

## LA TEORIA DE RESPUESTA A LOS ITEMES: UNA MODERNA ALTERNATIVA PARA EL ANÁLISIS PSICOMÉTRICO DE INSTRUMENTOS DE MEDICIÓN

EILIANA MONTERO \*

*Recibido: 1 Agosto 2000*

---

### Resumen

La Teoría de Respuesta a los Ítemes, TRI, (o “Item Response Theory”, en inglés) es un enfoque más poderoso que la Teoría Clásica de los Tests para el análisis de calidad técnica de instrumentos de medición psicométricos. Comparada con la Teoría Clásica, la TRI permite obtener más información sobre los ítemes de la prueba y sobre el constructo o variable latente que interesa medir en los examinados. Además, con la TRI se pueden generar pruebas adaptadas al nivel de los individuos. También se logran indicadores de las propiedades psicométricas del instrumento, y de los ítemes que lo componen, según los diferentes valores del constructo de interés. Estas características de la TRI le dan ventajas esenciales en relación con la Teoría Clásica. Sin embargo, hay todavía problemas de estimación no resueltos, que provocan, algunas veces, la imposibilidad de ajustar ciertos modelos en conjuntos específicos de datos. Para lograr comprender cabalmente la naturaleza, potencialidades y limitaciones de la TRI se requiere de conocimientos en matemática y estadística. Es quizá por esta razón que en nuestra región latinoamericana estos enfoques son poco conocidos entre los científicos sociales. En general, las dos disciplinas que pueden beneficiarse más con la aplicación de estos modelos son la Psicología y la Educación.

**Palabras clave:** Medición, Psicometría, Teoría de Respuesta a los Ítemes, Análisis de Ítemes, Estimación de Modelos

### Abstract

Compared to Classical Test Theory, Item Response Theory (IRT) represents a more powerful approach for the technical quality analysis of psychometric measurement instruments. Item Response Theory allows us to obtain more information about

---

\*Escuela de Estadística, Universidad de Costa Rica, 2060 San José, Costa Rica.  
E-Mail: emontero@cariari.ucr.ac.cr.

the test items and the construct or latent variable that is intended to be measured in the examinees. Likewise, with IRT, it is possible to generate tests adapted to the examinees levels. Furthermore, indicators of the instrument and the individual items psychometric properties are calculated according to the different values of the construct. These characteristics give IRT essential advantages over Classical Test Theory. However, there are still unsolved estimation problems, which sometimes make it impossible to adjust certain models on specific sets of data. To correctly understand the nature, potentials and limitations of IRT, knowledge in mathematics and statistics is required. Perhaps due to this fact, IRT approaches are not well known among social scientists in the Latin American region. In general, the two disciplines that can benefit the most from the application of these models are Psychology and Education.

**Keywords:** Measurement, Psychometrics, Item Response Theory, Item Analysis, Model Estimation.

**Mathematics Subject Classification:** 92J45, 62P15

## 1. Justificación

Cuando el objeto de estudio en una disciplina científica involucra a personas o grupos sociales, los retos para la medición se presentan desde el momento mismo en que se intentan definir las características o fenómenos a ser investigados. Los constructos son entonces conceptualizaciones, construcciones mentales que usan los científicos sociales para lograr describir o explicar aspectos que se desean estudiar en los individuos. El constructo está asociado a ciertos referentes teóricos que se emplean para definirlo y describirlo. Al hablar, por ejemplo, de medir constructos como inteligencia, actitud hacia la matemática, aptitud académica, grado de socialización o nivel de motivación en una persona, los desafíos para la medición se presentan con su definición, para luego abarcar todo su proceso de “operacionalización”. Operacionalizar un constructo significa representarlo por medio de indicadores. Los indicadores se refieren usualmente a hechos, conductas o acciones observables. Una vez que los indicadores están correctamente definidos, se usan entonces para medir empíricamente la intensidad con que se manifiesta el constructo en los individuos objeto de estudio. (Babbie, 1995)

Para lograr inferencias y decisiones correctas acerca de los constructos por medio de los indicadores se requiere, por un lado, que la medición sea confiable o precisa, esto significa que aplicada en las mismas circunstancias a los mismos sujetos produzca los mismos resultados. Por otro lado, la medición debe ser además válida, es decir, debe estar representando correctamente al constructo que interesa medir. Es a estas dos propiedades, validez y confiabilidad, a las que nos referimos usualmente cuando hablamos de la calidad técnica de un instrumento de medición.

La psicometría nos brinda un cuerpo de teoría y métodos para la medición de constructos en ciencias sociales. Uno de sus propósitos principales es el desarrollo de técnicas de aplicación empírica que permitan construir instrumentos de medición, indicadores, de alta confiabilidad y validez. Estas técnicas y métodos se basan en enfoques cuantitativos y utilizan conceptos, procedimientos y medidas derivadas de la estadística y la matemática.

