FEBRERO 2016
VOL. 3, NO. 1

revie

Revista de Investigación y Evaluación Educativa

ISSN 2409-1553

ideice
Instituto Dominicano de Evaluación e Investigación de la Calidad Educativa

revie
Revista de Investigación y Evaluación Educativa

Revista Digital de suscripción gratuita del Instituto Dominicano de Evaluación e Investigación de la Calidad Educativa (IDEICE)

Periodicidad Semestral

Edición
Febrero 2016, Vol.3, No. 1

Dirección Ejecutiva
Julio Leonardo Valeirón Ureña

Consejo Editorial
Dinorah de Lima Jiménez
Julián Álvarez Acosta
Luis Camilo Matos De León

Correción de estilos
Ramón Fari Rosario

Coordinación General
Dilcia Armesto Núñez

Diseño y Diagramación
Natasha Mercedes Arias

ISSN: 2409-1553

IDEICE
Ave. César Nicolás Penson No. 30, Gazcue
Santo Domingo, D.N.
Teléfono: +1 (809) 732-7152
www.ideice.gob.do

Santo Domingo, Rep. Dom.

EDITORIAL

La presente edición de **REVIE** contiene cuatro investigaciones, las tres primeras auspiciada por el IDEICE y la cuarta corresponde a un artículo de un reconocido investigador internacional. Con este nuevo número reafirmamos el firme compromiso con la investigación de calidad como soporte científico, aportando evidencias y conocimientos pertinentes en la toma de decisiones en el ámbito de la educación.

Domínguez Ruiz y colaboradoras estudian la tasa de retorno de la educación en la población dominicana entre 18 y 65 años que reciben ingresos por remuneraciones laborales; ofrecen informaciones relevantes para la elaboración de políticas públicas relacionadas a la oferta laboral, el nivel de salarios y la equidad en la distribución del ingreso.

Oscar Amargós en el artículo *Evaluación de resultados e impacto de la política de educación secundaria en República Dominicana*, realiza una comparación entre los egresados de la modalidad general y los titulados de la modalidad técnico profesional, con el propósito de aportar evidencias objetivas para sustentar las decisiones de las autoridades educativas nacionales evaluando las variables que permiten determinar los efectos e imparto de las políticas de educación secundaria en el desarrollo económico y social del país.

El estudio de González y su colaborador sobre *Un modelo predictivo de deserción escolar para la República Dominicana* nos ofrece una importante herramienta para la predicción del riesgo de deserción en los estudiantes de nivel básico y medio del sistema educativo nacional.

Este número concluye con la entrega del investigador Díaz Esteve, profesor de la Universidad de Valencia, del artículo sobre la *Importancia de utilizar la teoría de la respuesta al Item (TRI) en la construcción de pruebas de aptitud y conocimiento* donde nos presenta los fundamentos teóricos y principios básicos sobre los que se ha construido esta teoría y elementos, así como también utiliza los datos obtenidos en la aplicación de una prueba de aptitud donde se pueden visualizar los valores paramétricos de los ítems y su interpretación.

Finalmente, el IDEICE reafirma su vocación investigativa para con ello no solo conocer la realidad educativa, sino propiciar su transformación, estando seguros de hacer presente nuestra total convicción de que las investigaciones y reflexiones presentadas, serán un aporte que nutrirá el conocimiento acerca de la educación y sus procesos.

Julio Leonardo Valeirón Ureña
Director Ejecutivo

CONTENIDO

JOSÉ V. DÍAZ ESTEVE

jose.v.diaz@uv.es

Profesor de la Facultad de Psicología
Universidad de Valencia, España.

IMPORTANCIA DE UTILIZAR LA TEORÍA DE LA RESPUESTA AL ÍTEM (TRI) EN LA CONSTRUCCIÓN DE PRUEBAS DE APTITUD Y CONOCIMIENTO

RESUMEN

El objetivo de este artículo consiste en introducir al lector en los conceptos básicos relacionados con la Teoría de la Respuesta al Ítem (TRI), por tal razón, ha sido dividido en dos secciones:

En la primera sección, se presentan los fundamentos teóricos de la TRI y los principios básicos sobre los que se ha construido esta teoría y elementos.

En la segunda, se utilizan los datos obtenidos en la aplicación de una prueba de aptitud donde se pueden visualizar los valores paramétricos de los ítems y su interpretación.

A través de la TRI se conoce la estructura de un ítem por medio de su curva característica (CCI), así como también, nos permite realizar estimaciones de los parámetros de los ítems y del nivel de aptitud de los sujetos.

PALABRAS CLAVE

Teoría de respuesta al ítem (TRI); Test; Pruebas de aptitud y conocimiento; psicometría.

ABSTRACT

The purpose of this article is to introduce the reader into the basic contents associated to the Item Response Theory (IRT) therefore, it has been structured in two sections:

In the first section we present the theoretical fundaments of the IRT and the basic principles and elements that this theory is based upon.

In the second, we use the data obtained from the application of an aptitude test in which we can visualize the parametric values of the items and their meaning.

Through the IRT we know the structure of one item by it's characteristic curve (CCI), and we can also make estimations of the parameters of the items and level of aptitude of the subjects.

KEYWORDS

Item response theory (IRT); Test; Aptitude and knowledge test; Psychometrics.

INTRODUCCIÓN

Los llamados "tests estandarizados", basados en las llamadas *teorías débiles de los tests* (TCT y TGT) han sido objeto de duras críticas, tales como: "el carácter tautológico de sus supuestos, la fuerte dependencia del valor de los parámetros y de las características de los ítems de la muestra de sujetos; estas condiciones y otras han dificultado considerablemente la posibilidad de colocar los valores obtenidos sobre una métrica común, lo que hace muy difícil: la estimación de los parámetros de los ítems, la comparación entre puntuaciones de los sujetos, el análisis diferencial de los ítems, la construcción de bancos de ítems y de tests computarizados"(Barbero, I., 1996, p.143). De ahí que en las últimas décadas del siglo XX, sobre todo a partir de los años ochenta, gracias a la generalización del uso de ordenadores personales, se ha extendido una nueva teoría de los tests, llamada *Teoría de la Repuesta a los Ítems* (TRI), que debe ser considerada como una alternativa más elaborada y eficaz en la construcción de tests.

Entre los autores que más han contribuido a la formulación, implantación y generalización de esta teoría se pueden citar a Rasch (1960, 1968,1980), Wright y Col. (1969, 1976, 1979,1992), Lord (1975, 1980), Hulin, Drasgow y Parson (1983), Baker (1985), Hambleton y Swaminatham (1985), Hambleton, Swaminhatham y Rogers (1991), y en lengua española Muñiz Fernández (1990,1997), Santisteban (1990), y Martínez Arias (1995), Diaz Esteve (1997).

Este artículo tiene dos secciones:

I. En la primera sección se presentan los fundamentos teóricos de la TRI y los principios básicos sobre los se ha construido esta teoría y elementos
II. En la segunda, utilizando los datos obtenidos en la aplicación de una prueba de aptitud, se puede visualizar los valores paramétricos de los ítems y su interpretación.

I. FUNDAMENTOS TEÓRICOS DE LA TRI

La TRI, aunque presenta una carácter eminentemente tecnológico, permite efectuar una mejor construcción de tests y realizar un análisis de los resultados de "tipo teórico-fundante de la medición de los rasgos psicológicos tales como: la transitividad, la representatividad, la unicidad, la significación, la presencia del cero absoluto en sus escalas, los de la estructura de los rasgos (status, estabilidad, reificación, circularidad, etc.), así como los problemas relativos a la validez" (Muñiz, J., 1996, p. 24).

I.1 SUPUESTOS FUNDAMENTALES DE LA TRI

La TRI, al igual que lo que lo hace la TCT, se fundamenta en algunos supuestos y procedimientos, como estos:

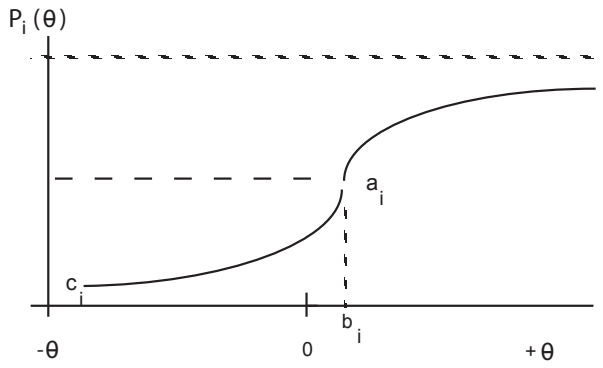
1º Existe en los sujetos, que van ser evaluados con un test, un *rasgo latente o atributo*; rasgo que suele ser simbolizado por θ en la TRI.

2º Luego se elige un conjunto de comportamientos observables capaces de medir este rasgo. A este conjunto total de ítems se le llama *dominio del atributo*; si se extrae del mismo una muestra representativa de **n** de ítems se obtiene un test.

3º Aplicada esta muestra de ítems a los sujetos, éstos dan respuestas diversas a los ítems, condicionadas al nivel de aptitud (θ) que tiene el sujeto).

La TRI especifica el último supuesto de esta forma: "la relación existente entre las respuestas dadas a los ítems y la aptitud se expresa a través de una función creciente que indica la probabilidad de acertar el ítem, condicionada al nivel de aptitud del sujeto, cuya representación gráfica tiene la forma de ojiva denominada Curva Característica del Ítem (CCI), como puede apreciarse en la Figura 1.

FIGURA I.1
CCI DE UN TEST DE APTITUDES



En esta CCI se pueden apreciar tres parámetros:

- El *parámetro* b_i, que equivale al valor en la abscisa Ø del punto de inflexión de la CCI, es decir, la proyección de dicho punto sobre la misma. Este parámetro indica el grado de *dificultad* del ítem, ya que en ese punto, el ítem presenta el valor máximo en la función información. Los valores de este parámetro, teóricamente, pueden estar entre: $-\infty \leq b_i \leq \infty$. Aunque en la práctica se tomen valores que están entre -3.50 y 3.50.

