porcentaje de respuestas correctas e índice de discriminación, medido a través del coeficiente de correlación punto-biserial entre un ítem y el puntaje total de la prueba.

Los resultados del “índice de dificultad” calculado por los evaluadores, mostraron grandes diferencias entre los países. En algunos países los ítems fueron mayoritariamente fáciles, y en otros fueron mayoritariamente difíciles (Casassus & Cusato 2001, pp. 182). Los resultados del “índice de discriminación clásico” calculado por los evaluadores, mostraron que algunos ítems tuvieron discriminaciones especialmente bajas

en Brasil, pero en general, el índice de discriminación fue alto. Adicionalmente, el informe indicó que una confiabilidad individual alta y una buena discriminación constituyen un requisito para un programa de pruebas individuales, más esto no es necesario en un estudio con fines de investigación, donde el contenido de la prueba tiene prioridad sobre la confiabilidad individual (Casassus & Cusato 2001, pp 182-183).

La estrategia usada por los evaluadores de la prueba PERCE para generar escalas comunes para todos los países, habiendo encontrado diferencias entre ellos, fue el uso de un modelo de Rasch. Extrajeron una muestra de datos, para las formas A y B, de los grados tercero y cuarto en cada país. El tamaño de la muestra fue de aproximadamente 300 estudiantes para cada combinación de formas y grados para cada país, en lenguaje y en matemática. La meta era aplicar la Escala de Rasch a un grupo representativo de estudiantes y dar a cada país igual participación en la definición de la escala. Las muestras (aproximadamente 7.200 para cada prueba) se consideraron lo suficientemente grandes para generar una escala de Rasch precisa (Casassus & Cusato 2001, pp. 183-184).

En este estudio, se desarrollaron análisis de dimensionalidad de los datos previas al ajuste del modelo TRIMJ, que sugirieron el uso de tres dimensiones de los trazos latentes de los individuos. Los parámetros de los items tuvieron diferentes valores en cada país, resultado coherente con lo encontrado por los evaluadores de la prueba PERCE en el informe Casassus & Cusato (2001). Estos resultados confirmaron los supuestos de este trabajo: los trazos latentes son multidimensionales y los parámetros de los items varían entre países, lo que mostró que los supuestos del modelo de Rasch no se cumplen en la prueba PERCE, por lo que los análisis desarrollados en los informes Casassus & Cusato (1998b) y Casassus & Cusato (2001) con el modelo de Rasch, no son válidos desde el punto de vista teórico.

Hasta ahora, se ha demostrado que las condiciones requeridas para la aplicación de un modelo de Rasch, no se cumplieron en los resultados de la prueba PERCE. A continuación, se describen varias alternativas de análisis de los resultados de la prueba, que resultan más adecuadas que el uso del modelo de Rasch.

Una de las posibles alternativas para el análisis de resultados de la prueba PERCE, es el uso del modelo propuesto por Sheng & Wikle (2008), donde se estima una dimensión del trazo latente para todos los items de la prueba, y varias dimensiones para cada subprueba. Con esta estrategia, se estima la matriz de correlaciones entre los trazos latentes de cada subprueba.

Otra alternativa para el análisis de los resultados de la prueba PERCE es el uso de un modelo TRIM clásico (Reckase 2009). En este modelo se asume que los trazos latentes de las subpruebas son multidimensionales, por lo que se supera la limitación del modelo de Rasch al análisis de pruebas con trazos latentes unidimensionales. Adicionalmente, Los resultados de análisis multidimensionales pueden ayudar a aclarar las razones para grandes FDI estadísticos (Reckase 2009, pp. 75).

El modelo unidimensional de TRI con efectos aleatorios propuesto por Fox (2010), es otra alternativa para el análisis de resultados de la prueba PERCE, porque fue diseñado para el análisis de resultados de pruebas con items no invariantes. En este modelo, se estima un conjunto de parámetros de los items en cada población, para tener en cuenta las características específicas de los items en cada una, y se estima un conjunto de parámetros globales de los items, para permitir las comparaciones entre poblaciones. Sin embargo, el autor asumió que el trazo latente de los individuos es unidimensional, supuesto que no se cumple en esta prueba, y que es poco probable que se cumpla en la mayoría de los casos, porque es posible que los individuos requieran más de un trazo latente para responder a los items de la prueba (Reckase 2009).

La alternativa propuesta en este trabajo para el análisis de pruebas internacionales como la prueba PERCE, es el uso del modelo TRIMJ, que fue diseñado para modelar datos de pruebas aplicadas a poblaciones divididas en clusters, donde los trazos latentes son multidimensionales y hay items que no son invariantes entre poblaciones. Con el uso del modelo TRIMJ se tienen en cuenta las dimensiones de los trazos latentes, a partir de un análisis de dimensiones previo al ajuste del modelo. Además, se estima un conjunto de parámetros de los items en cada población, que permite tener en cuenta las características específicas de los items en cada una, y se estima un parámetro global, que permite hacer comparaciones entre poblaciones, porque se establece una escala de medición común entre países, sin necesidad de que los items sean invariantes. En la sección 6, se demostró que la aplicación de este modelo a una muestra de datos de la prueba PERCE produjo mejores resultados que la aplicación del modelo TRIM clásico.

El modelo propuesto TRIMJ, fue diseñado con el uso de la estrategia de los modelos jerárquicos, y con el supuesto que los trazos latentes provienen de una única distribución, siguiendo a Fox (2010). Este modelo

se usó para ajustar varios conjuntos de datos simulados, obteniendo resultados satisfactorios, de acuerdo con lo recomendado por la literatura acerca del tema, ver por ejemplo Patz et al. (2002) y (Bolt 2003). Se usó el modelo TRIM clásico con datos simulados con subpoblaciones, con lo que se verificó que ese modelo no ajusta adecuadamente esta clase de datos, por lo que es necesario modelarlos teniendo en cuenta que los items no son invariantes.

Se desarrolló una aplicación del modelo a una muestra de datos reales de la prueba PERCE, el ajuste fue adecuado de acuerdo a los criterios de evaluación que se usaron. El porcentaje de valores atípicos y valores extremos fue aceptable de acuerdo con la literatura. La evaluación de las predicciones resultó en una compatibilidad adecuada del modelo con los datos observados. Con el criterio del DIC, se seleccionó el modelo TRIMJ como mejor modelo, comparado con el modelo TRIM clásico. Estos resultados pueden leerse en la sección 6.

En la aplicación del modelo TRIMJ, se encontró que los individuos que obtuvieron puntaje clásico igual a 1 en una subprueba, es decir, que contestaron todas las preguntas correctamente en una subprueba, obtuvieron valores diferentes del trazo latente en la dirección de referencia, como se puede observar en la figura 10. En cambio, los estudiantes que contestaron correctamente todos los 51 items, obtuvieron valores iguales del trazo latente en las tres dimensiones. Es decir, los estudiantes que contestaron igual las preguntas en una subprueba, y diferentes en la otra, tuvieron diferentes valores en los trazos latentes en la subprueba en la cual contestaron igual. Este resultado es coherente con la correlación de las dos subpruebas, que se calculó en 0.638 en los puntajes clásicos y 0.87 en  $G$ , en los datos analizados conjuntamente. Así, la estimación de los trazos latentes fue influenciada por la correlación que existe entre ellos.

En la figura 12 se observa que en Bolivia, los valores más grandes del trazo latente corresponden a dos estudiantes que obtuvieron un puntaje clásico inferior a uno en la prueba. Existen otros estudiantes que obtuvieron puntaje clásico de uno en la prueba, y tienen un trazo latente de menor valor. Lo anterior puede deberse a que esos estudiantes tuvieron mejores resultados en la prueba de matemática. Los que tuvieron mayor puntaje clásico en la prueba de matemática, tuvieron un trazo latente más grande en la prueba de lenguaje, aunque hayan tenido menor cantidad de respuestas correctas que otros en esta prueba. Lo anterior puede ser el resultado de la influencia mutua entre las dos subpruebas, posiblemente debido a la correlación entre ellas. Sin embargo, algunos autores afirman que estos resultados son fallas de los modelos.

Los dos valores más grandes del trazo latente calculados en la muestra, tanto en matemática como en lenguaje, son de estudiantes de Bolivia. Aunque los estudiantes de Cuba tuvieron mayor cantidad de respuestas correctas, los parámetros más altos no pertenecen a estudiantes de Cuba, sino de Bolivia. Lo anterior puede deberse a que el modelo permite evaluar a los estudiantes de cada país de acuerdo a sus propios parámetros de los items. En contraste, los análisis con el modelo TRIM clásico (Reckase 2009, pp. 86) dieron como resultado que los trazos latentes más altos en las dos subpruebas, son de estudiantes de Cuba.

Los resultados de los estudiantes de Cuba, fueron superiores en puntaje. Adicionalmente, los resultados de los análisis desarrollados mostraron un comportamiento atípico en Cuba en la mayoría de los casos. Debido a esto, en el ICFES se sugirió retirar a Cuba del análisis en este trabajo, pero precisamente el comportamiento atípico de los resultados de los estudiantes de Cuba en la prueba, fue el motivo por el cual se tomaron datos de este país para la aplicación del modelo propuesto. Sin los datos de Cuba los resultados habrían sido más homogéneos, pero este trabajo se realizó para tratar de dar respuesta a interrogantes que se encuentran sin solucionar, antes que para facilitar el desarrollo de la investigación. La correlación del puntaje clásico con los trazos latentes estimados fue menor en Cuba, especialmente en la subprueba de Lenguaje (ver tabla 6). Sin embargo, la correlación del puntaje clásico de cada país con el trazo latente  $G$  global que se puede leer en la tabla 8, fue similar en los tres países en la subprueba de matemática, aunque inferior en la subprueba de lenguaje. Además, en la evaluación del modelo se encontró que la mayor parte de los valores atípicos y extremos estuvieron en Cuba. Las direcciones principales  $w$ , tuvieron diferencias en Cuba en la subprueba de matemática, en la subprueba de lenguaje no tuvieron diferencias significativas.

TABLA 8: Correlación del puntaje clásico con el trazo latente unidimensional  $G$  global por país

Correlación	Bolivia	Colombia	Cuba
Matemática	0.97	0.96	0.95
Lenguaje	0.96	0.955	0.87

Los resultados atípicos de Cuba pueden estar explicados por las diferencias en las densidades de los trazos latentes de Cuba. Las gráficas de densidad de los trazos latentes  $\theta$  y del trazo latente unidimensional  $G$ , mostraron que Bolivia y Colombia tuvieron distribuciones similares, pero la distribución de Cuba parece ser asimétrica, lo que muestra que hay clusters. Sin embargo, la densidad del trazo latente  $G$  internacional en cada subprueba fue simétrica, por lo que se cumple el supuesto de normalidad internacionalmente (ver figura 8). Por el anterior resultado, es posible que el uso de distribuciones asimétricas sea más apropiado para modelar los datos de toda la población, así se podrían para mejorar las estimaciones, sin afectar la comparabilidad de los resultados entre países. Otros autores han desarrollado modelos con distribuciones asimétricas, como Bazan (2005) y Da Silva (2008).

De acuerdo con Carnoy (2007), existen tres posibles explicaciones principales para el desempeño atípico de los estudiantes de Cuba respecto a otros países en la región, a saber: diferencias en las familias del estudiante, especialmente en la educación de los padres y el énfasis del hogar en el éxito académico de los niños; las diferencias en la cantidad y calidad de los recursos académicos de la escuela y las diferencias en el entorno social del alumno fuera de la escuela.