Un instrumento psicométrico se caracteriza porque intenta representar el constructo por medio de un puntaje numérico derivado de la aplicación de un conjunto de reactivos o estímulos a los sujetos de interés. En su forma más usual, el instrumento está compuesto por una serie de ítems o preguntas, cada uno de los cuales es calificado o respondido por el individuo de acuerdo con una cierta escala de medición. En muchas ocasiones las opciones de respuesta se pueden categorizar en dos tipos, correcto o incorrecto, como en el caso de las pruebas de conocimientos y habilidades intelectuales. En otras ocasiones interesa medir más bien sentimientos o dimensiones afectivas como motivación o actitud. En ese contexto generalmente se usan categorías ordinales para las opciones de respuesta (una de las más comunes es la escala Likert de cinco puntos). El puntaje total en el instrumento se genera sumando las puntuaciones individuales para cada ítem y es el indicador que representa el nivel que toma el constructo de interés en cada uno de los sujetos estudiados.

Determinar el grado de calidad técnica de un instrumento es de mucha importancia y puede tener grandes implicaciones prácticas, ya que muchas veces las decisiones que se toman a partir de los puntajes derivados de su aplicación pueden afectar grandemente la vida de las personas. Al respecto, se puede pensar en las consecuencias derivadas de los resultados de las pruebas estandarizadas de conocimientos que realizan en muchos países los estudiantes en primaria y secundaria, las pruebas de aptitud académica para ingreso a universidades y los instrumentos psicométricos para medir factores de personalidad, que se aplican frecuentemente en la selección de personal para puestos de trabajo.

Tradicionalmente, los métodos basados en la llamada Teoría Clásica de los Tests han sido utilizados para el análisis de calidad técnica de instrumentos de medición psicométricos. Uno de los resultados más importantes que se deriva a partir de los postulados de la Teoría Clásica es el conocido coeficiente de confiabilidad  $\alpha$  de Cronbach, medida que se usa para juzgar la confiabilidad de un instrumento desde el punto de vista de su consistencia interna.

En Costa Rica, por ejemplo, la Teoría Clásica de los Tests se ha aplicado rutinariamente para el análisis y la construcción del banco de ítems de las pruebas de aptitud académica de dos de las principales universidades estatales, la Universidad de Costa Rica y el Instituto Tecnológico. Más recientemente, también se ha incorporado a las pruebas sumativas que deben presentarse al término de la educación secundaria (llamadas pruebas de bachillerato). Sin embargo, existe ahora un enfoque más poderoso que la Teoría Clásica para el análisis de la calidad técnica de un instrumento de medición psicométrico. Este se llama Teoría de Respuesta a los Ítems o TRI (“Item Response Theory”, IRT, en inglés). En América Latina no es hasta hace relativamente pocos años que las técnicas basadas en la Teoría de Respuesta a los Ítems se han comenzado a estudiar. Esta autora conoce de aplicaciones del modelo de Rasch (uno de los más simples) en México, Argentina y Colombia. En nuestra región existe también el problema de la relativamente escasa formación en metodología cuantitativa y estadística que tiene la mayoría de los educadores y psicólogos, cuyas disciplinas son las que más se benefician con la aplicación de estos modelos. Aún más, también es opinión de esta autora que muchos de nuestros matemáticos y estadísticos tampoco se han interesado especialmente por estudiar estas temáticas. Estas circunstancias, ciertamente, no han ayudado mucho a la generación de investigaciones y foros de discusión sobre la TRI en los países latinoamericanos, por lo que actualmente

se presenta una especie de vacío en cuanto a ese aspecto. Este artículo pretende, precisamente, contribuir a llenar esa carencia y se deriva en gran parte de la experiencia que esta autora ha tenido con el uso de modelos TRI para el análisis de diversos instrumentos de medición en el campo educativo y psicológico en Costa Rica, Panamá y Nicaragua. En el documento se expondrán los fundamentos y conceptos clave asociados a estos modelos, los procedimientos para su estimación, su utilidad práctica, y finalmente, algunos de los problemas aún no resueltos en su aplicación.

## 2. Antecedentes

Aunque los desarrollos básicos de los modelos de TRI se generaron con el trabajo pionero de psicómetras como A. Birnbaum, G. Rasch, F.M. Lord y M.R. Novick que datan de los años 50, no es sino hasta los años 70 que estos enfoques comienzan a divulgarse y a aplicarse en los Estados Unidos y otros países industrializados. La difusión de estos modelos ha dependido, en gran medida, del advenimiento de los computadores personales, al igual que procedimientos de análisis cuantitativo multivariable, tales como el análisis de factores o la regresión múltiple. (Muñiz, 1990; Hambleton, & Swaminathan, 1989).