- El *parámetro* a_i, es un valor proporcional a la pendiente de la recta tangente a la CCI en el punto de inflexión de la curva. Este parámetro, también llamado *poder discriminativo del ítem*, indica el grado en que la respuesta al ítem varía con el nivel de aptitud del sujeto. Toma siempre valores positivos, ya que si tomara valores negativos la curva sería decreciente, por lo que sus valores pueden estar entre: $0 \leq a_i \leq \infty$.

-Y el *parámetro* c_i, denominado de *asintoticidad*, que representa la probabilidad de acertar el ítem cuando la aptitud es nula, es decir, cuando el sujeto contesta el ítem al azar. Este parámetro, también llamado *índice de conjetura o pseudo-azar*, equivale a la altura del punto donde la curva corta al eje P(Ø). Generalmente su valor se toma aproximadamente, como **c=1/k** siendo k el número de alternativas del ítem. Sus valores, por tanto estarán entre: $0 \leq c_i \leq 1.0$.

La CCI expresa, pues, la relación entre las respuestas al ítem y el nivel de aptitud, que se expresa por medio de una ecuación exponencial, que indica la *probabilidad* de acertar condicionada al nivel de aptitud de los sujetos:

$$P(\underline{\overline{X}} \mid Ø)$$

Donde:

$\underline{\overline{X}}$- representa el sistema de puntuaciones que pueden obtener los sujetos en el ítem o test.

Ø- los valores dados al rasgo que se pretende medir (escala métrica de Ø)

P- la relación de probabilidad existente entre ambos elementos.

Visto todo lo anterior se podría afirmar: que el *corazón* de la TRI es un modelo matemático, capaz de expresar la probabilidad de acertar el ítem condicionada a la aptitud del sujeto y que el *objetivo del test* sería estimar el valor verdadero de la aptitud en cada sujeto. Esta estimación se hará, naturalmente, a través de las puntuaciones observadas.

I.2 CONDICIONES ESENCIALES EN LA TRI

Aceptados los supuestos antes señalados, resulta necesario, averiguar si se cumplen en el *modelo de medición* adoptado ciertas condiciones, como serían: la *independencia local, la unidimensionalidad del rasgo y la falta de presión temporal:*

- El concepto de *independencia local* viene a expresar, que no hay ningún tipo relación de dependencia entre los elementos que están definidos en un mismo campo. Aplicado este concepto a los tests se puede hablar tanto de la *independencia local de los ítems*, como la de *independencia local de sujetos*, así como de la *independencia local de las dimensiones del rasgo.*

- La segunda condición, la *unidmensionalidad* del rasgo, es indispensable en la mayoría de los modelos, ya que la probabilidad de que los sujetos acierten a los ítems debe depender solamente del nivel de la aptitud de los sujetos, y no de otros factores. De modo que el concepto de *unidimensionalidad* está incrustrado en la misma concepción de la CCI. Dada la dificul-

tad de encontrar tests plenamente unidimensionales se ha optado por considerar como tales, aquellos que tengan un primer factor *dominante*. Se han propuestos muchos métodos para demostrar la unidimensionalidad de los tests, Hattie (1985) cita ochenta y siete, casi todos ellos basados en el análisis factorial, pero éstos no resultan totalmente satisfactorios, sobre todo cuando se trata de ítems dicotómicos (ver: Martínez Arias, 1995). De ahí que en la actualidad se esté trabajando intensamente en la línea de modelos multidimensionales para estimar los parámetros de los ítems (ver: Reckase, 1979, 1983, 1997; Maydeu, 1996).

- Existe una tercera condición, señalada por Hambleton & Swaminathan (1985) que exige ausencia de presión temporal en la ejecución del test (Speededness), ya que si el sujeto no responde a un ítem no se sabrá con certeza, si el fracaso es debido a la falta de capacidad del sujeto u a otros condicionantes externos como la falta de tiempo para resolverlo. Esta condición debe ser considerada como una consecuencia de la anterior, *la unidimensioalidad*. Por lo que es preferible dar tiempo suficiente y controlar las respuestas de los sujetos que no alcanzan el final del test. Más tarde Hambleton, Swaminathan y Rogers (1991), ante la dificultad de obtener tests donde todos los ítems sean contestados por todos los sujetos, recomiendan utilizar aquellos protocolos que al menos un 75 % de sujetos completan el test y aquellos ítems que sean completados al menos por el 80% de los examinados.

I.3. VENTAJAS DE LOS ESTIMADORES DE LA TRI

La TRI permite hacer estimaciones de los parámetros de los ítems y del nivel de aptitud de los sujetos. Estos estimadores deben estar adornados de estas cualidades:

- aunque se obtienen a través de las respuestas de los sujetos a los ítems, los *parámetros estimados* (de los ítems y de los sujetos) no se expresan como dependientes directamente de las características particulares de la muestra, *cualidad* que es identificada como *invarianza de los parámetros,*

- el nivel de aptitud estimada en cada sujeto, no depende de la cantidad de ítems contestados, sino de sus peculiaridades, de ahí que sus resultados sean comparables entre sí, cualidad que es conocida como *(invarianza muestral de los ítems),*

- los supuestos de *homocedasticidad* y *paralelismo* no se requieren con los procedimientos estadísticos de estos modelos; la precisión de los tests se expresa mediante la llamada *función de información*,

- los modelos de la TRI rompen el concepto unitario del error típico de medida, ya que ofrecen un estimador del mismo para cada ítem y sujeto,

- cualquier aplicación de la TRI permite evaluar el ajuste de los datos al modelo, ya que pueden existir diversos modelos.

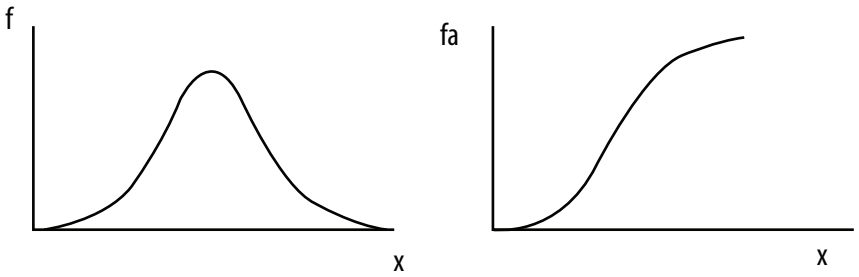
I.4 MODELOS MATEMÁTICOS DE LA TRI

Se ha dicho que la relación entre la probabilidad de acertar un ítem y el nivel de aptitud se expresa a través de una *función monótona creciente*, cuya representación gráfica tiene la forma de una S invertida acharada, que se suele denominar *Curva Característica del Ítem* (CCI), que es definida formalmente por medio de una ecuación exponencial. Los modelos más utilizado en la TRI son los *logísticos* y de la *ojiva normal*. Los modelos de la ojiva normal (donde la CCI es una curva normal acumulada) preceden temporalmente a los logísticos, (Lawley, 1943; Tucker, 1946; Lord, 1952), sin embargo dado que los modelos logísticos resultan más manejables desde el punto de vista matemático y que sus resultados son muy similares, estos se suelen utilizar preferentemente.

I.4.1 MODELOS DE LA OJIVA NORMAL

Estos modelos fueron propuestos por Lord y Novick (1968) para ser aplicados a los tests con ítems dicotómicos. La curva característica de estos ítems sería una curva acumulada de distribución normal, que recibe el nombre de ojiva normal, en la que se asume el cumplimiento de los supuestos básicos de la TRI.

FIGURA.2
DISTRIBUCIÓN NORMAL



Curva Normal Simple Curva Normal Acumulada

La ecuación que expresa la probabilidad de acertar el ítem en función de la aptitud del sujeto es:

$$P_i(\theta) = \int_{-\infty}^{L_i} f(x)dx$$

Donde:

- f (x) es la función densidad de la distribución normal, con fórmula:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ para puntuaciones directas

$$f(x) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ para puntuaciones típicas

$L_i = a_i(\theta - b_i)$, donde ai y bi son los parámetros de la CCI

Los parámetros a_i y b_i son, los que determinan la forma de la CCI. Los modelos de la ojiva normal son básicamente tres:

Un parámetro: $Pi(\theta) = \int_{-\infty}^{(\theta - b_i)} \frac{1}{\sqrt{2\pi}}\ e^{(-z^2/2)dz}$

Dos parámetros: $Pi(\theta) = \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}}\ e^{(-z^2/2)dz}$

Tres Parámetros:

$$Pi(\theta) = c_i + (1 - c_i)\int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}}\ e^{(-z^2/2)dz}$$

El cálculo de estos valores se facilita tremendamente si busca en la tabla de la curva normal, el valor hallado al efectuar los cálculos de los datos de límite superior de la integral, como si fuera Z. El valor hallado corresponderá a la probabilidad acumulada en la puntuación Z, es decir, la probabilidad de acertar el ítem.

1.4.2 MODELOS LOGÍSTICOS

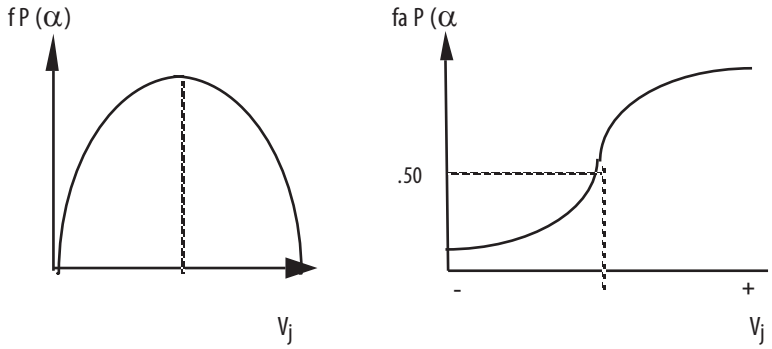
Los modelos logísticos están construidos sobre una distribución de puntuaciones, que toma la forma de una curva logística, que es una curva monótona creciente donde la razón de crecimiento es proporcional a la inversa de la función y aceleradamente negativa.

$$y = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}}$$

Si se sustituye las variables: x é y por valores correspondientes a los parámetros de los ítems utilizados y por la probabilidad de acertar el ítem, se tendrán los distintos sub modelos de la TRI.