Según el informe Casassus & Cusato (1998b, pp. 31), en los países donde hay poblaciones con lenguas maternas diferentes al español los resultados fueron más bajos, comparados con los países donde la mayoría de la población habla español. Un ejemplo de ello es Bolivia, donde hay grupos de población que hablan idiomas diferentes al español. Además, lo anterior puede ser la causa de las diferencias de los parámetros de Bolivia en los ítems de lenguaje, que son más grandes que en los demás países.

El uso del modelo TRIMJ, permitió mejorar la estimación de los trazos latentes de los países, porque se estimó un conjunto de parámetros de los ítems específico para cada país, que fue usado para estimar los trazos latentes de los estudiantes. Lo anterior implica el uso una escala diferente para cada uno de los países, entonces, para hacer comparaciones entre países, se pusieron los ítems en dirección del parámetro internacional.

La aplicación del modelo TRIMJ al análisis de pruebas internacionales supera la necesidad de ítems invariantes, porque se hacen comparaciones significativas a través de la estimación de un conjunto de parámetros de los ítems en cada población, a pesar de las diferencias en los parámetros de los ítems entre países. El parámetro global permite comparar los parámetros de los ítems entre países, y los parámetros de los ítems en cada población permiten evaluar a los estudiantes de acuerdo con los parámetros de su propio país, lo que evita que se presente sesgo en los trazos latentes a causa de las diferencias en los parámetros de los ítems.

## 8. Conclusiones

Se propuso un modelo multidimensional jerárquico de teoría de respuesta al ítem TRIMJ, para modelar datos provenientes de pruebas aplicadas a poblaciones con clusters. En el modelo propuesto se estima un conjunto de parámetros de los ítems para cada cluster de población, y un conjunto de parámetros globales de los ítems, para hacer comparaciones de resultados entre clusters. El modelo TRIMJ se aplicó a datos simulados y reales.

El modelo propuesto se usó con varios conjuntos de datos simulados, con diferente número de ítems, personas y dimensiones. El algoritmo funcionó satisfactoriamente en las simulaciones. Se usó el modelo TRIM clásico para ajustar datos con clusters, lo que resultó en una recuperación deficiente de los parámetros del modelo, de acuerdo con lo sugerido por la literatura acerca del tema. Este resultado muestra la necesidad de modelar los datos teniendo en cuenta que hay diferencias entre clusters.

Se aplicó el modelo TRIMJ a una muestra de datos reales de la prueba PERCE (Casassus & Cusato 1998b) de tres países: Bolivia, Colombia y Cuba. El ajuste del modelo TRIMJ fue adecuado de acuerdo con los resultados de los criterios de evaluación aplicados, y se verificó el cumplimiento de los supuestos. Se usó además el modelo TRIM clásico, para comparar los resultados de los dos modelos con los datos reales. Al

comparar el DIC de los dos modelos resultó seleccionado el modelo TRIMJ como mejor modelo. Por los tiempos de estimación con Winbugs, se recomienda para grandes aplicaciones considerar el uso de otras herramientas más veloces.

En la aplicación a datos reales, los parámetros de los ítems mostraron diferencias entre países. Este resultado es coherente con lo encontrado por autores que han analizado pruebas similares, como Fox (2010, pp. 216) en la prueba PISA (Program for International Student Assessment). Las diferencias más marcadas en los parámetros de los ítems en las dos subpruebas se presentaron en Cuba. En la subprueba de matemática, varios ítems tomaron valores más grandes de MDISC en Cuba respecto al parámetro internacional, lo que hace más discriminantes estos ítems en Cuba. En la subprueba de lenguaje, varios ítems tuvieron valores de MDIFF más bajos en Cuba respecto al parámetro internacional, lo que indica que la subprueba fue en general más fácil para los estudiantes de Cuba seleccionados en la muestra. Por lo anterior, aunque cuando se aplicó la prueba los autores asumieron invarianza de los ítems (Casassus & Cusato 1998b), estas diferencias en los parámetros entre países mostraron que los ítems tienen diferentes características en cada país observado.

En el análisis de resultados de la prueba PERCE el LLCE usó un modelo de Rasch (Casassus & Cusato 1998b), pero en la aplicación a datos reales se cumplieron los supuestos del modelo propuesto, a saber: multidimensionalidad de los trazos latentes y diferencias de los parámetros de los ítems de los países con los parámetros internacionales. Lo anterior, implica que no se cumplen los supuestos del modelo de Rasch en los datos de la prueba PERCE por lo que los resultados de los análisis del LLCE no son válidos desde el punto de vista teórico. Por todo lo anterior, los resultados de los análisis de la prueba PERCE pueden ser ampliamente mejorados con la aplicación del modelo TRIMJ.

Los resultados de Cuba fueron diferentes y atípicos en la mayoría de los análisis desarrollados. Los parámetros de los ítems de Cuba tuvieron diferencias más grandes con el parámetro internacional, en comparación con los otros dos países. La correlación entre los trazos latentes estimados y el puntaje clásico de los estudiantes de Cuba, fue menor respecto a los otros países. Las distribuciones de los trazos latentes de Cuba fueron asimétricas, evidencia de que hay clusters entre países. Por lo anterior, es posible mejorar las estimaciones con el uso de distribuciones asimétricas para modelar los datos de toda la población.

Los individuos de la muestra de datos reales que respondieron igual a los ítems de una subprueba, tuvieron diferentes trazos latentes. Lo anterior puede deberse a que en la muestra de datos reales seleccionada, los trazos latentes estimados tuvieron una alta correlación entre las dos subpruebas. Los estudiantes que tuvieron mayor cantidad de respuestas correctas en una subprueba, tuvieron un trazo latente más pequeño respecto a otros estudiantes que tuvieron un puntaje clásico mayor en la otra subprueba. Es posible que este resultado se deba también a la correlación entre las dos subpruebas.

## 9. Propuestas para trabajos posteriores

El análisis de pruebas internacionales es un tema con amplias posibilidades de investigación por desarrollar. Una de ellas es la interpretación de los trazos latentes. El análisis de dimensiones de los datos reales mostró que el número de dimensiones de los trazos latentes no coincidió con el de la prueba. Cuando hay más dimensiones que subpruebas, existen dimensiones que no tienen interpretación directa, pero se puede desarrollar una interpretación a través de la teoría subyacente y se tienen en cuenta en el análisis (Montenegro 2011, pp. 41). Esta interpretación no era objetivo del modelo propuesto, pero es un tema de importancia para el diseño y análisis de las pruebas, por lo cual es un análisis por desarrollar.

Otro tema por desarrollar es la propuesta de modelos para el análisis de datos con diferencias en las medias de las distribuciones de los trazos latentes entre poblaciones o diferencias en los trazos latentes y en los parámetros de los ítems. En datos agrupados jerárquicamente, la invarianza puede presentarse de varias maneras: en los parámetros de los ítems o en las medias de los parámetros de las personas (Fox 2010, pp. 195). Esta investigación se enfocó en el análisis de datos con diferencias entre subpoblaciones, cuando estas se encuentran en los parámetros de los ítems. Una nueva investigación puede abordar el tema del análisis de datos a partir del supuesto de que las diferencias se encuentran en las medias de las distribuciones de los parámetros de las personas, o diferencias en los parámetros de los ítems y en los trazos latentes simultáneamente.

El diseño del modelo TRIMJ no tuvo como objetivo detectar el FDI ni sus causas. La información que proviene de variables examinadas en la prueba, como información sociodemográfica de los estudiantes e información de los sistemas de educación, puede ser utilizada para identificar las razones del FDI, porque permite indagar acerca de las razones del desempeño de cada país en las pruebas, de manera que es posible relacionar los resultados con las características particulares de los estudiantes y de los sistemas de educación. Estos son temas de interés para la formulación de políticas para mejorar los sistemas educativos, por lo que pueden ser el objetivo de otros proyectos de investigación.

Existen varios métodos para detectar el FDI con el uso de TRI unidimensional (Magis & De Boeck 2010), sin embargo, se mostró en este trabajo que muchas pruebas pueden ser multidimensionales, como PERCE, por lo que para detectar el FDI es necesario el desarrollo de una metodología que tenga en cuenta la multidimensionalidad. Para detectar el FDI en datos multidimensionales es necesario tener en cuenta que los datos tienen varias dimensiones, y los ítems pueden tener diferentes direcciones, por lo que el método que se desarrolle para detectar el FDI, debe tener en cuenta las dimensiones y sus direcciones.

Los resultados de Cuba fueron atípicos desde varios puntos de vista, según los resultados de los análisis desarrollados en esta investigación. Uno de los hallazgos fue que la distribución de los trazos latentes en Cuba es asimétrica, por lo que un trabajo por desarrollar es el uso de una distribución asimétrica para el modelamiento de los datos, para mejorar las estimaciones conservando las ventajas del uso de modelamiento jerárquico.

## 10. Agradecimientos

Esta investigación fue financiada por el Instituto Colombiano para la Evaluación de la Educación ICFES y parcialmente por el grupo de investigación SICS Research Group, del departamento de estadística de la Universidad Nacional de Colombia y Colciencias, en el marco del proyecto Multidimensional Item Response Theory Models for Practical Application in Large Tests Designed to Measure Multiple Constructs.

## Referencias

- Bazan, J. (2005), Uma família de modelos de teoria de resposta ao ítem normal assimétrica, Phd tesis, Instituto de Matemática e Estatística da Universidade de São Paulo.
- Bolt, D. (2003), 'Estimation of compensatory and noncompensatory multidimensional item response models using markov chain monte carlo', *Applied Psychological Measurement* **26**(6), 395–414.
- Carlin, B. (2009), *Bayesian methods for data analysis*, 3 edn, CRC Press: Taylor y Francis Group.
- Carnoy, M. (2007), *Cuba's Academic Advantage, Why Students in Cuba Do Better in School*, Stanford University Press.
- Casassus, J., F. J. P. J. & Cusato, S. (1998a), *Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado*, Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación.
- Casassus, J., F. J. P. J. & Cusato, S. (1998b), *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados en tercero y cuarto grado - primer informe de laboratorio, informe 1*, Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación, Oficina Regional de Educación Para América Latina y EL Caribe.
- Casassus, J., F. J. P. J. & Cusato, S. (2001), *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos del tercer y cuarto grado de la educación básica - informe técnico, informe Técnico*, Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación, Oficina Regional de Educación Para América Latina y EL Caribe.
- Da Silva, G. (2008), *Modelos multidimensionais da TRI com distribuições assimétricas para os tracos latentes*, Phd tesis, Instituto de Matemática e Estatística da Universidade de São Paulo.
- Dinno, A. (2012), 'Package paran', *CRAN R*.
- Duong, T. (2013), 'Package ks', *Cran R*.
- Fox, J. (2010), *Bayesian item response modeling*, Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA.

