

# Comparación de diferentes modelos de la Teoría de Respuesta al Ítem

Carlos Berejnoi<sup>1</sup> & María A. Barros<sup>2</sup>

(1) *Facultad de Ingeniería, Universidad Nacional de Salta.*  
*berejnoi@unsa.edu.ar*

(2) *Colegio Secundario N°5081 "Raúl Ricardo Alfonsín", Salta.*  
*marile74ar@gmail.com*

**RESUMEN:** La Teoría de Respuesta al Ítem (Item Response Theory o IRT), usada en psicología para medir o estimar rasgos o habilidades latentes de los individuos, tiene un importante campo de aplicación en lo que respecta a la evaluación en educación. Esta teoría permite, mediante modelos estadísticos, caracterizar los reactivos o ítems de una evaluación basándose en el nivel de dificultad, y además comparar el nivel de habilidad de cada estudiante frente a la dificultad de cada ítem, independientemente del nivel de los otros individuos que realizan la evaluación. En esta teoría, para reactivos cuyas respuestas son dicotómicas (correctas o incorrectas), es común el uso del modelo de Rasch, y de los modelos logísticos de uno (1PLM), dos (2PLM) o tres parámetros (3PLM). En este trabajo se presenta un análisis del ajuste del modelo de Rasch y de los modelos logísticos a los ítems de un examen parcial de Análisis Matemático I de la Facultad de Ingeniería de la UNSa. Por la complejidad que implica la estimación de los parámetros de los modelos, resulta necesario utilizar herramientas informáticas. En este trabajo se usó el software estadístico libre R con el paquete ltm.

## 1 INTRODUCCIÓN

Si bien los modelos de la Teoría de Respuesta al Ítem (IRT) son encontrados mayormente en la literatura referida a evaluaciones psicológicas, ellos están siendo aplicados con mucho éxito en otros campos (Johnson, 2007). Una de las principales aplicaciones de estos modelos es en educación (Rizopoulos, 2006).

Uno de los aspectos básicos en el proceso de enseñanza-aprendizaje es conocer cuánto aprendió el estudiante, es decir, tener una medida de las habilidades por él desarrolladas. Es común utilizar para ello algún instrumento (prueba, evaluación, examen), compuesto por reactivos (ítems).

Pero al hablar de cuánto, se debe establecer una escala de medición. La IRT se basa más en los ítems del examen que en el puntaje del mismo. Según la IRT, la habilidad (representada con el símbolo  $\theta$ ), es considerada como una variable continua real que puede ser medida en una escala que tiene un punto medio igual a cero (Baker, 2001). De esta manera se pueden comparar habilidades entre individuos, comparando el parámetro  $\theta$  de cada uno de ellos.

Un aspecto muy interesante en esta teoría es que la habilidad de un individuo y el nivel de dificultad de un ítem se miden en la misma

escala, siendo posible predecir mediante modelos estadísticos la probabilidad de éxito de un individuo (conocida su habilidad) frente a un ítem.

Existen diversos modelos estadísticos usados en la IRT. En este trabajo se comparan cuatro modelos en el ajuste de los resultados obtenidos en un examen parcial de Análisis Matemático I de la Facultad de Ingeniería de la UNSa, determinándose cuál es la mejor opción.

## 2 MODELOS ESTADÍSTICOS

### 2.1 Curva característica del ítem

La IRT permite construir para cada reactivo la Curva Característica del Ítem (ICC), una función que indica la probabilidad de éxito en la respuesta, dependiendo del nivel de habilidad latente del individuo (Rizopoulos, 2008).

Para reactivos donde la respuesta es dicotómica (se admite sólo la posibilidad de correcta o incorrecta) la probabilidad de éxito será:

$$p = P(X = 1) = \frac{\exp(f(\theta))}{1 + \exp(f(\theta))} \quad (1)$$

El símbolo  $X$  es una variable aleatoria que indica que la respuesta es correcta ( $X=1$ ) o incorrecta ( $X=0$ ).

Según la expresión que tome la función  $f(\theta)$  se tiene (Rizopoulos, 2010):

- Modelo de Rasch: supone que todos los reactivos tienen el mismo grado de discriminación.

La probabilidad de éxito en la respuesta está dada por:

$$p = \frac{\exp(\theta - \delta_i)}{1 + \exp(\theta - \delta_i)} \quad (2)$$

El símbolo  $\delta_i$  es el parámetro de dificultad del ítem  $i$ . Tanto él como  $\theta$  se refieren a la misma escala de medición.