## 3. Qué es un modelo de teoría de respuesta a los ítemes

Se puede decir que un modelo TRI es una conceptualización, que partiendo de ciertos conceptos básicos de medición y usando las herramientas de la estadística y la matemática, busca encontrar una descripción teórica para explicar el comportamiento de datos empíricos derivados de la aplicación de un instrumento psicométrico. Los parámetros estimados por el modelo permiten entonces evaluar la calidad técnica de cada uno de los ítemes o reactivos por separado y del instrumento como un todo y a la vez estimar el nivel que cada examinado presenta en el constructo de interés. En un modelo de TRI se asume que hay una variable latente o constructo  $\theta$ , no observable directamente y que se desea estimar para cada examinado a partir de las respuestas suministradas por éste en el instrumento de medición. Además se asume que para cada ítem o pregunta el comportamiento de las respuestas dadas por los examinados puede ser modelado mediante una función matemática que se denomina Curva Característica del Ítem o *CCI*. Otros conceptos fundamentales en TRI son la Función de Información y el Error Estándar de Medición, que serán explicados en la sección titulada “Conceptos Fundamentales en TRI”.

## 4. Condiciones de aplicación

En cuanto a los requisitos que se deben cumplir para ajustar un modelo de TRI es necesario mencionar que debe contarse con los datos derivados de la aplicación del instrumento en muestras relativamente grandes de sujetos (como mínimo 200 personas). En general las respuestas a los ítemes deben ser dicotomizadas, es decir clasificadas usando los códigos 0 ó 1. Esta codificación dicotómica resulta obvia cuando se están analizando los resultados de pruebas para medir conocimientos o habilidades intelectuales, pues la mayoría de ellas

utilizan ítems de escogencia única (“multiple choice” en inglés). Pero es también posible dicotomizar las respuestas a ítems que se contestan en una escala ordinal, tipo Likert, por ejemplo, asignando el 1 a las categorías que representen mayor intensidad o nivel en el constructo de interés, y el 0 a las que representen niveles más bajos. Así, por ejemplo, en una prueba psicométrica para medir motivación se asignará un 1 para las respuestas que indiquen mayores niveles de motivación y 0 para las que indiquen niveles más bajos.

Además, al igual que en el caso de la Teoría Clásica de los Tests, se deben cumplir aquí dos supuestos básicos de la teoría psicométrica. El primero se denomina supuesto de unidimensionalidad, y establece que el test debe medir fundamentalmente un solo constructo subyacente. El grado de cumplimiento de este supuesto se puede explorar por medio de la aplicación del análisis de factores, técnica estadística multivariable. El segundo supuesto se llama independencia local y establece que la probabilidad de respuesta correcta en un ítem no está afectada por la forma en que se han respondido los otros. Esto se cumple usualmente si los ítems no están “encadenados” entre sí. Este supuesto también implica la independencia entre las respuestas de los sujetos entre sí, es decir que no debe haber influencia de lo contestado por una persona sobre lo contestado por otra, lo cual se logra evitando la copia o plagio entre los sujetos al responder la prueba.

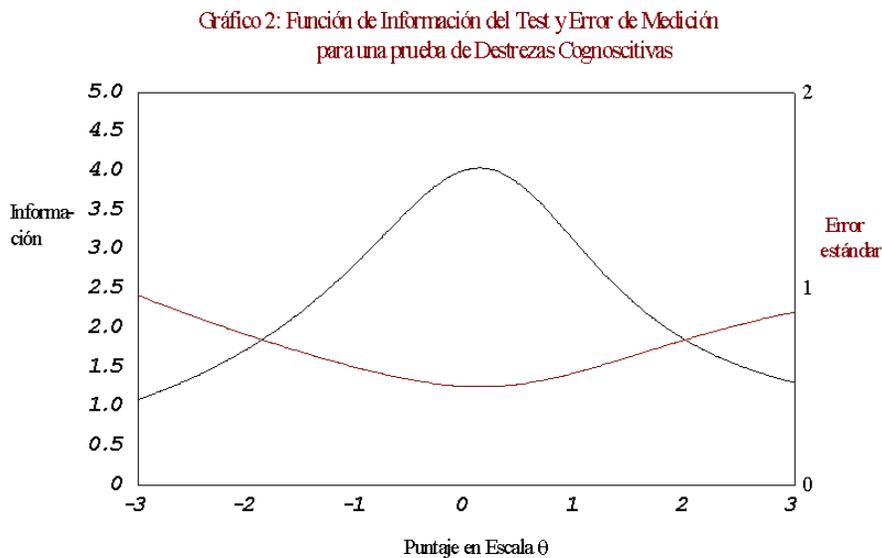
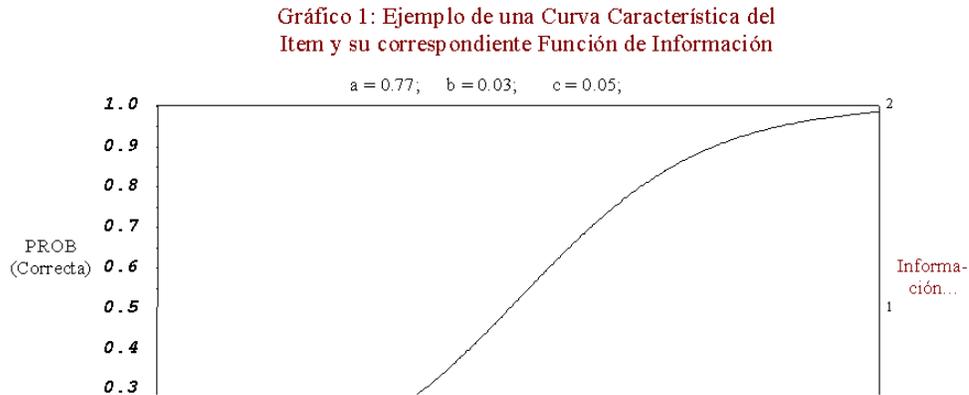
## 5. Conceptos fundamentales en TRI