- Glas, B. (2001), 'Mcmc estimation and some model-fit analysis of multidimensional irt models', *Psychometrika* **66**(4), 541–562.
- Horn, J. (1965), 'A rationale and test for the number of factors in factor analysis', *Psychometrika* **30**(2), 179–185.
- Kromrey, D., Parshall, C. & Chason, W. (1999), 'Generating item responses based on multidimensional item response theory', *SUGI 24* (241), Posters.
- Ledesma, RD, V.-M. P. (2007), 'Determining the number of factors to retain in efa: an easy-to-use computer program for carrying out parallel analysis', *Practical Assessment, Research and Evaluation* **12**, 1–11.
- Magis, D. Béland, S. T. F. & De Boeck, P. (2010), 'A general framework and an r package for the detection of dichotomous differential item functioning', *Behavior Research Methods* **42**(3), 847–862.
- McKay, S. (2010), 'Bugs code for item response theory', *Journal of Statistical Software* **36**.
- Montenegro, A. (2011), Multidimensional Item Response Theory Models where the Ability has a Latent Linear Structure, PhD thesis, Universidad Nacional de Colombia.
- Montenegro, A. & Cepeda, E. (2010), 'Synthesizing the ability in multidimensional item response theory models', *Revista Colombiana de Estadística* **33**(1), 127–147.
- Ntzoufras, I. (2009), *Bayesian Modelling Using WINBUGS*, John Wiley and Sons, Inc.
- OCDE, O. p. l. c. y. e. d. e. (2006), 'El programa pisa de la ocde, qué es y para qué sirve', *OCDE*.
- Patz, R., Yao, L. & Hanson, B. (2002), 'More efficient markov chain monte carlo estimation in irt using marginal posteriors', *Unknown*.
- Reckase, M. (2009), *Multidimensional item response theory*, Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA.
- Ripley, B. (2013), 'Package mass', *Cran R*.
- Sheng, Y. & Wickle, C. (2008), 'Bayesian multidimensional irt models with a hierarchical structure', *Educational and Psychological Measurement* **68**(3), 413–430.
- Stout, W. (1990), 'A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation', *Psychometrika* **55**, 293–325.
- Team, R. D. C. (2008), *R: A Language and Environment for Statistical Computing*, Vol. Vienna of ISBN 3-900051-07-0, R Foundation for Statistical Computing.
- Wang, M. (1986), 'Fitting a unidimensional model to multidimensional item response data', *University of Iowa*.
- Yen, W. (1985), 'Increasing item complexity: A possible cause of scale shrinkage for unidimensional item response theory', *Psychometrika* pp. 399–410.

## Apéndice A.

### Código WinBUGS TRIMJ aplicación

```

model
{
# Specify Tree-dimensional Two-Parameter Logit Model
for (j in 1:252) {
theta1[j,1:3] ~ dnorm(mu[1:3],tau[1:3,1:3])
for (k in 1:T) {
p[j,k] <-exp(a1[k]*theta1[j,1]+b1[k]*theta1[j,2]+c1[k]*theta1[j,3]+d1[k])/
(1+exp(a1[k]*theta1[j,1]+b1[k]*theta1[j,2]+c1[k]*theta1[j,3]+d1[k]))
response[j,k] ~ dbern(p[j,k])
}
}
for (j in 253:446) {
theta2[j,1:3] ~ dnorm(mu[1:3],tau[1:3,1:3])
for (k in 1:T) {
p[j,k] <-exp(a2[k]*theta2[j,1]+b2[k]*theta2[j,2]+c2[k]*theta2[j,3]+d2[k])/

```

```

(1+exp(a2[k]*theta2[j,1]+b2[k]*theta2[j,2]+c2[k]*theta2[j,3]+d2[k]))
response[j,k] ~ dbern(p[j,k])
}
}

for (j in 447:N) {
theta3[j,1:3] ~ dnorm(mu[1:3],tau[1:3,1:3])
for (k in 1:T) {
p[j,k] <-exp(a3[k]*theta3[j,1]+b3[k]*theta3[j,2]+c3[k]*theta3[j,3]+d3[k])/
(1+exp(a3[k]*theta3[j,1]+b3[k]*theta3[j,2]+c3[k]*theta3[j,3]+d3[k]))
response[j,k] ~ dbern(p[j,k])
}
}

mu[1]<-0
mu[2]<-0
mu[3]<-0

# Specify Priors for Discrimination and Difficulty Parameters
a1[15]<-1; b1[15]<-0; c1[15]<-0;
a1[51]<-0; b1[51]<-1; c1[51]<-0;
a1[1]<-0; b1[1]<-0; c1[1]<-1;
a2[15]<-1; b2[15]<-0; c2[15]<-0;
a2[51]<-0; b2[51]<-1; c2[51]<-0;
a2[1]<-0; b2[1]<-0; c2[1]<-1;
a3[15]<-1; b3[15]<-0; c3[15]<-0;
a3[51]<-0; b3[51]<-1; c3[51]<-0;
a3[1]<-0; b3[1]<-0; c3[1]<-1;
d1[1]<-0; d2[1]<-0; d3[1]<-0;

#sigma[1:2,1:2]<-inverse(tau[1:2,1:2])
for (k in 2:14) {
a1[k] ~ dnorm(a[k],tau1[k])I(0,);
b1[k] ~ dnorm(b[k],tau1[k])I(0,);
c1[k] ~ dnorm(c[k],tau1[k])I(0,);
a2[k] ~ dnorm(a[k],tau1[k])I(0,);
b2[k] ~ dnorm(b[k],tau1[k])I(0,);
c2[k] ~ dnorm(c[k],tau1[k])I(0,);
a3[k] ~ dnorm(a[k],tau1[k])I(0,);
b3[k] ~ dnorm(b[k],tau1[k])I(0,);
c3[k] ~ dnorm(c[k],tau1[k])I(0,);
}
for (k in 16:50) {
a1[k] ~ dnorm(a[k],tau1[k])I(0,);
b1[k] ~ dnorm(b[k],tau1[k])I(0,);
c1[k] ~ dnorm(c[k],tau1[k])I(0,);
a2[k] ~ dnorm(a[k],tau1[k])I(0,);
b2[k] ~ dnorm(b[k],tau1[k])I(0,);
c2[k] ~ dnorm(c[k],tau1[k])I(0,);
a3[k] ~ dnorm(a[k],tau1[k])I(0,);
b3[k] ~ dnorm(b[k],tau1[k])I(0,);
c3[k] ~ dnorm(c[k],tau1[k])I(0,);
}
}

```



```

for (k in 2:T){
d1[k] ~ dnorm(d[k],tau2[k]);
d2[k] ~ dnorm(d[k],tau2[k]);
d3[k] ~ dnorm(d[k],tau2[k]);
}

for (k in 2:14) {
a[k] ~ dnorm(0,1)I(0,);
b[k] ~ dnorm(0,1)I(0,);
c[k] ~ dnorm(1,1)I(0,);
tau1[k]~dgamma(0.1,0.001)
}

for (k in 16:50) {
a[k] ~ dnorm(1,1)I(0,);
b[k] ~ dnorm(0,1)I(0,);
c[k] ~ dnorm(0,1)I(0,);
tau1[k]~dgamma(0.1,0.001)
}
for (k in 2:T){
d[k] ~ dnorm(0,.001);
tau2[k]~dgamma(0.1,0.001)
}

}

```

## Apéndice B.

### Código para generar datos simulados en R

```

#utilities
#####
# norm of a matrix of vectors.
# x is vector or a matrix, the vectors must be rows of x
# Sigma is the symetric matrix positive definite that define the product.
# Simmetry and pd are not tested
# if Sigma is missing, the common escalar product is used (Sigma =Identity)
#####
norm<-function(x,Sigma){
#control section
if(!is.numeric(x))
  stop("'x' must be numeric")
if(!missing(Sigma)){
  if(!is.matrix(Sigma))
    stop("'Sigma' must be a matrix")
  if(!is.numeric(Sigma))
    stop("'Sigma' must be numeric")
  if(any(is.na(Sigma)))
    stop("'Sigma' contain NA values")
  if(!identical((dim(Sigma)[1]), (dim(Sigma)[2])))

```

```

    stop("'Sigma' must be a squared matrix")
  if(is.matrix(x) & !identical(dim(Sigma)[1], dim(x)[2]))
    stop(" 'x' and 'Sigma' can not be operate")
  if(!is.matrix(x) & !identical(dim(Sigma)[1],length(x)))
    stop(" 'x' and 'Sigma' can not be operate")
  # warning. simmetry test
  if((sum(abs(Sigma-t(Sigma)))!=0))
    warning("'Sigma' is not symmetric")

} #end if !missing(Sigma)
#work section
if(!is.matrix(x)){
  if (missing(Sigma)) return(sqrt(sum(x*x)))
  else return(sqrt(matrix(x,1,length(x))%%Sigma %% matrix(x,length(x),1)))
}

if(missing(Sigma)) return(sqrt(apply(x*x,1,sum)))
else return(sqrt(apply((x%%Sigma)*x,1,sum)))
} # end norm

#normalize funtion

normalize<-function(x,Sigma){
  #control section
  if(!is.numeric(x))
    stop("'x' must be numeric")
  #work section
  if(!is.matrix(x))
    return(x/norm(x,Sigma))
  else
    return(x/matrix(norm(x,Sigma),dim(x)[1],dim(x)[2]))
} # end normalize

#####
# functions to generate the item directions, ref directions and sigma coeff
# to simulate lsmirt models
#####
get.itemDirections<-function(beta_v,size,noise,seed){
  #control section
  if(!is.numeric(beta_v))
    stop("'beta_v' must be numeric")
  if(is.matrix(beta_v) & !identical(dim(beta_v)[1],length(noise)))
    stop("'beta_v' and 'noise have non-compatible lengths")
  if(!is.matrix(beta_v) & !identical(length(beta_v),length(noise)))
    stop("'beta_v' and 'noise have non-compatible lengths")
  if(is.matrix(beta_v) & !identical(dim(beta_v)[1],length(size)))
    stop("'beta_v' and 'size have non-compatible lengths")
  if(!is.matrix(beta_v) & !identical(length(beta_v),length(size)))
    stop("'beta_v' and 'size have non-compatible lengths")

  #work section
  if(missing(beta_v)){

```

```

    beta1 <- normalize(c(1,1,1))
    beta2 <- normalize(c(1,0.2,0.1))
    beta3 <- normalize(c(.3,1,0.1))
    beta4 <- normalize(c(.25,.25,1))
    beta_v <- rbind(beta1,beta2,beta3,beta4)
  }
else{
  if(is.matrix(beta_v)){
    m <- dim(beta_v)[1]
    d <- dim(beta_v)[2]
  }
  else{
    m <- 1
    d <- length(beta_v)
  }
}
if(missing(noise))
  noise<-seq(0.1,by=.05, length.out=m)
if(!missing(seed))
  set.seed(seed)
if(missing(size))
  size <- rep(15,m)

# now generate the directions
mv <- matrix(0,sum(size),d)
#index
idx_last <- 0
for(i in 1:m){
  idx_first <-idx_last+1
  idx_last <- sum(size[1:i])
  idx <- idx_first:idx_last
  mv[idx,] <- matrix(beta_v[i,],idx_last-idx_first+1,d,byrow=TRUE)+
    matrix(runif((idx_last-idx_first+1)*d,-noise[i],noise[i]),idx_last-idx_first+1,d)
}
#change negative values by size
mv<-ifelse(mv<0,0,mv)
#normalize the vectors
mv<-normalize(mv)
mv
} #end get.ItemDirections

#####
# end utilities
#####

#####
#
# simulation section: generate the data
#####
# generate item directions, ref directions and s_j and sigma_j coefficient_j
# test, reproducible
# these are the initial reference directions