Según la ecuación (2), en este modelo la probabilidad de éxito es una función de la diferencia entre la habilidad de una persona y la dificultad del ítem (Wu y Adams, 2007). Si ambas son iguales, la probabilidad de éxito es 0.5. En la Fig. 1 se observa esta propiedad en una ICC correspondiente a un ítem de dificultad igual a -0.4.

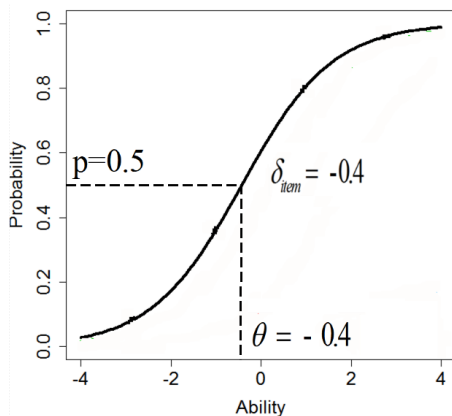


Figura 1. Ejemplo de ICC de un ítem según el modelo de Rasch.

Cuanto mayor sea  $\delta_i$ , más a la derecha se posicionará la curva ICC, y mayor será la dificultad del ítem: para un individuo con determinada habilidad, su probabilidad de éxito resultará menor. Las curvas de los diferentes ítems no se cruzan.

Trabajando la ecuación (1), se llega a:

$$\ln\left(\frac{p}{1-p}\right) = \theta - \delta_i \quad (3)$$

La ecuación (3) muestra que la diferencia entre los parámetros del individuo y del ítem, en el modelo Rasch, es el logaritmo natural de la

razón de probabilidad (*odd* en inglés) de éxito de una persona sobre el ítem. La unidad de medida en la escala para  $\theta$  y  $\delta_i$  se conoce como *logit*. Este término proviene de la contracción de la expresión en inglés *log odds unit* (Wu & Adams, 2007).

- Modelo Logístico de un parámetro (1PLM): surge el parámetro de discriminación ( $\alpha$ ), pero supone que es el mismo para todos los ítems.

$$p = \frac{\exp[\alpha(\theta - \delta_i)]}{1 + \exp[\alpha(\theta - \delta_i)]} \quad (4)$$

El modelo de Rasch es un caso particular del 1PLM, donde  $\alpha = 1$ .

- Modelo Logístico de dos parámetros (2PLM): la función de probabilidad tiene en cuenta los parámetros de dificultad ( $\delta_i$ ) y discriminación ( $\alpha_i$ ) de cada ítem

$$p = \frac{\exp[\alpha_i(\theta - \delta_i)]}{1 + \exp[\alpha_i(\theta - \delta_i)]} \quad (5)$$

El parámetro  $\alpha_i$  se puede interpretar como la “pendiente” de la ICC. Cuanto más elevado sea su valor, el ítem nos permitirá diferenciar más los estudiantes según sus habilidades. En este caso las curvas ICC pueden cruzarse, a diferencia de lo que ocurriría con los dos modelos anteriores. En la Fig 2 se observan dos ICC de igual nivel de dificultad pero con diferentes valores de  $\alpha$ .

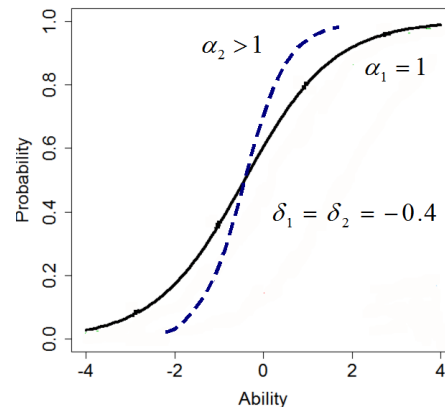


Figura 2. ICC de dos ítems de diferentes niveles de discriminación.

