

## RELACIÓN TIMSS–SIMCE: Calidad psicométrica e invarianza de puntajes y parámetros

### INTRODUCCIÓN

El año 2010 los cuartos básicos rindieron la prueba nacional SIMCE 2010 y una muestra probabilística de ellos también fue medida en la prueba internacional TIMSS 2011<sup>1</sup>. Ambas pruebas tuvieron un área de aprendizaje en común, Matemática. En el Apunte 4 (abril, 2013) se hizo el ejercicio de llevar ambas a una escala común y comparar sus estándares<sup>2</sup>. En este Apunte se presenta un estudio de la calidad psicométrica, en términos de dificultad, discriminación y posibilidad de adivinación, de las preguntas SIMCE en relación con las preguntas TIMSS. Para esto se investigó la estabilidad o invarianza de los valores de los parámetros de las preguntas de ambas pruebas, como asimismo de los puntajes de los alumnos bajo distintos escenarios.

Existen dos características de la Teoría de Respuesta al Ítem (TRI) que la hacen más atractiva que la Teoría Clásica (TC) para analizar una prueba. Una es la invarianza en los parámetros de dificultad, discriminación y pseudoazar<sup>3</sup> de las preguntas y la otra es la invarianza entre los puntajes de los examinados. Con la primera, se garantiza que la determinación de las características de las preguntas será estable en el constructo o rasgo latente<sup>4</sup> medido por la prueba independiente del nivel de habilidad de la población examinada. Esto no sucede en la TC, ya que la dificultad de una pregunta se estima con la proporción de examinados que la contestan correctamente. Con la segunda característica, TRI garantiza que la estimación de la habilidad de los sujetos en el constructo medido, sea independiente del conjunto de preguntas a las que se ven sometidos los examinados<sup>5</sup>. Esto se conoce como invarianza de las puntuaciones. En la TC, las puntuaciones generalmente se calculan con el número de respuestas correctas<sup>6</sup>, por lo que la puntuación de un examinado depende de la conformación de la prueba<sup>7</sup>.

La invarianza en los parámetros de las preguntas permite, entre otras cosas, la construcción de pruebas comparables en el tiempo, lo que simplifica el monitoreo del avance educativo. Por su parte, la invarianza entre las puntuaciones permite garantizar que la determinación del nivel de habilidad en el constructo sea comparable entre pruebas distintas.

---

<sup>1</sup> Los alumnos matriculados en 4.º básico en 2010 rindieron la prueba SIMCE de Lectura y Matemática y, además, una muestra probabilística de ellos fue sometida a la prueba TIMSS, que evalúa las áreas de Matemática y Ciencias.

<sup>2</sup> Ver Apunte 4 en [http://www.agenciaeducacion.cl/wpcontent/uploads/2013/04/alineamiento\\_mat\\_SIMCE\\_2010.pdf](http://www.agenciaeducacion.cl/wpcontent/uploads/2013/04/alineamiento_mat_SIMCE_2010.pdf)

<sup>3</sup> Dependiendo del tipo de modelo que se utilice.

<sup>4</sup> El constructo o rasgo latente corresponde a la habilidad que se quiere medir.

<sup>5</sup> Siempre que las preguntas estén diseñadas para medir el constructo deseado.

<sup>6</sup> O una función de las respuestas correctas.

<sup>7</sup> En cuanto el tipo de preguntas.

Para que efectivamente las pruebas cumplan con estas dos características, deben ser construidas y analizadas en el marco de la TRI. Para poder emplear esta teoría se espera que se cumplan supuestos bastante más exigentes que los de la TC. Básicamente se deben cumplir estos tres:

- que las preguntas o ítems de la prueba midan un solo constructo o rasgo latente,
- que para cada pregunta exista una relación funcional entre el nivel de habilidad en el constructo y la probabilidad de acertar al ítem, y
- que la respuesta del examinado a un ítem cualquiera no se vea afectada por su respuesta a las demás preguntas.

Estos tres supuestos aseguran que quien responda cada una de las preguntas de una prueba lo haga solamente de acuerdo a su nivel de habilidad y no por otra razón<sup>8</sup>. Existen técnicas estadísticas que permiten determinar el grado de cumplimiento de estos supuestos (Hambleton y Swaminathan, 1985). Específicamente, para examinar el grado de cumplimiento del segundo supuesto se utilizan diversas familias de funciones que se diferencian en el número y tipo de parámetros que las caracterizan.

Tanto las pruebas SIMCE como las pruebas internacionales, se construyen y analizan en el marco de la Teoría de Respuesta al Ítem. Para satisfacer el segundo supuesto, la prueba TIMSS y SIMCE utilizan una función logística de tres parámetros. Con ella se modela la probabilidad de acertar correctamente una pregunta en función de la habilidad que mide:

$$P(\theta) = c + \frac{(1-c) e^{1,7\alpha(\theta-b)}}{1+e^{1,7\alpha(\theta-b)}}$$

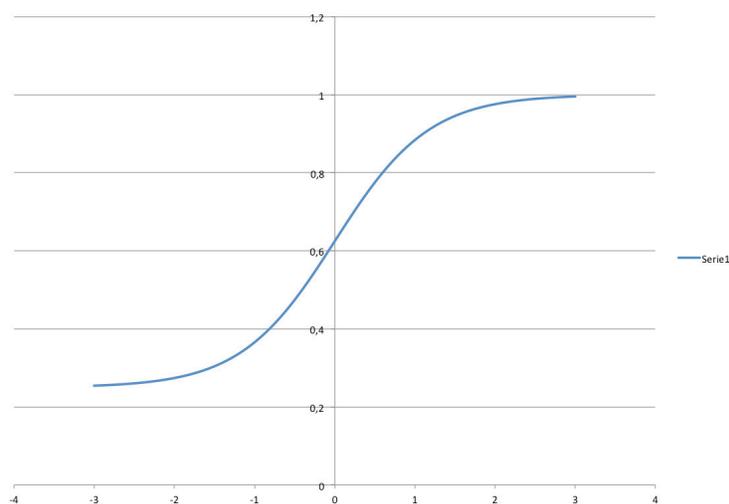
$P(\theta)$  es la probabilidad de acertar correctamente al ítem en función de la **habilidad** o rasgo latente  $\theta$  que se pretende medir. Los parámetros del modelo son  $\alpha$ ,  $b$  y  $c$ , donde  $\alpha$  corresponde a la **discriminación**,  $b$  es la **dificultad** y  $c$  es el **pseudoazar**. La curva representada en las Figuras 1 y 2 recibe el nombre de **Curva Característica del Ítem** (CCI). Los valores de los parámetros determinan las características de esta curva. Asimismo, los parámetros se usan para estimar la habilidad de cada examinado.

El **pseudoazar** ( $c$ ) corresponde a la probabilidad de acertar correctamente al ítem cuando la habilidad  $\theta$  es muy pequeña. La **dificultad** ( $b$ ) es el punto medio entre  $c$  y 1, y corresponde al punto en la escala de habilidades donde se tiene más posibilidades de acertar al ítem que de errarlo, descontando el valor de  $c$ . Este parámetro también recibe el nombre de **parámetro de localización** ya que indica la habilidad mínima que se requiere para tener más probabilidades de éxito que de fracaso al afrontar el ítem (descontando el **pseudoazar**). La **discriminación** ( $\alpha$ ) es un valor proporcional al valor de la pendiente o inclinación máxima de la curva que se alcanza justamente en el nivel de dificultad  $b$ . Este valor da una idea de la tasa de crecimiento de la habilidad necesaria para aumentar sustantivamente la probabilidad de acertar al ítem.

<sup>8</sup> Para un examen comparativo entre los supuestos de ambas teorías, revisar Hambleton y Swaminathan, 1985.