```

```

# 3 dimensions
#####

#reference directions of each cluster of items
beta1 <- normalize(c(1,0.2,0.1,0.1))
beta2 <- normalize(c(.3,1,0.1,0.2))
beta3 <- normalize(c(.25,0.25,1,.10))
beta4 <- normalize(c(.25,0.25,.25,1))

beta_w <- rbind(beta1,beta2,beta3,beta4)

K1 <- 10
K2 <- 10
K3 <- 10
K4 <- 10
K <- K1+K2+K3+K4

# size of each initial subtest
size <- c(K1,K2,K3,K4)

# alpha params
# seed for reproductibility
seed_alpha <- 400L
set.seed(seed_alpha)
sd.alpha <- 1/4
Alpha <- rlnorm(K,0,sd.alpha)
# seed for reproductibility
seed_gamma <- 250L
set.seed(seed_gamma)
sd.gamma <- 1
#difficulty
b <- rnorm(K,0,sd.gamma)

noise <- seq(0.1,by=.05, length.out=4)

# obtain the directions    beta_vj's
# direction of the items
#seed for reproductibility
seed.beta <- 100L
# generate the item directions
beta_vj <- get.itemDirections(beta_w,size,noise,seed=seed.beta)

# matrix of a's (slopes)
a <- Alpha*beta_vj

# latent traits
#size of the sample
N <- 1000

dim.trait <- 4

```

```

seed.5 <-500L
set.seed(seed.5)

# Generate the normal standard
# it is not necessary the inclusion of the covariance matrix, because
# the direction of the items are not orthogonal, so the correlations are included
Theta <- matrix(rnorm(N*dim.trait,0,1),N,dim.trait)
Theta <- Theta %*% solve((chol(cov(Theta))))

# predictor

eta <- Theta%*%t(a) - matrix(b,N,K,byrow=TRUE)

# probaility of response, using the normal link
pr <- pnorm(eta)
# generate the binary test
U <-matrix(runif(N*K),N,K)
test<- ifelse(pr>U,1,0)

```

## Apéndice C.

### Código para estimar los criterios de evaluación y de bondad de ajuste del modelo TRIMJ

```

# El siguiente código fue tomado de Fox pág. 111
for ( j in 1 :N){
for ( k in 1 :T){
residn[ j , k ] <- 0.3989*exp(-.5*(pow(-p [ j , k ],2))) # es un outlier si está entre
+- 1.5 según Fox y entre +-2 según Montenegro. #Residual Bayesiano
residual[ j , k ] <- (residn[ j , k ] /phi(p[ j , k ])) *response[ j , k ] +
(-residn[ j , k ] /phi(-p[ j , k ]))* (1-response[j,k]) # residual debe ser
<0 para y=0, >0 para y=1. Residual Bayesiano latente
pval[j,k]<-step(residual[j,k]) # P-valor según Carlin y Ioannis
}
}

# To compute a CPO pág. 169 Taehoon Kang (Codigo winbugs TRIM)
#for (j in 1:N){
#for (k in 1:T){
#prob[j,k]<-pow(p[j,k],response[j,k])*pow((1-p[j,k]),(1-response[j,k])) # CPO
#inprob[j,k]<-pow(prob[j,k],-1) # inverso CPO
#}}

# Log Likelihood. Tomado de la pág. 154 y 155 de cho (code winbugs UIRT)
#for (j in 1:N) {
#for (k in 1:T) {
#l[j,k]<-log(p[j,k])*response[j,k]+log(1-p[j,k])*(1-response[j,k])
#}}
#loglik <-sum(l[1:N,1:T])

```

# Apéndice D.

## Resultados de algunas simulaciones

TABLA 9: Resultados simulación número 1

Parámetro	Correlación	Varianza	Sesgo	Error cuadrático
$\theta_1$	0.91	0.8335093	0.0393308	0.8350562
$\theta_2$	0.66	0.8457419	0.0498217	0.8482241
Subpoblación 1	Correlación	Varianza	Sesgo	Error cuadrático
$a_1$	0.97	0.7128598	-0.1002940	0.7229187
$a_2$	0.93	0.7128598	0.1189425	0.7270071
$d$	0.7	0.3770546	-0.8070060	1.0283130
Subpoblación 2	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0.97	0.6565572	-0.3034235	0.7486230
$a_2$	0.6	0.6565572	0.6229015	1.0445630
$d$	0.89	0.3770546	0.3929940	0.5314989

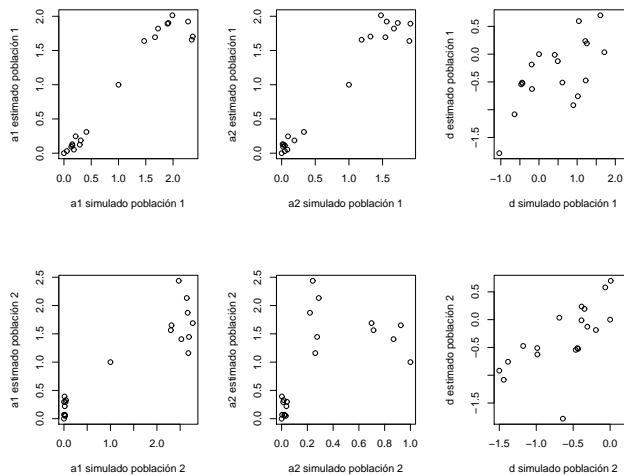


FIGURA 14: Parámetros de los items simulados vs. recuperados, simulación 1.

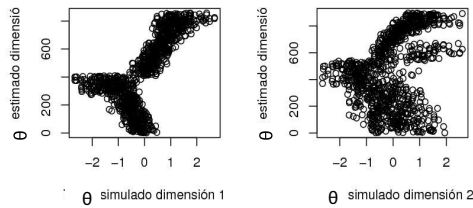


FIGURA 15:  $\theta$  simulados vs. estimados, simulación 1.

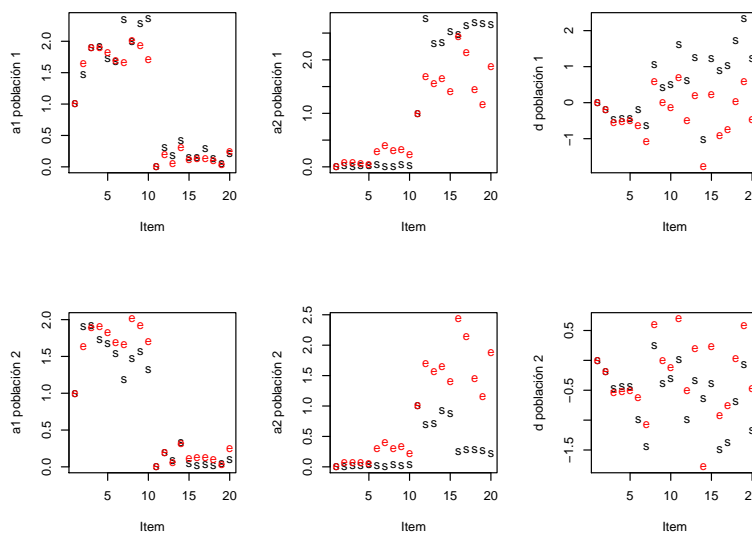


FIGURA 16: Parámetros de los items simulados vs. recuperados, simulación 1. s: parámetros simulados, e: parámetros recuperados. Primera fila, parámetros de la primera subpoblación; segunda fila, parámetros de la segunda subpoblación.

## Apéndice E. Resultados simulaciones con recuperación de parámetros con el modelo TRIMJ

TABLA 10: Resultados recuperación de parámetros simulación 4

Subpoblación 1	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,969	0,105	0,284	0,143
$a_2$	0,964	0,097	0,188	0,129
$d$	0,990	0,088	0,621	0,110
$\theta_1$	0,804	0,476	0,674	0,597
$\theta_2$	0,801	0,479	0,629	0,598
Subpoblación 2	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,983	0,079	0,244	0,102
$a_2$	0,963	0,099	0,201	0,129
$d$	0,984	0,078	0,348	0,104
$\theta_1$	0,817	0,464	0,662	0,579
$\theta_2$	0,793	0,478	0,619	0,608

TABLA 11: Resultados recuperación de parámetros simulación 13

Subpoblación 1	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,952	0,112	0,207	0,145
$a_2$	0,966	0,110	0,324	0,172
$a_3$	0,946	0,150	0,180	0,190
$d$	0,988	0,109	0,682	0,140
$\theta_1$	0,789	0,490	0,633	0,614
$\theta_2$	0,804	0,474	0,651	0,603
$\theta_3$	0,782	0,484	0,582	0,623
Subpoblación 2	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,919	0,141	0,183	0,188
$a_2$	0,931	0,096	0,131	0,138
$a_3$	0,938	0,096	0,237	0,227
$d$	0,994	0,097	1,119	0,132
$\theta_1$	0,773	0,511	0,607	0,647
$\theta_2$	0,739	0,536	0,522	0,674
$\theta_3$	0,771	0,505	0,581	0,646

TABLA 12: Resultados recuperación de parámetros simulación 16

Subpoblación 1	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,937	0,100	0,143	0,136
$a_2$	0,936	0,134	0,269	0,179
$d$	0,989	0,106	0,878	0,142
$\theta_1$	0,755	0,519	0,588	0,656
$\theta_2$	0,802	0,479	0,635	0,599
Subpoblación 2	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,947	0,117	0,228	0,170
$a_2$	0,918	0,128	0,164	0,168
$d$	0,995	0,082	0,747	0,100
$\theta_1$	0,768	0,517	0,586	0,641
$\theta_2$	0,767	0,505	0,573	0,643
Subpoblación 3	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,943	0,108	0,169	0,146
$a_2$	0,947	0,140	0,289	0,173
$d$	0,994	0,097	0,803	0,109
$\theta_1$	0,768	0,512	0,579	0,642
$\theta_2$	0,813	0,464	0,639	0,584

TABLA 13: Resultados recuperación de parámetros simulación 17

Subpoblación 1	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,953	0,112	0,207	0,144
$a_2$	0,967	0,111	0,322	0,171
$a_3$	0,946	0,146	0,178	0,187
$d$	0,988	0,107	0,681	0,137
$\theta_1$	0,789	0,490	0,269	0,615
$\theta_2$	0,804	0,474	0,301	0,603
$\theta_3$	0,782	0,484	0,355	0,623
Subpoblación 2	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,919	0,137	0,181	0,184
$a_2$	0,931	0,095	0,132	0,136
$a_3$	0,938	0,095	0,238	0,226
$d$	0,994	0,095	1,120	0,130
$\theta_1$	0,773	0,502	0,337	0,635
$\theta_2$	0,739	0,536	0,451	0,674
$\theta_3$	0,773	0,504	0,394	0,644
Subpoblación 3	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,973	0,072	0,169	0,097
$a_2$	0,946	0,110	0,138	0,162
$a_3$	0,943	0,139	0,233	0,182
$d$	0,993	0,083	1,128	0,099
$\theta_1$	0,750	0,536	0,635	0,661
$\theta_2$	0,781	0,504	0,674	0,624
$\theta_3$	0,757	0,526	0,644	0,654

## Apéndice F.

### Resultados datos simulados y estimación con el modelo TRIM con parámetros globales



TABLA 14: Resultados recuperación de parámetros simulación 20

Media global	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,985	0,071	0,164	0,096
$a_2$	0,975	0,077	0,177	0,099
$d$	0,981	0,064	0,197	0,086
Subpoblación 1	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,959	0,083	0,179	0,119
$a_2$	0,960	0,104	0,223	0,130
$d$	0,983	0,146	0,461	0,182
$\theta_1$	0,793	0,480	0,638	0,137
$\theta_2$	0,82	0,473	0,695	0,117
Subpoblación 2	Correlación	Media de las diferencias	Varianza	Error cuadrático
$a_1$	0,960	0,112	0,187	0,151
$a_2$	0,957	0,102	0,171	0,134
$d$	0,995	0,082	0,709	0,100
$\theta_1$	0,765	0,517	0,578	0,173
$\theta_2$	0,763	0,508	0,549	0,175

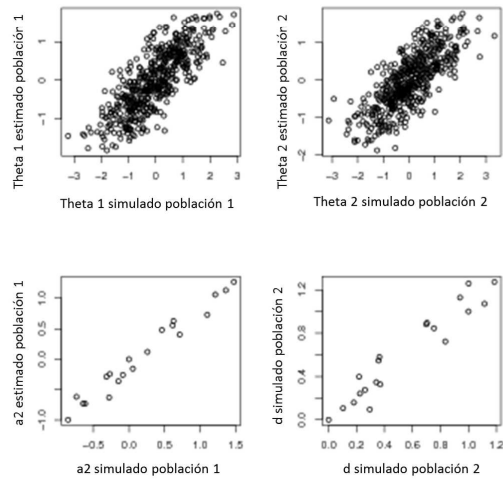


FIGURA 17: Resultados simulación 20

Identificación de los gráficos en cada fila de izquierda a derecha: Primera fila:  $\theta$  dimensión 1, población 1,  $\theta$  dimensión 2, población 2 Segunda fila:  $a_2$  subpoblación 1,  $d$  subpoblación 2.