En la práctica para maximizar el acuerdo entre el modelo logístico y el normal suele colocarse una constante D con valor de 1.7, que hace que los resultados en ambos modelos difieran en menos de 4 centésimas para cualquier valor de x.

FIGURA. I.3
MODELO LOGÍSTICO



Curva logística simple — Curva logística acumulada

Los modelos de la ojiva normal son básicamente tres:

- De un parámetro:

$$P_i(\theta) = \frac{e^{D\,(\theta - b_i)}}{1 + e^{D\,(\theta - b_i)}} = \frac{1}{1 + e^{-D\,(\theta - b_i)}}$$

- De dos parámetros:

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

- De tres parámetros:

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

I.4.2.1. EL MODELO LOGÍSTICO DE UN PARÁMETRO

El modelo logístico de Rasch (1960, 1968) es el que más se utilizó en las primeras investigaciones en la TRI. Este modelo supone que todos los ítems tienen el mismo poder discriminativo y que varían sólo en función de su dificultad; por lo tanto las CCI son funciones logísticas de un sólo parámetro, como lo indica la ecuación siguiente:

$$P_i(\theta) = \frac{e^{D\,(\theta - b_i)}}{1 + e^{D\,(\theta - b_i)}} = \frac{1}{1 + e^{-D\,(\theta - b_i)}}$$

Donde:

$P_i(Ø)$: es la probabilidad de acertar el ítem en un determinado nivel de aptitud.

Ø : representa los valores que puede tomar la variable medida, es decir la aptitud.

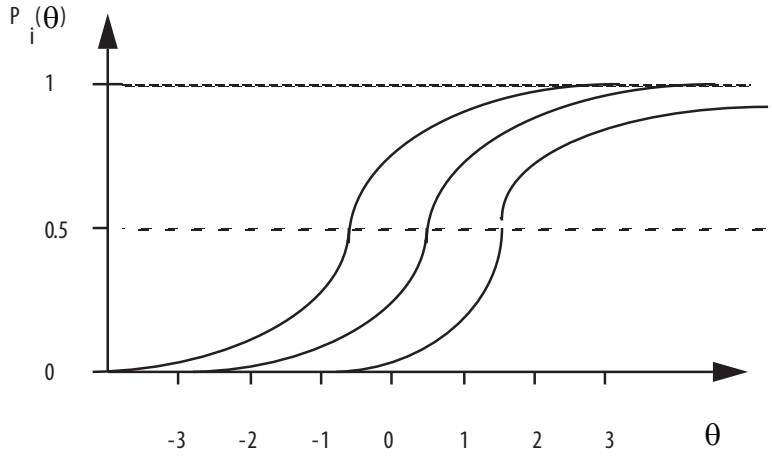
b_i: es el parámetro de la dificultad del ítem

D: una constante convencional (D=1.7).

e: la base de los logaritmos neperianos.

Estos modelos suponen que todas las CCI en los distintos niveles de dificultad presentan la misma pendiente, como puede observarse en la figura 2, donde los valores del parámetro dificultad son respectivamente: -1, 0, 1:

FIGURA 4
CCI EN EL MODELO RASCH PARA $b_1=-1, b_2=0, b_3=1$



Esta forma restringida (la de Rasch) de los modelos logísticos resulta más fácil de calcular, más parsimoniosa y más robusta según algunos autores, pero presenta el grave inconveniente de ser más dificultoso el ajuste de los datos reales al modelo, sobre todo cuando el número de ítems no es muy grande (n<50) y los índices de discriminación y los aciertos al azar son evidentes.

I.4.2.2. MODELO LOGÍSTICO DE DOS PARÁMETROS

El modelo logístico de dos parámetros presenta una forma de distribución logística, definida por los parámetros (a_i, b_i). Este modelo fue propuesto inicialmente por Birnbaum (1968), y viene definido por esta ecuación:

$$P_i(\theta) = \frac{e^{D a_i (\theta - b_i)}}{1 + e^{D a_i (\theta - b_i)}} = \frac{1}{1 + e^{-D a_i (\theta - b_i)}}$$

Donde:

$P_i(Ø)$: es la probabilidad de acertar el ítem en un determinado nivel de aptitud.

Ø: representa los valores que puede tomar la variable medida.

a_i: es el parámetro de la discriminación del ítem

b_i: es el parámetro de la dificultad del ítem

D: una constante convencional (D=1.7).

e: la base de los logaritmos neperianos.

Estos modelos suponen que todas las CCI presentan la misma asintoticidad, siendo diferentes los otros parámetros. Es el modelo más utilizado para los tests de rendimiento.

I.4.2.3 MODELO LOGÍSTICOS DE TRES PARÁMETROS

El modelo logístico de tres parámetros, fue propuesto por Lord & Novick (1968) y Lord (1980), y tiene las mismas propiedades que el anterior, pero añade un tercer parámetro c_i, por lo que la ecuación toma esta forma:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{D a_i (\theta - b_i)}}{1 + e^{D a_i (\theta - b_i)}} = c_i + (1 - c_i) \frac{1}{1 + e^{-D a_i (\theta - b_i)}}$$

Donde:

- a_i y b_i y son parámetros con el mismo significado que antes

- c_i es el parámetro que representa la asintoticidad de las CCI e indica la probabilidad de que un examinado

con nivel bajo de aptitud responda correctamente al elemento. Este parámetro intenta controlar el nivel de conjetura en los niveles bajos del atributo medido, por lo que se llama "parámetro de conjetura"

-la función P_i (Ø) toma valores mínimos, equivalentes a c_i cuando Ø decrece.

-D es la constante.

Este modelo suele ser utilizado sobre todo en los tests de Inteligencia y Aptitudes y es considerado el más general de todos los modelos logísticos, de modo que:

-si c=0, el mismo adopta la forma del modelo de dos parámetros

-si además se asume que a_i es constante se tiene el modelo de un parámetro.

EJEMPLO I.2
MODELOS LOGÍSTICOS

Dado un test, que ha sido aplicado a 500 sujetos, si los ítems son localmente independientes, calcular la probabilidad de acertar, en los tres modelos logísticos, el ítem 14 en el nivel de aptitud Ø=2, si se sabe que sus parámetros tienen estos valores: a=2, b=0.5, y c=0.25.

a. Modelo logístico de un parámetro

$$P_i(\theta) = \frac{e^{D\ (\theta - b_i)}}{1 + e^{D\ (\theta - b_i)}} = \frac{2.72^{(1.7)\ (2-0.5)}}{1 + 2.72^{(1.7)\ (2-0.5)}} = 0.9276$$

b. Modelo logístico de dos parámetros

$$P_i(\theta) = \frac{e^{D\ a_i\ (\theta - b_i)}}{1 + e^{D\ a_i\ (\theta - b_i)}} = \frac{2.72^{(1.7)\ (2)\ (2-0.5)}}{1 + 2.72^{(1.7)\ (2)\ (2-0.5)}} = 0.9939$$

c. Modelo logístico de tres parámetros

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{D\ a_i\ (\theta - b_i)}}{1 + e^{D\ a_i\ (\theta - b_i)}} = 0.25 + (1 - 0.25)\frac{2.72^{(1.7)\ (2-0.5)}}{1 + 2.72^{(1.7)\ (2-0.5)}} = 0.9954$$

Como puede observarse los valores son similares en los tres modelos. Se puede observar, también, que a medida que se utilizan más parámetros, el valor de la probabilidad de acertar se incrementa, aunque la ganancia obtenida entre el uso de modelos de dos a tres parámetros es casi imperceptible.

I.5 ESTIMACIÓN DE LOS PARÁMETROS DE LOS ÍTEMS EN LA TRI:

La estimación de los parámetros en TRI resulta muy compleja. Esta solamente puede hacerse mediante un programa de ordenador adecuado. Se han propuesto varios métodos: estimación por la Máxima Verosimilitud (ML), Estimación Bayesiana (MAP), y Esperado a Posteriori (EAP). Todos ellos efectúan procesos de aproximaciones sucesivas (iteraciones) hasta que los valores alcanzan los niveles que hacen más verosímiles los datos obtenidos. De modo que se toman los valores que maximizan la probabilidad de ocurrencia de los datos obtenidos al aplicar los ítems a los sujetos.

Los programas más utilizados son:
BICAL de Wright et Al (1970)
BILOG de Mislevy y Bock (1986)
MULTILOG de Mislevy y Bock (1988)
LOGIST de Wingersky, Barton y Lord (1984)
DATAGEN de Hambleton y Rovinelli (1983)
RASCAL y XCALIBRE de Assessment System Corporation (1996)

En este texto se trabajará preferentemente con el Assessment System Corporation, que permite trabajar con el mismo fichero de datos tanto en la TCT como en la TRI.

Sea cual sea el modelo escogido para trabajar, la probabilidad de dar una respuesta correcta depende de la aptitud del sujeto y del valor de los parámetros que caracterizan al ítem. Pero resulta que estos valores no se conocen. Sólo se tiene información de los valores

atribuidos a las respuestas de los sujetos. Para poder aplicar los modelos de la TRI hace falta conocer los valores de estos elementos (el nivel de aptitud del sujeto y el valor de los parámetros), que al no ser observables, sólo pueden ser estimados.

Para efectuar estas estimaciones se parte de la matriz de datos en donde están las respuestas de los **N** sujetos a los **n** ítems. Partiendo de ella se estiman los valores de los parámetros, de los ítems primero, y de la aptitud de cada sujeto después. Se han propuesto varios métodos de estimación: por máxima verosimilitud marginal conjunta (MLC), por máxima verosimilitud marginal (MLM), los basados en el análisis factorial no lineal y los basados en el *Xi-cuadrado* mínimo.