- Modelo Logístico de tres parámetros (3PLM): aparece el parámetro  $c_i$  (guessing parameter) que tiene en cuenta la probabilidad de que el ítem haya sido respondido al azar. De esta manera, aún siendo bajo el nivel de habilidad de un individuo, existirá una probabilidad mayor que cero de que responda correctamente el ítem.

$$p = c_i + (1 - c_i) \frac{\exp[\alpha_i(\theta - \delta_i)]}{1 + \exp[\alpha_i(\theta - \delta_i)]} \quad (6)$$

### 2.2 Curvas de información

En la IRT se pretende estimar el nivel de habilidad de cada examinado (Baker, 2001). La cantidad de información que disponemos se puede calcular como la inversa de la desviación estándar de la estimación del parámetro de habilidad  $\sigma$ :

$$I = \frac{1}{\sigma^2} \quad (7)$$

Si la cantidad de información es grande entonces los valores de habilidad estimados serán similares a los valores reales. Por el contrario, si la cantidad de información es pequeña entonces habrá una gran dispersión en las estimaciones de habilidad. La curva de información de un ítem (IIC) en general es una función que tiene un máximo en algún nivel de habilidad. Ese valor para el cual la información es máxima será el centro de un intervalo donde la estimación es precisa.

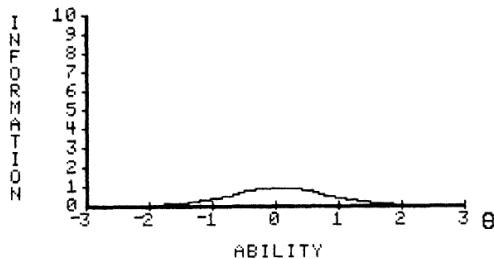


Figura 3. Función de información de un ítem (Baker, 2001).

Para cada ítem o reactivo se puede obtener una curva de este tipo, donde el máximo será muy bajo porque se usa sólo un ítem para obtener la curva.

Como los ítems son parte de un examen, se obtiene información del test a un nivel de habilidad determinado sumando la información de todos los ítems, y la correspondiente curva de la función de información del test TIF, como se muestra en la Fig. 4.

$$I(\theta) = \sum_{i=1}^N I_i(\theta) \quad (8)$$

La forma ideal de la curva de información del test depende de la intencionalidad del mismo. Para un examen de propósito general se buscará una recta horizontal constante, a un nivel muy alto de información. Si la intención es seleccionar

candidatos por habilidad, el punto donde se presenta la máxima información tendría que estar cerca del nivel de habilidad de corte (Baker, 2001).

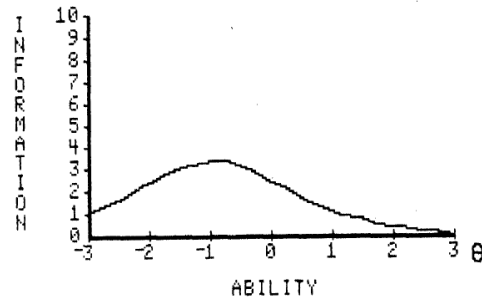


Figura 4. Función de información del test (Baker, 2001).

### 3 LENGUAJE R Y EL PAQUETE LTM

El lenguaje de programación R (R Core Team, 2013) es la implementación open source de software estadístico S, distribuido en forma gratuita bajo licencia GPL (General Public License) del proyecto de software libre GNU. Permite definir funciones nuevas a partir de aquellas que trae el programa, las que pueden agruparse en lo que se llaman paquetes.

En este trabajo se utilizó el paquete ltm (Rizopoulos, 2006), que permite estimar los parámetros de los modelos estadísticos descriptos. Las técnicas de estimación asumen que los individuos son independientes entre sí y que los reactivos funcionan de la misma manera para todos los individuos, es decir, no existen factores de diferenciación en la capacidad de respuesta de los participantes.

De las cuatro técnicas existentes para la estimación de los parámetros de los modelos de IRT (joint maximum likelihood, conditional maximum likelihood, marginal maximum likelihood, y Bayesian) (Johnson, 2007) el paquete ltm utiliza la Marginal Maximum Likelihood Estimation (Rizopoulos, 2006).

Los parámetros de los modelos 1PLM, 2PLM y 3PLM se obtienen con las funciones rasch(), ltm() y tpm() respectivamente. En el caso del modelo de Rasch, también se utiliza rasch() pero restringiendo en el argumento de la función el valor del parámetro de discriminación a 1.

Para cualquiera de los cuatro modelos, con la función factor.scores() se accede al parámetro de habilidad de los individuos.

El paquete también dispone de la función plot() que permite graficar las curvas características de los ítems, las curvas de información del ítem y del test.

La función anova() permite decidir, entre dos modelos, cuál ajusta mejor los datos

experimentales, realizando el test de razón de probabilidad (LRT). La misma función anova() devuelve los valores BIC y AIC, correspondientes al criterio de información bayesiano y al criterio de información de Akaike respectivamente. En ambos casos, cuanto menor es el valor obtenido, mejor resulta el modelo.

