

## TEORÍA DE RESPUESTA A LOS ÍTEMES VERSUS TEORÍA CLÁSICA DE LOS TESTS: ANÁLISIS EMPÍRICO COMPARATIVO<sup>1</sup>

*Eiliana Montero Rojas*

**Resumen:** *La Teoría de Respuesta a los Ítems (TRI, o "Item Response Theory", en inglés) representa un enfoque psicométrico más poderoso que la Teoría Clásica de los Tests para el análisis de calidad técnica de instrumentos de medición en educación y psicología. Comparada con la Teoría Clásica, la TRI permite obtener más información sobre las propiedades de los ítems de la prueba y sobre el constructo que está siendo evaluado en los examinados. El estudio presentado pretende evidenciar empíricamente algunas de las ventajas esenciales que presenta la TRI en relación con la Teoría Clásica, por medio de la comparación entre los resultados obtenidos aplicando una y otra a datos derivados de la administración de una prueba diseñada para medir habilidades cognitivas en estudiantes de noveno año. Partiendo de esos resultados se logran establecer algunas de las bondades relativas de la TRI, especialmente en lo que se refiere a la estimación del nivel de habilidad de los examinados. En cuanto a la mayor precisión para estimar los parámetros de calidad técnica de los ítems (discriminación y dificultad) los beneficios no parecen ser tan contundentes ya que se presentan altas correlaciones entre la estimación clásica y la estimación bajo el modelo TRI.*

### Base teórica conceptual

La utilidad práctica que tiene un instrumento de medición está directamente relacionada con su calidad técnica, es decir, con las propiedades psicométricas que presenta. Este asunto cobra especial importancia cuando se trata de pruebas tales como las de admisión a las universidades, o las pruebas sumativas que se administran al finalizar la educación secundaria (pruebas de Bachillerato). Las decisiones que se toman de acuerdo con los resultados de estos exámenes pueden afectar de manera profunda la vida de las personas.

Por muchos años la Teoría Clásica de los Tests ha servido de marco conceptual para el análisis psicométrico de pruebas estandarizadas (Anastasi, 1988; Nunnally & Bernstein, 1995). Por ejemplo, en nuestro país se ha aplicado rutinariamente para el análisis y la construcción del banco de ítems de la Prueba de Aptitud Académica de la Universidad de Costa Rica. Más recientemente también se ha incorporado a las pruebas de Bachillerato de la educación secundaria, como parte regular del análisis de ítems que se lleva a cabo posterior a su administración.

Sin embargo, la Teoría Clásica de los Tests presenta ciertos problemas fundamentales desde el punto de vista de la medición (Muñiz, 1990; Hambleton, Swaminathan & Rogers, 1991; Bock, 1997). Primero, los parámetros de los ítems dependen del grupo de

examinados a los que se administra la prueba. Los tan conocidos índices de discriminación (correlación entre la puntuación del ítem y el puntuación total de la prueba) y dificultad (porcentaje de respuestas correctas) variarán dependiendo del grupo examinado. Asimismo, las puntuaciones totales o calificaciones en el constructo de interés para cada examinado dependerán del nivel de dificultad del instrumento de medición utilizado. Así, pruebas más fáciles producirán puntuaciones más elevadas, y pruebas más difíciles resultarán en puntuaciones más bajas. Además, los índices clásicos de confiabilidad (el más conocido es el Alfa de Cronbach) tienen el inconveniente de que son medidas globales y no permiten establecer qué precisión tienen la prueba o los ítems según las diferentes categorías de puntuación de los examinados. (Hambleton & Swaminathan, 1989; Nunnally & Bernstein, 1995).

La Teoría de Respuesta a los Ítems (TRI) busca subsanar estas debilidades de la Teoría Clásica, pues intenta obtener estimaciones de los parámetros del ítem que sean independientes de la muestra de examinados, así como puntuaciones o calificaciones que sean independientes del instrumento de medición utilizado. Además, se logran estimaciones individuales de la precisión de la prueba y de los ítems para cada uno de las diferentes puntuaciones totales o calificaciones que reciben los examinados. (Hambleton & Swaminathan, 1989; Bock, 1997; McDonald, 1999). Estas características de la TRI le dan ventajas esenciales en relación con la Teoría Clásica, que permiten, entre otras cosas, la construcción de pruebas adaptadas al nivel del examinado, una mayor precisión en las estimaciones de las puntuaciones totales en el constructo o variable de interés y una comparación (equiparación) más fácil de puntuaciones resultantes de diferentes instrumentos de medición. (Hambleton & Swaminathan, 1989; Muñiz, 1990).