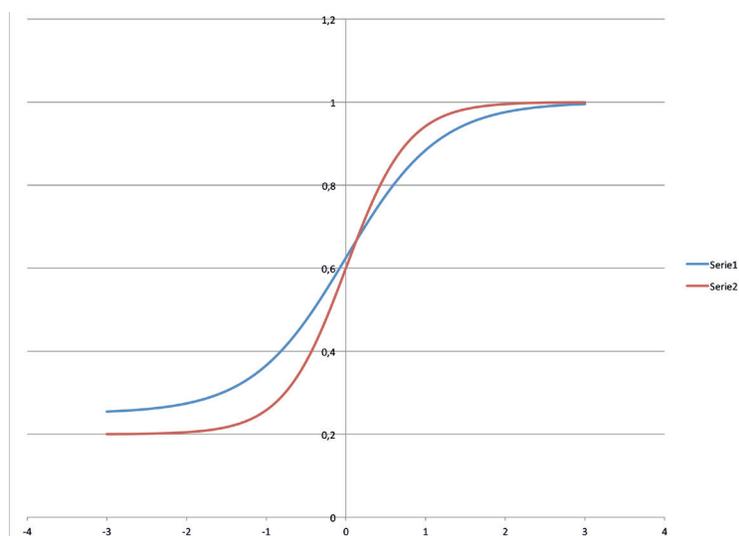
Las Figuras 1 y 2 grafican esta función. La Figura 1 corresponde a la CCI de un ítem con parámetros  $\alpha=1$ ,  $b=0$  y  $c=0,25$ .

**Figura 1**



La Figura 2 describe el mismo tipo de curva pero para una pregunta con el parámetro de discriminación más alto ( $\alpha=1,5$ ) y el de pseudoazar más bajo ( $c=0,2$ ), manteniendo la dificultad ( $b=0$ ). Se observa que la inclinación de la curva es más pronunciada que en la Figura 1, pues tiene un parámetro de discriminación más alto. Al aumentar la habilidad, la probabilidad de acertar al ítem aumenta más que en el caso anterior. También se observa que la curva tiende a un valor de  $c$  más pequeño pues el ítem presenta menos pseudoazar. Esto quiere decir que los sujetos con poca habilidad tienen menos probabilidad de acertar a la respuesta correcta en este ítem que en el anteriormente graficado.

**Figura 2**



Para este Apunte se estudió la determinación de los parámetros y las puntuaciones en las pruebas SIMCE y TIMSS<sup>9</sup>. En el proceso también se pudo examinar la calidad psicométrica de las preguntas SIMCE en relación a las de la prueba TIMSS. Para ello se dispuso de la determinación de los parámetros de las preguntas para la prueba SIMCE en la población chilena, como asimismo la de la prueba TIMSS para la población que evalúa<sup>10</sup>. Asimismo, dado que tenemos una muestra de estudiantes chilenos que rindió ambas pruebas, con ella se pudo determinar conjuntamente los parámetros de las pruebas SIMCE y TIMSS. Además, se pudo determinar los puntajes de los alumnos en una prueba usando los parámetros de las preguntas de la otra como ancla y así estudiar el efecto en las puntuaciones de configuraciones de pruebas distintas<sup>11</sup>.

Concretamente se responden las siguientes interrogantes:

- ¿Es posible verificar la invarianza de los parámetros de las preguntas de las pruebas TIMSS y SIMCE?
- ¿Es posible verificar la invarianza de los puntajes de los examinados en las pruebas TIMSS y SIMCE?
- ¿Tienen las preguntas SIMCE calidades psicométricas similares a las de la prueba TIMSS?

## METODOLOGÍA

Para establecer o verificar invarianza tanto en los parámetros de las preguntas como entre las puntuaciones de los examinados, se determinaron las características de las preguntas SIMCE y TIMSS de manera conjunta y libre<sup>12</sup> usando la muestra de alumnos que rindió ambas pruebas. También se calibró cada prueba haciendo importación de los parámetros de una para llevarla a la escala de la otra. De esta manera se obtuvo un conjunto de parámetros para cada pregunta según:

- Calibración original (usando la población propia para cada caso)
- Calibración conjunta (usando la muestra de estudiantes y ajustando simultáneamente ambas pruebas)
- Calibración con importación de los parámetros de la otra prueba.

Todas las calibraciones se llevaron a cabo utilizando el software Parscale 4.1 (2003).

<sup>9</sup> La determinación de los parámetros se hace empleando técnicas de maximización de la verosimilitud y bayesianas.

<sup>10</sup> Participaron 66 países, pero solamente se estima las propiedades de las preguntas para los países pertenecientes a la OECD (Organización para la Cooperación y el Desarrollo Económico, en español).

<sup>11</sup> Cabe mencionar que bajo TRI la escala de medición que se puede utilizar para medir la habilidad está basada en una transformación de la razón entre probabilidad de acierto ( $p$ ) a la probabilidad de no acierto ( $1-p$ ), es decir  $p/(1-p)$ . En la literatura estadística esta razón es llamada *odds*. Por ejemplo, un *odds* de 2 significa que la probabilidad de acertar es el doble de la de fallar. En general, esta escala se somete a una transformación lineal para entregar los resultados en una métrica más comprensible.

<sup>12</sup> Esto sin imponer ninguna métrica.

## RESULTADOS

Para analizar la invarianza de los parámetros, se comparó el valor estimado de cada parámetro obtenido al calibrar las pruebas en su métrica original y en una calibración conjunta. Para esto último se hizo el supuesto que hubo una sola gran prueba rendida por la muestra de estudiantes. Las calibraciones originales se obtuvieron al ajustar los parámetros de las preguntas de la prueba a la población y constructo para la cual fueron especialmente diseñadas. Al modificar la población, es decir, al utilizar solo la muestra de alumnos que rindió ambas pruebas, se está variando las características de la muestra de examinados. Si los valores de las estimaciones no difieren mayormente entre uno y otro caso se está en presencia de invarianza de los parámetros, ya que se podría identificar la CCI independientemente del grupo de examinados considerados.

Alguna variabilidad entre dichas estimaciones es esperable en este caso, pues los constructos o rasgos latentes de ambas pruebas no son idénticos ya que provienen de marcos de referencia distintos. La prueba TIMSS mide la habilidad para un currículo internacional definido ad hoc, en cambio la prueba SIMCE mide la habilidad para el currículo chileno vigente.

## INVARIANZA DE LOS PARÁMETROS

El parámetro de dificultad en las pruebas analizadas mediante TRI se mueve entre  $-2,5$  y  $2,5$ <sup>13</sup> en una escala centrada en 0 y con desviación estándar de 1. A su vez, el parámetro de discriminación se mueve en rangos de 0,3 a 2,0<sup>14</sup> y el de pseudoazar entre 0,01 y 0,4<sup>15</sup>. Si dividimos estas escalas en 6 categorías, tal como muestra la Tabla 1, se observa que un gran porcentaje de los ítems mantiene su categoría de pertenencia.

**Tabla 1**  
**Porcentaje de preguntas que mantienen su categoría de pertenencia entre calibración original y conjunta**

Prueba	Dificultad	Discriminación	Azar
TIMSS	88%	81%	84%
SIMCE	94%	92%	89%

<sup>13</sup> La escala en la que se mide habilidad se suele centrar en 0, con 1 de desviación estándar, aunque en general se reporta efectuando después una transformación lineal.

<sup>14</sup> Recordemos que es proporcional a una pendiente.

<sup>15</sup> Recordemos que se identifica con la probabilidad de acertar con habilidad nula.

En la Tabla 2 se reportan las diferencias absolutas para cada prueba y parámetro:

**Tabla 2**  
**Diferencias absolutas en los valores de los parámetros**  
**entre calibración original y conjunta**

	Parámetros	Mínimo	Máximo	Media	Desviación Est.
TIMSS	Diferencia abs. dificultad	0,01	1,85	0,50	0,37
	Diferencia abs. discriminación	0,00	1,40	0,19	0,18
	Diferencia abs. azar	0,00	0,22	0,05	0,05
SIMCE	Diferencia abs. dificultad	0,06	0,60	0,32	0,12
	Diferencia abs. discriminación	0,00	0,65	0,12	0,12
	Diferencia abs. azar	0,00	0,22	0,04	0,05

Se puede concluir que los valores de los parámetros de las preguntas no cambian significativamente al ser estas calibradas con distintas muestras de alumnos. Lo anterior es un comportamiento individual de las preguntas por lo que también se debe estudiar si la relación entre las preguntas dentro de la prueba cambia. Podría darse el caso que una pregunta no modifique mucho sus parámetros pero pase a ser comparativamente más fácil o difícil.