#### 4 METODOLOGÍA

En el presente trabajo se aplican los cuatro modelos descriptos a los ítems del primer examen parcial de Análisis Matemático I de la Facultad de Ingeniería de la UNSa, correspondiente al 1<sup>er</sup> Cuatrimestre de 2013. El mismo consistió de 14 reactivos del tipo elección múltiple (5 respuestas posibles para cada ítem). La cantidad de alumnos examinados fue de 715 (448 alumnos ingresantes y 267 recursantes).

Se utilizó el paquete ltm con el software R, para estimar los parámetros correspondientes a los cuatro modelos, así como también se obtuvieron las curvas ICC, IIC y TIF de cada uno de ellos, para finalmente decidir cuál de ellos ajusta mejor los resultados experimentales.

#### 5 RESULTADOS

La función `descript()` arroja la siguiente información:

Tabla 1. Proporciones de respuestas correcta e incorrectas para cada ítem.

	Incorrecta	Correcta	logit
X1	0.47	0.53	0.11
X2	0.41	0.59	0.36
X3	0.67	0.33	-0.70
X4	0.83	0.17	-1.57
X5	0.65	0.35	-0.64
X6	0.31	0.69	0.78
X7	0.55	0.45	-0.18
X8	0.62	0.38	-0.51
X9	0.81	0.19	-1.43
X10	0.51	0.49	-0.05
X11	0.71	0.29	-0.88
X12	0.53	0.47	-0.11
X13	0.94	0.06	-2.68
X14	0.40	0.60	0.41

Cada ítem se identifica con el prefijo X, seguido del número de ejercicio.

De la Tabla 2 se desprende que 389 alumnos respondieron incorrectamente menos de 6 preguntas, cantidad mínima exigida para aprobar el examen.

Tabla 2. Frecuencias de puntaje total.

Total de ejercicios correctos	Frecuencia	Frecuencia acumulada
0	7	7
1	24	31
2	57	88
3	89	177
4	94	271
5	118	389
6	82	471
7	68	539
8	49	588
9	58	646
10	35	681
11	19	700
12	8	708
13	6	714
14	1	715

Tabla 3. Coeficiente de correlación biserial puntual de cada ítem con el puntaje total.

Ejercicio	$r_{nbi}$
X1	0.465
X2	0.505
X3	0.306
X4	0.487
X5	0.335
X6	0.454
X7	0.405
X8	0.547
X9	0.269
X10	0.503
X11	0.424
X12	0.546
X13	0.424
X14	0.360

##### 5.1 Estimación de parámetros

En las Tablas 4 y 5 se presentan los parámetros estimados para cada ítem según los diferentes modelos analizados.

Tabla 4. Parámetro  $\delta$  en modelos Rasch y 1PLM.

Ítem	Rasch	1PLM
X1	-0.135	-0.153
X2	-0.431	-0.488
X3	0.819	0.928
X4	1.834	2.08
X5	0.753	0.853
X6	-0.932	-1.055
X7	0.211	0.24
X8	0.595	0.675
X9	1.673	1.898
X10	0.058	0.066
X11	1.033	1.171
X12	0.131	0.149
X13	3.053	3.469
X14	-0.493	-0.558

En el modelo de Rasch, el parámetro de discriminación  $\alpha$  de todos los ítems fue 1, mientras que el estimado al usar el modelo 1PLM resultó igual a 0.8629.

Tabla 5. Parámetros para los modelos 2PLM y 3PLM.

	2PLM		3PLM		
	$\delta$	$\alpha$	$\delta$	$\alpha$	$c$
X1	-0.150	0.865	-0.143	0.885	0
X2	-0.405	1.128	0.140	1.681	0.236
X3	1.906	0.377	1.984	1.753	0.274
X4	1.369	1.661	1.386	1.621	0
X5	1.620	0.410	1.765	2.040	0.280
X6	-0.984	0.948	-0.958	0.978	0
X7	0.283	0.712	0.885	1.188	0.215
X8	0.502	1.337	0.687	1.801	0.095
X9	3.927	0.375	2.317	2.043	0.159
X10	0.056	1.084	0.063	1.113	0
X11	1.241	0.803	1.403	1.586	0.147
X12	0.108	1.354	0.306	1.700	0.095
X13	1.923	2.355	1.855	2.526	0
X14	-0.770	0.576	-0.750	0.589	0

5.2 Comparación de modelos mediante anova()

Se utilizó la función anova() para decidir qué modelo ajusta mejor los datos. Las tablas 6 a 11 muestran los resultados obtenidos al comparar los modelos de a pares, siempre considerando como hipótesis nula que aquel que mejor ajusta los datos es el de menor número de parámetros. Evaluando el valor de LRT (Likelihood ratio test) y el valor de p.values, se puede descartar la hipótesis nula a favor de la hipótesis alternativa.