Un concepto fundamental en la TRI es el de Curva Característica del Ítem (CCI). Esta consiste en el ajuste de un modelo mate-

mático al comportamiento del ítem. En el eje horizontal se representa la puntuación total que obtiene cada examinado en la variable de interés, o sea, su calificación. Generalmente esta escala se ajusta para que tenga el mismo rango de una variable normal estándar, es decir, que varíe entre -3 y +3. (Muñiz, 1990; Hambleton, Swaminathan & Rogers, 1991). En el eje vertical se representan las probabilidades de respuesta correcta. Así, cada punto en la CCI representa la probabilidad de respuesta correcta para la puntuación correspondiente. El modelo matemático que se ajusta es usualmente la función logística en uno, dos y tres parámetros. Estos parámetros se definen así:

- a: Índice de discriminación del ítem. Es proporcional a la pendiente de la CCI en el punto de inflexión. Entre mayor sea su valor mayor es la discriminación.
- b: Índice de dificultad. Es el valor de la puntuación o calificación 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 al azar. Su valor está entre cero y uno.

Otro concepto fundamental en la TRI es el de la Función de Información (Muñiz, 1990; Hambleton, Swaminathan & Rogers, 1991). Este es un indicador de la precisión de la prueba. Usualmente su forma es de curva normal. Su valor está en función de las puntuaciones o calificaciones, lo que implica que la precisión del test no es uniforme a lo largo de la escala (no es la misma para todos los examinados).

Cuando se ajusta un modelo de dos o tres parámetros se tiene que las calificaciones de los examinados serán diferentes en la medida en que haya patrones diferentes de respuesta. Esto quiere decir que aunque dos examinados tengan el mismo número de respuestas correctas, bajo la TRI tendrán diferentes calificaciones si no han respondido bien exactamente las mismas preguntas. Esto sucede

porque el modelo de TRI le da más peso o más importancia en la calificación a los ítemes que tienen mayor poder discriminatorio, o sea que aportan mayor información sobre el constructo de interés. Con esto se logra una mayor precisión en la medición, de tal forma que se puede tener menos ítemes en la prueba. (Muñiz, 1990).

La CCI, la Función de Información y el error de medición se representan gráficamente. Estos gráficos son una herramienta sumamente útil para el análisis de las propiedades y calidad técnica del test como un todo y de los ítemes que lo componen.

## Objetivos

De acuerdo con el planteamiento teórico descrito en la sección anterior, la investigación se planteó los siguientes objetivos:

1. Ajustar un modelo de TRI a una prueba estandarizada, construida en nuestro país, para medir destrezas cognitivas en estudiantes de noveno año.
2. Comparar la Teoría Clásica de los Tests y la TRI en cuanto a la precisión en la estimación de los parámetros de los ítemes y en las calificaciones de los examinados.
3. De acuerdo con los resultados obtenidos, concluir en términos de las bondades que presenta la TRI sobre la Teoría Clásica para el análisis de pruebas estandarizadas y la asignación de calificaciones.

## Procedimiento

La base de datos utilizada correspondió a los resultados obtenidos en una prueba para medir destrezas cognitivas que se desarrolló en el IIMEC. Esta fue administrada a una muestra aleatoria nacional de 1933 estudiantes de noveno año. Se analizaron los 37 ítemes de escogencia única en la prueba. Pa-

ra hacer el ajuste del modelo de TRI se utilizó el paquete psicométrico BILOG (Mislevy y Bock; 1990). Este programa de cómputo permite realizar la estimación de los modelos logísticos de uno, dos y tres parámetros y es uno de los más utilizados en la actualidad (Muñiz, 1990). Se ajustó un modelo logístico de dos parámetros, ya que el de tres parámetros no fue estimable. Usando los mismos datos se obtuvieron igualmente los índices clásicos de discriminación y dificultad. Además, también por medio del BILOG se obtuvieron las estimaciones de las puntuaciones totales o calificaciones de los estudiantes según la TRI, las que después se compararon con el porcentaje de respuestas correctas, que es la estimación de la calificación del examinado según la Teoría Clásica.