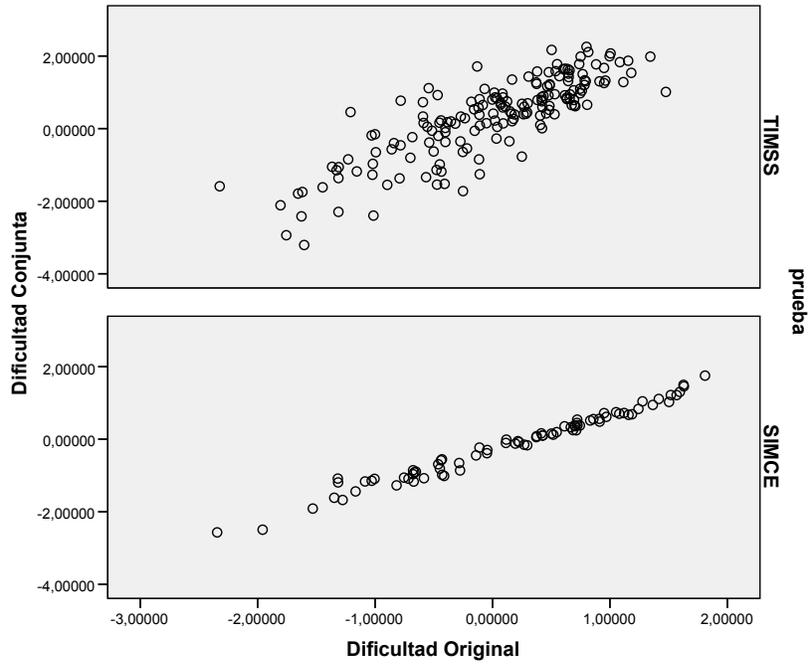
Para estudiar el comportamiento relativo, se debe analizar las correlaciones entre los valores de los parámetros obtenidos en cada calibración. En la Tabla 3 se observan correlaciones altas, por lo que se deduce que los parámetros mantienen su comportamiento relativo entre sí:

**Tabla 3**  
**Correlación entre las calibraciones conjuntas y originales**

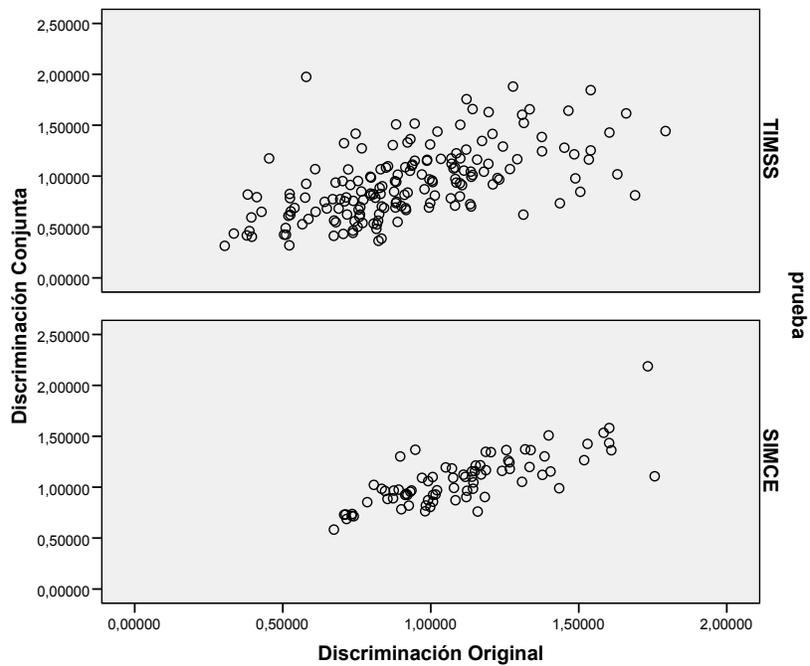
Prueba	Dificultad	Discriminación	Pseudoazar
TIMSS	0,82	0,64	0,54
SIMCE	0,99	0,77	0,61

Las Figuras 3, 4 y 5 muestran los valores de los distintos parámetros (dificultad, discriminación y pseudoazar), estimados para sus valores originales y provenientes de la calibración conjunta. Con ellas, se ratifica la alta correlación entre las calibraciones conjuntas y originales.

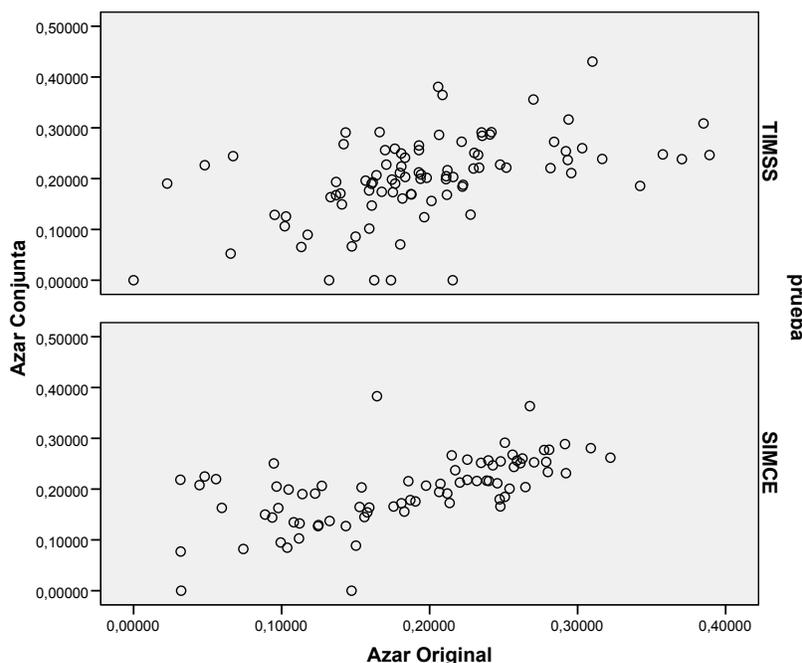
**Figura 3**  
Correlación entre las calibraciones conjuntas y originales para el parámetro de dificultad



**Figura 4**  
Correlación entre las calibraciones conjuntas y originales para el parámetro de discriminación



**Figura 5**  
**Correlación entre las calibraciones conjuntas y originales para el parámetro de pseudoazar**



En general se observa que, para los tres tipos de parámetros, existe una relación lineal bastante fuerte entre los parámetros obtenidos originalmente y los de la calibración conjunta<sup>16</sup>, lo que indica que la posición relativa de los parámetros dentro de las pruebas originales y la conjunta no se ve mayormente afectada.

### INVARIANZA DE LOS PUNTAJES

Considerando que se tiene tres tipos de calibraciones para cada prueba (la original, la conjunta y la con importación del comportamiento de parámetros de la otra prueba), las respuestas de cada alumno tanto para TIMSS como para SIMCE se pueden analizar de esas tres maneras. Para el análisis de la prueba TIMSS no se obtuvieron los puntajes asociados a la calibración conjunta<sup>17</sup>, por lo que solamente se presentan los resultados con la importación de parámetros SIMCE calculados usando la metodología de dicha prueba<sup>18</sup>.

Bajo TRI se puede definir la métrica de reporte usando transformaciones lineales, luego lo relevante es comparar el grado de asociación lineal existente entre las puntuaciones de los alumnos de la

<sup>16</sup> Esto en una muestra de alumnos de habilidades distintas a las poblaciones utilizadas para las calibraciones originales.