Tabla 6. Comparación Rasch vs. 1PLM.

Modelo	AIC	BIC	LRT	p.values
Rasch	11636.67	11700.68		
1PLM	11627.50	11696.08	11.18	0.001

Tabla 7. Comparación Rasch vs. 2PLM.

Modelo	AIC	BIC	LRT	p.values
Rasch	11636.67	11700.68		
2PLM	11519.29	11647.31	145.38	<0.001

Tabla 8. Comparación Rasch vs. 3PLM.

Modelo	AIC	BIC	LRT	p.values
Rasch	11636.67	11700.68		
3PLM	11524.59	11716.62	168.09	<0.001

Tabla 9. Comparación 1PLM vs. 2PLM.

Modelo	AIC	BIC	LRT	p.values
1PLM	11627.50	11696.08		
2PLM	11519.29	11647.31	134.21	<0.001

Tabla 10. Comparación 1PLM vs. 2PLM.

Modelo	AIC	BIC	LRT	p.values
1PLM	11627.50	11696.08		
3PLM	11524.59	11716.62	156.91	<0.001

Tabla 11. Comparación 2PLM vs. 3PLM.

Modelo	AIC	BIC	LRT	p.values
2PLM	11519.29	11647.31		
3PLM	11524.59	11716.62	22.7	0.065

5.3 Comparación de modelos mediante margins()

Se utilizó la función margins() para obtener los valores residuales de chi-cuadrado en la tablas de contingencia de 2x2 para todos los ejercicios tomados de a pares. Así se tienen cuatro tablas por modelo: (0,0), (0,1) (1,0) y (1,1).

Se acepta como máximo valor residual a 3.5. Es así como el modelo de Rasch arroja hasta doce pares de ejercicios con problemas, el modelo 1PLM hasta 7, mientras que los modelos 2PLM y 3PLM sólo un par en una de las tablas.

5.4 Curvas características y de información

Las figuras 5 a 16 muestran las curvas características de ítem (ICC), las curvas de información del ítem (IIC) y las funciones de información del test (TIF) de los cuatro modelos. Se presentan sólo las ICC de tres ítems, para dar claridad a los gráficos, y a los efectos de apreciar la diferencia en los valores de los parámetros de esos ítems.

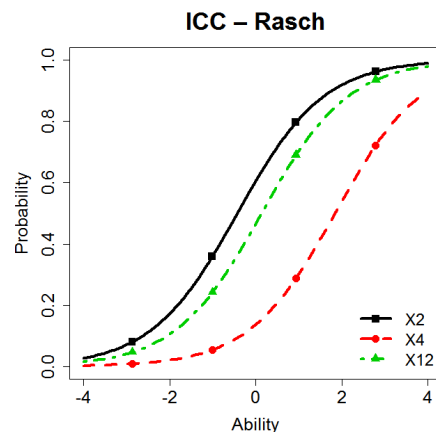


Figura 5. Curvas características de tres ítems según modelo de Rasch.

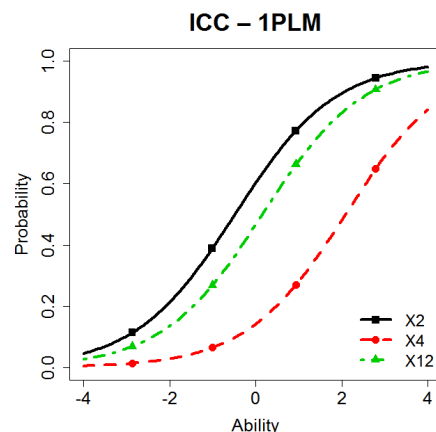


Figura 6. Curvas características de tres ítems según 1PLM.