## Apéndice G.

### Parámetros estimados con el modelo TRIMJ con la muestra de datos reales

TABLA 15: Parámetros  $a_1$  con TRIMJ datos reales

Matemática	Internacional	Bolivia	Colombia	Cuba	Lenguaje	Internacional	Bolivia	Colombia	Cuba
Item 2	0.6821	0.662	0.7146	0.6966	Item 33	0.2008	0.1849	0.1987	0.2065
Item 3	0.6675	0.6761	0.6365	0.698	Item 34	0.3852	0.2788	0.3254	0.5568
Item 4	0.5452	0.5424	0.5257	0.5808	Item 35	0.3023	0.2492	0.3234	0.3281
Item 5	0.3956	0.4191	0.3327	0.4302	Item 36	0.5712	0.761	0.575	0.4039
Item 6	0.2052	0.2071	0.2104	0.1936	Item 37	0.4096	0.2916	0.5443	0.4007
Item 7	0.3832	0.3762	0.3661	0.4105	Item 38	0.3473	0.3095	0.3789	0.3191
Item 8	0.2488	0.2296	0.265	0.2488	Item 39	0.5443	0.5827	0.57	0.4915
Item 9	0.6469	0.7368	0.5809	0.6416	Item 40	0.3231	0.3524	0.2934	0.3215
Item 10	0.4084	0.3831	0.4309	0.4145	Item 41	0.4469	0.4932	0.4788	0.3719
Item 11	0.4991	0.469	0.5015	0.5341	Item 42	0.4565	0.5998	0.3711	0.4077
Item 12	0.616	0.6414	0.589	0.6268	Item 43	0.1639	0.1275	0.1571	0.1864
Item 13	0.6821	0.6464	0.6547	0.7668	Item 44	0.2207	0.1764	0.2653	0.2014
Item 14	0.4908	0.5324	0.4166	0.5306	Item 45	0.2091	0.1739	0.1892	0.2467
Item 16	1.598	1.877	1.394	1.988	Item 46	0.3692	0.3702	0.3395	0.3972
Item 17	1.162	1.165	1.131	1.209	Item 47	0.4851	0.848	0.2613	0.3695
Item 18	1.554	1.355	1.759	1.732	Item 48	0.4716	0.5738	0.4181	0.4312
Item 19	0.8546	0.7729	0.8317	0.9902	Item 49	0.5201	0.5159	0.5291	0.52
Item 20	1.56	1.559	1.565	1.593	Item 50	0.2436	0.2595	0.2306	0.2179
Item 21	1.364	1.319	1.408	1.391	Promedio	0.37	0.397	0.358	0.354
Item 22	0.7041	0.7114	0.6984	0.7175					
Item 23	0.8726	0.8683	0.8817	0.8795					
Item 24	0.8658	0.8932	0.8042	0.9267					
Item 25	0.5714	0.5354	0.5794	0.6062					
Item 26	0.9184	0.8912	0.9156	0.9852					
Item 27	0.936	0.9861	0.8858	0.9569					
Item 28	0.6977	0.1867	0.7397	1.316					
Item 29	0.6801	0.6306	0.7679	0.6623					
Item 30	1.143	1.131	1.166	1.165					
Item 31	0.7764	0.7363	0.767	0.8415					
Item 32	0.8684	0.8449	0.7964	0.9987					
Promedio	0.78	0.766	0.77	0.857					

TABLA 16: Parámetros  $a_2$  con TRIMJ datos reales

Matemática	Internacional	Bolivia	Colombia	Cuba	Lenguaje	Internacional	Bolivia	Colombia	Cuba
Item 2	0.6989	0.5689	0.779	0.7827	Item 33	1.413	1.425	1.427	1.39
Item 3	0.3948	0.4075	0.3959	0.3848	Item 34	0.8406	0.6404	0.8797	0.9884
Item 4	0.9486	0.9632	0.982	0.922	Item 35	0.8814	0.8737	0.8192	0.9503
Item 5	0.4043	0.5458	0.2698	0.4001	Item 36	1.623	1.826	1.394	1.709
Item 6	0.1313	0.1062	0.1267	0.1486	Item 37	1.097	1.081	1.059	1.166
Item 7	0.1202	0.09415	0.09606	0.1445	Item 38	0.9056	0.6907	0.9562	1.063
Item 8	0.4787	0.4801	0.4754	0.4854	Item 39	1.73	1.646	1.821	1.752
Item 9	0.2652	0.1886	0.3248	0.2723	Item 40	1.437	1.431	1.456	1.434
Item 10	0.3741	0.3563	0.3765	0.3954	Item 41	1.238	1.283	1.19	1.257
Item 11	0.3476	0.3235	0.3311	0.3853	Item 42	1.521	1.381	1.692	1.531
Item 12	0.4903	0.5218	0.4396	0.5153	Item 43	0.5368	0.5608	0.4738	0.5702
Item 13	0.4358	0.4399	0.5137	0.3622	Item 44	0.8292	0.9215	0.7987	0.7615
Item 14	0.1743	0.1191	0.126	0.2622	Item 45	1.016	1.056	0.9566	1.032
Item 16	0.5952	0.4973	0.7193	0.3832	Item 46	0.8187	0.8625	0.6891	0.9037
Item 17	0.2362	0.224	0.2282	0.2396	Item 47	1.258	1.215	1.261	1.328
Item 18	0.5841	0.5553	0.6542	0.4868	Item 48	0.9637	1.018	0.8627	1.008
Item 19	0.2041	0.1621	0.1787	0.2115	Item 49	1.662	1.702	1.662	1.636
Item 20	0.484	0.4799	0.5023	0.4565	Item 50	0.7616	0.7413	0.6442	0.8964
Item 21	0.2911	0.2692	0.2971	0.2836	Promedio	1.14	1.13	1.11	1.18
Item 22	0.3965	0.482	0.3414	0.35					
Item 23	0.7134	0.7341	0.6854	0.719					
Item 24	1.003	1.006	1.02	0.9866					
Item 25	0.8424	0.8191	0.8496	0.8547					
Item 26	0.5382	0.5989	0.4006	0.5982					
Item 27	0.5951	0.59	0.5531	0.6317					
Item 28	0.4492	0.1759	0.1994	0.6311					
Item 29	0.606	0.5342	0.6423	0.6246					
Item 30	0.6193	0.6033	0.5676	0.6747					
Item 31	0.3537	0.3827	0.3234	0.3346					
Item 32	0.2896	0.3085	0.2293	0.2823					
Promedio	0.46	0.451	0.454	0.47					

TABLA 17: Parámetros  $a_3$  con TRIMJ datos reales

Matemática	Internacional	Bolivia	Colombia	Cuba	Lenguaje	Internacional	Bolivia	Colombia	Cuba
Item 2	1.188	1.201	1.18	1.987	Item 33	0.8745	0.9011	0.8524	0.8836
Item 3	1.023	1.024	1.023	2.24	Item 34	0.3347	0.3031	0.288	0.3966
Item 4	1.843	1.824	1.886	1.335	Item 35	0.3436	0.3798	0.2817	0.3681
Item 5	0.5492	0.6496	0.4191	2.781	Item 36	0.8373	0.8534	1.012	0.7314
Item 6	0.557	0.5289	0.5821	0.58	Item 37	0.4333	0.4868	0.3443	0.4775
Item 7	0.6211	0.6726	0.564	7.861	Item 38	0.7253	0.6385	0.6144	0.9971
Item 8	1.09	1.085	1.098	3.823	Item 39	0.5764	0.5864	0.5473	0.6089
Item 9	0.8657	0.836	0.8942	1.32	Item 40	0.6299	0.5978	0.6525	0.6509
Item 10	1.732	1.767	1.714	2.14	Item 41	0.5924	0.4439	0.6724	0.6817
Item 11	1.7	1.707	1.707	1.185	Item 42	0.8036	0.7201	0.9135	0.8321
Item 12	0.4648	0.4299	0.488	1.767	Item 43	0.5625	0.5159	0.5868	0.596
Item 13	0.9563	0.9805	0.9326	1.986	Item 44	0.7204	0.784	0.6542	0.7401
Item 14	0.2465	0.2594	0.1959	1.679	Item 45	0.6672	0.6313	0.6657	0.7142
Item 16	0.7012	0.5319	0.3797	1.335	Item 46	0.2695	0.2399	0.2479	0.3087
Item 17	0.5046	0.522	0.4642	0.5333	Item 47	0.5654	0.5879	0.5823	0.559
Item 18	0.8989	0.7448	1.125	0.937	Item 48	0.7666	0.7954	0.7636	0.7683
Item 19	0.5455	0.485	0.5385	0.6255	Item 49	0.9769	0.9598	1.03	0.9607
Item 20	0.6836	0.6748	0.6542	0.7343	Item 50	0.4016	0.3525	0.4312	0.4244
Item 21	0.9498	0.9304	0.9477	0.9904	Promedio	0.615	0.598	0.618	0.649
Item 22	0.5158	0.5356	0.4664	0.5558					
Item 23	0.5867	0.5698	0.642	0.5559					
Item 24	0.8733	0.9203	0.8462	0.8751					
Item 25	0.5501	0.538	0.5433	0.5763					
Item 26	0.2339	0.1674	0.1912	0.3227					
Item 27	0.5778	0.5886	0.5414	0.6124					
Item 28	0.6103	0.3721	0.267	1.268					
Item 29	0.6897	0.7189	0.6302	0.743					
Item 30	0.6922	0.6798	0.6334	0.7834					
Item 31	0.2173	0.1534	0.2123	0.2764					
Item 32	0.3835	0.389	0.2639	0.5074					
Promedio	0.768	0.749	0.734	1.43					