<sup>17</sup> La razón es que este estudio se realizó cuando aún no se disponía del reporte Técnico de TIMSS 2012 donde se detallan los modelos y las variables condicionantes que permiten la obtención de los plausible values, que es la metodología con la que se calculan y reportan los resultados TIMSS. El reporte todavía no está disponible pues se tendrá durante el segundo semestre del año 2013. Esa información es necesaria, ya que los puntajes TIMSS no son obtenidos directamente de la aplicación de puntuación bajo TRI (como lo son los puntajes SIMCE) sino que sufren un ajuste bajo variables condicionantes, debido a que cada alumno que rinde TIMSS responde una porción menor del total de la prueba.

<sup>18</sup> Se ajusta un estimador EAP (*expected a posteriori*).

muestra bajo las distintas calibraciones. En la Tabla 4 se entregan las correlaciones obtenidas por los puntajes de los alumnos según prueba y método de calibración. Se observan correlaciones altas, todas superiores a 0,80, entre las puntuaciones de las distintas pruebas. Esto es señal de una gran estabilidad en la estimación de la habilidad subyacente a los constructos o rasgos latentes que miden cada prueba, independiente de la manera de calibrar las preguntas. Se puede comprobar, además, que más del 64% de la variabilidad en una prueba puede ser explicado por las otras<sup>19</sup>.

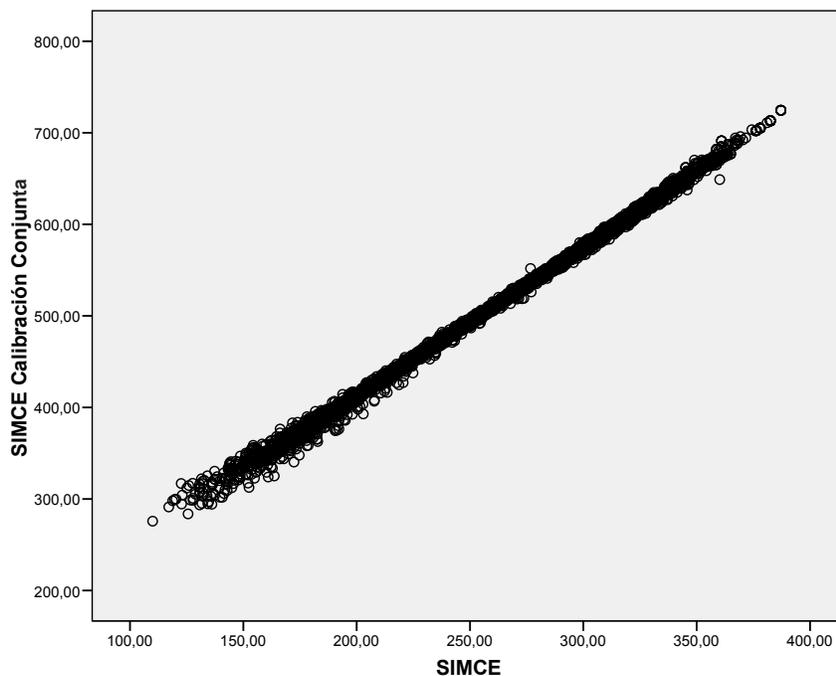
**Tabla 4**  
**Correlaciones entre pruebas SIMCE y TIMSS según tipo de calibración**

	SIMCE	TIMSS	SIMCE Conjunta	SIMCE Importación TIMSS	TIMSS Importación SIMCE
TIMSS	0,82	1,00	0,82	0,82	0,94
SIMCE	1,00	0,82	1,00	1,00	0,80
SIMCE Calibración Conjunta	1,00	0,82	1,00	1,00	0,82
SIMCE Importación TIMSS	1,00	0,82	1,00	1,00	0,80
TIMSS Importación SIMCE	0,80	0,94	0,82	0,80	1,00

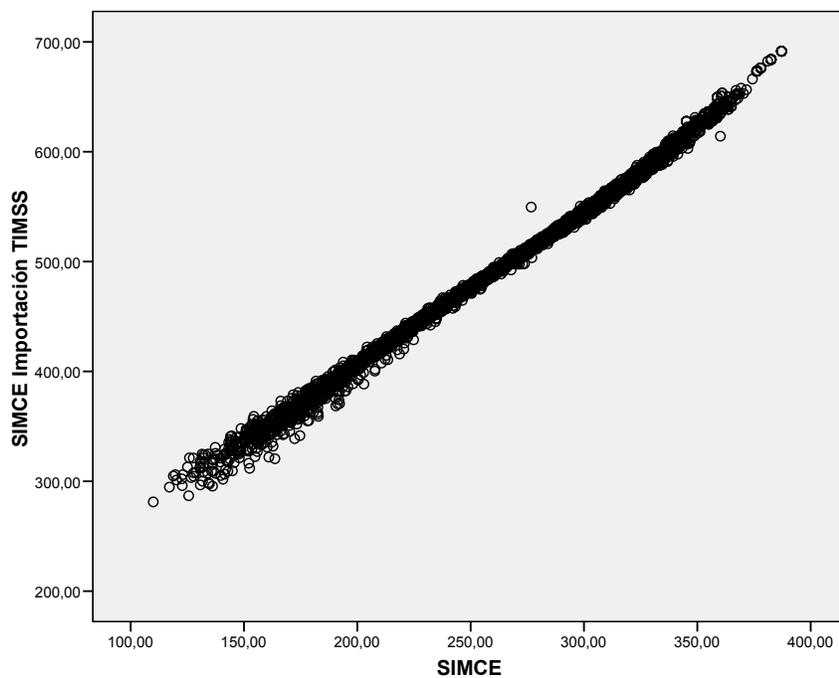
En las Figuras 6, 7 y 8 se presentan algunas de estas asociaciones en forma gráfica.

<sup>19</sup> A modo de referencia comparativa para las pruebas PISA y SIMCE se han encontrado correlaciones de entre 0,74 y 0,76 (Matus et al. 2012) entre las puntuaciones, asimismo en el plano internacional se han observado correlaciones entre 0,66 y 0,67 entre PISA y PLAN (ACT, Inc. 2011).

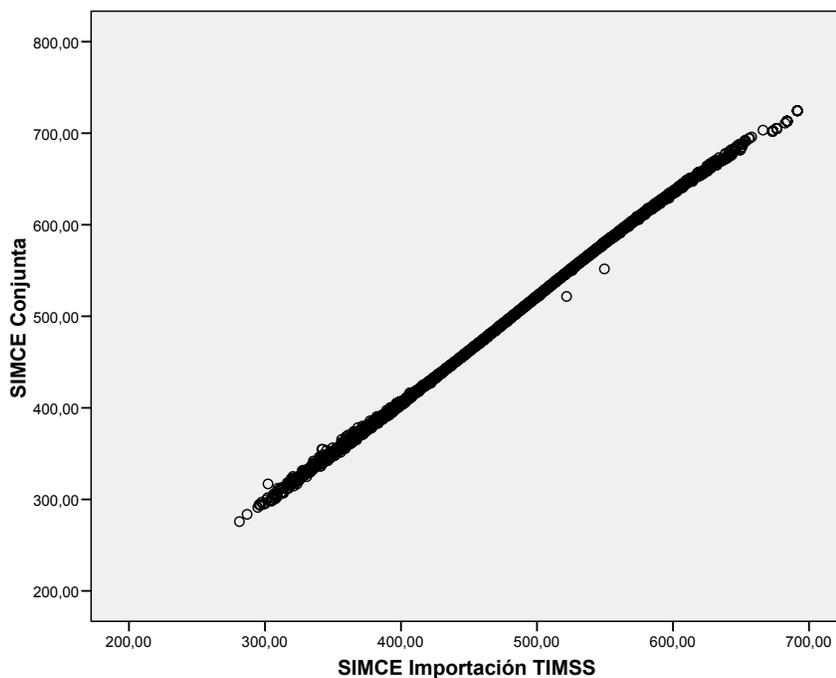
**Figura 6**  
**Correlación entre SIMCE y SIMCE calibración conjunta**



