*Cuadernos de la Facultad de Ingeniería*

Universidad Cátolica de Salta, vol. 8, 2014: 5-16

**Comparación de diferentes modelos de la Teoría de Respuesta al Ítem**1

**Carlos Berejnoi**2 **& María A. Barros**3

**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.

Palabras clave: evaluación-Teoría de Respuesta al Item- Rasch- Modelos logísticos

**1. Introducción**

Si bien los modelos de la Teoría de Respues- ta 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 princi- pales aplicaciones de estos modelos es en educa- ción (Rizopoulos, 2006).

Uno de los aspectos básicos en el proceso de enseñanza-aprendizaje es conocer cuánto apren- dió 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).

1. IX Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, Santiago del Estero, octubre de 2013. Publicado en CD, ISSN 1853-7871
2. Facultad de Ingeniería, Universidad Nacional de Salta.
3. Colegio Secundario Nº5081 «Raúl Ricardo Alfonsín», Salta.

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 mis- mo. Según la IRT, la habilidad (representada con el símbolo ** ), 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 habi-

incorrecta (X=0).

Según la expresión que tome la fun- ción *f* **  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:

 exp**  ** 

*p* 

lidades entre individuos, comparando el parámetro ** de cada uno de ellos.

*i*

1 exp**  *i* 

(2)

Un aspecto muy interesante en esta teoría es

que la habilidad de un individuo y el nivel de difi- cultad de un ítem se miden en la misma escala, siendo posible predecir mediante modelos esta- dísticos 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, deter- minándose cuál es la mejor opción.

# Modelos estadadísticos

## Curva característica del ítem

La IRT permite construir para cada reactivo la Curva Característica del Ítem (ICC), una fun- ción que indica la probabilidad de éxito en la res- puesta, dependiendo del nivel de habilidad laten- te del individuo (Rizopoulos, 2008).

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

El símbolo *i* es el parámetro de dificultad

del ítem *i.* Tanto él como** se refieren a la mis- ma escala de medición.

Según la ecuación (2), en este modelo la pro- babilidad 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 igua- les, la probabilidad de éxito es 0.5. En la Fig. 1 se observa esta propiedad en una ICC correspon- diente a un ítem de dificultad igual a -0.4.



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

 exp *f* ** 

(1)

Cuanto mayor sea *i* , más a la derecha se

*p*  *P*  *X*  1 

1 exp *f* ** 

posicionará la curva ICC, y mayor será la dificultad del ítem: para un individuo con determinada habili-

El símbolo X es una variable aleatoria

que indica que la respuesta es correcta (X=1) o

dad, su probabilidad de éxito resultará menor. Las curvas de los diferentes ítems no se cruzan.

Trabajando la ecuación (1), se llega a:

ln 

*p*   **  **

(3)

 1 *p*  *i*

 

La ecuación (3) muestra que la diferencia en- tre 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 ** y *i* se conoce como *logit.* Este tér-

mino 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 (** ), pero supone que es el mismo para todos los ítems.

Figura 2. ICC de dos ítems de diferentes niveles de discri- minación.

* Modelo Logístico de tres parámetros (3PLM): aparece el parámetro *ci* (guessing

*p*  exp ** **  *i* 

1 exp ** **  *i* 

(4)

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

El modelo de Rasch es un caso particular

que cero de que responda correctamente el ítem.

del 1PLM, donde **  1 .

*p*  *c*  1  *c*  exp *i* **  *i* 

(6)

*i i* 1  exp ** **  ** 

* + Modelo Logístico de dos parámetros (2PLM): la función de probabilidad tiene en cuen-

ta los parámetros de dificultad ( *i* ) y discrimina- ción (*i* ) de cada ítem

 *i i* 

## Curvas de información

En la IRT se pretende estimar el nivel de ha- bilidad de cada examinado (Baker, 2001). La canti-

*p*  exp *i* **  *i* 

1 exp *i* **  *i* 

(5)

dad de información que disponemos se puede cal- cular como la inversa de la desviación estándar de la estimación del parámetro de habilidad ** :

El parámetro *i* se puede interpretar como la «pendiente» de la ICC. Cuanto más elevado

sea su valor, el ítem nos permitirá diferenciar más

*I*  1

** 2

(7)

los estudiantes según sus habilidades. En este caso las curvas ICC pueden cruzarse, a diferencia de lo que ocurría con los dos modelos anteriores. En la Fig 2 se observan dos ICC de igual nivel de dificultad pero con diferentes valores de ** .

Si la cantidad de información es grande en- tonces 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 habili- dad 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.

*N*



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

# Lenguaje R y Paquete LTM

El lenguaje de programación R (R Core Team, 2013) es la implementación open source del 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 aque- llas que trae el programa, las que pueden agru- parse en lo que se llaman paquetes.

En este trabajo se utilizó el paquete ltm (Rizopoulos, 2006), que permite estimar los pará- metros 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 to- dos los individuos, es decir, no existen factores

*I* **    *Ii* **

*i*1

(8)

de diferenciación en la capacidad de respuesta de los participantes.

De las cuatro técnicas existentes para la esti-

La forma ideal de la curva de información del test depende de la intencionalidad del mis- mo. Para un examen de propósito general se bus- cará una recta horizontal constante, a un nivel muy alto de información. Si la intención es selec- cionar candidatos por habilidad, el punto donde se presenta la máxima información tendría que estar cerca del nivel de habilidad de corte (Baker, 2001).

mación 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 característi- cas 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 experimen- tales, 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 infor- mación de Akaike respectivamente. En ambos casos, cuanto menor es el valor obtenido, mejor resulta el modelo.

