

# **Análisis de ítemes de estadística en el marco de la Teoría de Respuesta al ítem.**

María Silvia Galibert; Carlos O.Pano; Horacio F. Attorresi y María Ester Aguerri.

Cita:

María Silvia Galibert; Carlos O.Pano; Horacio F. Attorresi y María Ester Aguerri (2004). *Análisis de ítemes de estadística en el marco de la Teoría de Respuesta al ítem. VI Jornadas de Sociología. Facultad de Ciencias Sociales, Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-045/448>

## **Análisis de ítems de estadística en el marco de la Teoría de Respuesta al ítem.**

María Silvia Galibert; Carlos O.Pano; Horacio F. Attorresi y María Ester Aguerri

Instituto de Investigaciones. Facultad de Psicología - Universidad de Buenos Aires

[galibert@psi.uba.ar](mailto:galibert@psi.uba.ar)

Objetivos: Mostrar cómo los modelos de la Teoría de Respuesta al Ítem permiten analizar las respuestas a ítems de tipo multiple choice. Temática principal: Se presentan modelos para ítems politómicos y en particular se aplica el modelo de Thissen & Steinberg al estudio de ítems de una prueba objetiva.

Metodología: Se analizan los ítems de un examen de estadística respondidos por 699 estudiantes de carrera de Psicología. Se muestran las curvas características (CC) de las opciones del ítem al ajustar el modelo de Thissen & Steinberg. Estas curvas representan la probabilidad de elegir la opción en función del nivel de habilidad. El modelo presenta una curva adicional llamada DK (Don't Know) que representa la probabilidad de contestar sin saber.

Resultados y Conclusiones: El análisis de las CC resultó útil para orientar posibles cambios tanto para mejorar los ítems como para reforzar algunos aspectos en el dictado de la materia. Se observó que la CC de la respuesta correcta resultó creciente en la mayoría de los ítems. Se deberán reformular algunas de las opciones en ítems donde la respuesta correcta resultó, por efecto del azar, ser la más elegida en niveles bajos de habilidad y aquellas que presentan CC constantes.

## INTRODUCCIÓN

La buena calidad de una prueba comienza con la buena calidad de sus ítemes; por ello es tan importante la fase del análisis de ítemes. Los principales criterios por tener en cuenta en esta fase son que la opción correcta (clave) haya sido más elegida que cualquier distractor y que la correlación ítem – test sea positiva. En cuanto a la distribución de frecuencias de los distractores, se prefiere que sea aproximadamente uniforme. Esta preferencia es consistente con el hecho de que las respuestas a los ítemes suelen puntuarse de manera dicotómica con 1 ó 0 según sean correctas o incorrectas. La puntuación 0 no discrimina entre los distractores. Si entre ellos hubiera alguno más próximo a la opción correcta, elegirlo no contribuiría más al puntaje del sujeto que el haber elegido un distractor más lejano. La dicotomización, pues, pierde la información que podría estar contenida en los distractores. Esta pérdida no sería tal si todos ellos fueran “igualmente” lejanos porque no aportarían información diferente. La equidistancia se refleja en la homogeneidad de las frecuencias con las que son elegidos. La adopción de un modelo de distribución uniforme para los distractores hace que un sistema de puntuación dicotómica sea prácticamente equivalente a un sistema de puntuación que ponderara los distractores; ya que distractores igualmente lejanos tendrían el mismo peso. Sin embargo, a la hora de inventar distractores, surgen naturalmente algunos más cercanos a la clave que otros y esto no sólo no debería ser un defecto del ítem sino que puede aportar información de interés desde el punto de vista cognitivo, si se discriminaran las diferentes respuestas erróneas. La Teoría de Respuesta al Ítem (TRI) ha hecho importantes contribuciones en este sentido a través de los modelos de respuesta para ítemes politómicos (van der Linden y Hambleton, 1997). Dichos modelos describen la probabilidad de elección de cada alternativa en

función del nivel del rasgo que se desea medir. De la representación gráfica de las funciones de respuesta se obtienen curvas cuyo estudio enriquece y complementa el análisis de ítems clásico como también puede contribuir a la comprensión de los procesos cognitivos involucrados en la respuesta al ítem.