Una de las principales nociones en Teoría de Respuesta a los Ítems es la de Curva Característica del Ítem o *CCI*. La *CCI* es el ajuste de una función matemática al comportamiento del ítem, partiendo de las respuestas suministradas por los examinados en el instrumento. Cada punto en la *CCI* representa la probabilidad de que un sujeto con un nivel específico en el constructo de interés  $\theta$  responda correctamente al ítem. La escala  $\theta$  que se mide en el eje horizontal generalmente se ajusta para que, al igual que la distribución normal estándar, tenga un promedio de 0 y una desviación estándar igual a 1. Por tanto, la gran mayoría de sus valores estarán incluidos en el intervalo  $[-3, 3]$ .

Otros conceptos básicos son el Error Típico de Medida o Error Estándar de Medición y la Función de Información. El Error Estándar de Medición se define como la variabilidad de las estimaciones del constructo  $\theta$  en cada individuo. Su valor está en función del valor de  $\theta$ , de manera que en general, los errores de estimación no son uniformes a lo largo de la escala. El recíproco del Error de Medición al cuadrado se denomina Función de Información, *FI*, y es un indicador de la precisión que tiene el ítem para estimar cada uno de los diferentes valores de  $\theta$  en los sujetos. La *CCI* y la *FI* son herramientas fundamentales en la construcción y análisis de calidad técnica de pruebas, pues permiten evaluar las propiedades psicométricas de los ítems y a partir de ahí seleccionar aquellos que cumplan con los propósitos específicos establecidos para esa prueba o instrumento. Esto se mostrará con mayor detalle en la sección “Construyendo y Analizando Pruebas con TRI”.

En el gráfico 1 se presentan ejemplos de una *CCI* y su correspondiente Función de Información. La línea en forma de S alargada es la *CCI*, mientras que la Función de Información es la que se asemeja a una curva normal. Entre mayor sea el valor de la *FI*

mayor precisión se logrará en la estimación de  $\theta$ . Por su parte el gráfico 2 muestra la función de información del test y el error de medición para una prueba de Destrezas Cognoscitivas.



## 6. Modelos para TRI: la función logística en uno, dos y tres parámetros

La *CCI* no es otra cosa que un modelo matemático que se usa para describir el comportamiento de la probabilidad de respuesta correcta a un ítem, según los diferentes niveles en el constructo,  $\theta$ , que presenten los examinados. La función logística ha probado ser muy útil como modelo para la *CCI*. La idea es encontrar una función matemática que pueda usarse válidamente para representar a la *CCI* a partir de los datos empíricos derivados de la aplicación del instrumento. Usualmente se trabajan modelos logísticos de uno, dos o

tres parámetros. El modelo de tres parámetros es el más general y el más realista desde el punto de vista empírico. La expresión matemática para la *CCI* en el modelo logístico de tres parámetros es:

$$CCI = P_i(\theta) = c_i + (1 - c - i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

en donde los parámetros se definen así:

$\theta$  : valor del constructo o variable latente que se desea estimar en cada examinado

$a$  : Índice de discriminación del ítem. Es proporcional a la pendiente de la *CCI* en el punto de inflexión. El poder discriminatorio de un ítem es su capacidad para diferenciar sujetos que exhiben diferentes niveles en el constructo de interés. Entre mayor sea el valor de  $a$  mayor es el poder discriminatorio del ítem.