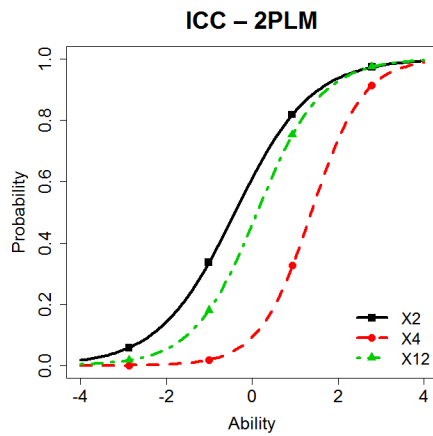


Figura 7. Curvas características de tres ítems según 2PLM.

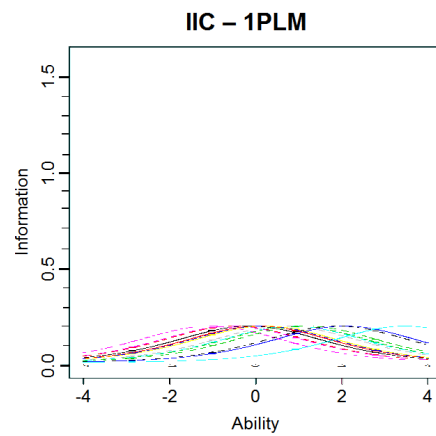


Figura 10. Curvas de información de ítems según 1PLM.

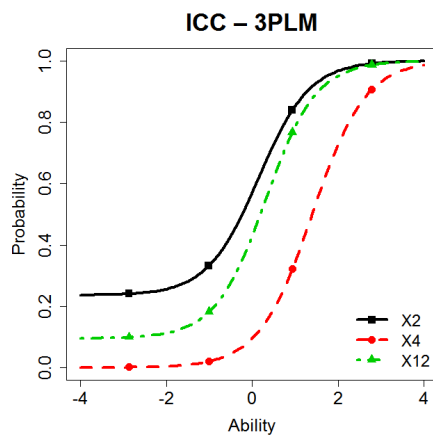


Figura 8. Curvas características de tres ítems según 3PLM.

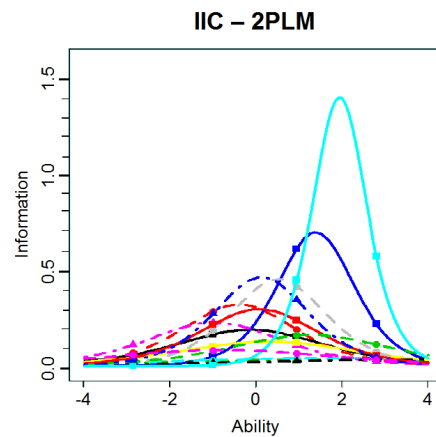


Figura 11. Curvas de información de ítems según 2PLM.

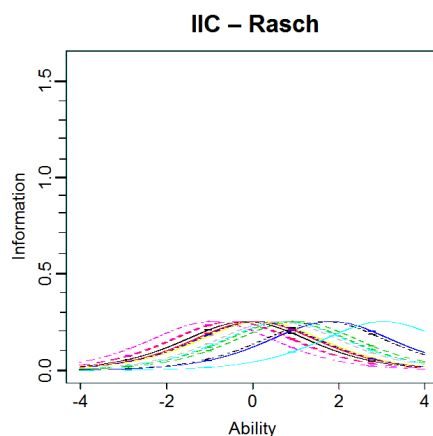


Figura 9. Curvas de información de ítems según modelo de Rasch.

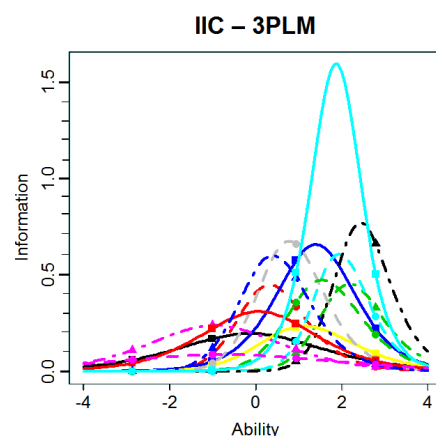


Figura 12. Curvas de información de ítems según 3PLM.

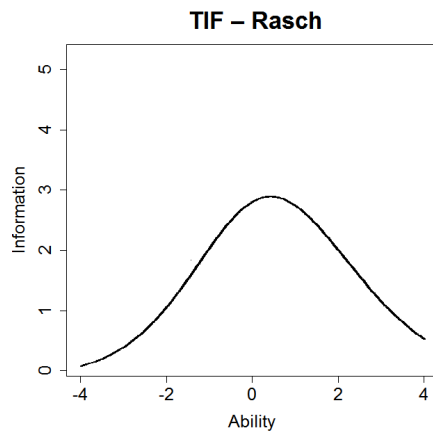


Figura 13. Función de información del test según modelo de Rasch.