Bock (1972) propuso un modelo para la respuesta  $X = h$  con  $h = 1, 2, \dots, m_j$  a un ítem  $j$  de elección múltiple con  $m_j$  alternativas:

$$P(x_j = h / \theta; \vec{a}, \vec{c}) = \frac{\exp(a_h \theta + c_h)}{\sum_{k=1}^{m_j} \exp(a_k \theta + c_k)}$$

donde  $\vec{a}$  y  $\vec{c}$  son vectores  $m_j$  dimensionales sujetos a las restricciones

$$\sum_{h=1}^{m_j} a_h = \sum_{h=1}^{m_j} c_h = 0$$

con el fin de salvar la indeterminación de la escala y hacerlos estimables.

Samejima (1979) modificó el modelo de Bock (1972) incluyendo una categoría latente a la que denominó categoría cero; es la correspondiente a las respuestas de quienes eligen una opción por azar. En este modelo  $d_h = 1/m_j$  indica la proporción de sujetos que, habiendo respondido por azar, eligen la alternativa  $h$ .

$$P(x_j = h / \theta; \vec{a}, \vec{c}) = \frac{\exp(a_h \theta + c_h)}{\sum_{k=0}^{m_j} \exp(a_k \theta + c_k)} + \frac{d_h \exp(a_0 \theta + c_0)}{\sum_{k=0}^{m_j} \exp(a_k \theta + c_k)}$$

donde  $\vec{a}$  y  $\vec{c}$  son vectores  $m_j+1$  dimensionales sujetos a

$$\sum_{h=0}^{m_j} a_h = \sum_{h=0}^{m_j} c_h = 0$$

Thissen y Steinberg (1984, 1997) llaman DK: "don't know" a la categoría 0 del modelo de Samejima y proponen una proporción  $d_h$  no constante para cada alternativa; por lo

que se agrega la restricción  $\sum_{h=1}^{m_j} d_h = 1$ . Así, la curva DK representa la proporción de

sujetos que contestan el ítem sin saber, en función de su nivel de habilidad y  $d_h$  la proporción de los mismos que eligió la alternativa  $h$ .

La respuesta “sin saber” puede pensarse como una respuesta “al azar” pero que no otorga, a diferencia del modelo de Samejima, la misma probabilidad a cada opción.

Caracteriza a aquellos sujetos que, dada su falta de criterio o conocimiento, pueden ser atraídos más por una opción que otra pero por una razón no pertinente a la habilidad que mide el ítem; de allí que la probabilidad de elegir una u otra alternativa no es necesariamente homogénea. El parámetro  $a$  se relaciona con el crecimiento y monotonía de las funciones. La alternativa a la que le corresponde el mayor valor de  $a$  suele ser la clave cuya curva se espera que sea monótona creciente y asintótica a 1 por la derecha. La curva DK comúnmente está asociada al valor más negativo de  $a$ , lo que implica, como es de desear, que sea decreciente y asintótica a cero por la derecha.

Valores intermedios de  $a$  producen curvas no monótonas, que suelen ser las de los distractores. El parámetro  $c$  refleja la frecuencia relativa de elección de cada alternativa. Alternativas con similares valores de  $a$  serán más frecuentemente elegidas para mayores valores de  $c$ .

El objetivo de este trabajo es presentar una aplicación del modelo “Multiple Choice” de Thissen y Steinberg al análisis de ítems de una prueba de estadística. Se exhiben algunos ejemplos que ilustran su utilidad para orientar los cambios en la mejora de los ítems como también para conocer el perfil cognitivo de los estudiantes.