$b$  : Índice de dificultad. Es el valor del puntaje  $\theta$  que corresponde al punto de inflexión de la *CCI*. Valores bajos indican ítems fáciles y valores altos ítems difíciles.

$c$  : Probabilidad de acertar un ítem cuando el examinado contesta al azar.

$D$  : es una constante a la cual se le asigna generalmente el valor de 1.7

Cuando se asume que el parámetro de azar,  $c$ , es igual a cero se obtiene el modelo de dos parámetros, en el que las *CCI* difieren en términos de su dificultad y de su discriminación (parámetros  $b$  y  $a$  respectivamente). Si además de asumir  $c = 0$  se supone también que el parámetro de discriminación,  $a$ , es constante para todos los ítems, se obtiene el modelo de un parámetro o modelo de Rasch, en el que las *CCI* se diferencian únicamente en términos de sus niveles de dificultad (parámetro  $b$ ).

Es importante en este punto también mencionar que en los modelos de dos y tres parámetros pueden existir diferentes estimaciones de  $\theta$ , el constructo, para individuos con la misma puntuación empírica en la prueba. Por ejemplo, dos personas pueden lograr el mismo número de respuestas correctas en la prueba, pero aún así obtener estimaciones muy diferentes para su nivel en el constructo  $\theta$ , probado que hayan tenido un patrón diferencial de respuesta, es decir, que no hayan respondido correctamente los mismos ítems. Esto sucede así porque el procedimiento para estimar  $\theta$  en los modelos de dos y tres parámetros le da mayor ponderación a aquellos ítems que tienen mayor discriminación. Por tanto, dado el mismo número de respuestas correctas, examinados que han respondido correctamente ítems de mayor poder discriminatorio obtendrán estimaciones de  $\theta$  superiores a las estimaciones obtenidas para examinados que contestaron correctamente ítems de menor calidad técnica. La Teoría Clásica de los Tests no disfruta de esta propiedad y en ella no es posible diferenciar las puntuaciones de individuos que hayan obtenido el mismo número de respuestas correctas en el instrumento.

Una vez estimados los parámetros de la *CCI*, estos mismos pueden ser usados para el cálculo de sus correspondientes Funciones de Información, mediante la siguiente fórmula:

$$I(\theta) = \sum_{i=1}^n \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

en donde  $P'_i(\theta)$  es la derivada de  $P_i(\theta)$ , la Curva Característica del Item,  $Q_i(\theta)$  es la probabilidad complementaria  $1 - P_i(\theta)$ , y  $n$  es el número de ítems.

Finalmente, se puede demostrar que en los modelos de uno y dos parámetros, la información es máxima en el valor que corresponde al parámetro de dificultad,  $b$ .

## 7. Estimación del modelo

Una vez escogido el modelo de TRI, el proceso de estimación involucra, fundamentalmente, la obtención de los parámetros que definen la *CCI*, partiendo de las respuestas dadas por los individuos al aplicar el instrumento. En el modelo de Rasch, de un parámetro, debe estimarse únicamente el valor de  $b$ , la dificultad. En el modelo de dos parámetros deben estimarse  $a$  y  $b$  (la discriminación y la dificultad). Finalmente en el modelo de tres parámetros debe estimarse también el valor de  $c$  (el parámetro de azar). En todos ellos además debe también estimarse el valor de  $\theta$  para cada uno de los examinados.

El procedimiento de estimación más utilizado involucra encontrar los valores que maximizan la Función de Verosimilitud, es decir, hallar los valores para los parámetros que maximizan la probabilidad de que ocurran los datos observados dado el modelo teórico planteado. A las expresiones así encontradas se les denomina estimadores máximo verosímiles. Los procedimientos de estimación basados en la Función de Verosimilitud son muy comunes en Estadística Inferencial.

A continuación se presenta la función de verosimilitud para el caso más general del modelo de tres parámetros. La letra  $L$  es por la palabra inglesa que significa verosimilitud, “likelihood”:

$$L(u|\theta, a, b, c) = \prod_{a=1}^N \prod_{i=1}^n [P_{ia}(\theta)]^{u_{ia}} [Q_{ia}(\theta)]^{(1-u_{ia})}$$

donde  $P_{ia}(\theta)$  representa la expresión para la *CCI* que fue presentada en la sección anterior, o sea el modelo matemático para la probabilidad de respuesta correcta,  $Q_{ia}(\theta) = 1 - P_{ia}(\theta)$  es igual a su complemento, o sea la probabilidad de no responder correctamente el ítem,  $N$  es el número de examinados,  $n$  es el número de ítems y  $u_{ia}$  es una variable aleatoria que toma los valores 0 ó 1 (variable aleatoria de Bernoulli).