TABLA 18: Parámetros  $d$  con TRIMJ datos reales

Matemática	Internacional	Bolivia	Colombia	Cuba	Lenguaje	Internacional	Bolivia	Colombia	Cuba
Item 2	1.519	1.346	0.9497	2.277	Item 33	2.524	2.25	2.17	3.173
Item 3	-0.04247	-0.05357	-0.08052	0.008656	Item 34	0.4201	-0.4673	0.2041	1.444
Item 4	3.215	3.219	3.229	3.191	Item 35	0.5615	-0.166	0.1503	1.687
Item 5	0.215	0.2379	0.03801	0.3752	Item 36	2.303	2.217	2.325	2.37
Item 6	-0.7502	-0.6804	-1.14	-0.4448	Item 37	0.4872	-0.2117	0.1858	1.461
Item 7	-1.696	-1.628	-1.926	-1.546	Item 38	-0.005581	-0.3691	-0.2511	0.553
Item 8	1.827	1.842	1.822	1.815	Item 39	1.73	1.417	1.471	2.262
Item 9	-0.2279	-0.3119	-0.4532	0.07191	Item 40	1.972	1.595	0.9542	3.39
Item 10	1.951	1.776	1.774	2.315	Item 41	0.2882	0.08217	0.18	0.5968
Item 11	0.8501	0.8201	0.907	0.807	Item 42	0.816	0.9172	0.1567	1.362
Item 12	0.3371	0.1505	-0.06416	0.9068	Item 43	0.4896	-0.2501	0.5802	1.185
Item 13	0.1408	0.1653	0.1474	0.1103	Item 44	-0.1816	-0.1806	-1.064	0.7132
Item 14	-0.4285	-0.482	-0.4319	-0.3721	Item 45	-1.119	-1.429	-1.089	-0.8188
Item 15	0.8515	0.4093	0.6871	1.45	Item 46	1.224	0.8981	0.9143	1.841
Item 16	-1.009	-1.057	-1.383	-0.6096	Item 47	1.802	1.697	1.774	1.92
Item 17	0.1125	0.2514	-0.2594	0.3585	Item 48	2.461	2.239	2.616	2.52
Item 18	0.3291	0.4203	0.2843	0.2884	Item 49	1.699	1.587	1.677	1.852
Item 19	-0.9685	-1.187	-1.396	-0.3121	Item 50	0.407	0.1204	0.01751	1.101
Item 20	-0.4806	-0.7099	-0.8056	0.07261	Item 51	0.901	0.4669	-0.4301	2.69
Item 21	-0.7679	-1.208	-0.8312	-0.2812	Promedio	0.988	0.65	0.66	1.64
Item 22	0.1772	0.2243	0.1317	0.1758					
Item 23	0.3776	0.3769	0.4209	0.3489					
Item 24	1.549	1.549	1.387	1.708					
Item 25	0.6218	0.3552	0.5864	0.918					
Item 26	1.03	1.031	0.925	1.129					
Item 27	0.864	0.9259	0.676	1.006					
Item 28	-1.081	-1.135	-0.9768	-1.124					
Item 29	0.002807	-0.4356	-0.3481	0.7385					
Item 30	0.7571	0.4314	0.7129	1.126					
Item 31	-0.2639	-0.7646	-0.6799	0.684					
Item 32	-0.2928	-0.4552	-0.322	-0.1063					
Promedio	0.28	0.17	0.11	0.55					

TABLA 19: Discriminación multidimensional (MDISC) internacional y por país con TRIMJ

Matemática	Bolivia	Colombia	Cuba	Internacional	Lenguaje	Bolivia	Colombia	Cuba	Internacional
Item 2	1.48	1.58	2.25	1.54	Item 33	1.70	1.67	1.66	1.67
Item 3	1.29	1.27	2.38	1.28	Item 34	0.76	0.98	1.20	0.98
Item 4	2.13	2.19	1.72	2.14	Item 35	0.98	0.92	1.07	0.99
Item 5	0.95	0.60	2.84	0.79	Item 36	2.15	1.82	1.90	1.91
Item 6	0.58	0.63	0.63	0.61	Item 37	1.22	1.24	1.32	1.25
Item 7	0.78	0.68	7.87	0.74	Item 38	0.99	1.20	1.49	1.21
Item 8	1.21	1.23	3.86	1.22	Item 39	1.84	1.99	1.92	1.90
Item 9	1.13	1.11	1.49	1.11	Item 40	1.59	1.62	1.61	1.60
Item 10	1.84	1.81	2.22	1.82	Item 41	1.44	1.45	1.48	1.44
Item 11	1.80	1.81	1.36	1.81	Item 42	1.67	1.96	1.79	1.78
Item 12	0.93	0.88	1.94	0.91	Item 43	0.77	0.77	0.85	0.79
Item 13	1.25	1.25	2.16	1.25	Item 44	1.22	1.07	1.08	1.12
Item 14	0.60	0.48	1.78	0.58	Item 45	1.24	1.18	1.28	1.23
Item 16	2.01	1.61	2.43	1.84	Item 46	0.97	0.81	1.03	0.94
Item 17	1.30	1.24	1.34	1.29	Item 47	1.59	1.41	1.49	1.46
Item 18	1.64	2.19	2.03	1.89	Item 48	1.41	1.23	1.34	1.32
Item 19	0.93	1.01	1.19	1.03	Item 49	2.02	2.03	1.97	2.00
Item 20	1.77	1.77	1.81	1.77	Item 50	0.86	0.81	1.02	0.89
Item 21	1.64	1.72	1.73	1.69					
Item 22	1.01	0.91	0.97	0.96					
Item 23	1.27	1.29	1.26	1.27					
Item 24	1.63	1.55	1.61	1.59					
Item 25	1.12	1.16	1.20	1.16					
Item 26	1.09	1.02	1.20	1.09					
Item 27	1.29	1.18	1.30	1.25					
Item 28	0.45	0.81	1.93	1.03					
Item 29	1.10	1.18	1.18	1.14					
Item 30	1.45	1.44	1.56	1.47					
Item 31	0.84	0.86	0.95	0.88					
Item 32	0.98	0.87	1.16	0.99					

TABLA 20: Dificultad multidimensional por país (MDIFF) con TRIMJ

Matemática	Bolivia	Colombia	Cuba	Internacional	Lenguaje	Bolivia	Colombia	Cuba	Internacional
Item 2	-0.91	-0.60	-1.01	-0.99	Item 33	-1.33	-1.30	-1.91	-1.51
Item 3	0.04	0.06	0.003	0.03	Item 34	0.61	-0.21	-1.20	-0.43
Item 4	-1.51	-1.47	-1.85	-1.50	Item 35	0.17	-0.16	-1.58	-0.57
Item 5	-0.25	-0.06	-0.13	-0.27	Item 36	-1.03	-1.28	-1.25	-1.20
Item 6	1.18	1.80	0.71	1.23	Item 37	0.17	-0.15	-1.11	-0.39
Item 7	2.10	2.84	0.20	2.29	Item 38	0.37	0.21	-0.37	0.00
Item 8	-1.52	-1.49	-0.47	-1.50	Item 39	-0.77	-0.74	-1.18	-0.91
Item 9	0.28	0.41	-0.05	0.20	Item 40	-1.00	-0.59	-2.11	-1.23
Item 10	-0.96	-0.98	-1.04	-1.07	Item 41	-0.06	-0.12	-0.40	-0.20
Item 11	-0.46	-0.50	-0.60	-0.47	Item 42	-0.55	-0.08	-0.76	-0.46
Item 12	-0.16	0.07	-0.47	-0.37	Item 43	0.32	-0.75	-1.40	-0.62
Item 13	-0.13	-0.12	-0.05	-0.11	Item 44	0.15	1.00	-0.66	0.16
Item 14	0.80	0.90	0.21	0.74	Item 45	1.15	0.92	0.64	0.91
Item 16	0.53	0.86	0.25	0.55	Item 46	-0.93	-1.13	-1.78	-1.31
Item 17	-0.19	0.21	-0.27	-0.09	Item 47	-1.06	-1.26	-1.29	-1.23
Item 18	-0.26	-0.13	-0.14	-0.17	Item 48	-1.58	-2.13	-1.88	-1.87
Item 19	1.28	1.39	0.26	0.94	Item 49	-0.79	-0.83	-0.94	-0.85
Item 20	0.40	0.46	-0.04	0.27	Item 50	-0.14	-0.02	-1.08	-0.45
Item 21	0.74	0.48	0.16	0.46					
Item 22	-0.22	-0.15	-0.18	-0.18					
Item 23	-0.30	-0.33	-0.28	-0.30					
Item 24	-0.95	-0.89	-1.06	-0.98					
Item 25	-0.32	-0.50	-0.77	-0.54					
Item 26	-0.95	-0.91	-0.94	-0.95					
Item 27	-0.72	-0.57	-0.77	-0.69					
Item 28	2.51	1.20	0.58	1.05					
Item 29	0.40	0.29	-0.63	0.00					
Item 30	-0.30	-0.49	-0.72	-0.51					
Item 31	0.91	0.79	-0.72	0.30					
Item 32	0.46	0.37	0.09	0.30					