## **CARACTERÍSTICAS DEL INSTRUMENTO Y DE LA MUESTRA**

La prueba de Estadística es una prueba de rendimiento, ya que constituye el primer parcial de la materia para evaluar a los alumnos de la Cátedra I de Estadística de la Facultad de Psicología, UBA. Cubre los contenidos correspondientes los temas de estadística descriptiva y modelos de probabilidad. Consiste de 20 ítems de cuatro opciones con una sola verdadera, que comprenden la evaluación de conceptos (14) y la resolución de ejercicios (6). Fue administrada a 699 estudiantes, 20% de varones y 80% de mujeres.

## **METODOLOGÍA**

**Ajuste del modelo.** Se estimaron los parámetros bajo el modelo Multiple Choice de Thissen y Steinberg mediante el programa MULTLOG<sup>TM</sup> (Thissen y Steinberg, 1991). Con el fin de aumentar la precisión de las estimaciones, se llevó a cabo una primera corrida del programa, se buscó qué estimaciones eran similares y se restringieron a la igualdad los correspondientes parámetros. Bajo estas restricciones se realizó una nueva corrida para obtener las estimaciones definitivas y se graficaron las curvas. A partir de ellas, se llevó a cabo el análisis de ítems.

## **Ejemplos de análisis del ítem sobre la base de las curvas de respuesta**

### *Ítem 1*

La distribución binomial requiere que:

- a) la variable tome sólo dos valores: éxito y fracaso.
- b) las observaciones sean estables.
- c) el éxito en una observación sea independiente del fracaso.
- d) las observaciones sean independientes.

*La clave es d).*

La distribución de frecuencias de la Tabla 1 muestra que el distractor a) resultó mucho más atractivo que la opción correcta, habiendo sido elegido por el 59% de los sujetos sobre un 15% que eligió la opción correcta d). El 21% habría emitido su respuesta sin saber, del cual la mitad habría sido atraído por b). La Figura 1 muestra que los sujetos que eligieron el distractor a) se hallan distribuidos entre casi todos los niveles de conocimiento, excepto en los extremadamente altos donde se percibe una tendencia creciente de la curva que representa la opción correcta. Este resultado no es sorprendente porque es un error común que los alumnos confundan la variable Bernoulli con la Binomial. La curva DK, por otra parte, no tiene el comportamiento esperado (eventualmente alta en los niveles bajos y luego decreciente llegando a ser nula en los niveles de medio a altos) sino que se mantiene más o menos constante a lo largo de todo el rango de variación del rasgo latente. Éste es un ejemplo de ítem que no funciona “estadísticamente bien” pero la solución no está en modificarlo sino en reforzar la enseñanza para que los alumnos puedan relacionar al mismo tiempo que diferenciar ambos modelos.

#### *Ítem 4*

*Completar la siguiente frase:*

La moda.....la partición adoptada para las categorías de la variable.

- a) determina
- b) representa a
- c) influye en
- d) depende de

*La clave es d).*

La Tabla 2 muestra que el 47% eligió la respuesta correcta; es decir, es un ítem de dificultad media. Sin embargo, la mitad de los sujetos habría hecho su elección sin saber, de los cuales el 48% habría elegido la opción correcta. La Figura 2 corrobora el efecto de la respuesta azarosa a través de la curva que la representa, que decrece demasiado lentamente. Tal vez por ello, la curva correspondiente a la clave es demasiado alta en los niveles bajos y casi constante hasta el nivel medio después del cual comienza a hacerse creciente pero sin llegar a ser asintótica a 1, como sería de desear; en efecto la probabilidad de respuesta correcta no superaría 0,8 aún en los niveles altos de conocimiento. A juzgar por las preguntas de los alumnos en parcial, tanto la respuesta azarosa como la falta de convergencia a 1, podrían explicarse por la desorientación que a muchos de ellos les provoca la palabra “partición”, concepto no muy explicitado en la bibliografía y que habría que destacar más, o bien cambiar dicha palabra por la de “clasificación”.

#### *Ítem 7*

La condición de estabilidad necesaria para que una variable tenga distribución binomial significa que en cada ensayo ...

- a) debe ocurrir el mismo suceso.
- b) la probabilidad de éxito es la misma.
- c) el resultado no es afectado por el de ningún otro ensayo.
- d) sólo puede ocurrir éxito o fracaso.