Dada la expresión matemática que define  $P_{ia}(\theta)$ , la *CCI*, resulta claro que la estimación de los parámetros del modelo no es directa. El procedimiento debe, en general, realizarse utilizando métodos numéricos y un proceso iterativo de aproximaciones sucesivas. Para realizar estas operaciones existen algunos paquetes de software que se dedican exclusivamente a la estimación de modelos TRI. El preferido de esta autora es el BILOG (Mislevy & Bock, Scientific Software Corporation, Chicago, Illinois). Generalmente se utilizan métodos de Newton- Raphson en ese proceso para hallar los máximos de la Función de Verosimilitud.

El modelo de Rasch, de un parámetro, es estimable en la gran mayoría de los casos. No sucede lo mismo con los modelos de dos y de tres parámetros. La configuración específica y consistencia que presenten entre sí los datos empíricos derivados de las respuestas al instrumento es un elemento fundamental para lograr una estimación exitosa en los modelos de dos y tres parámetros, pues son estos datos empíricos los que se usan como “semilla” en

el proceso iterativo de aproximaciones sucesivas. Así, especialmente en el caso del modelo de tres parámetros, sucede con cierta frecuencia que la solución no converge y, por tanto, el modelo no es estimable.

Una vez que la estimación de los parámetros ha sido llevada a cabo, existen pruebas de bondad de ajuste que se deben realizar para establecer si el modelo teórico reproduce de manera aceptable el comportamiento de los datos observados. Entre ellas se encuentran las conocidas pruebas basadas en la distribución Chi-cuadrada y las asociadas al comportamiento de los residuos, en donde se analiza la magnitud y conducta de las diferencias entre los valores estimados con base en el modelo y los valores empíricos observados.

## 8. Construyendo y analizando pruebas con TRI: algunos ejemplos

A partir de la exposición que se ha dado en este documento se pueden deducir algunas de las propiedades más útiles que presenta la TRI cuando se desea construir o evaluar la calidad técnica de un instrumento psicométrico. Por medio de la Función de Información se pueden construir pruebas adaptadas al nivel de los examinados, esto gracias a que con ella se puede establecer para qué rangos de  $\theta$  tiene el ítem (o la prueba) mayor precisión. Esta característica permite a su vez disminuir dramáticamente el número de ítems en la prueba, sin perder información.

Se pueden además evaluar los ítems utilizando la *CCI* en términos de su dificultad y discriminación. Por ejemplo, si lo que interesa es construir una prueba para seleccionar candidatos a una beca académica, la *CCI* y la Función de Información nos permitirán escoger aquellos ítems que brinden mayor precisión en la estimación de los niveles altos de  $\theta$ , estos ítems serán difíciles. Por el contrario, si el instrumento más bien intenta identificar estudiantes con problemas de aprendizaje, se deberán seleccionar ítems fáciles y que den información óptima para niveles bajos de  $\theta$ .

Los gráficos 3, 4, 5 y 6 muestran *CCI* y Funciones de Información para ítems con diferentes características. El gráfico 3 ilustra el caso de un ítem de alta calidad técnica que brinda mayor información para niveles bajos de  $\theta$ . El gráfico 4, por el contrario, presenta un ítem de pobre calidad técnica, pues ofrece muy poca información. El ítem representado en el gráfico 5 es de mediana dificultad y da mayor información en niveles intermedios de  $\theta$ . Por su parte, el ítem mostrado en el gráfico 6 tiene algún poder discriminatorio pero resulta demasiado difícil. Finalmente, en el gráfico 7 se presentan la Función de Información para un test y su correspondiente Error de Medición (los ítems presentados antes no tienen relación con este instrumento). Dado que la Función de Información para la prueba es la suma de las FI individuales para cada ítem, el examen del comportamiento de esta función permite concluir, en este caso, que la prueba está compuesta por ítems que en su mayoría brindan mayor precisión en la medición de individuos con niveles relativamente altos en el constructo  $\theta$ .