Como es lógico prever, la estimación de los parámetros resulta muy compleja. Esta solamente puede hacerse mediante un programa de ordenador adecuado. Existen muchos programas. Nosotros trabajamos preferentemente, a pesar de sus limitaciones, con el MAC-BILOG y/o PC-BIOG (Mislevy y Bock, 1990, versión 3.0) para ítems dicotómicos en los tres submodelos señalados y el MULTILOG (Thissen, 1988) para items graduados tipo likert y politómicos o categóricos.

El BILOG puede trabajar para estimar la aptitud con el MML (marginal maximum likelihood), con el MMAP (marginal maximum a posteriori likelihood). También permite estimar la distribución latente de la aptitud y probar el buen ajuste del modelo (esto lo hace en la tercera fase del programa). Puede calcular los estimadores de los items por tres métodos: ML (maximum likelihood), EAP (expected a posteriori) que es una estimación bayesiana, y el MAP (maximum a posteriori) llamado estimación modal bayesiana que utiliza la puntuación Fisher estimada sobre la función información a posteriori (Bock, Mislevy y Woodsen, 1982).

I.6 REFLEXIONES SOBRE LA TRI

No cabe la menor duda que la TRI resuelve mucho mejor que la TCT algunos problemas prácticos, tales como:

- la construcción de tests, ya que si se conocen sus parámetros la selección y sustitución de ítems es mucho más fácil, así como la determinación del error y la precisión del test (Theunissen, 1985),

- la elaboración de los tests adaptativos, ya que se pueden escoger los ítems más próximos a la aptitud del sujeto (Weis, 1984),

- la evaluación educacional en gran escala, ya que puede medir mejor la eficiencia al nivel poblacional, así como evaluar programas o estrategias al poder disponer de instrumentos capaces de hacer estimaciones, invariadas y estables (Bock, Mislevy y Wodson, 1982; Messick, Beaton y Lord, 1983).

- las técnicas de la TRI parecen tener más potencial que el demostrado en la actualidad, ya que se han utilizado generalmente a resultados de tests, concebidos, construidos y administrados dentro del marco teórico de la TCT, donde los ítems no contestados no se tienen en cuenta (Mislevy, 1991),

- la TRI al trabajar sobre los patrones de respuestas, puede efectuar análisis de respuestas inesperadas o raras que pueden revelar concepciones defectuosas o imprecisas en el aprendizaje (Tatsouka, 1983),

- finalmente, entre otras cosas, la TRI permite definir mejor los mecanismos intelectuales que el sujeto utiliza para adquirir conocimientos, las estrategias mentales que utilizan los sujetos (destrezas meta cognitivas) Pero todas estas y otras posibilidades tienen que ver con la nueva teoría de los tests que está naciendo al intentar aplicar las técnicas de la TRI a las nuevas aportaciones de la Teoría Cognitiva de los tests.

II. UN EJEMPLO DE TRATAMIENTO DE DATOS EN UNA PRUEBA DE APTITUD SIGUIENDO EL MODELO TRI

Para observar el tratamiento de los datos en el modelo TRI, se ha utilizado la prueba IAUB.2 de aptitud académica, que pretende evaluar el nivel de desarrollo de la inteligencia académica en los estudiantes que están en el bachillerato. El instrumento que se analiza consta de 60 ítems, divididos en cuatro sub pruebas: la **CV**

con 15 ítems buscan evaluar la aptitud verbal, la **CM** con 15 ítems que busca evaluar la aptitud matemática, la **CE** con 15 ítems que busca evaluar la aptitud espacio-estructural y la **CC** con 15 ítems que buscan evaluar la madurez psicosocial a través de la capacidad de emitir juicios acerca de algunas situaciones referentes a la vida práctica. Esta prueba se aplicó a 446 alumnos de bachillerato.

Para efectuar el análisis psicométrico se ha utilizado el programa XCALIBRE de Assessment System Corporation, versión 1996, siguiendo estos pasos y obteniendo estos resultados:

II.1 UNIDIMENSIONALIDAD DE LA PRUEBA IAUB.2

Una de las condiciones que se exige para aplicar el modelo TRI, es que la prueba sea unidimensional, por lo que se inicia este ejercicio estudiando la posible unidemensionalidad de la prueba IAUB.2, utilizando el programa SPSS y partiendo de la matriz de puntuaciones obtenidas en las cuatro sub pruebas obtenida en el archivo IAUB2 score del ITEMANW, v. 1996.

Procesados los datos de esta matriz con el SPSS, se han obtenido estos resultados:

MATRIZ DE CORRELACIONES

		CV	CM	CE	CC
Correlación	VAR00002	1.000	.302	.247	.229
	VAR00003	.302	1.000	.396	.318
	VAR00004	.247	.396	1.000	.401
	VAR00005	.229	.318	.401	1.000

Nota: aunque los valores de las correlaciones no son muy altos, hay que señalar que estos son significativas al nivel de significancia del 0.05.

En el cuadro siguiente se presenta el *por ciento* de lo común (comunalidad) que tienen las sub pruebas entre sí, extraído por el Método de extracción: Análisis de los componentes principales.

COMUNALIDADES

	INICIAL	EXTRACCIÓN
CV	1.000	.355
CM	1.000	.539
CE	1.000	.567
CC	1.000	.495

Avanzando el proceso del AFE, se han encontrado cuatro *componentes o factores* que subyacen detrás de las cuatro puntuaciones de la prueba IAUB2, que explican la proporción de varianza total de este modo:

VARIANZA TOTAL EXPLICADA

COMPONENTE	AUTOVALORES INICIALES			SUMAS DE LAS SATURACIONES AL CUADRADO DE LA EXTRACCIÓN		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	1.955	48.885	48.885	1.955	48.885	48.885
2	.810	20.249	69.134			
3	.664	16.607	85.742			
4	.570	14.258	100.000			

Finalmente en el cuadro anterior se ve que *existe un factor común* que explica cerca del 49% de la varianza total de la prueba. Factor que se ha llamado en esta investigación "inteligencia académica" y que presenta las siguientes saturaciones con este primer factor… De acuerdo con la mayoría de los autores, los datos encontrados justifican la unidimensionalidad de la prueba.

MATRIZ DE COMPONENTES[a]

	COMPONENTE
	1
CV	.596
CM	.734
CE	.753
CC	.703

II.2 PREPARACIÓN DE LA MATRIZ DE DATOS QUE PERMITEN APLICAR EL XCALIBRE:

La matriz de datos debe presentarse de este modo:

60 0 N 6
abbcbdbaabaddcabddcdacaddaaabbbaddccdbccbdababadaadacabcdbcb
444444444444444444444444444444444444444444444444444444444444
111111111111111111111111111111111111111111111111111111111111
A01 abaabdbacddddaddbacacbadcacbbdcbcbbcbddc0cbdbaacdacabcc0cbb
A02 ccbcbabaacbddcabc00dbcbdaa0000bacaccbbcccdcccbadacdacabcdccb
……....
……
Ver "User's Manual for the ASC Item and Tests Analysis Package, 1996.

II.3 EL OUT POUT

El programa tras un breve intervalo de tiempo saca estos datos:

II.3.1 DATOS GENERALES DE LA APLICACIÓN DEL XCALIBRE SOBRE LOS DATOS DEL IAUB2.DAT

Datos del programa:

```
          XCALIBRE (tm) for Windows -- Version 1.10      Page  1
  Copyright (c) 1995 by Assessment Systems Corporation, All Rights
Reserved
         Marginal Maximum-Likelihood IRT Parameter Estimation Program
XCALIBRE Analysis from Data File:
C:\USERS\JOSEVI~1\DESKTOP\ITEMAN\XCALWIN\IAUB2.DAT
Date: Jul 02, 2014                                          Time: 16:17
********************ANALYSIS SUMMARY INFORMATION  ********************
Data (Input) File: C:\USERS\JOSEVI~1\DESKTOP\ITEMAN\XCALWIN\IAUB2.DAT
Analysis Output File
C:\USERS\JOSEVI~1\DESKTOP\ITEMAN\XCALWIN\IAUB2.OUT
Score Output File: NONE
Item Name File: NONE
Statistics Output File: NONE

               *****  CONFIGURATION INFORMATION  *****
            Item Parameter Priors:  COMMON
            Allow Priors to Float:  YES
                   IRT Model Used:  3-parameter
          Maximum Number of Loops: 12
       Starting Prior Distribution Moments:
              Mean      SD
         a    0.7500   0.1200
         b    0.0000   1.0000
         c    0.2500   0.0250
NOTE:  *** will be printed when the c standard error value > 0.10
The number of items was:   60
There was no item linkage (independencia local de los items)
The key was:
abbcbdbaabaddcabddcdacaddaaabbbaddccdbccbdababadaadacabcdbcb
The numbers of alternatives were:
444444444444444444444444444444444444444444444444444444444444
The inclusion specifications were:
YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
```

```
Cálculos iterativos (loops) para alcanzar el ajuste deseado (0.05)
The maximum parameter change on loop 1 was      1.658
The maximum parameter change on loop 2 was      0.365
The maximum parameter change on loop 3 was      0.185
The maximum parameter change on loop 4 was      0.123
The maximum parameter change on loop 5 was      0.054
The maximum parameter change on loop 6 was      0.050

Estadísticos generales de la prueba
Mean Number-Correct Score =   43.289
Number-Correct Standard Deviation =    6.201
K-R 21 Reliability = 0.763
The number of examinees was 446

 Final Parameter Summary Information:
          Mean      SD
Theta     0.00    1.00
  a       0.44    0.05
  b      -1.00    1.53
  c       0.26    0.01
```

COMENTARIOS

De esta sección del output es importante fijarse:

- en el modelo de TRI en el que se ha trabajado: *3 Parámetros*,

- el número de iteraciones (loops) que ha necesitado el programa para *ajustar la CCI* de los ítems reales al ideal (fit),

- en los *estadísticos básicos de la prueba* (Media y Des. Típica), que son iguales que antes,

- en el *coeficiente de fiabilidad*, estimado por la formula K-R-21 (0.763),

- en el *número de sujetos* de la muestra (446)

- y finalmente en el *valor de los intervalos de la abscisa* (theta) de la CCT, en los valores de la *Media y D.T. de los tres parámetros* que permiten de ítems identificar los ítems y en las *puntuaciones de los sujetos.*

En este caso concreto se puede afirmar que la prueba IAUB2 en conjunto presenta una *baja capacidad discriminativa* (a=0.44), una *dificultad baja* (b=-1,00) y un valor del parámetro c (conjetura) cercano al ideal (0.26).