*La clave es b).*

La opción correcta fue elegida por el 55% de los alumnos, del 21% que contestó sin saber, un 28% habría elegido esta opción y un 71% la c) siendo atraídos, tal vez, por su mayor longitud (Tabla 3). El comportamiento de las curvas (Figura 3) es satisfactorio



para la opción correcta b) y para los distractores c) y d) excepto por el hecho de una ligera tendencia decreciente para la opción correcta en los niveles de conocimiento más altos, coincidente con el crecimiento de la curva de c). La opción c) se refiere al otro supuesto de la distribución binomial, la condición de independencia. Los alumnos de mayor nivel de conocimiento conocerían ambas condiciones y probablemente un grupo de ellos las haya confundido. De ser así, se plantearía la necesidad de modificar el distractor c). Sin embargo, debe tenerse en cuenta que la estimación de las curvas suele ser más imprecisa en los extremos de la escala y más confiables en los niveles intermedios. En particular en este ítem, puede observarse en la fila I(THETA) de la Tabla 3, que la mayor información se da en los niveles -0.5, 0 y 0.5; por lo que antes de aventurarse a aceptar una interpretación habría que verificar que el comportamiento indeseado no se debe a problemas en la estimación. En cuanto a la curva del distractor a), puede observarse que es prácticamente nula a lo largo de toda la escala, por lo que no funciona como tal.

### *Ítem 15*

Se estima que el 80% de los alumnos de Estadística aprobarán este parcial. La probabilidad de que en una muestra de cinco aprueben cuatro y desaprobe uno es

- a) 0,0064
- b) 0,4000
- c) 0,4096
- d) 0,5000

*La clave es c).*

A pesar de que un análisis clásico nos diría que este ítem es fácil (67% de respuesta correcta, Tabla 4), una inspección de la curva correspondiente a la clave (Figura 4) nos muestra que la probabilidad de responder correctamente no supera 0,8 en los niveles más altos de conocimiento; es decir la curva no es asintótica a 1 como es de esperar. El ítem no discrimina mayormente entre los alumnos de conocimientos medios y altos, como muestra la lentitud de su crecimiento en estos niveles. Sin embargo es útil para discriminar a los sujetos de peores rendimientos; nótese en el renglón I(THETA) de la Tabla 4, que arroja mayor información en los niveles entre -1,5 y -1. Por otra parte la respuesta azarosa, 15%, afortunadamente no favoreció a los sujetos de estos niveles los cuales se vieron atraídos mayoritariamente por los distractores a) y b) (43% y 40% respectivamente); por lo que es muy conveniente mantener este ítem como está con el fin de tener estimaciones precisas en los niveles bajos de habilidad. La opción d) no contribuye significativamente.

#### *Ítem 19*

Dada una distribución de frecuencias, para hallar la mediana es necesario encontrar dos conjuntos de valores de la variable, digamos A y B, tales que se verifique entre otras, la siguiente condición:

- a) Exactamente un elemento queda fuera de A U B.
- b) La frecuencia, tanto de A como de B, es a lo sumo  $n/2$ .
- c) Por lo menos un elemento queda fuera de A U B.
- d) La frecuencia, tanto de A como de B, es igual a  $n/2$ .

*La clave es b).*

Con un 52% de respuesta correcta (Tabla 5), es un ítem de dificultad media. Del 28% de aquéllos que habrían contestado sin saber, un elevado porcentaje, 71%, eligió la

respuesta correcta. Esto puede visualizarse en la Figura 5, donde la curva de la clave representa una probabilidad bastante elevada de respuesta correcta (del orden de 0,4) desde los niveles bajos hasta medios. Por esta razón, el ítem resulta más útil para discriminar a los sujetos de niveles de medios a altos, donde la curva crece más rápidamente, lo cual puede verificarse con la función  $I(\theta)$  que alcanza el máximo en 0,5. Con el fin de que los sujetos que contestaron sin saber se vieran más atraídos por un distractor que por la clave, se sugiere permutar el orden de las opciones b) y d). Tal vez, estando la opción d) antes que la correcta, estos sujetos se vieran más atraídos hacia ella y se disminuyera así la probabilidad de respuesta correcta por azar.