Gráfico 3  
Curva Característica del Ítem y Función de Información

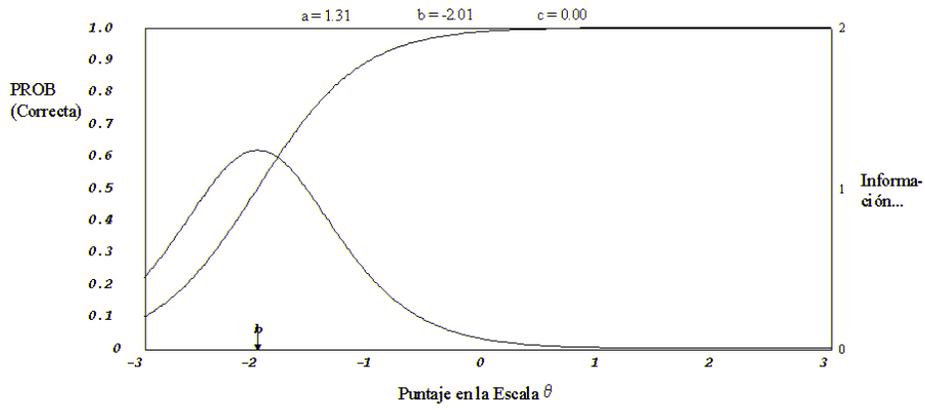


Gráfico 4  
Curva Característica del Ítem y Función de Información

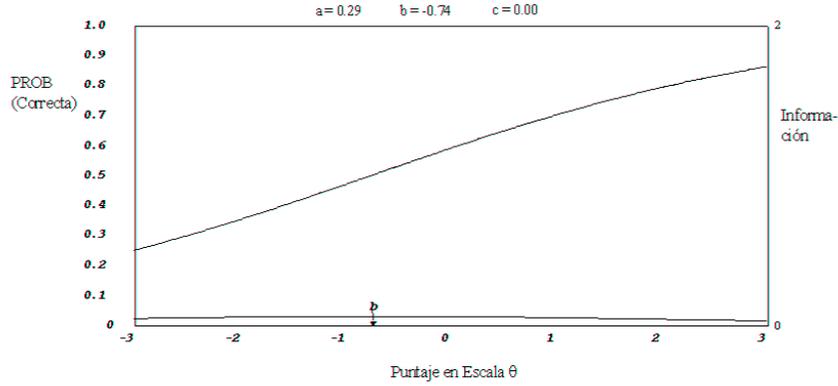


Gráfico 5  
Curva Característica del Ítem y Función de Información

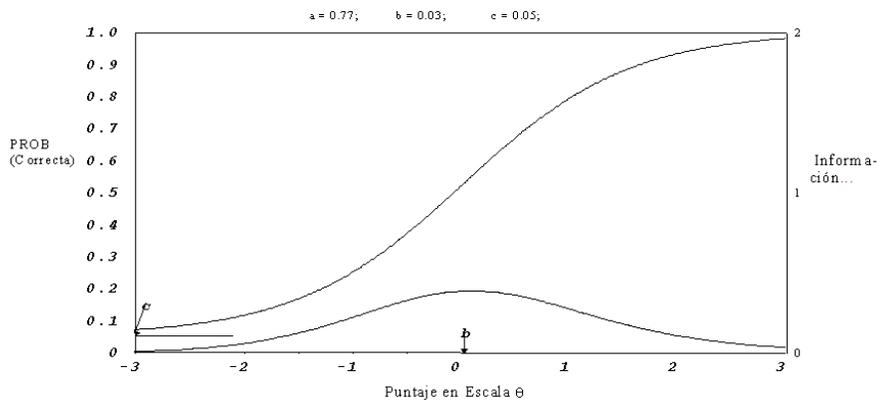


Gráfico 6  
Curva Característica del Item y Función de Información

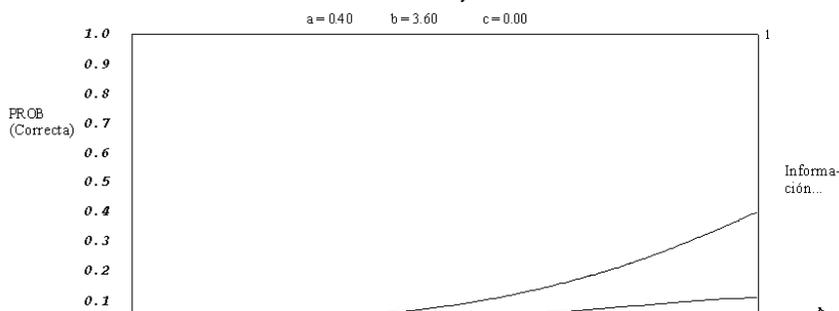
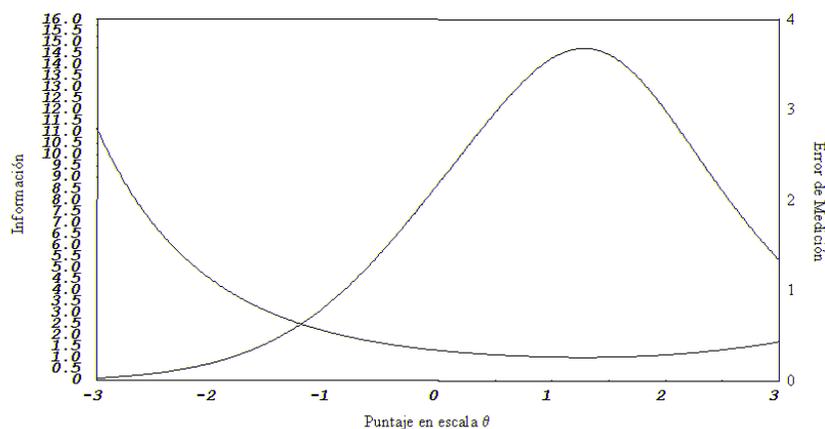