Los datos encontrados exigen revisar los valores paramétricos de los ítems para eliminar en primer lugar los que el *sistema rechaza* y en segundo lugar mejorar los aceptables, pero que caen en zonas que se han llamado de aceptación critica (AC), pues aportan poco a los objetivos de la prueba.

II.3.2 ANÁLISIS DE LOS PARÁMETROS DE LOS ÍTEMS

El XCALIBRE después de presentar los datos generales del input saca las siguientes tablas:

1ª **FINAL ITEM PARAMETER ESTIMATES**, en la que se ofrecen estos datos de cada ítem:

- los valores de los parámetros **a, b y c,**

- el valor residual promedio entre el ajuste (fit) de las CCI real y teórica,

- el **PC**, el porciento de sujetos que acierta el ítem,

- el **PBs** el porciento de que sujetos que aceptaron el ítem en el *grupo superior* (los mejores resultados en la puntuación total de la prueba) y el **PBt** el porciento de que sujetos que aceptaron el ítem en el grupo total.

- Finalmente **N**, que indica el número de sujetos de la muestra.

FINAL ITEM PARAMETER ESTIMATES

Item	Lnk	Flg	a	b	c	Resid	PC	PBs	PBt	N
01			0.39	-1.49	0.27	0.61	0.80	0.23	0.22	446
02			0.48	-2.52	0.27	0.16	0.91	0.28	0.29	446
03			0.43	-2.74	0.27	0.71	0.90	0.19	0.21	446
04			0.41	-2.61	0.27	1.11	0.89	0.20	0.20	446
05			0.36	-1.71	0.27	0.69	0.82	0.14	0.17	446
06			0.45	-2.19	0.27	0.28	0.88	0.25	0.27	446
07			0.43	-2.44	0.27	0.71	0.89	0.20	0.21	446
08			0.39	-2.13	0.27	0.79	0.85	0.19	0.19	446
09			0.46	-1.95	0.26	0.49	0.89	0.29	0.32	446
10			0.47	2.50	0.26	0.74	0.35	0.15	0.13	446
11			0.42	1.69	0.26	0.73	0.47	0.22	0.16	446
12		P	0.57	-3.00	0.27	1.43	0.98	0.08	0.09	446
13			0.45	-2.42	0.27	0.30	0.93	0.29	0.25	446
14			0.43	1.12	0.27	0.85	0.54	0.23	0.21	446
15			0.35	-2.24	0.27	1.41	0.85	0.11	0.11	446
16			0.42	-2.45	0.27	1.05	0.88	0.24	0.23	446
17			0.39	-0.50	0.27	0.61	0.73	0.25	0.23	446
18			0.40	1.04	0.27	1.04	0.58	0.24	0.21	446
19			0.50	-0.61	0.26	0.28	0.79	0.48	0.42	446
20			0.45	-1.71	0.26	0.64	0.87	0.36	0.34	446
21			0.39	-1.69	0.27	0.44	0.82	0.17	0.20	446
22			0.46	0.41	0.27	0.52	0.64	0.29	0.29	446
23			0.43	-1.07	0.26	0.36	0.77	0.32	0.33	446
24			0.42	-1.35	0.27	0.57	0.79	0.27	0.27	446
25			0.44	-1.77	0.27	0.34	0.86	0.29	0.28	446
26			0.48	0.15	0.26	0.65	0.69	0.35	0.35	446
27			0.48	1.81	0.25	0.54	0.46	0.28	0.25	446
28			0.54	1.63	0.23	0.91	0.40	0.39	0.37	446
29			0.44	-0.42	0.27	0.77	0.75	0.30	0.30	446
30			0.39	0.31	0.27	0.69	0.67	0.19	0.23	446
31		P	0.53	-3.00	0.26	0.39	0.94	0.31	0.31	446
32			0.45	-1.55	0.26	0.58	0.85	0.35	0.32	446
33			0.37	-0.97	0.27	0.79	0.75	0.21	0.22	446
34			0.44	-1.05	0.26	0.25	0.77	0.34	0.35	446
35			0.42	-1.16	0.26	1.28	0.76	0.32	0.31	446
36			0.42	1.89	0.26	0.83	0.43	0.19	0.15	446
37			0.53	2.43	0.25	0.93	0.34	0.16	0.16	446
38			0.40	-0.14	0.26	0.58	0.65	0.27	0.26	446
39			0.45	0.25	0.25	0.39	0.61	0.36	0.34	446
40			0.37	0.83	0.26	0.99	0.55	0.23	0.20	446
41			0.40	-0.52	0.26	0.70	0.70	0.27	0.28	446
42			0.48	-1.73	0.26	0.44	0.85	0.37	0.37	446
43			0.40	1.48	0.27	0.82	0.48	0.19	0.16	446
44			0.47	0.47	0.26	0.54	0.58	0.30	0.31	446
45			0.42	-1.56	0.27	0.36	0.83	0.24	0.27	446
46			0.45	-2.55	0.27	0.36	0.90	0.24	0.26	446
47			0.43	-2.47	0.27	0.82	0.89	0.23	0.23	446
48			0.43	-2.44	0.27	0.50	0.88	0.23	0.24	446
49			0.35	-1.68	0.27	0.89	0.80	0.11	0.13	446
50			0.42	-0.17	0.27	0.49	0.66	0.24	0.25	446
51		P	0.52	-3.00	0.27	1.38	0.96	0.18	0.17	446
52		P	0.45	-3.00	0.27	0.82	0.95	0.11	0.14	446
53			0.40	-0.90	0.27	0.48	0.74	0.29	0.27	446
54			0.43	-0.61	0.26	0.68	0.70	0.33	0.32	446
55			0.50	0.11	0.25	0.43	0.61	0.39	0.39	446
56			0.38	-0.48	0.27	0.50	0.69	0.22	0.21	446
57		P	0.54	-3.00	0.26	1.00	0.95	0.29	0.32	446
58		P	0.45	-3.00	0.27	0.49	0.93	0.17	0.18	446
59			0.36	-1.76	0.27	1.22	0.82	0.15	0.15	446
60			0.43	-2.10	0.27	0.29	0.86	0.26	0.25	446

COMENTARIOS

El método TRI facilita muchísimo la selección de los ítems, ya que suelen aparecer unas letras en la *columna Flag*, que definen automáticamente las características de los ítems. Las letras que suelen aparecer son estas:

- **P:** que indican si el *ítem* es *potencialmente problemático*, por alguna de estas razones:
a valor <0.30 **b** valor >2.95 **b** valor <-2.95 **c** valor >0.40
-**K:** que indica un posible *error en la selección de la clave*, ya que uno de los distractores obtiene una correlación más alta que la clave, -**R:** que indica que el *valor residual promedio del ajuste* excede al **valor 2**, que es el margen promedio de diferencia máximo permitido entre la CCI real y la CCI teórica.

Observando las letras que aparecen en este análisis de ítem de la prueba IAUB2, se constata que sólo seis ítems tienen la letra **P** (12, 31, 51, 52, 57 y 58) y que todos ellos presentan el mismo problema (**b=** -3.00) es decir, que más del 95 % de sujetos aciertan el ítem, por lo que su capacidad de discriminación es demasiado baja y que a la *diferencia* entre los sujetos que aciertan el ítem que forman el grupo superior (27%) y el inferior (27%) es nula o muy pequeña.

Se he señalado, que la prueba IASUB2 es demasiado fácil. A esta misma conclusión se llega, si nos fijamos en los valores del parámetro **b**, ya que se encuentra que el 73% (44) de los ítems tienen valores negativos y sólo el 27% (16) los tienen positivos. A esta misma conclusión se llega si se observa en el cuadro "Final Parameter Sumary Information", visto antes, en el que se indica que la **Media** de los valores del parámetro **b** es -1.00 y la **Desv Tip.** 1.53, resultados que permite inferir que el *nivel de aptitud* que debe tener el sujeto para responder en general a la prueba es bajo.

La tabla ITEM PARAMETER ESTIMATES W/STANDARD ERRORS, que ofrece estos datos de cada ítem:

- los valores de los parámetros **a, b** y **c** con sus respectivos **errores**,