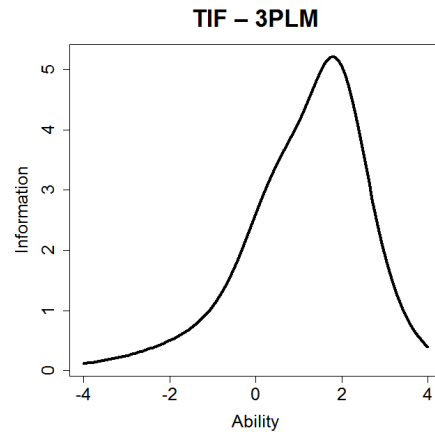


Figura 16. Función de información del test según 3PLM

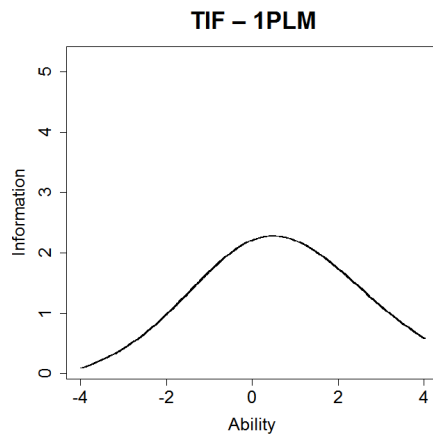


Figura 14. Función de información del test según 1PLM.

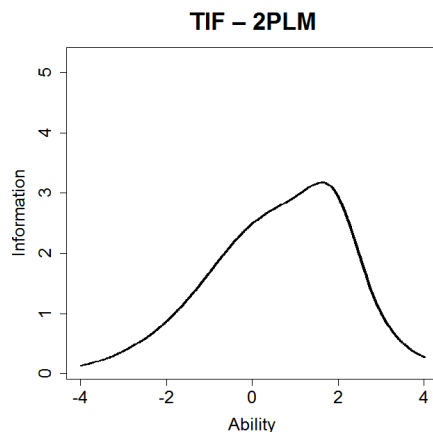


Figura 15. Función de información del test según 2PLM.

## 6 ANÁLISIS Y DISCUSIÓN

En la Tabla 3 se observa que los coeficientes de correlación biserial puntual de cada ítem con el puntaje total no son iguales. Cuanto más alto es el valor de  $r_{pbi}$ , mayor será la discriminación del ítem (Brown, 2001). De esta manera, no se cumpliría el supuesto de los modelos de Rasch y 1PLM de considerar todos los reactivos con igual nivel de discriminación.

Los valores de dificultad de cada ítem son similares para Rasch y para 1PLM, al igual que el parámetro de discriminación común a todos los ítems,  $\alpha = 1$  y  $\alpha = 0.863$  respectivamente. Según estos modelos el ítem de menor dificultad es el número 6, mientras el de mayor dificultad es el número 6, coincidiendo con lo presentado en Tabla 1.

La IRT se focaliza en los ítems y no en la calificación total del examen para medir la habilidad que posee un examinado. En el caso analizado, con los modelos de Rasch y 1PLM, el nivel de habilidad de un alumno que aprobó el examen (con 6 o más ejercicios correctos) es de  $\theta \geq 0.14$ . Mientras que con los modelos de 2PLM y 3PLM, si se quisiera establecer una relación directa entre un puntaje mínimo y habilidad, no se podría hacer. Así es como en 2PLM se tienen individuos con habilidad  $\theta = 0.41$  con 5 respuestas correctas, que no aprobarían el examen, y otros con menor habilidad ( $\theta = -0.424$ ) que sí lo harían al tener 6 respuestas correctas. Esto es debido a que el nivel de habilidad estimada al usar estos dos modelos, no depende de la cantidad de respuestas correctas, sino que también influye cuáles fueron los ítems respondidos correctamente.

Los modelos 2PLM y 3PLM indican que el ejercicio de mayor dificultad es el número 9. Esto

en principio no coincide con lo indicado en la Tabla 1, pero hay que tener en cuenta que el nivel de discriminación del ítem 13 en ambos modelos es superior al correspondiente al ítem 9. De ahí, si bien su dificultad es menor resulta un reactivo que discrimina más a los alumnos que pueden responderlo bien de aquellos con menor habilidad.