TABLA 21: Parámetros  $\theta$  con TRIMJ datos reales Bolivia

Persona	$\theta_1$	$\theta_2$	$\theta_3$	Persona	$\theta_1$	$\theta_2$	$\theta_3$
2	0.868	0.5493	1.165	128	0.8825	-0.9005	0.5198
4	-0.2343	0.04956	-1.265	130	-0.436	0.9986	0.7094
6	-1.628	-0.009586	0.1618	132	-0.5388	0.2131	0.4075
8	-0.9566	-0.2168	-0.8266	134	0.7341	-0.3106	0.2632
10	-0.4564	-0.3655	-0.2493	136	0.2146	0.2288	-0.592
12	-1.7	-0.3719	-0.04418	138	0.362	-0.1458	1.17
14	-0.4397	-0.6742	-0.7512	140	0.07687	0.9295	0.844
16	-0.6875	-0.5431	-0.5206	142	0.3674	-1.209	-0.9429
18	-0.1725	0.6866	-0.8283	144	-0.8395	-0.763	-0.5028
20	-0.2332	1.871	0.07545	146	-0.3841	-0.4024	0.2595
22	-0.3298	1.744	0.8585	148	-1.6	1.096	-0.5254
24	-0.6133	-0.435	-0.3191	150	0.2273	-0.7186	-0.4366
26	-0.2469	-0.2655	0.3737	152	-0.8535	-0.1149	0.5086
28	-0.2905	-0.7911	0.1638	154	-1.026	0.515	-0.1793
30	-0.4942	1.116	-0.28	156	0.1382	0.1673	0.3748
32	-0.5014	-0.8096	-0.2935	158	-1.089	-0.7255	-0.08924
34	-0.6717	-1.213	0.1167	160	-0.4836	-1.357	0.05991
36	1.03	0.5704	0.662	162	0.2428	-1.484	-0.2833
38	1.14	0.2669	-0.08119	164	-0.9136	-1.152	-0.775
40	-0.3446	0.5441	-0.2243	166	-0.3347	-1.206	-0.1366
42	-1.244	-0.1487	0.2704	168	-0.5758	-0.6614	0.04615
44	-1.299	-0.06616	-0.3113	170	-0.4855	-0.8129	0.08822
46	-0.803	0.2513	-0.2667	172	-0.2267	-0.8715	-0.2587
48	0.2758	0.6856	0.687	174	-0.3966	-0.3512	0.6942
50	0.4821	-0.06045	0.5539	176	-0.2016	0.1463	0.4883
52	0.2342	-0.1593	-0.971	178	-0.834	0.3842	0.4529
54	-0.2935	-0.3436	0.6035	180	-0.1853	-0.3797	-1.247
56	-0.02927	-0.3765	-1.358	182	-0.728	0.2118	-0.07906
58	-0.7294	-0.2809	-0.1479	184	0.7465	0.3709	0.9088
60	-1.077	-0.2874	-1.136	186	0.505	0.9656	0.692
62	0.3595	-0.4993	-0.1225	188	-0.02149	-1.295	-0.5964
64	1.439	0.4314	0.9571	190	-1.133	-0.6707	-1.084
66	1.394	1.188	1.743	192	-0.6141	-0.03275	-0.588
68	0.6022	-1.689	1.423	194	-0.5194	-0.3945	-0.8352
70	0.5162	-1.656	1.227	196	-0.7655	-1.106	-0.1351
72	-0.8375	1.394	0.07247	198	0.5483	0.2996	-0.2418
74	0.5198	-0.1294	0.3145	200	-0.26	0.8959	-0.8979
76	0.2907	0.5119	-1.022	202	0.1721	0.2268	0.7768
78	-0.5117	-0.6924	-0.6566	204	-0.7998	-1.099	-0.2002
80	0.9259	0.403	1.022	206	-0.7715	-1.702	-0.4821
82	0.2144	-0.214	0.6456	208	0.4277	-0.3669	0.576
84	-0.7559	0.2215	0.6611	210	0.1357	0.5559	-0.01934
86	0.1131	-0.03523	-0.4928	212	-0.7977	-1.717	0.5909
88	0.9587	-0.8673	0.4937	214	-1.056	0.9266	-0.1966
90	0.9622	0.7698	0.9764	216	-0.9835	-0.2477	-0.09346
92	-0.5342	-0.5432	0.7809	218	-1.316	-0.4919	0.04485
94	0.1066	0.5873	0.7205	220	-0.431	0.7103	0.753
96	1.193	0.9939	0.4942	222	0.8133	0.8798	0.5155
98	-0.9662	-0.3511	0.08713	224	1.967	0.06698	-1.074
100	0.7616	0.5985	1.397	226	1.742	0.1728	-1.215
102	0.9228	0.03391	1.426	228	1.817	0.339	-0.998
104	1.697	0.05655	-0.7812	230	-0.4846	-0.732	-0.4977
106	-0.7392	1.442	-0.15	232	-0.6948	-0.71	-0.2264
108	-1.721	-1.305	0.9836	234	-0.9527	-0.2571	0.5528
110	-0.4612	0.6269	0.1532	236	0.9736	0.8748	0.95
112	-1.647	-1.984	-1.904	238	0.6393	0.5965	0.1493
114	-0.3302	-0.5847	-1.306	240	0.3729	0.5985	-0.7807
116	-0.4688	-1.228	-0.9736	242	0.3924	-0.07583	-0.8829
118	-0.8008	-1.426	-0.8658	244	0.7529	0.7628	-0.4218
120	0.3424	-0.9043	0.6148	246	-0.658	-0.4103	0.5083
122	0.1863	-0.09213	-0.3089	248	-0.9697	0.09972	0.4308
124	-0.1951	-0.6187	0.08478	250	0.3608	-0.5671	0.3032
126	-0.2459	0.5957	-1.292	252	-0.3564	0.6011	-0.171

TABLA 22: Parámetros  $\theta$  con TRIMJ datos reales Colombia

Persona	$\theta_1$	$\theta_2$	$\theta_3$	Persona	$\theta_1$	$\theta_2$	$\theta_3$
1	0.01771	-0.1763	-0.3924	98	-0.8118	0.514	0.08642
3	-0.09122	-0.2219	-0.982	100	-0.4432	0.3406	0.1755
5	-0.5705	-0.01059	-0.9933	102	0.4096	1.435	-0.8596
7	-0.2991	-0.803	-0.9059	104	-0.6066	-0.5565	-0.9746
9	-0.1623	-0.5907	-0.07435	106	-0.5768	-0.4853	0.843
11	-0.1508	-1.106	-1.308	108	-1.774	0.5719	0.2633
13	-0.4952	-1.199	-0.07074	110	-0.03343	0.5805	0.2363
15	-0.9738	-0.203	-0.5251	112	-0.6066	0.1225	0.6517
17	0.7977	-0.9325	-0.3606	114	-1.241	0.412	0.3148
19	-0.8434	-0.1829	-0.3092	116	0.09158	-0.8258	-0.6917
21	-0.009321	0.5372	1.235	118	-0.897	0.2752	0.3662
23	-0.1744	-0.328	-0.8583	120	-1.33	0.1504	0.973
25	-0.3327	-0.6559	-0.9338	122	0.584	1.071	0.714
27	-0.7179	-1.156	-0.4052	124	1.081	-1.164	0.7828
29	-0.5872	-0.4082	-0.8473	126	0.5606	-0.1928	-0.5692
31	0.05028	0.01385	-0.2949	128	-0.07681	-0.5957	-0.1511
33	0.2312	0.2188	0.2619	130	-0.8818	-1.005	-0.4222
35	-0.677	0.1563	-1.181	132	-0.563	-0.7462	-1.371
37	-0.3783	-0.9522	0.07911	134	-0.3469	-0.1361	-0.5864
39	-1.249	-0.6831	0.3887	136	0.3277	-0.8895	0.423
41	0.3575	-0.1108	-1.855	138	-0.6533	-0.1215	0.4164
43	-0.6768	-1.06	0.4907	140	-1.373	-0.7409	-0.2086
45	0.5489	-1.268	-1.317	142	0.1257	0.769	0.4007
47	0.1217	0.09138	0.4549	144	0.2497	-0.8408	0.6635
49	-1.543	-0.4495	0.04529	146	0.01821	-0.02402	-0.385
51	-0.3269	-0.647	-0.9299	148	-0.254	1.034	1.205
53	0.2395	0.554	0.3371	150	0.06244	0.5314	0.4422
55	-0.03182	-1.147	0.06401	152	0.7045	0.1597	0.8018
57	-0.9819	-0.8762	0.0323	154	0.329	-0.4	0.1845
59	0.2614	0.2541	-0.4746	156	-1.079	0.499	-0.008182
61	-1.088	0.3565	-0.6728	158	-0.1014	-0.8999	-0.02531
63	-0.4813	0.3678	0.8365	160	0.6954	-0.4005	0.7007
65	-0.6398	-0.07344	0.2263	162	-0.582	0.4822	0.7375
67	-0.6608	-0.8326	0.3245	164	-0.1717	-0.4868	-0.6609
69	-0.4167	-0.8123	1.062	166	-0.6634	-0.1534	-0.00566
71	0.8259	0.01439	0.5461	168	-0.6771	-0.2143	-0.8855
73	-0.5692	-0.3742	0.09187	170	1.757	0.774	0.7585
75	0.9091	0.6212	0.1036	172	0.4721	0.6604	0.9146
77	0.0783	-0.1551	-0.8053	174	-1.527	0.6521	0.9759
79	-0.651	0.1734	0.6477	176	1.238	1.45	-0.1671
81	-1.166	0.1837	-1.383	178	0.5089	1.07	-0.03722
83	0.9461	-1.948	1.365	180	-0.1037	-0.9506	-0.2959
85	-0.1642	-0.5597	0.3407	182	0.6236	-0.9402	-0.9863
87	-0.7585	0.1941	0.3889	184	-0.1995	-0.3458	-0.3286
89	0.2523	0.7893	0.2419	186	-0.1208	0.747	-0.1371
91	-0.1429	-1.286	0.002858	188	-0.4511	-0.3199	0.6069
93	0.1241	-1.094	-0.63	190	6.78E-02	-1.395	-1.121
95	-0.1754	-0.9594	-1.479	192	-0.2355	-0.7534	0.2323
97	1.133	-0.9851	-0.06588	194	-0.1262	-0.5455	-0.05201

TABLA 23: Parámetros  $\theta$  con TRIMJ datos reales Cuba

Persona	$\theta_1$	$\theta_2$	$\theta_3$	Persona	$\theta_1$	$\theta_2$	$\theta_3$
1	-0.7085	0.5355	1.06	79	1.025	-0.04525	1.257
3	0.331	0.3212	0.7241	81	0.2722	-0.3236	0.34
5	0.4237	0.3689	0.817	83	1.04	0.2193	1.576
7	-0.2423	-0.6636	0.6838	85	1.294	1.082	1.422
9	-0.01381	0.655	1.134	87	-0.5384	0.7702	1.106
11	1.315	1.07	1.415	89	0.9024	0.9832	1.185
12	0.9969	0.6633	1.217	90	1.322	0.4908	0.8952
14	0.9966	0.6763	1.21	92	1.365	-0.6229	0.2298
16	-0.9078	1.397	0.5223	94	1.133	0.1506	-0.3922
18	0.4734	1.12	0.8209	96	0.398	0.09483	0.8194
20	0.4897	1.293	-0.2687	98	-0.004734	0.1837	0.189
22	-0.9317	1.256	0.9935	100	-0.8707	-0.4183	-0.1484
23	0.7175	0.6466	1.557	101	0.2867	0.2515	1.253
25	1.171	1.122	1.221	103	-0.3323	0.149	1.207
27	0.5721	-0.08279	0.8679	105	0.9741	-0.6031	1.212
29	1.182	-0.01493	0.4439	107	0.5727	0.8053	0.5272
31	0.3521	-0.1117	1.164	109	0.8759	0.4911	1.196
33	-0.2372	0.1789	-0.2509	111	0.6616	1.076	1.584
34	0.9624	0.7427	-0.6645	112	1.461	-1.11	1.225
36	1.478	-0.1124	1.251	114	-0.2324	-0.5933	-0.1857
38	1.173	1.132	1.204	116	-0.09732	1.218	-0.4056
40	0.5811	-0.4728	0.8896	118	1.621	-1.975	1.239
42	0.6256	-0.1872	1.161	120	0.7861	1.157	1.254
44	1.527	1.471	-0.1231	122	0.7902	0.7717	1.328
45	1.289	0.5013	1.616	123	1.037	1.005	1.34
47	0.532	0.7795	0.7582	125	-0.2311	-0.3479	0.4669
49	-0.6978	1.226	-0.3589	127	1.344	0.837	0.8264
51	-1.805	1.085	-0.1655	129	-0.5397	0.376	0.6352
53	-0.2588	1.333	1.161	131	0.8788	0.05679	1.338
55	0.1357	1.23	-0.5723	133	0.5425	-1.027	1.424
56	0.4539	-0.1056	-0.1011	134	0.3434	-1.927	1.441
58	0.7783	0.7347	0.8324	136	-1.138	-0.04366	0.3984
60	0.007603	1.26	-0.467	138	-0.09872	0.5101	1.409
62	-0.6476	0.7519	0.4255	140	0.4484	0.5145	0.7561
64	-0.3225	0.4884	-0.2439	142	0.2527	0.6027	1.36
66	-1.024	0.8544	1.002	144	-0.6116	-0.3169	0.6382
67	-0.2332	0.2187	1.287	145	-0.08507	-1.153	-0.0175
69	0.2232	-1.357	0.711	147	1.354	0.08158	0.831
71	1.229	0.8795	-0.1003	149	-0.7455	-0.568	0.9384
73	0.1978	1.414	-0.01708	151	1.218	1.149	1.01
75	0.3468	0.3607	0.8867	153	1.076	-0.5006	1.409
77	0.8732	0.701	1.305	155	1.306	1.086	1.649