- el **valor residual promedio** entre el ajuste (fit) de la CCI real y teórica

ITEM PARAMETER ESTIMATES W/STANDARD ERRORS

Item	Lnk	Flg	a	error	b	error	c	error	Resid
1			0.39	0.125	-1.49	0.205	0.27	***	0.61
2			0.48	0.093	-2.52	0.216	0.27	***	0.16
3			0.43	0.092	-2.74	0.237	0.27	***	0.71
4			0.41	0.095	-2.61	0.237	0.27	***	1.11
5			0.36	0.121	-1.71	0.223	0.27	***	0.69
6			0.45	0.098	-2.19	0.207	0.27	***	0.28
7			0.43	0.096	-2.44	0.223	0.27	***	0.71
8			0.39	0.105	-2.13	0.224	0.27	***	0.79
9			0.46	0.102	-1.95	0.196	0.26	***	0.49
10			0.47	0.159	2.50	0.301	0.26	***	0.74
11			0.42	0.170	1.69	0.242	0.26	***	0.73
12		P	0.57	0.093	-3.00	0.239	0.27	***	1.43
13			0.45	0.095	-2.42	0.218	0.27	***	0.30
14			0.43	0.194	1.12	0.207	0.27	***	0.85
15			0.35	0.107	-2.24	0.244	0.27	***	1.41
16			0.42	0.096	-2.45	0.227	0.27	***	1.05
17			0.39	0.190	-0.50	0.195	0.27	***	0.61
18			0.40	0.205	1.04	0.218	0.27	***	1.04
19			0.50	0.153	-0.61	0.156	0.26	***	0.28
20			0.45	0.107	-1.71	0.189	0.26	***	0.64
21			0.39	0.117	-1.69	0.211	0.27	***	0.44
22			0.46	0.218	0.41	0.176	0.27	***	0.52
23			0.43	0.138	-1.07	0.183	0.26	***	0.36
24			0.42	0.126	-1.35	0.193	0.27	***	0.57
25			0.44	0.108	-1.77	0.195	0.27	***	0.34
26			0.48	0.207	0.15	0.163	0.26	***	0.65
27			0.48	0.161	1.81	0.227	0.25	***	0.54
28			0.54	0.158	1.63	0.197	0.23	***	0.91
29			0.44	0.182	-0.42	0.177	0.27	***	0.77
30			0.39	0.242	0.31	0.203	0.27	***	0.69
31		P	0.53	0.092	-3.00	0.241	0.26	***	0.39
32			0.45	0.114	-1.55	0.187	0.26	***	0.58
33			0.37	0.159	-0.97	0.208	0.27	***	0.79
34			0.44	0.136	-1.05	0.177	0.26	***	0.25
35			0.42	0.134	-1.16	0.188	0.26	***	1.28
36			0.42	0.164	1.89	0.254	0.26	***	0.83
37			0.53	0.163	2.43	0.283	0.25	***	0.93
38			0.40	0.217	-0.14	0.192	0.26	***	0.58
39			0.45	0.216	0.25	0.173	0.25	***	0.39
40			0.37	0.225	0.83	0.222	0.26	***	0.99
41			0.40	0.184	-0.52	0.190	0.26	***	0.70
42			0.48	0.105	-1.73	0.182	0.26	***	0.44
43			0.40	0.180	1.48	0.236	0.27	***	0.82
44			0.47	0.213	0.47	0.173	0.26	***	0.54
45			0.42	0.117	-1.56	0.197	0.27	***	0.36
46			0.45	0.093	-2.55	0.223	0.27	***	0.36
47			0.43	0.095	-2.47	0.223	0.27	***	0.82
48			0.43	0.095	-2.44	0.222	0.27	***	0.50
49			0.35	0.126	-1.68	0.230	0.27	***	0.89
50			0.42	0.209	-0.17	0.184	0.27	***	0.49
51		P	0.52	0.092	-3.00	0.242	0.27	***	1.38
52		P	0.45	0.091	-3.00	0.249	0.27	***	0.82
53			0.40	0.154	-0.90	0.191	0.27	***	0.48
54			0.43	0.168	-0.61	0.178	0.26	***	0.68
55			0.50	0.200	0.11	0.157	0.25	***	0.43
56			0.38	0.197	-0.48	0.202	0.27	***	0.50
57		P	0.54	0.092	-3.00	0.240	0.26	***	1.00
58		P	0.45	0.091	-3.00	0.249	0.27	***	0.49
59			0.36	0.120	-1.76	0.227	0.27	***	1.22
60			0.43	0.101	-2.10	0.208	0.27	***	0.29

COMENTARIO

También se ha indicado antes que el método TRI es más preciso que el TCT, ya que ofrece la estimación del error que subyace en la estimación de los tres parámetros **a** y **b** y el valor residual media del ajuste (fit) en la CCI real y de la CCI teórica, como puede verse en la tabla "ITEM PARAMETER ESTIMATES W/STANDARD ERROR". Esta precisión se evidencia en el hecho que la TCT engloba todos los errores sistemáticos en el constructo "error típico de medida" (SEM), en cambio la TRI especifica un error propio para los parámetros a y b del ítem.

II.4 LA TABLA ITEM ANALYSIS: EN LA QUE SE OFRECEN DOS SECCIONES PARA CADA UNA DE LAS ALTERNATIVAS

El porciento de sujetos que responden en cada alternativa:	*La correlación ítem-theta en cada una de las alternativa*

ITEM ANALYSIS

	Endorsement Rate					Item-Theta Corr.				
Item	**1**	**2**	**3**	**4**	**Oth**	**1**	**2**	**3**	**4**	**Oth**
1	79~	10	9	1	1	22~	-10	-17	-11	2
2	7	90~	2	1	1	-20	29~	-24	-1	-2
3	4	90~	2	1		-24	21~	-06	-3	0
4	2	7	89~	2		-26	-2	20~	-16	-3
5	10	80~	6	2	2	-9	17~	-8	-6	-8
6	5	5	3	87~	1	-8	-14	-26	27~	-2
7	1	88~	6	4		-11	21~	-13	-11	-5
8	85~	3	3	9	1	19~	-18	-13	-5	-3
9	83~	7	2	2	6	32~	-17	-30	-2	-12
10	1	35~	61	2		-4	13~	-3	-27	-6
11	41~	7	23	17	12	16~	-6	-7	-20	1
12	01~	1	0	98~	00	-3	-7	2	09	0
13	2	2	2	87~	6	-18	-12	-18	25~	-6
14	19	19	47~	3	13	-4	-13	21~	-13	-6
15	84~	11	1	3	1	11~	-10	3	-7	2
16	7	88~	4	2		0	23~	-22	-29	
17	15	5	4	67~	9	-17	-18	-1	23~	-2
18	11	9	13	46~	20	-11	-9	-4	21~	-8
19	4	10	66~	4	16	-29	-31	42~	-5	-10
20	2	5	4	80~	8	-5	-11	-36	34~	-12
21	81~	13	1	3	1	20~	-14	-16	-7	-4
22	6	13	54~	11	16	-19	-16	29~	-4	-10
23	74~	10	9	4	3	33~	-26	-6	-16	-11
24	20	0	1	78~		-25	-13	-1	27~	-3
25	2	3	9	82~	4	-9	-11	-22	28~	-7
26	56~	7	8	11	19	35~	-24	-17	-5	-13
27	33~	9	14	16	28	25~	-13	-5	-7	-9
28	34~	3	5	42	16	37~	-9	-32	-12	-8
29	3	65~	13	4	14	-22	30~	-11	-15	-10
30	17	55~	2	8	18	0	23~	-18	-9	-17
31	3	94~	2	1		-22	31~	-13	-19	-3
32	79~	5	2	6	7	32~	-14	-25	-19	-5
33	1	3	20	73~	3	-16	-17	-7	22~	-15
34	14	4	4	75~	3	-11	-24	-22	35~	-13
35	20	2	76~	2	1	-16	-34	31~	-17	-3
36	50	5	42~	2	1	-3	-16	15~	-22	2
37	4	50	11	33~	2	-15	1	-13	16~	-5
38	13	64~	5	16	2	-6	26~	-26	-12	-4
39	7	17	57~	12	6	-11	-25	34~	-12	-4
40	3	37	54~	4	2	-29	0	20~	-23	-5
41	9	68~	9	11	3	-18	28~	-19	-3	-9
42	5	2	7	83~	4	-20	-18	-19	37~	-12
43	47~	11	38	2	2	16~	-8	-9	-8	
44	9	56~	22	9	4	-9	31~	-14	-15	-12
45	79~	5	4	7	5	27~	-10	-7	-20	-11
46	2	89~	1	7		-3	26~	-19	-22	
47	89~	5	6	0		23~	-20	-13	6	3
48	4	2	6	88~		-11	-13	-16	24~	
49	80~	3	6	10	1	13~	-10	-13	0	-6
50	65~	1	24	8	2	25~	-5	-17	-12	-4
51	1	2	0	96~		-14	-12	1	17~	
52	94~	1	4	0	1	14~	-8	-11	0	-5
53	25	0	73~	1		-21	4	27~	-29	
54	70~	20	5	4		32~	-27	-10	-7	-3
55	2	60~	5	31	2	-13	39~	-20	-26	-5
56	2	27	69~	2		-9	-15	21~	-12	-5
57	0	3	1	95~	1	-11	-11	-31	32~	-11
58	1	93~	3	3		2	18~	-16	-13	-2
59	2	15	80~	1	1	-20	-8	15~	-3	1
60	2	86~	6	6	1	-23	25~	-16	-7	-4

COMENTARIOS

Los datos que aparecen en la tabla ITEM ANALYSIS resulta esencial para analizar la fortaleza de los ítems de la prueba y para mejorar la capacidad información de algunos de acuerdo a esos criterios referentes:

-al **porciento de respuestas** dadas a las alternativas:

-el mayor porcentaje debe estar en la clave (~)

-luego el resto de alternativas deben tener porcientos armónicamente distribuidos, de modo que el poder de atracción de los distractores dependa de la cercanía conceptual que se quiera dar a los mismos.

-sería recomendable modificar aquellas alternativas que tienen igual porciento dentro de cada ítems

-finalmente se deben cambiar aquellas alternativas que tengan porcentajes o muy bajos, ya que su poder de atracción es irrelevante.