## **RESULTADOS Y CONCLUSIONES**

El análisis de las curvas de las opciones resultó útil para orientar posibles cambios tanto para mejorar los ítems como para reforzar algunos aspectos en el dictado de la materia. Se observó que la curva de la respuesta correcta resultó creciente en la mayoría de los ellos aunque no siempre asintótica a 1. Se deberían reformular algunas de las opciones en aquellos ítems donde la respuesta correcta resultó, por efecto del azar, ser la más elegida en niveles bajos de habilidad. En la mayoría de los ítems uno de los distractores no contribuyó significativamente; por lo que habría que buscar alternativas más atractivas. Con respecto a los sujetos, fue útil para precisar su perfil. Se corroboró lo que se sabía por experiencia acerca de los conceptos que suelen confundir los alumnos de Estadística. Esta información resultará muy útil a los docentes para reorientar su enseñanza desde un diagnóstico más objetivo.

## **BIBLIOGRAFIA**

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are

scored in two or more latent categories. *Psychometrika* 37, 29-51.

Samejima, F (1979). *A New Family of Models for the Multiple Choice Item* (Research Report #79-4). Knoxville, TN: University of Tennessee, Department of Psychology.

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

Thissen, D. y Steinberg, L. (1991). *MULTILOG<sup>TM</sup> User's Guide. Multiple, Categorical Item analysis and Test Scoring Using Item Response Theory*. Scientific Software, Inc.

Thissen, D. y Steinberg, L. (1997). A Response Model for Multiple-Choice Items. En , W. J. van der Linden. & R. K. Hambleton (Eds.): *Handbook of Modern Item Response Theory*. NewYork, Springer.

van der Linden, W. y Hambleton, R.H. (1997). *Handbook of Modern Item Response Theory*. Springer-Verlag New York Inc.

**Tabla 1:** Fragmento de salida traducida de MULTLOG exhibiendo la estimación de los parámetros para la obtención de las curvas de respuesta, el máximo de información que aporta el ítem para estimar cada nivel de rendimiento y la distribución de frecuencias de las opciones.

ITEM 1: CLAVE: d

CATEGORÍA(K):	DK		a	b	c	d			
A(K)	-0.59	-0.50	-0.50	-0.27		1.85			
C(K)	1.17	2.22	-2.10		0.80	-2.10			
D(K)		0.02	<b>0.50</b>		0.02		0.47		
THETA:	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
I(THETA):	0.01	0.01	0.01	0.01	0.01	0.03	0.16	<b>0.61</b>	1.11

#### FRECUENCIAS OBSERVADAS Y ESPERADAS

CATEGORÍA(K):	DK		a	b	c	d
FREC. OBS.	0	413		78	101	104
PROP. OBS.		0.00	<b>0.59</b>	0.11	0.15	0.15
PROP. ESP.		<b>0.21</b>	0.59	0.11	0.15	0.15

**Tabla 2:** Fragmento de salida traducida de MULTLOG exhibiendo la estimación de los parámetros para la obtención de las curvas de respuesta, el máximo de información que

aporta el ítem para estimar cada nivel de rendimiento y la distribución de frecuencias de las opciones.

ITEM 4:

CLAVE: d

CATEGORÍA(K):	DK	a	b	c	d	
A(K)	0.46	-4.23		0.92	0.92	1.94
C(K)	3.15	-8.91		2.05	1.66	2.05
D(K)			0.16	0.33	0.04	<b>0.48</b>

THETA:	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
I(THETA):	0.34	0.02	0.01	0.01	0.04	0.10	0.17	0.21	0.19

#### FRECUENCIAS OBSERVADAS Y ESPERADAS

CATEGORÍA(K):	DK	a	b	c	d		
FREQ. OBS.	0	61	221		86	327	
PROP. OBS.	0.00	0.09	0.32	0.12	<b>0.47</b>		
PROP. ESP.		<b>0.50</b>		0.09	0.32	0.12	0.47

**Tabla 3:** Fragmento de salida traducida de MULTLOG exhibiendo la estimación de los parámetros para la obtención de las curvas de respuesta, el máximo de información que aporta el ítem para estimar cada nivel de rendimiento y la distribución de frecuencias de las opciones.