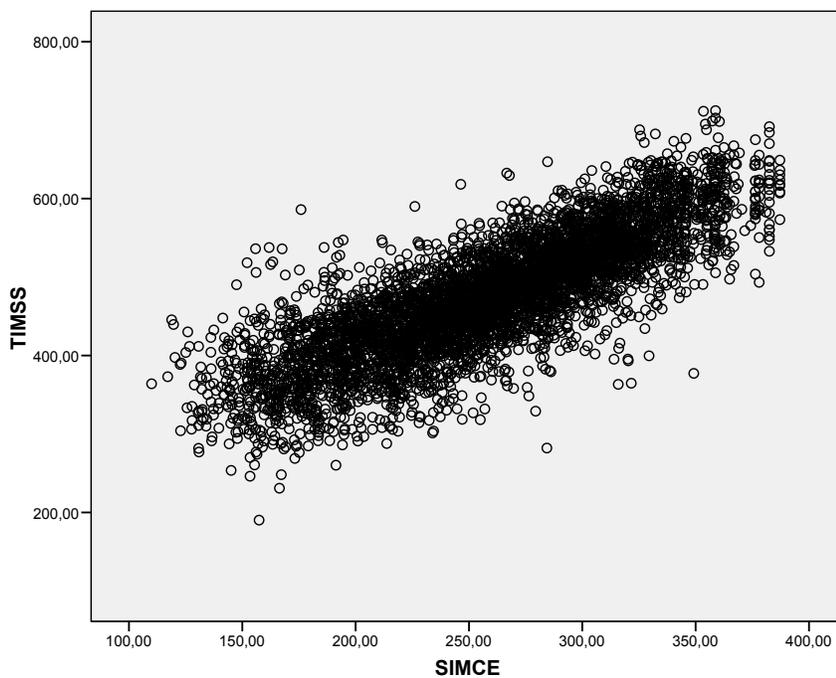
**Figura 7**  
**Correlación entre SIMCE y SIMCE importación TIMSS**



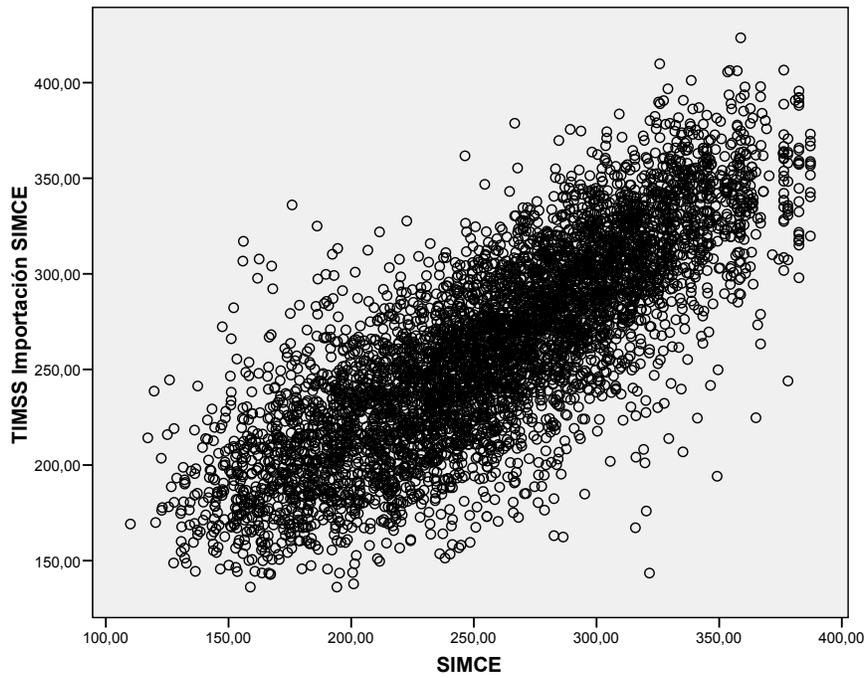
**Figura 8**  
**Correlación entre SIMCE importación TIMSS y SIMCE calibración conjunta**



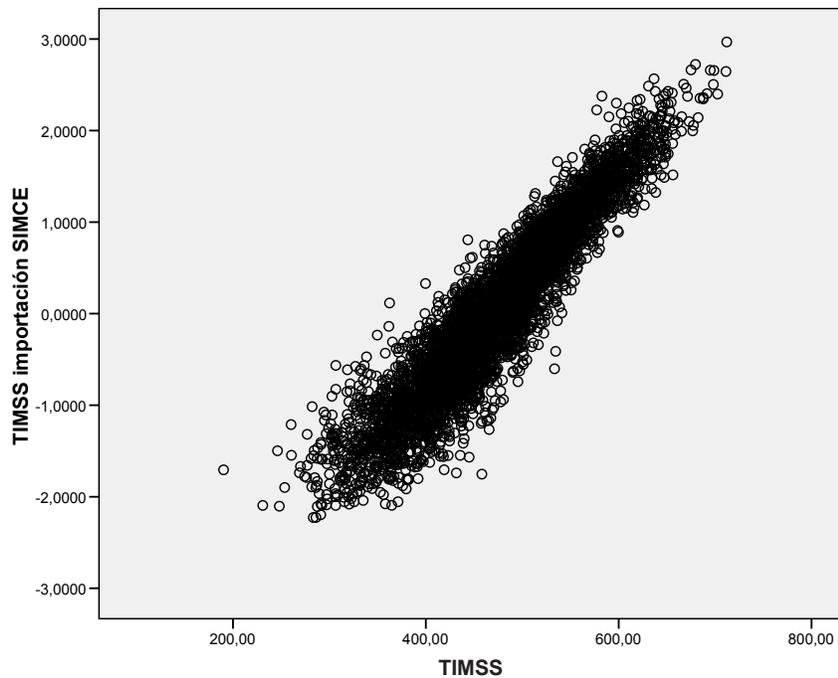
**Figura 9**  
**Correlación entre SIMCE y TIMSS**



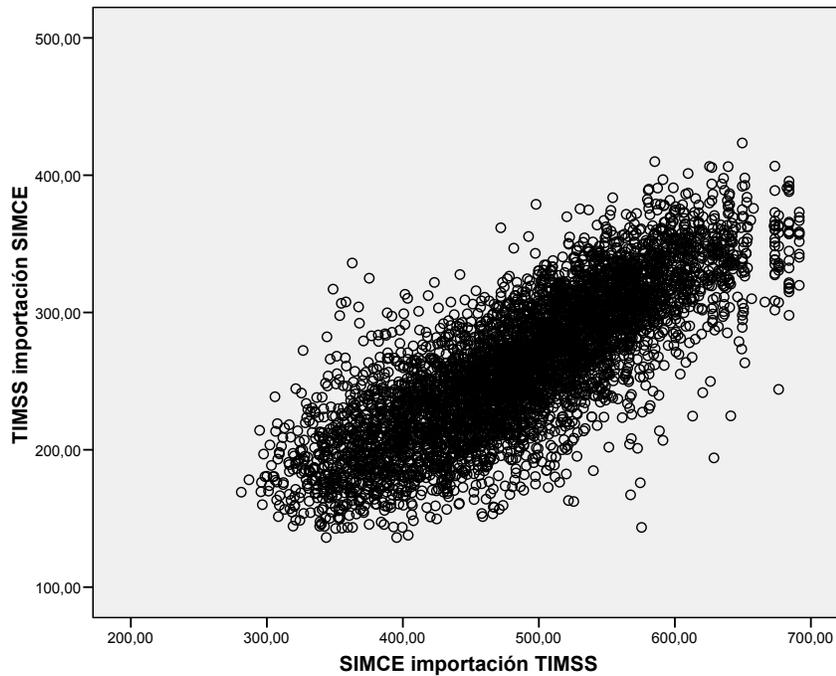
**Figura 10**  
**Correlación entre SIMCE y TIMSS importación SIMCE**