Gráfico 7: Información del Test y Error de Medición



## 9. Problemas no resueltos aún en TRI

Hay todavía algunos aspectos problemáticos asociados con la teoría y la aplicación de estos modelos que han generado controversia entre sus estudiosos o que requieren de mayor investigación para ser resueltos con un grado aceptable de consenso. Entre ellos se encuentra el debate entre los proponentes del modelo de Rasch, de un parámetro, versus los que prefieren el uso de los modelos de dos y tres parámetros. El modelo de Rasch tiene la gran ventaja de su simplicidad y de que, en la gran mayoría de las condiciones, es estimable, quizá por eso es el preferido de las grandes compañías en los Estados Unidos y otros países que se dedican a la construcción y evaluación de instrumentos psicométricos. Sin embargo, en opinión de esta autora, los supuestos que establece el modelo de Rasch para su aplicación son demasiado restrictivos para muchas situaciones reales. Esto es especialmente crítico cuando nos referimos a la suposición que hace Rasch de que todos los ítems tienen el mismo poder discriminatorio (el parámetro  $a$  es constante). Esta autora ha encontrado, de acuerdo con diferentes aplicaciones que ha realizado usando TRI, que en muchas ocasiones el modelo de dos parámetros resulta estimable, pero no así el de tres y parece ser, en realidad, más poderoso que del de Rasch para el análisis de calidad técnica.

Este es, en fin, un punto de controversia y se requieren de más estudios que evidencien las bondades y limitaciones de cada uno de estos enfoques.

Los problemas asociados a la estimación de los modelos también pueden enfocarse desde la perspectiva de las técnicas que se usan para estimar los parámetros de interés. Entre ellos los métodos numéricos, basados en Newton- Raphson, son los más usados, pero hay todavía espacio para la experimentación con otros procedimientos de aproximación numérica que pueden ser más eficientes o exhaustivos. Y es en este aspecto precisamente donde se requiere ayuda de matemáticos (y matemáticas) especializados en métodos numéricos y optimización.

Otro asunto que se debe mencionar es el hecho de que la mayor parte de los desarrollos en TRI se han concentrado en ítemes de respuesta dicotómica y aunque en secciones anteriores se argumentó que, en general, es posible dicotomizar respuestas dadas originalmente en más de dos categorías, es necesario que más investigación se realice en términos de cómo usar más eficientemente los resultados derivados de instrumentos que usan otras escalas de respuesta.

Finalmente se debe indicar la complejidad asociada a la aplicación de la TRI que ya se mencionó al principio de este documento y que se hace evidente a partir de la discusión anterior. Se requiere de un buen nivel de conocimientos en Matemática y Estadística para lograr comprender cabalmente la naturaleza de estos modelos, así como sus potencialidades y limitaciones. Es opinión de esta autora que en América Latina la mayoría de los psicólogos y educadores no reciben una formación suficientemente rigurosa en cuanto a aplicaciones cuantitativas. Por tanto, la popularización de estos modelos, su estudio e investigación, dependerá en gran medida de cuán bien puedan ser entendidos estos enfoques por los investigadores en educación y psicología y de la habilidad que tengan los conocedores para "mercadear el producto".

## Referencias

- Anastasi, A. (1988) *Psychological Testing*. MacMillan, New York. (Una edición anterior, de 1977, está traducida bajo el nombre de *Tests Psicológicos*).
- Babbie, E. (1995) *The Practice of Social Research*. Wadsworth, Belmont CA.
- Hambleton, R.K.; Swaminathan, H. (1989) *Item Response Theory: Principles and Applications*. Kluwer, Dordrecht.
- McDaniel, E. (1994) *Understanding Educational Measurement*. Brown & Benchmark, Madison WN.
- Muñiz-Fernández, J. (1990) "Teoría de Respuesta a los Items: Un Nuevo Enfoque en la Evolución Psicológica y Educativa".
- Neuman, W.L. (1994) *Social Research Methods: Qualitative and Quantitative Approaches*. Allyn and Bacon, Boston.
- Nunnally, J.C.; Bernstein, I.J. (1995) *Teoría Psicométrica*. McGraw-Hill Interamericana, México D.F.