ITEM 7:

CLAVE: b

CATEGORÍA(K):	DK		a	b	c	d			
A(K)	-0.97	-0.65		0.79	1.80	-0.97			
C(K)	0.59	-0.66		1.52	-2.05		0.59		
D(K)			0.00		<b>0.28</b>	<b>0.71</b>	0.00		
THETA:	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
I(THETA):	0.02	0.07	0.22	<b>0.43</b>	<b>0.52</b>	<b>0.37</b>	0.19	0.14	0.17

#### FRECUENCIAS OBSERVADAS Y ESPERADAS

CATEGORÍA(K):	DK	a	b	c	d	
FREQ. OBS.	0	36	385	133	143	
PROP. OBS.		0.00	0.05	<b>0.55</b>	0.19	0.21
PROP. EXP.		<b>0.21</b>	0.05	0.55	0.18	0.21

**Tabla 4:** Fragmento de salida traducida de MULTLOG exhibiendo la estimación de los parámetros para la obtención de las curvas de respuesta, el máximo de información que aporta el ítem para estimar cada nivel de rendimiento y la distribución de frecuencias de las opciones.

ITEM 15:

CLAVE: c

CATEGORÍA(K):	DK	a	b	c	d	
A(K)	-1.82	0.34	1.02		0.60	-0.14
C(K)	-0.61	0.74	-1.69		2.18	-0.61

D(K)				<b>0.43</b>	<b>0.40</b>		0.09	0.09	
THETA:	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
I(THETA):	0.25	<b>0.60</b>	<b>0.61</b>	0.29	0.09	0.03	0.03	0.02	0.02

#### FRECUENCIAS OBSERVADAS Y ESPERADAS

CATEGORÍA(K):	DK	a	b	c	d	
FREQ. OBS.	0	143		50	454	35
PROP. OBS.	0.00	0.21	0.07	<b>0.67</b>	0.05	
PROP. EXP.		<b>0.15</b>		0.21	0.07	0.66 0.05

**Tabla 5:** Fragmento de salida traducida de MULTLOG exhibiendo la estimación de los parámetros para la obtención de las curvas de respuesta, el máximo de información que aporta el ítem para estimar cada nivel de rendimiento y la distribución de frecuencias de las opciones.

ITEM 19:

CLAVE: b

CATEGORÍA(K):	DK		a	b	c	d	
A(K)		-1.47	-0.54		3.08	0.39	-1.47
C(K)		0.55		0.15	-0.42	-0.42	0.15
D(K)			0.01	<b>0.71</b>	0.28	0.01	
THETA:	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0 1.5 2.0



I(THETA): 0.04 0.07 0.09 0.12 0.34 **2.14** 1.19 0.27 0.05

#### FRECUENCIAS OBSERVADAS Y ESPERADAS

CATEGORÍA(K):	DK	a	b	c	d
FREQ. OBS.	0	91	364	105	135
PROP. OBS.	0.00	0.13	<b>0.52</b>	0.15	0.19
PROP. EXP.	<b>0.28</b>	0.13	0.52	0.15	0.19