# Metodología

En el presente trabajo se aplican los cuatro modelos descriptos a los ítems del primer exa- men parcial de Análisis Matemático I de la Facul- tad de Ingeniería de la UNSa, correspondiente al 1er Cuatrimestre de 2013. El mismo consistió de 14 reactivos del tipo elección múltiple (5 respues- tas posibles para cada ítem). La cantidad de alum- nos examinados fue de 715 (448 alumnos ingre- santes 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 obtu- vieron las curvas ICC, IIC y TIF de cada uno de ellos, para finalmente decidir cuál de ellos ajusta mejor los resultados experimentales.

# Resultados

La función descript() arroja la siguiente in- formación:

**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, se- guido del número de ejercicio.

De la Tabla 2 se desprende que 389 alum- nos respondieron incorrectamente menos de 6 preguntas, cantidad mínima exigida para apro- bar el examen.

**Tabla 2. Frecuencias de puntaje total**

|  |  |  |
| --- | --- | --- |
| Total de ejercicioscorrectos | Frecuencia | Frecuenciaacumulada |
| 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 | *rpbi* |
| 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 |

### *Estimación de parámetros*

En las Tablas 4 y 5 se presentan los paráme- tros estimados para cada ítem según los diferen- tes modelos analizados.

**Tabla 4. Parámetro** ** **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 dis- criminación ** de todos los ítems fue 1, mien- tras que el estimado al usar el modelo 1PLM re- sultó igual a 0.8629.

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

|  |  |  |
| --- | --- | --- |
|  | 2PLM | 3PLM |
|  | ** | ** | ** | ** | *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 |

### *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 hipó- tesis 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. 3PLM.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

### *Comparación de modelos mediante margins()*

Se utilizó la función margins() para obtener los valores residuales de chi-cuadrado en la ta- blas de contingencia de 2x2 para todos los ejer- cicios 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 mo- delo 1PLM hasta 7, mientras que los modelos 2PLM y 3PLM sólo un par en una de las tablas.

### *Curvas características y de información*

Las figuras 5 a 16 muestran las curvas ca- racterísticas de ítem (ICC), las curvas de infor- mación del ítem (IIC) y las funciones de infor- mación 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 apre-

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

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



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

####

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



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

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



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

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, **  1 y **  0.863 respectivamente. Se-

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



Figura 16. Función de información del test según 3PLM

# 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 *rpbi* , mayor será la discriminación del ítem (Brown, 2001). De esta manera, no se cumpliría el supuesto de los modelos de Rasch y

gún estos modelos el ítem de menor dificultad es el número 6, mientras el de mayor dificultad es el número 13, 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 habi- lidad que posee un examinado. En el caso anali- zado, con los modelos de Rasch y 1PLM, el ni- vel de habilidad de un alumno que aprobó el exa- men (con 6 o más ejercicios correctos) es de

**  0.14 . Mientras que con los modelos de 2PLM y 3PLM, si se quisiera establecer una relación di- recta entre un puntaje mínimo y habilidad, no se podría hacer. Así es como en 2PLM se tienen individuos con habilidad **  0.41 con 5 respues- tas correctas, que no aprobarían el examen, y otros con menor habilidad (**  0.424 ) que sí lo ha- rían al tener 6 respuestas correctas. Esto es debi- do a que el nivel de habilidad estimada al usar es- tos dos modelos, no depende de la cantidad de respuestas correctas, sino que también influye cuá- les 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 ni- vel de discriminación del ítem 13 en ambos mo- delos 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 pre- senta 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 en- contraron que el modelo que brinda la mayor cantidad de información es el 2PLM, pero próxi- mo al 3PLM. En las Fig. 9 a 12 se aprecia que la cantidad de información por ítem es mucho me- nor 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 **  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 **  2 , siendo más precisas las medi- ciones en esa zona, para **  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 da- tos 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 mo-

delo 3PLM, el hecho de obtenerse en algunos

# 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 acce- der a toda la información necesaria para deter- minar los niveles de habilidad de los examina- dos, y los parámetros de varias distribuciones es- tadísticas de IRT.
* Resulta interesante ahondar en otros as- pectos no analizados, entre ellos la medición de las habilidades de los estudiantes y el seguimien- to del proceso de aprendizaje.
* Se pueden conformar bancos de reactivos, conociendo sus niveles de discriminación y difi- cultad, para mejorar los instrumentos evaluativos.
* La IRT podría resultar ser la mejor op- ción 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* & *Evlution SIG Newsletter*,

casos valores de *c*  *c*

modelo es el 2PLM.

max

indicaría que el mejor

5 (3), 13 – 17, 2001.

Johnson, M.S., Marginal Maximum Likelihood

El paquete ltm resulta fácil de usar, brindando muchas herramientas para el análisis de la informa- ción, tanto en forma gráfica (Fig. 5 a Fig. 16) como analítica.

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 Soft- ware*, 17 (5), 1-25, 2006.

Rizopoulos, D., Item Response Theory Using the ltm Package, *The R User Conference 2008*, Technische UniversitÄat Dortmund, August 14th, 2008.

Rizopoulos, D., Item Response Theory in R using Package ltm, *Seminar WU WirtschaftsuniversitÄat*, 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, Aus- tralia, 2007.