-al **valor de las correlaciones ítem-theta:**

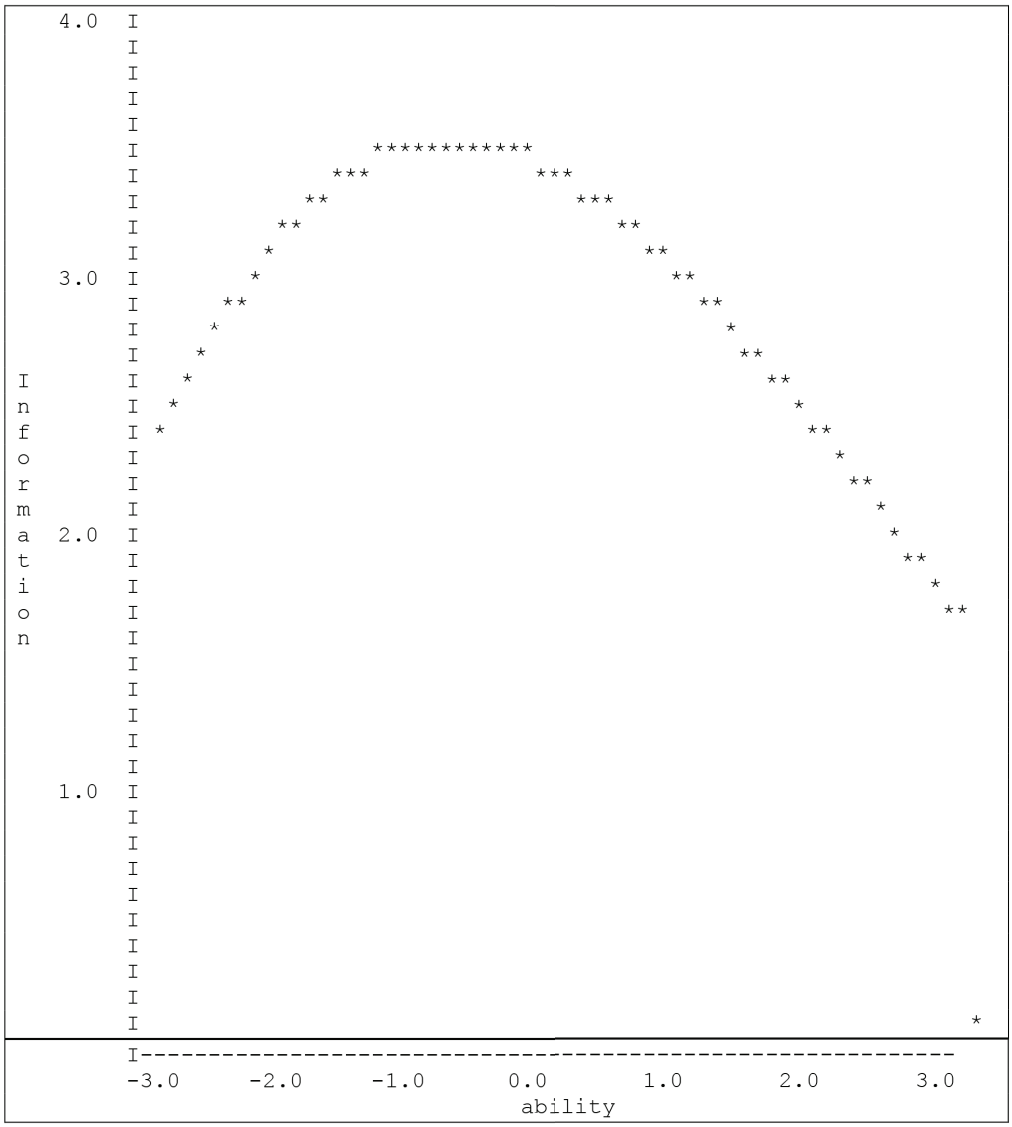
- la correlación en la clave debe ser positiva y significativa de acuerdo al grado de libertad adoptados

- las correlaciones de los distractores deben ser negativas y en cuanto sea posible significativa de acuerdo a los grados de libertad adoptados

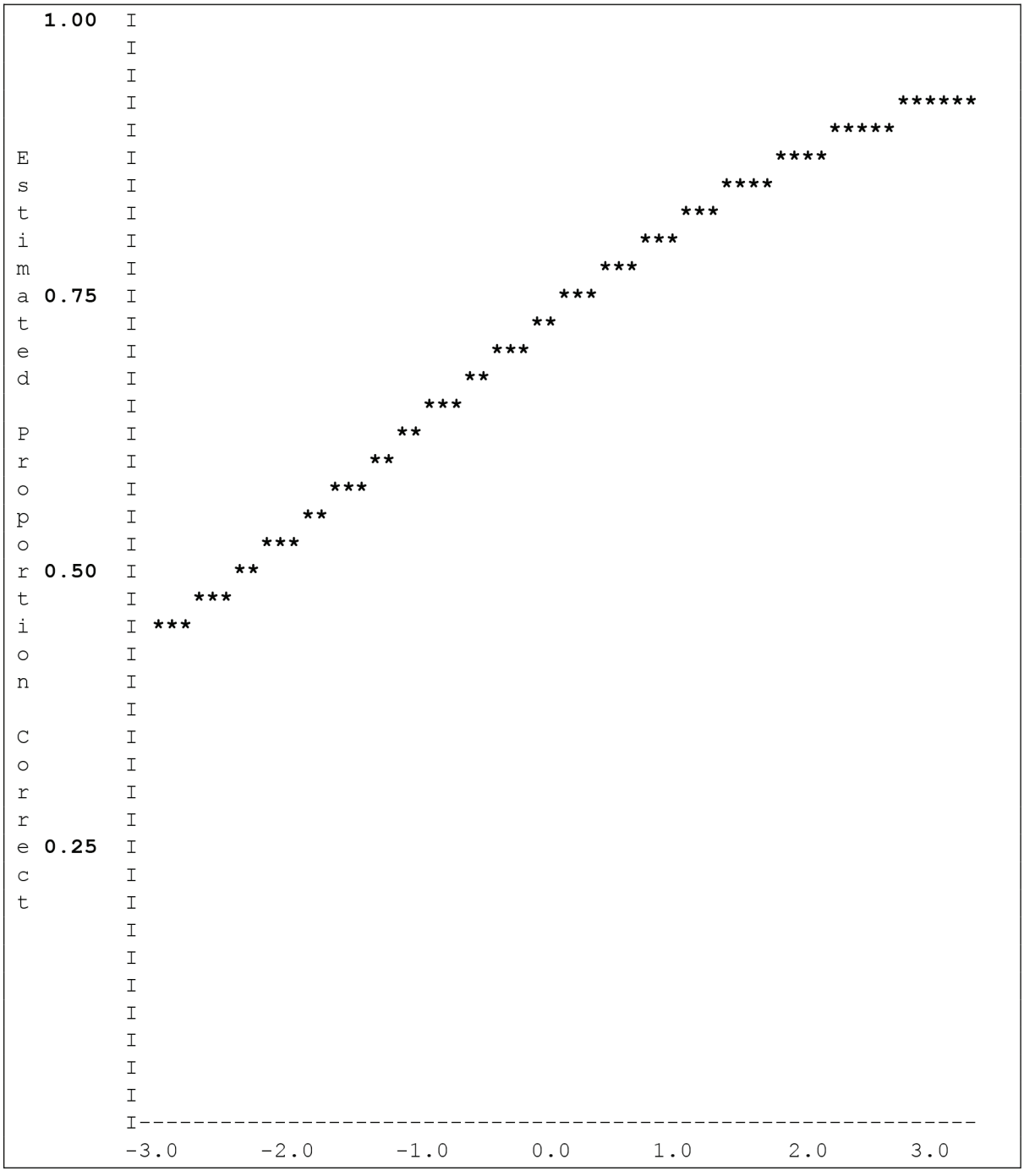
X-B.3.4 GRAFICAS DEL "TEST INFORMATION CURVE"

Test characteristics:	K-R 21 Reliability	Expected Information	Average Information
	0.763	3.223	2.869

Test Information Curve



TEST CHARACTERISTIC CURVE



REFERENCIAS

Ackerman, P.L. (1992). A didactic explanation of item bias, item impact and item validity from multidimensional perspective. *Journal of Educatinal Measurement* , 27, 241-253.

Angoff, W.H. (1982). Uses of difficulty and discrimination indices for detecting item bias. En R.A. Berk (Ed.). *Handbook of Methods for detecting item bias* (pp. 96-11). Baltimore: Johns Hopkins. University Press.

Barbero, 1. (1996). Los bancos de ítems. En J. Muñiz (Coord.), Psicometría, 139-170, Madrid: Universitas.

Baker, F.B. (1992). *Item response theory: Parameter Estimation. techniques*. New York: Marcel Dekker.

Binet, A. .(1911). Nouvelles Recherches sur la measure du niveau intellectuel chez les enfants d'echole. *Revue Philosophie*,11,191-244.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. En F.M. Lord & M.R. Novick. *Statistical theories of mental test scores* . Reading M.A.: Addison Wesley.

Bock, R.D, Mislevy, R. J. y Woodsen, C.E.M. (1982). The next stage in educational assessment. *Educational Researcher*, 11, 4-11, 16.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two o more categories. *Psychometrika*, 37, 29-51.

Bock, R.D. (1997). The Nominal Categories Model. En W.J. van der Linden y R.K Hambleton (Eds.).*Handbook of Modern Item Response Theory* (pp. 33-51). New York: Springer-Verlag.

Bock, R.D. Muraki E. y Pfeiffenbarger, W. (1988). Item pool miantenance in the presence of item parameter drift. *Journal of Educational Measurement*, 25, 275-285.

Bock, R.D. y Aitkin, M. (1981).Marginal maximum likelihood estimation of item parameters: Application of an EM algoritmo. *Psychometrika*, 46, 443-459.

Bock, R.D. y Wood, R.C. (1971). Test Theory,*Annual Review of Pshcology*, 22 193-224.

Bock, R.D. y Zimowski, M.F. (1997). Multiple Group IRT. En W.J. van der Linden y R.K Hambleton (Eds.).*Handbook of Modern Item Response Theory* (pp. 433-448). New York: Springer-Verlag.

Budgell, G.R. Raju. N.S. y Quartetti, D.A. (1995). Analysis of differential item functioning in traslated assessment instrument. *Applied Psychological Measurement*, 19, 309-321.

Camilli, G. (1993). *A critique of the chi-square method for assessing item bias*. Laboratory of Educational Research, University of Colorado.

Camilli, G. y Shepard, L.A. (1994). *Methods for identifing biased test items*. Thousand Oaks, CA: Sage Publications.

Candell, G.L y Drasgow, F. (1988). An iterative procedure for linking matrics and assessing item bias in item response theory. *Applied Psychological Measurement*, 12, 253-260.

Díaz, J.V. (1997): *La Teoría de las respuestas a los ítems aplicada a la construcción de tests aptitudes* .Valencia, Cristobal Serrano.