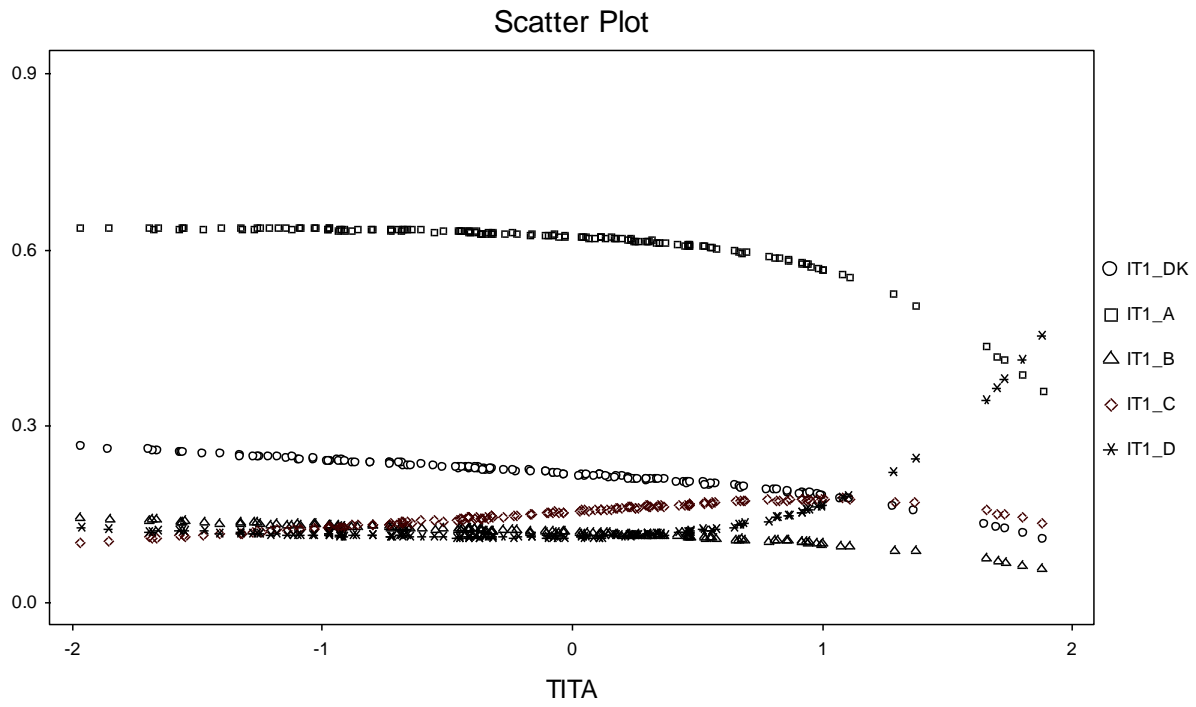
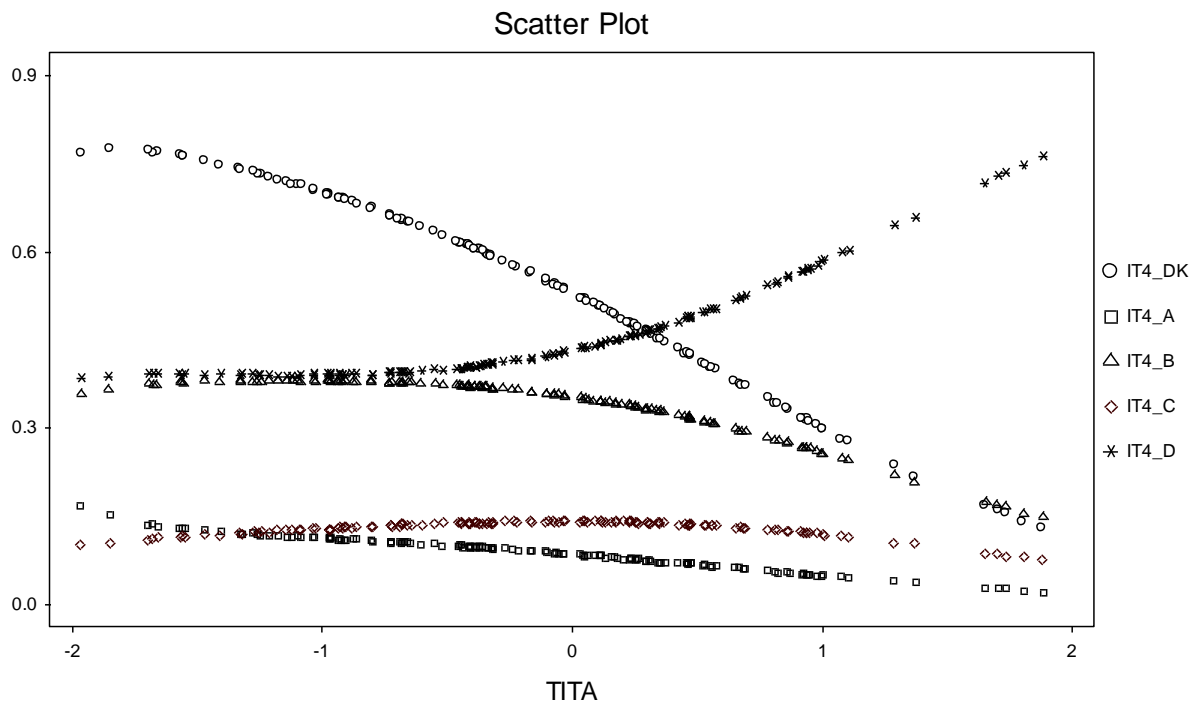