**Figura 11**  
**Correlación entre TIMSS y TIMSS importación SIMCE**



**Figura 12**  
**Correlación entre SIMCE importación TIMSS y TIMSS importación SIMCE**



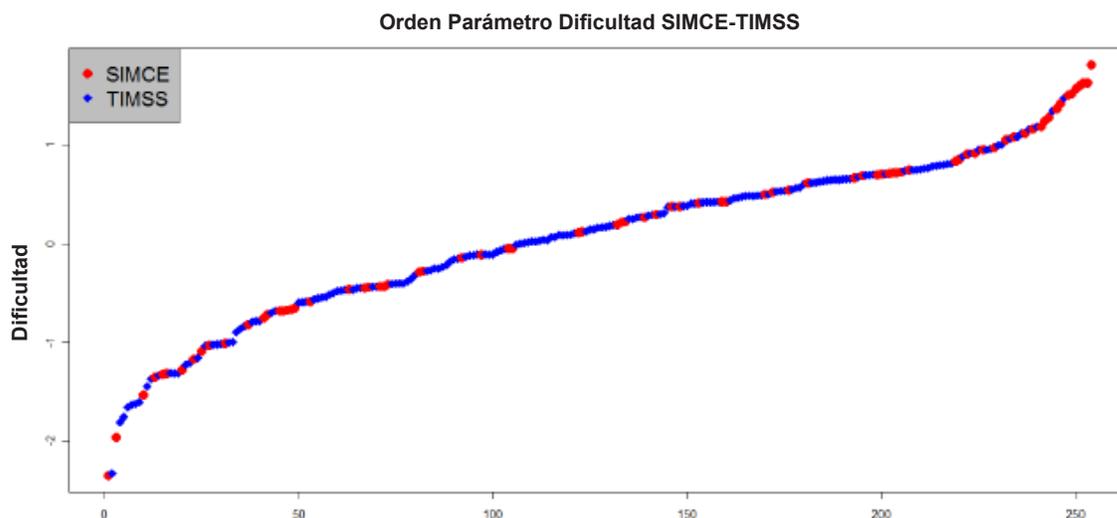
### CALIDAD PSICOMÉTRICA DE LAS PREGUNTAS

Para obtener los resultados anteriores se debió examinar las preguntas SIMCE como si pertenecieran a la prueba TIMSS (calibración mediante importación de parámetros TIMSS), y analizarlas conjuntamente (calibración conjunta). En estos ejercicios se pudo verificar que las preguntas SIMCE cumplían condiciones de ajuste similares a las preguntas TIMSS.

Primero examinemos las preguntas en sus métricas originales. Ordenando los ítems según su dificultad, obtenida al calibrar las preguntas en las poblaciones originales respectivas<sup>20</sup>, se obtiene la Figura 13.

<sup>20</sup> Para SIMCE, corresponde a los alumnos de cuarto básico 2010 y para TIMSS, la población internacional.

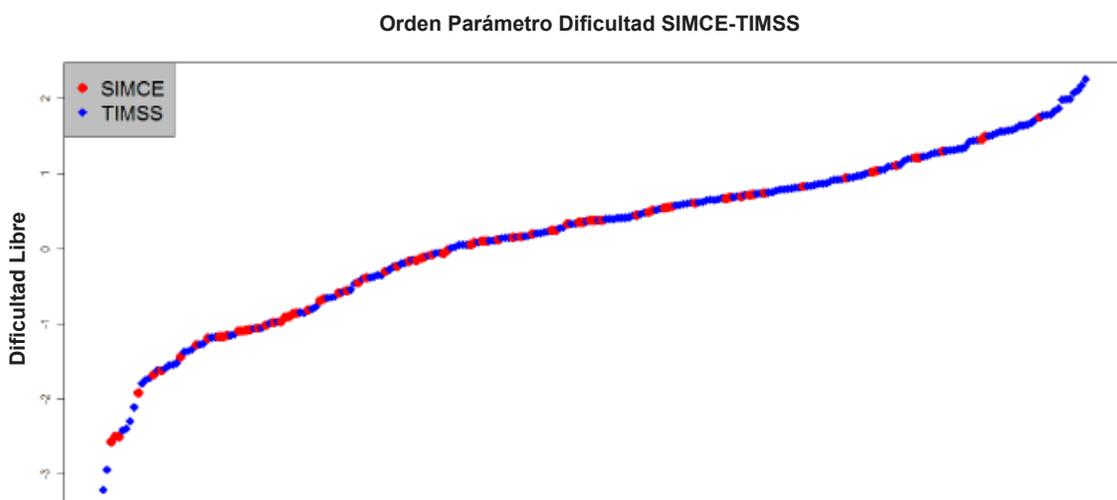
**Figura 13**  
**Ordenamiento del parámetro de dificultad SIMCE-TIMSS:**  
**calibración original**



Se observa que las preguntas SIMCE están distribuidas en todo el rango de dificultad que miden las pruebas. Esto implica que las preguntas SIMCE están diseñadas de manera de cubrir un rango de dificultad para su población objetivo similar al rango de dificultad que utiliza TIMSS para cubrir a su propia población.

En la Figura 14 se presenta como se ordenan por dificultad las preguntas si se calibran conjuntamente. De esta manera se examina si las preguntas SIMCE se hacen notoriamente más fáciles o difíciles que las TIMSS, aplicadas a una muestra representativa de la población chilena.

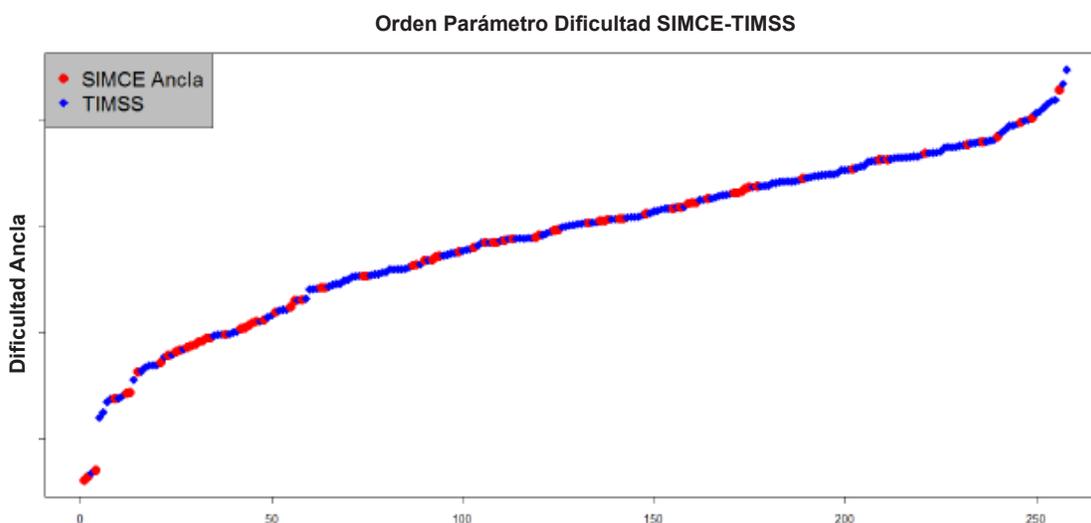
**Figura 14**  
**Ordenamiento del parámetro de dificultad SIMCE-TIMSS:**  
**calibración conjunta**



Observamos que si bien se observa una cierta tendencia a que las preguntas SIMCE se aglutinen en la parte de media hacia el rango de dificultad baja, existen preguntas SIMCE tan demandantes como las preguntas TIMSS.

Finalmente la Figura 15 presenta el ordenamiento de dificultad de las preguntas cuando se utiliza calibración con importación de parámetros TIMSS.