## Resultados

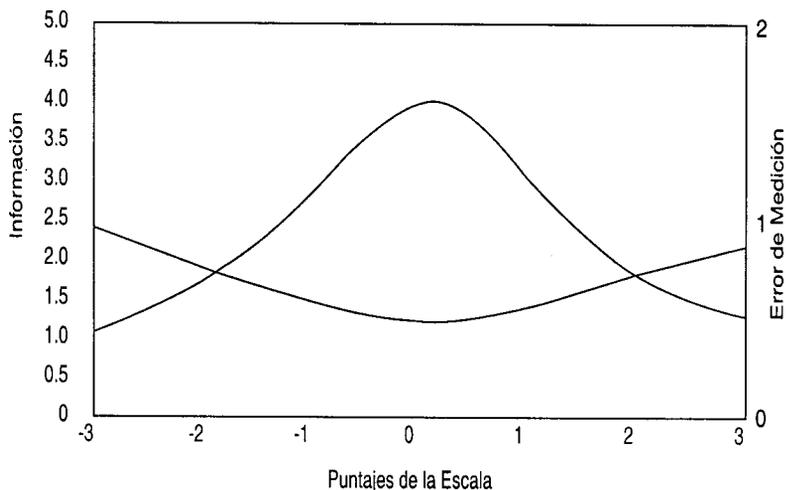
En el Gráfico 1 se observa la Función de Información de la prueba y el error de medición asociado. La Función de Información presenta la forma típica de curva normal y se maximiza para valores alrededor de 0, o sea valores intermedios de habilidad. Si la intención de la prueba es discriminar a niveles intermedios, parece estar cumpliendo su cometido.

Con los índices de dificultad y discriminación para cada ítem, calculados a partir de la Teoría Clásica y a partir de la TRI, se obtuvieron coeficientes de correlación de Pearson. Para el caso de las dificultades ese valor fue de 0,78 y para las discriminaciones de 0,90. Estos resultados evidencian que las estimaciones de los parámetros de los ítemes siguen tendencias similares en ambos enfoques, y concuerdan con hallazgos reportados por Hambleton y Swaminathan (1989).

A pesar de esto, en algunos ítemes se muestra claramente la mayor precisión del modelo de TRI. Por ejemplo, de acuerdo con los índices clásicos, los ítemes 4 y 10 resultan muy similares en cuanto a calidad técnica, ya que sus dificultades son, respectivamente, 0,68 y 0,64, y sus discriminaciones 0,30 y

Gráfico Nº 1

Función de información y error de medición  
Prueba de destrezas cognitivas



0,32. Sin embargo, observando sus curvas características del ítem (Gráficos 2 y 3) es claro que el ítem 10 ofrece bastante más información que el ítem 4. Aunque la literatura consultada no es específica en relación con este tipo de hallazgos, la hipótesis de esta autora, basada su experiencia al trabajar con estos modelos, es que en los ítems de calidad técnica media (según los criterios psicométricos)

es donde se notan con más intensidad las bondades de la TRI en cuanto a la estimación de los parámetros del ítem. Es decir, aquellos reactivos que, según la Teoría Clásica, presentan niveles apenas aceptables de calidad técnica (como los de nuestro ejemplo) son los que permiten diferenciarse más claramente por medio de la TRI en cuanto a sus parámetros de discriminación y dificultad.

Gráfico Nº 2

Curva característica del Ítem y función de información  
Ítem No. 4

$a=0.50$ ;  $b=-1.01$ ;

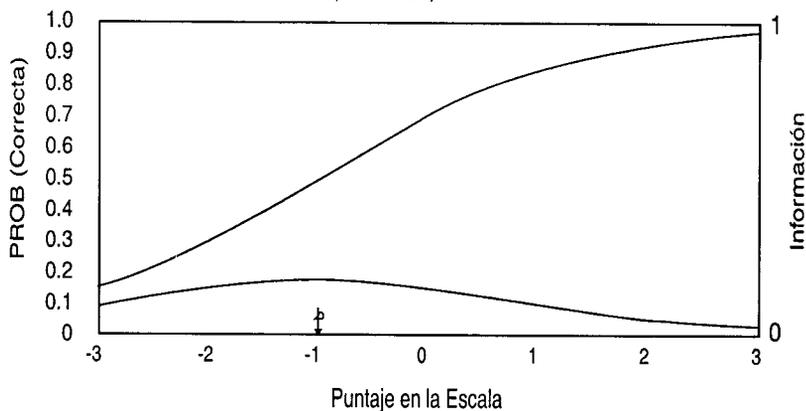
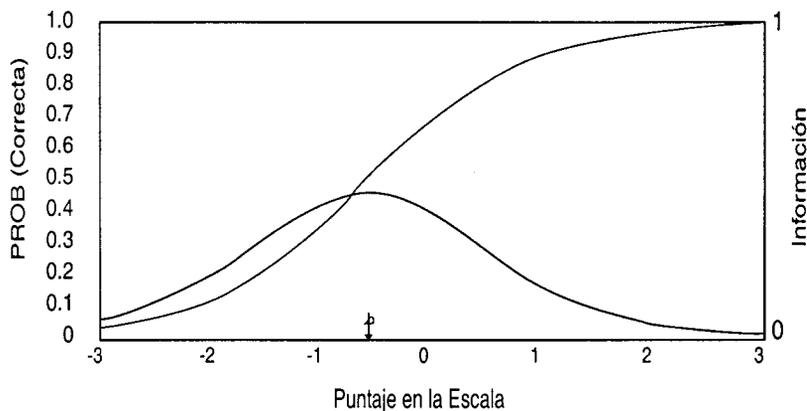