TABLA 24: Trazo latente unidimensional con TRIMJ datos reales Bolivia

Persona	Matemática	Lenguaje	Persona	Matemática	Lenguaje
2	1.538541765	1.224907088	128	0.580571166	-0.286012434
4	-0.963618227	-0.570603988	130	0.554179828	1.029583333
6	-0.96474706	-0.416644444	132	-0.006155097	0.198978703
8	-1.2499916	-0.821521842	134	0.536886549	0.063623273
10	-0.599825806	-0.553269287	136	-0.1611513	0.003702175
12	-1.283028357	-0.835683287	138	0.948926225	0.48508273
14	-1.033665394	-1.027775991	140	0.952564971	1.179091674
16	-0.995770086	-0.889437227	142	-0.831171703	-1.32977556
18	-0.397645342	0.179205578	144	-1.166380997	-1.114138485
20	0.599674147	1.561202459	146	-0.233001926	-0.344690602
22	1.00157964	1.76123052	148	-0.981084763	0.239806356
24	-0.774513313	-0.688695698	150	-0.407055433	-0.734499301
26	-0.016814033	-0.138413614	152	-0.269469978	-0.129951528
28	-0.380409102	-0.690192332	154	-0.596451177	0.061240486
30	-0.087976237	0.687146835	156	0.398978345	0.34461353
32	-0.825166563	-0.964612446	158	-1.044922929	-0.977506911
34	-0.819682082	-1.182555041	160	-0.787700312	-1.274676481
36	1.323160634	1.07413632	162	-0.58410548	-1.317412823
38	0.794370742	0.52768051	164	-1.539445873	-1.585058202
40	-0.168344018	0.266817115	166	-0.762080857	-1.186571827
42	-0.694177247	-0.375930835	168	-0.595748317	-0.713852237
44	-1.080264474	-0.571823949	170	-0.56593167	-0.798572451
46	-0.606626962	-0.135983909	172	-0.64556917	-0.921822002
48	0.888459523	0.96176515	174	0.062803361	-0.117676677
50	0.656136604	0.328209485	176	0.242702073	0.275724862
52	-0.542420034	-0.484872189	178	-0.105616888	0.277886126
54	0.07382551	-0.119926375	180	-1.081021061	-0.914921961
56	-1.050255302	-0.914116732	182	-0.449392125	-0.066971589
58	-0.680558034	-0.517598766	184	1.224029892	0.92673473
60	-1.558121573	-1.050227695	186	1.147194857	1.270248701
62	-0.032231428	-0.373365222	188	-0.891421489	-1.368348605
64	1.732330768	1.20250712	190	-1.704806884	-1.371519229
66	2.501977268	2.173221203	192	-0.800022809	-0.461200688
68	0.692182755	-0.652953968	194	-1.0358242	-0.848537112
70	0.519818519	-0.734341151	196	-1.005903251	-1.227061743
72	0.022357167	0.975266998	198	0.313596259	0.312762644
74	0.49811417	0.177446534	200	-0.42193981	0.302249066
76	-0.286561048	0.082769587	202	0.706896628	0.57827298
78	-1.025730501	-1.023765371	204	-1.068400903	-1.25915935
80	1.427841191	1.055508217	206	-1.461117697	-1.886868594
82	0.483029149	0.15796118	208	0.519793234	0.060131974
84	0.020820449	0.251479292	210	0.285184164	0.506087535
86	-0.261921236	-0.208830324	212	-0.781039885	-1.445840772
88	0.625898413	-0.246520889	214	-0.472772027	0.396362905
90	1.559605235	1.359682087	216	-0.798979322	-0.54045605
92	-0.042749671	-0.284696112	218	-1.018095324	-0.787070106
94	0.762555243	0.842573167	220	0.477677076	0.803690347
96	1.479233137	1.411357017	222	1.201419222	1.21161692
98	-0.708197573	-0.545969391	224	0.610972349	0.172824281
100	1.639262738	1.335453518	226	0.410889282	0.13644198
102	1.551755339	0.913286066	228	0.664657591	0.393683967
104	0.621875069	0.210577266	230	-0.918757797	-0.981265918
106	-0.040903358	0.949418445	232	-0.870553882	-0.907515928
108	-0.974179942	-1.196338702	234	-0.358978173	-0.261462302
110	0.033630093	0.465635374	236	1.589242278	1.441310263
112	-3.072413198	-2.996290136	238	0.741015977	0.761163553
114	-1.291687949	-1.157852331	240	-0.042068198	0.284625315
116	-1.406541609	-1.604747092	242	-0.349642807	-0.329266858
118	-1.627954361	-1.82489768	244	0.503864225	0.690842397
120	0.287345483	-0.406992523	246	-0.252527867	-0.324848634
122	-0.114855036	-0.156809203	248	-0.3159493	-0.014321168
124	-0.304852293	-0.54899343	250	0.222004588	-0.247757559
126	-0.783673041	-0.11939716	252	-0.119743415	0.334937398

TABLA 25: Trazo latente unidimensional con TRIMJ datos reales Colombia

Persona	Matemática	Lenguaje	Persona	Matemática	Lenguaje
2	-0.025562422	-0.736380078	98	-0.278679466	0.269436179
4	-1.160146312	-1.183211375	100	-0.045661986	0.256461798
6	-0.752436919	-1.059595772	102	0.260014448	0.940438008
8	-1.776452633	-1.978427155	104	-1.245093305	-1.0725537
10	-1.164055817	-1.249251177	106	-0.015982799	-0.180257536
12	0.066033062	0.210276895	108	-0.772123949	0.151935155
14	-0.28123998	-0.534521157	110	0.354542498	0.594025158
16	-1.026877456	-0.617484768	112	0.073283108	0.244183045
18	-0.632792287	0.64373091	114	-0.450607218	0.175668871
20	-0.045799399	0.020731714	116	-0.706768646	-0.994894948
22	-0.624686255	-0.908725284	118	-0.244179253	0.170538701
24	-0.665525431	-0.537659765	120	-0.18120413	0.227861456
26	-1.038727256	-0.635754312	122	1.258200764	1.387792684
28	0.997838737	1.359310732	124	0.771096936	-0.361135849
30	0.027281358	-0.231533108	126	-0.076749833	-0.278607588
32	-0.898061689	-1.371333698	128	-0.377218316	-0.596690911
34	-0.767320489	-1.276619487	130	-1.238279071	-1.275668178
36	-0.453961196	-0.168492051	132	-1.547145063	-1.402957998
38	-0.6364193	-0.542808753	134	-0.661038481	-0.47105684
40	0.604275694	0.639273899	136	0.148617489	-0.483367784
42	0.211703518	0.680255585	138	-0.204008718	-0.082737787
44	-1.003080804	-0.575631115	140	-1.319977801	-1.079562027
46	-0.536508489	0.138479606	142	0.638117458	0.870308012
48	-0.219409869	-0.414515063	144	0.272609649	-0.35281726
50	-0.492777459	-1.236884863	146	-0.24770338	-0.190336537
52	-0.587248047	-0.455244101	148	1.014071352	1.363564416
54	-0.902153551	-0.618814713	150	0.532470732	0.670063292
56	0.598382052	0.996549189	152	1.044604798	0.680669808
58	0.50189409	0.624894505	154	0.182172884	-0.173313265
60	-0.344271118	0.140563418	156	-0.521116412	0.145133626
62	-0.581287917	-0.21459786	158	-0.428247219	-0.805646874
64	0.732490774	0.573629424	160	0.7579146	0.154332828
66	1.216937488	1.441386633	162	0.283243933	0.596423124
68	-0.366209749	0.526311687	164	-0.729228516	-0.759183039
70	-0.904649871	-1.00045705	166	-0.497398465	-0.303868964
72	-0.194386919	-0.41128949	168	-1.102038476	-0.758177613
74	0.646560296	0.579768275	170	1.942069568	1.455683393
76	-0.342111584	-0.808512262	172	1.157772232	1.099503732
78	-1.070733357	-0.892689618	174	-0.115932876	0.606825185
80	-0.491978112	-0.869564267	176	1.259215028	1.479870597
82	-1.349686265	-0.992285827	178	0.719959642	1.027145436
84	-0.028598338	-0.743208799	180	-0.625214786	-0.972164998
86	0.590730403	0.693790815	182	-0.593554897	-1.089452837
88	-1.086302361	-0.509336692	184	-0.477201343	-0.495345234
90	0.677020356	-0.016000894	186	0.118285561	0.544456686
92	-1.390909171	-1.377301451	188	-0.023689713	-0.113813083
94	-0.749136234	-0.328052172	190	-1.220012246	-1.681448161
96	-1.248493707	-0.906408471	192	-0.292370269	-0.598267741
			194	-0.325876894	-0.521610128

TABLA 26: Trazo latente unidimensional con TRIMJ datos reales Cuba

Persona	Matemática	Lenguaje	Persona	Matemática	Lenguaje
2	0.09745081	-0.15134123	80	1.01302505	0.46581254
4	-0.35668548	0.01734659	82	0.57953082	1.17059036
6	0.52038228	0.93230307	84	2.09408435	1.99568024
8	2.11177814	2.01203697	86	0.28755954	0.5891962
10	1.44367311	1.15909207	88	2.09229515	1.9916046
12	1.53851312	1.35880086	90	1.29804906	1.14507263
14	1.5338906	1.36676696	92	0.50217232	-0.10082186
16	0.46223158	1.21513583	94	-0.0276131	0.22890431
18	1.08950228	1.4465948	96	0.90249752	0.54595399
20	0.09107591	1.11042744	98	0.20690628	0.24163582
22	0.8786451	1.29949712	100	-0.45279095	-0.63794395
24	1.47703618	1.14159711	102	1.01457409	0.91744678
26	1.53083252	1.69965744	104	1.53181952	1.67992449
28	0.75304478	1.04562454	106	1.26723841	0.67598288
30	-0.03318668	0.70869633	108	1.8761649	1.89109476
32	0.32574929	1.04463917	110	1.87550946	1.33226693
34	-0.23836295	0.57446214	112	1.39259004	-0.0498429
36	1.58189573	0.8238872	114	-0.33626605	-0.65004385
38	1.65097227	1.79884818	116	-0.21586679	0.84182742
40	0.92950748	0.13379518	118	1.31225636	-0.74857027
42	1.24457188	0.51200395	120	1.59325667	1.74874891
44	0.5496424	1.58094939	122	1.6026501	1.45159307
46	1.7427963	1.80320511	124	1.65712765	1.00849779
48	-0.27488702	0.77392341	126	1.4173494	1.10001005
50	-0.29634682	-0.03796609	128	1.54068335	1.39932348
52	1.53016576	1.598367	130	0.59854647	0.38530513
54	1.68512727	1.68526932	132	1.04112652	-0.63162946
56	0.01533454	-0.02586115	134	1.15103218	-0.92716736
58	1.1245323	1.19450307	136	0.04921226	-0.1353696
60	-0.2376521	0.87589788	138	1.38685511	1.04691499
62	0.34056788	0.68015832	140	0.92409316	0.89071503
64	-0.24331045	0.2322008	142	1.45437074	1.18997522
66	0.79629903	0.93550359	144	0.38059126	-0.13481633
68	0.03689074	0.46740352	146	1.89306622	1.69380863
70	1.29895272	0.89422876	148	0.75193258	0.40444383
72	1.88398008	1.88765488	150	1.00789467	1.14351995
74	1.11240158	1.02246648	152	1.43612601	0.98586496
76	0.78106974	0.35092627	154	1.92870553	1.25918307
78	0.8276707	0.46732574	156	1.58842304	1.40288527