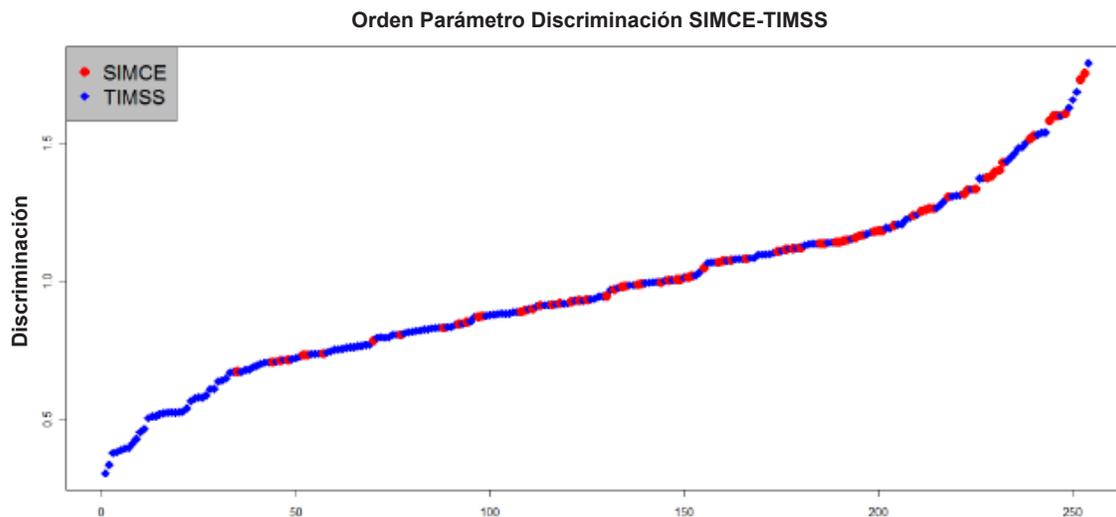
**Figura 15**  
**Ordenamiento del parámetro de dificultad SIMCE-TIMSS:**  
**calibración con importación de parámetros TIMSS**



Esta vez observamos que las preguntas SIMCE se distribuyen más homogéneamente en todo el rango de dificultad, es decir, si se analizan las preguntas SIMCE como preguntas TIMSS, las preguntas SIMCE son capaces de cubrir el rango de dificultad que mide TIMSS.

A continuación se analiza el parámetro de discriminación que, recordemos, representa la capacidad de las preguntas de distinguir entre personas de distinta habilidad, por lo que se debieran privilegiar valores altos. En la Figura 16 aparece el ordenamiento de los parámetros de discriminación para las preguntas en su calibración original.

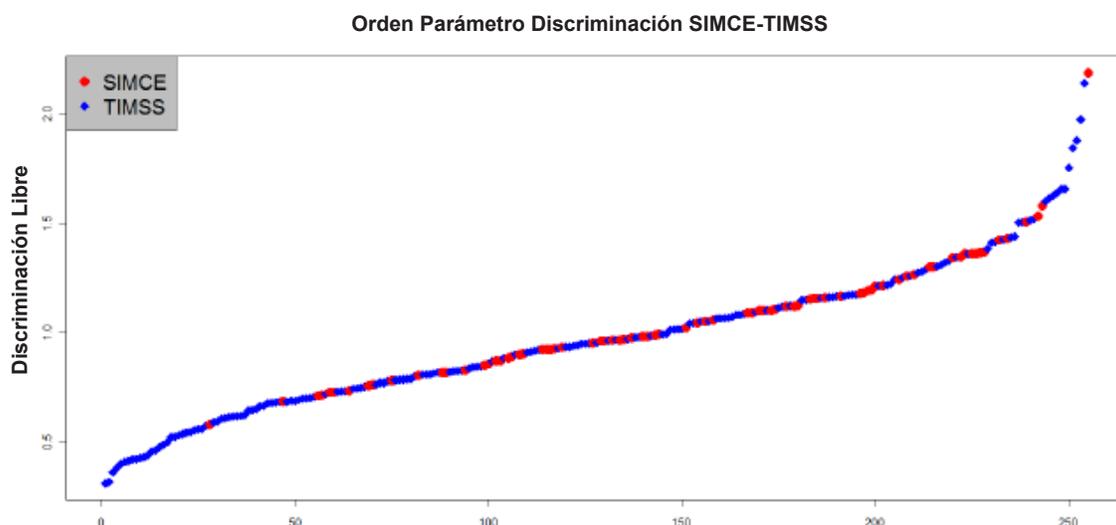
**Figura 16**  
**Ordenamiento del parámetro de discriminación SIMCE-TIMSS:**  
**calibración original**



Observamos que las preguntas SIMCE y TIMSS cubren de manera similar los valores de discriminación que van desde 0,3 a 2,5. Se observa que las preguntas SIMCE tienen todas discriminaciones superiores a 0,5, lo que indica que se comportan bastante bien en su población objetiva.

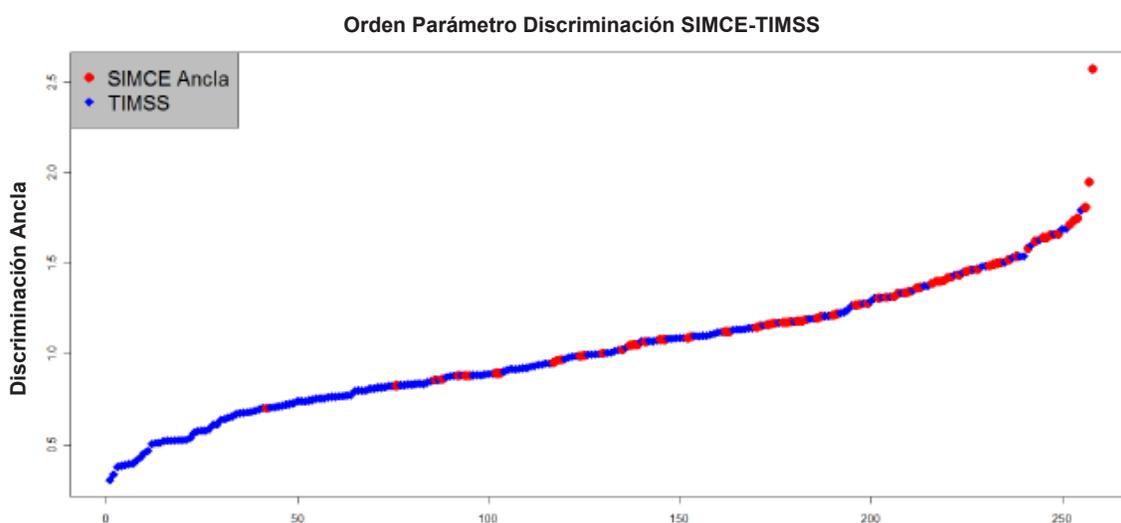
Las discriminaciones obtenidas de la calibración conjunta se grafican en la Figura 17.

**Figura 17**  
**Ordenamiento del parámetro de discriminación SIMCE-TIMSS:**  
**calibración conjunta**



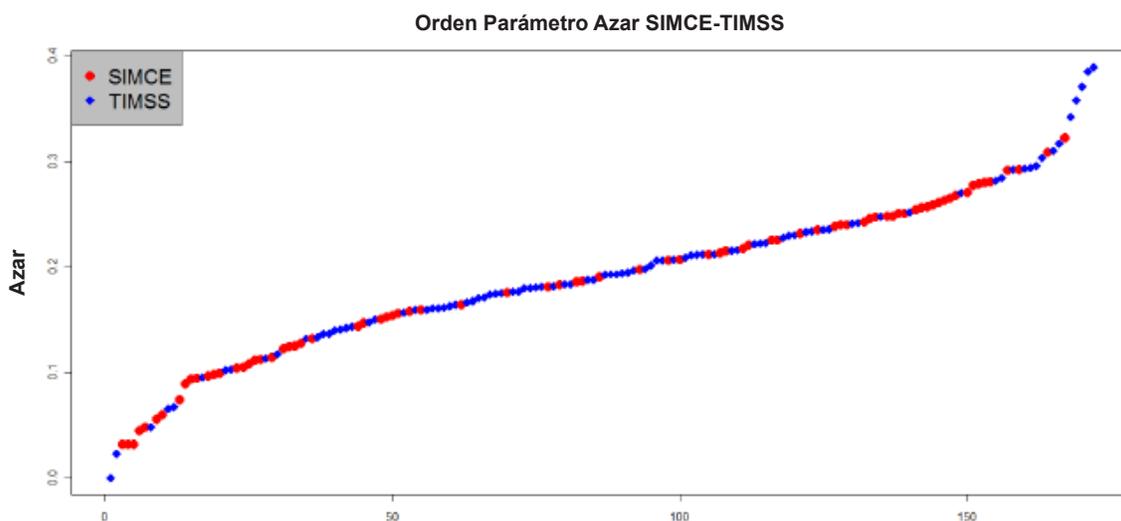
Observamos que el comportamiento conjunto de las preguntas en términos de discriminación se mantiene, siendo las preguntas SIMCE las que tienen mayor tendencia a discriminar (para la muestra de alumnos considerada), por sobre las preguntas TIMSS. Para la calibración con importación de parámetros TIMSS (Figura 18) se obtiene un comportamiento similar. Se puede entonces afirmar que las preguntas SIMCE tienen un buen poder discriminante.