Dorans, N.J. Schmitt, A.P. y Bleistein, C.A (1992). The standardization approach to asessing comprehensive differential item functioning. *Journal of Educational Measurement*, 29, 309-319.

Eells, K., Davis, A., Havighurs, R. J., Herrick,V.E., Tyler R.W. (1951): *Intelligence and Cultural Differences: A study of Cultural Learning and Problem-Solving*. University of Chicago Press-

Gómez, J. y Navas, M.J. (1996). Detección del funcionamiento differencial de los items mediante regresión logística: purificación paso a paso de la habilidad. Psicológica, 17, 397-411.

Hambleton, R. K. y Swaminathan, H. (1985). *Item response theory: Principles and Application.* Boston: Kluwer-Nijhoff Publishing.

Hambleton, R.K. Swaminathan, H. y Rogers, H.J. (1991). *Fundaments of Item response theory* Newbury: SAGE Publications.

Hattie, J.A. (1985). Methodological review: assessing unidimensonality of tests and items. *Applied Psychological Measurement*, 9, 139-164.

Hidalgo, M.D. y López Pina, J.A. (1997). Evluación del funcionamiento diferencial en items politómicos mediante el estadístico de Lord y las medidas de áreas. *Psicológica*, 18, 69-92.

Holland, P.W y Thayer, D.T. (1988). Differential item performance and the Mental-Haenszel procedure. En H. Wainer y H.I. Braun (Eds.). *Test Validity* (pp. 129-145).Hillsdale, NJ: Lawrence Erlbaum.

Holland, P.W y Wainer H. (Eds.) (1993) *Differential item functioning: Theory and practice*. Hillsdale, NJ: Lawrence Erlbaum.

Hulin, C.L. Drasgow, F. y Parsons, C.K. (1983). Item response theory. Homewood, I11: Dowjones Irwin.

Ironson, G.H. y Subkoviak, M.J. (1979). A comparison of several method of assessing item bias. *Journal of Educational Measurement*, 16, 209-225.

Jensen, A.R. (1979). g: Outmoded or unconquered frontier?*Creative Science and Technology*, 2 , 16-29.

Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.

Kim, S. Cohen, A.S. y Park, T (1995). Detection of differential item functioning in multiple groups.*Journal of Educational Measurement*, 32, 261-276.

Kim, S. y Cohen, A.S. (1991). A comparison of two area measures for detecting deifferential item functioning. *Applied Psychological Measurement*, 15, 269-278.

Kok, F.G, Mellenbergh, G.J.y van der Flier, H. (1985). Detecting experimentally induced item bias using th iterative logit method. *Journal of Educational Measurement*, 18, 1-11.

Lawley, D.N. (1943). On problem connected with item selection and test construction. *Preecedings of the Royal Society od Edinburgh*, 6, 273-287.

Linn, R.L. y Harnisch, D.L. (1981). Interactions between item content and group membership o achivement test item. *Journal of Educatonal Measurement*, 18, 109-118.

Lord, F. M. (1952). *A theory of test scores*. (Psychometric Monograph, nº 7) Iowa City, IA: Psychometric Society.

Lord, F. M. and Novick M. R. (1968). *Statistical of theories of mental test scores.* Reading Mass. Addison-Wesley.N.Y.

Lord, F.M.(1975). The "ability" scale in item characteristic curve theory. *Psychometrica*, 40 , 205-217.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, N. J. Lawrence Erlbaum Associates.

Martínez Arias, R.(1995): *Psicometría: Teoría de los tests psicológicos y educativos.* Madrid. Sintesis Psicología.

Maydeu, A. (1996). Modelos multidimensionales de la Teoría de respuesta a los itemes. En J. Muñiz (ED.). *Psicometría*. (pp. 811-868). Madrid. Universitas.

Messick, S, Beaton, A.E. y Lord, F.M. (1983). *National Assessment of Educational Progress reconsidered: A new desing for a new era.* (NAEP Rep-83-1). Princeton, NJ: National Assessment of Educational Progress.

Mislevy, R. J. y R. D. Bock (1990): BILOG 3. *Item Analysis and Test Scoring with Binary Logistic Models.* (2 eds), Moresville, IN: Scientific Software

Mislevy, R.J. (1993b): Foundation of a new test theory. En N.F. Frederiksen, R.J. Mislery and I.I. Bejar (Eds). *Test theory for a new generation of tests*. Hillsdale, N. J. Lawrence Erlbaum.

Múñiz Fernández , J. (1990). Teoría de la Respuesta a los ítems: Un nuevo enfoque en la evolución psicológica y educativa. Ediciones Pirámide Madrid.

Múñiz Fernández , J. (1996). Fiabilidad. En J. Muñiz (Ed),*Psicometría*. Madrid: Universitas.

Múñiz Fernández, J. (1997). Desarrollos y perspectivas en la psicometría actual. Sevilla. Ponencia presentada en el.*V Congreso de Metología de las CC. Humanas y Sociales.*.

Park, D. y Lautenschlager, G.J. (1990). Improving IRT item bias dtection wiyh iterative linking and ability scale purification. *Applied Psychological Measurement*, 14, 163-173.

Potenza, M.T. y Dorans, N.J (1995). DIF assessment for polythomously scored items: framework for classification and evaluation. *Applied Psychological Measurement*, 19, 23-37.

Prieto, G (1997). *Nuevas perspectivas en la construcción de tests*. V Congreso de Metodología de las Ciencias Humanas, Sevilla.

Prieto, G y Delgado, A (1996). Construcción de Items. En J..Muñiz (Ed.) *Psicometría*. Madrid: Universitas.

Raju, N.S. (1988). The area between two item charcteristic curves. *Psychometrika*, 53, 495-502

Raju, N.S. (1990). Determining the significance os estimated signed and unsigned areas between tow item response functions. *Applied Psychological Measurement,* 14 (2), 197-207.

Rasch, G.A. (1960, 1980). *Probabilistic models for some intelligence and attainment test*. Chicago University Press. (reedición del original Danmarks Pedagogiske Institute Copenhgen).

Rasch, G. A. (1968). Mathematical theory of objetivity and its consequences for model construction. Amsterdam: *a report from the European Meeting on Statistic, Econometrics, and Manegement Sciences.*

Reckase, M.D. (1979). Unifactor trait model applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230.

Reckase, M.D. (1983). *The definition of dificulty and discrimination for multidimensional item response theory models*. Paper presented at the meetting of American Educational research Association, Montreal, Canada.

Reckase, M. D. (1993). *The definition of dificulty and discrimination for multidimensional item response theory models*. Paper presented at the metting of American Educational research Association, Montreal, Canada.

Reckase, M. D. (1997). The past and future of multidimensional Item Response *Theory. Applied Psychological Measurement*, 21. 25-26.

Reynolds, C.R. y Brown, R.T. (1984). *Perspectives on bias in mental testing*. New York: Academic Press.

Santisteban, C (1990). Psicometría: teoría y práctica en la construcción de tests. Madrid: Norma.

Shealy, R. y Stout, W.F (1993). A model-based standardiztion approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58, 159-194.

Smith, P.L. (1978). Sampling errors of variance components in small sample generalizabiity indeces. J . *of Educational Statistics,* 3, 319-346.

Stern, W. (1914). *The Psychological Methods of Testing Intelligence*. Educational Psychol. Monograph. No 13. Baltimore: Warwick and York.

Stout, W. Li, H-N, Nandakumar, R. y Bolt, D. (1997). MULTISIB: A procedure to investigate DIF when a test i intentionally tow-dimensional. *Applied Psychological Measurement*, 21, 195-214.

Tatsuoka, K.K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354.

Theunissen, T.J.M. (1985). Binnary programing and test design. *Psychometrika*, 50, 411-420.

Thissen, D. (1988). MULTILOG. *Multipe, Categorical Item Analysis and Testing Scoring using Item Response Theory.* (Version 5.1). Moresville, IN: Scientific Software

Thissen, D y Steinberg, L. (1984). A response model for multiple-choice items. *Psychometrika*, 49, 501-519.

Thissen, D. Steinberg, L. y Wainer, H. (1993). Detection of differential item functioning using the paramaters of items response models. En P.W. Holland y H. Wainer (Eds.). *Differential item functioning*. Hillsdale, NJ: LEA:.

Tucker, L. R. (1946). Maximum validity of test with equivalent items. *Psychometrika*, 11, 1-13.

van der Flier, H, Mellenbergh, G.J, Ader, H.J. y Vijn, M. (1984). An iterative item bias detection method. *Journal of Educational Measurement*, 21, 131-145

Welch, C.J. y Miller, T.R. (1995). Assessing differential item functioning in direct wrinting assessments: problems and an example. *Journal of Educational Measurement*, 32, 163-178.

Wilson, M. R. (1989). SALTUS: A psychometric model of discontinuity in

Wright, B. y Panchapakesan, N. (1969). A procedure for sample-free item analysis.*Educationl and Psychological Measurement* 29 ,23-48.

Wright, B. D. y Mead, R.J. (1976). *BICAL: Calibration rating acores with the Rasch Model.* (Research Memorandum, nº 23) Chicago: Chicago Statistical Laboratory, Departament od Education, University of Chicago.

Wright, B. D. y Stones, M. H. (1979). Best test design. Chicago: MESA Press.

Wright, B.D. y Masters, G. N. (1982). *Rating Scale Analysis.Rasch Measurement*. Chicago: MESA.

Wright, B.D. y Linacre, J. M. (1992). *User's guide to BIGSTEPS; Rasch- Model computer program*. Chicago: MESA Press,

Zwick, R. (1997). The effect of adaptative administration on the variability of Mentel-Haenszel measurement of differential item functioning. *Educational and Psychological Measurement,* 57, 412-421.

Zwick, R. Donoghue, J.R. y Grima, A. (1993). Assessment of differentiall item functioning for performance tasks. *Journal of Educational Measurement*, 30 ,233-251.

Zwick, R., Thayer, D.T. y Wingersky, M. (1995). The effect of Rasch Calibration of Ability. and DIF on adaptative tests. Journal on Educational *Measurement*, 32, 341-363.