En la Tabla 5 se observa que el parámetro de respuesta al azar es en general bajo, pero en algunos casos superior al máximo esperado en ejercicios de respuestas múltiples con 5 opciones ( $c_{\max} = 0.2$ ).

El modelo que menor error de medición presenta es el 3PLM, seguido por el modelo 2PLM, y los menos precisos son los de Rasch y 1PLM. Esto concuerda parcialmente con lo informado por Moghadamzadeh et al (2011), quienes encontraron que el modelo que brinda la mayor cantidad de información es el 2PLM, pero próximo al 3PLM. En las Fig. 9 a 12 se aprecia que la cantidad de información por ítem es mucho menor en los modelos Rasch y 1PLM. En las Fig. 13 y 14 se observa que la cantidad de información se distribuye en forma casi simétrica alrededor de un valor de habilidad  $\theta = 0.5$ , mientras que las Fig. 15 y 16 muestran que si bien el valor máximo de la función se presenta para valores elevados de habilidad  $\theta \approx 2$ , siendo más precisas las mediciones en esa zona, para  $\theta \geq 0$  los valores de esta función superan el valor 2 (máximo alcanzado en el modelo 1PLM).

Las Tablas 6 a 11 muestran, en base a los valores AIC y BIC (cuanto menores, mejor es el modelo), y LRT (si es alto se descarta la hipótesis nula) que los modelos que mejor ajustan los datos son los modelos 2PLM y 3PLM, inclusive sin lograrse una definición contundente entre estos dos últimos (Tabla 11). Sin embargo, a pesar de que las mediciones son más precisas con el modelo 3PLM, el hecho de obtenerse en algunos casos valores de  $c > c_{\max}$  indicaría que el mejor modelo es el 2PLM.

El paquete ltm resulta fácil de usar, brindando muchas herramientas para el análisis de la información, tanto en forma gráfica (Fig. 5 a Fig. 15) como analítica.

## 7 CONCLUSIONES

- El modelo de IRT que mejor ajusta los datos analizados es el modelo logístico de dos parámetros.
- El paquete ltm es muy potente, versátil y fácil de utilizar, brindando la posibilidad de acceder a toda la información necesaria para determinar los niveles de habilidad de los

examinados, y los parámetros de varias distribuciones estadísticas de IRT.

- Resulta interesante ahondar en otros aspectos no analizados, entre ellos la medición de las habilidades de los estudiantes y el seguimiento del proceso de aprendizaje.
- Se pueden conformar bancos de reactivos, conociendo sus niveles de discriminación y dificultad, para mejorar los instrumentos evaluativos.
- La IRT podría resultar ser la mejor opción para la evaluación de los alumnos en base a competencias o habilidades desarrolladas.

## REFERENCIAS

- Baker F.B., *The Basics of Item Response Theory*, Carol Boston, Lawrence Rudner, USA, 2001.
- Brown, J.D., Statistics Corner: Questions and answers about language testing statistics: Point-biserial correlation coefficients, *Shiken: JLT Testing & Evolution SIG Newsletter*, 5 (3), 13 – 17, 2001.
- Johnson, M.S., Marginal Maximum Likelihood Estimation of Item Response Models in R, *Journal of Statistical Software*, Volume 20, Issue 10, 1-24, 2007.
- Moghadamzadeh, A., Salehi, K & Khodaie, E., A Comparison the Information Functions of the Item and Test in One, Two and Three Parametric Model of the Item Response Theory (IRT), International Conference on Education and Educational Psychology (ICEEPSY 2011), *Procedia - Social and Behavioral Sciences*, 29, 1359 – 1367, 2011
- R Core Team, R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Austria, 2013.
- Rizopoulos, D., ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses, *Journal of Statistical Software*, 17 (5), 1-25, 2006.
- Rizopoulos, D., Item Response Theory Using the ltm Package, *The R User Conference 2008*, Technische Universität Dortmund, August 14th, 2008.
- Rizopoulos, D., Item Response Theory in R using Package ltm, *Seminar WU Wirtschaftsuniversität*, Erasmus University Medical Center, the Netherlands, Department of Statistics and Mathematics, Wien, January 12th, 2010.
- Wu, M. & R. Adams, *Applying the Rasch model to psycho-social measurement: A practical approach*, Educational Measurement Solutions, Australia, 2007.