Gráfico Nº 3

Curva característica del Ítem y función de información  
Ítem No. 10

$a=0,80$ ;  $b=-0,58$



Sin embargo, es en las puntuaciones o calificaciones de los examinados en donde se evidencian con mayor fuerza las diferencias entre ambos enfoques. Se nota la mayor precisión proporcionada por la TRI, por cuanto estudiantes con la misma nota según el modelo clásico, tienen diferentes calificaciones según el modelo de TRI. Se debe recordar que en la Teoría Clásica la calificación del estudiante viene dada por el número (o porcentaje) de respuestas correctas obtenidas. Por ejemplo, hubo 134 estudiantes que obtuvieron 17 preguntas correctas, lo que equivale a una nota de 45,95% para todos ellos, de acuerdo con la Teoría Clásica. Pero según el enfoque de la TRI, las puntuaciones o calificaciones para este grupo de estudiantes van desde -0,58 para el más bajo, hasta 1,15 para el más alto. Esta diferencia tan sustancial se debe a que, como se dijo anteriormente, la puntuación depende del patrón de respuestas, pues los ítems que discriminan mejor, es decir que brindan más información, tienen mayor importancia en la calificación. Así, muy probablemente, el estudiante que tiene una calificación de 1,15 contestó bien 17 preguntas que tienen un alto poder discriminatorio, mientras que el que tiene una de -0,58

contestó bien 17 preguntas que no ayudan a discriminar, es decir, preguntas con una pobre calidad técnica.

Estos resultados logran evidenciar que, efectivamente, la TRI presenta importantes ventajas en relación con la Teoría Clásica, tanto para el análisis de los ítems como para la asignación de puntuaciones a examinados. La mayor precisión producto de estas ventajas permite, a su vez, tomar decisiones más certeras, menos sujetas a error, en relación con los examinados, por cuanto las estimaciones del constructo de interés en cada uno de ellos serán más confiables. Por tanto, se recomienda explorar las posibilidades de aplicación de la TRI en las diferentes pruebas estandarizadas que se administran regularmente en nuestro sistema educativo.

### Nota

- 1 La prueba de destrezas cognoscitivas que se utiliza aquí para ilustrar la aplicación de la TRI fue construida en el Instituto de Investigación para el Mejoramiento de la Educación Costarricense (IIMEC) por un equipo de investigación coordinado por la autora, se realizó como parte del convenio MEP-UCR y se financió con fondos del Programa para el Mejoramiento de

la Calidad de la Educación (PROMECE) durante los años 1995, 1996 y 1997.

### Referencias Bibliográficas

- Anastasi, A. *Psychological Testing*. New York: Macmillan, 1988.
- Bock, R.D. "A brief history of Item Response Theory". *Educational Measurement: Issues and Practice*. 16 (4), 1997.
- Hambleton, R.K. & Swaminathan, H. *Item Response Theory: Principles and Applications*. Boston, MA: Kluwer Academic Publishers, 1989.
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J. *Fundamentals of Item Response Theory*. Newbury Park, California: Sage, 1991.
- McDonald, R.P. *Test Theory: A Unified Treatment*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1999.
- Mislevy R.J. & Bock R.D. *BILOG 3: Item Analysis and Test Scoring with Binary Logistic Models*. Chicago, Illinois: Scientific Software Inc.
- Muñiz Fernández, José. *Teoría de Respuesta a los Ítems: Un Nuevo Enfoque en la Evolución Psicológica y Educativa*. Madrid: Ediciones Pirámide, 1990.
- Nunnally, J.C. & Bernstein, I.J. *Teoría Psicométrica*. México, D.F.: McGraw-Hill Interamericana, 1995.