**Figura 18**  
**Ordenamiento del parámetro de discriminación SIMCE-TIMSS:**  
**calibración con importación de parámetros TIMSS**



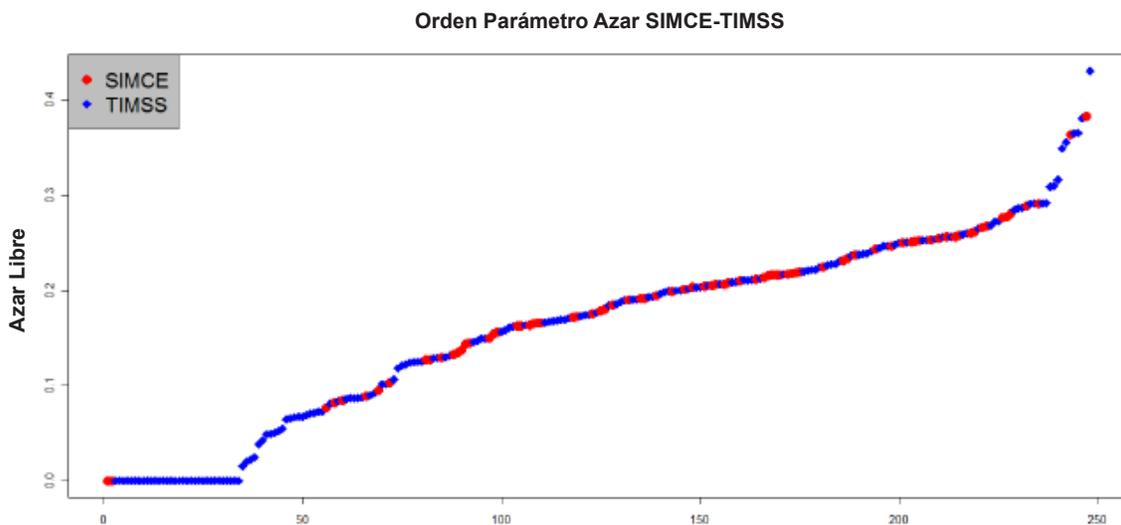
Recordando que el parámetro de pseudoazar se identifica con la probabilidad de acertar al ítem cuando se tiene una baja habilidad, se privilegia entonces valores pequeños. A continuación se presenta el ordenamiento de los parámetros de azar para las preguntas en sus calibraciones originales.

**Figura 19**  
**Ordenamiento del parámetro de pseudoazar SIMCE-TIMSS:**  
**calibración original**

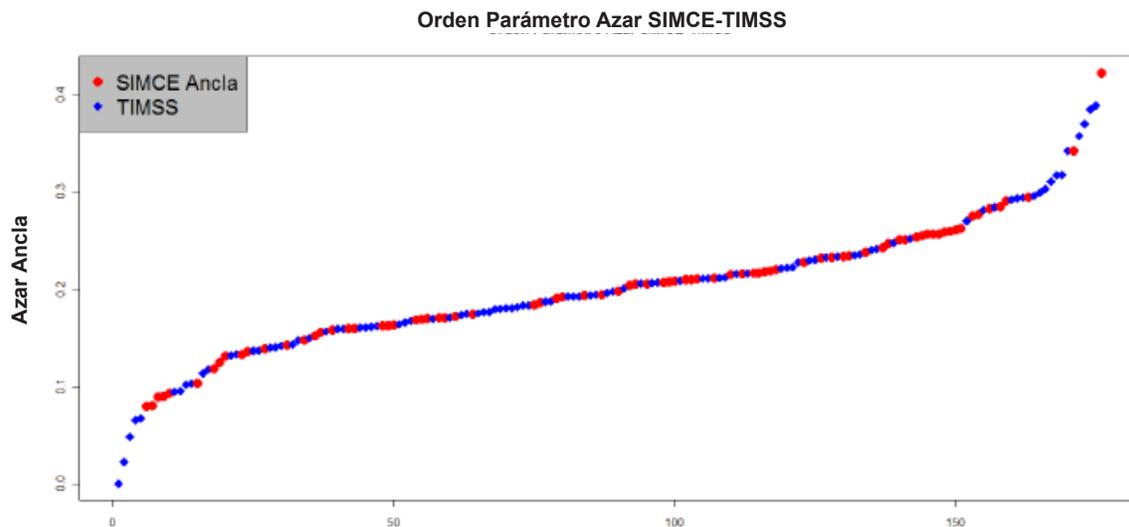


Observamos que las preguntas SIMCE tienen valores en el parámetro de pseudoazar similares a los de la prueba TIMSS, e inferiores a 0,3. Este comportamiento se mantiene tanto para las calibraciones conjuntas (Figura 20) como con la calibración bajo importación de parámetros TIMSS (Figura 21), salvo para un solo ítem que ve incrementado su parámetro hasta el valor 0,4.

**Figura 20**  
**Ordenamiento del parámetro de pseudoazar SIMCE-TIMSS:**  
**calibración conjunta**



**Figura 21**  
**Ordenamiento del parámetro de pseudoazar SIMCE-TIMSS:**  
**calibración con importación de parámetros TIMSS**



Podemos concluir entonces que el comportamiento psicométrico de las preguntas de la prueba SIMCE de Matemática 4.º Básico 2010 se asemeja bastante al comportamiento psicométrico de las preguntas de la prueba TIMSS 2011, aplicada a una muestra representativa de la población chilena<sup>21</sup>.

## CONCLUSIONES

Mediante este estudio hemos podido verificar la invarianza del comportamiento de los parámetros de las preguntas de la prueba SIMCE de Matemática 2010 como asimismo de la prueba TIMSS 2011 de Matemática. También se pudo verificar la estabilidad de la estimación de las puntuaciones de dichas pruebas, como asimismo se observó que el comportamiento psicométrico de las preguntas de las pruebas SIMCE de Matemática 2010 era similar al de las preguntas de la prueba TIMSS examinada.

<sup>21</sup> Para mayor información acerca de estos resultados se puede revisar Gaggero (2013) y Guzmán (2013).

## LISTA DE REFERENCIAS

- ACT, Inc. (2011). *Technical Manual for the PLAN-PISA International Benchmarking Linking Study* ACT, Inc.: Autor. Recuperado de [http://www.act.org/research/policymakers/pdf/PLAN\\_PISA\\_TechManual.pdf](http://www.act.org/research/policymakers/pdf/PLAN_PISA_TechManual.pdf)
- Gaggero, Miguel. (2013). *Análisis psicométrico en la comparación de test curriculares en el ámbito nacional e internacional*. [Seminario de Tesis] Ingeniería Estadística, USACH.
- Guzmán, Victor. (2013). *Rendimiento educacional Chileno Bajo criterios Internacionales: Alineación SIMCE - TIMSS Cuarto Básico Matemática*. [Seminario de Tesis] Ingeniería Estadística, USACH.
- Hambleton, R. & Swaminathan, H. (1985). *Item Response Theory, Principles and Applications*. Boston: Kluwer.
- Matus et al. (2012). *Alineamiento de las puntuaciones SIMCE 2008 y PISA 2009 en muestras de estudiantes de 2.º Medio. Lectura y Matemática*. Santiago de Chile: Ministerio de Educación, FONIDE. Recuperado de [http://www.agenciaeducacion.cl/wp-content/files\\_mf/semina2.pdf](http://www.agenciaeducacion.cl/wp-content/files_mf/semina2.pdf)