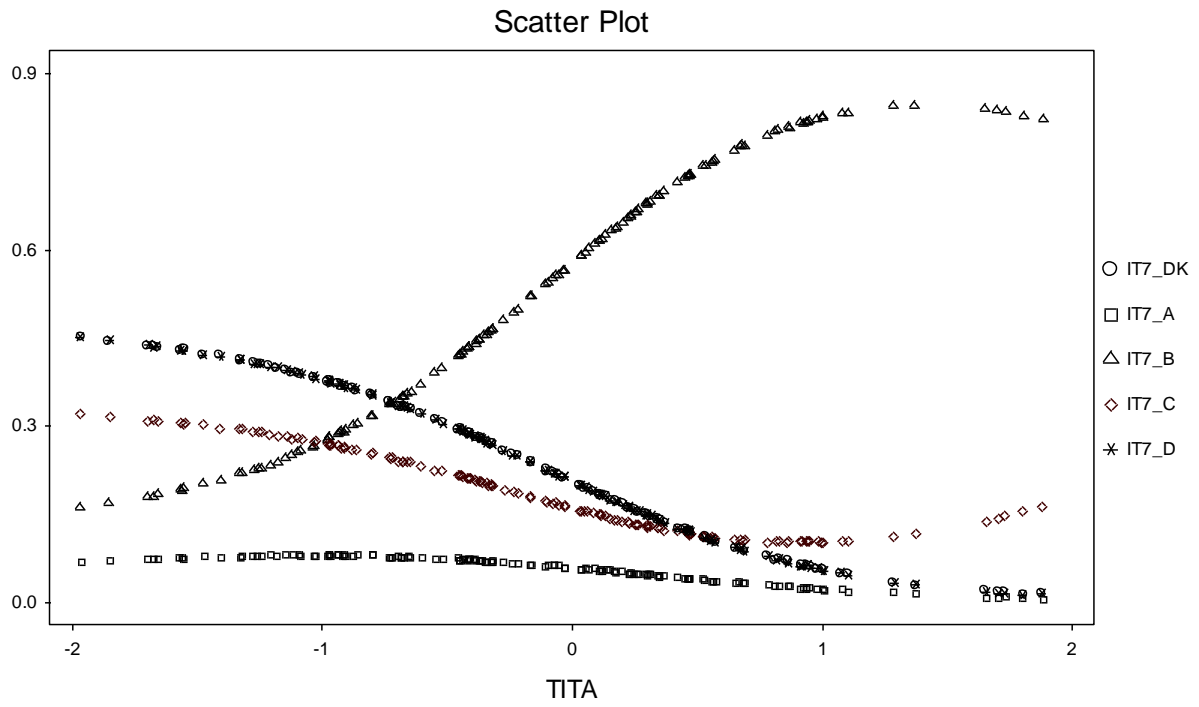


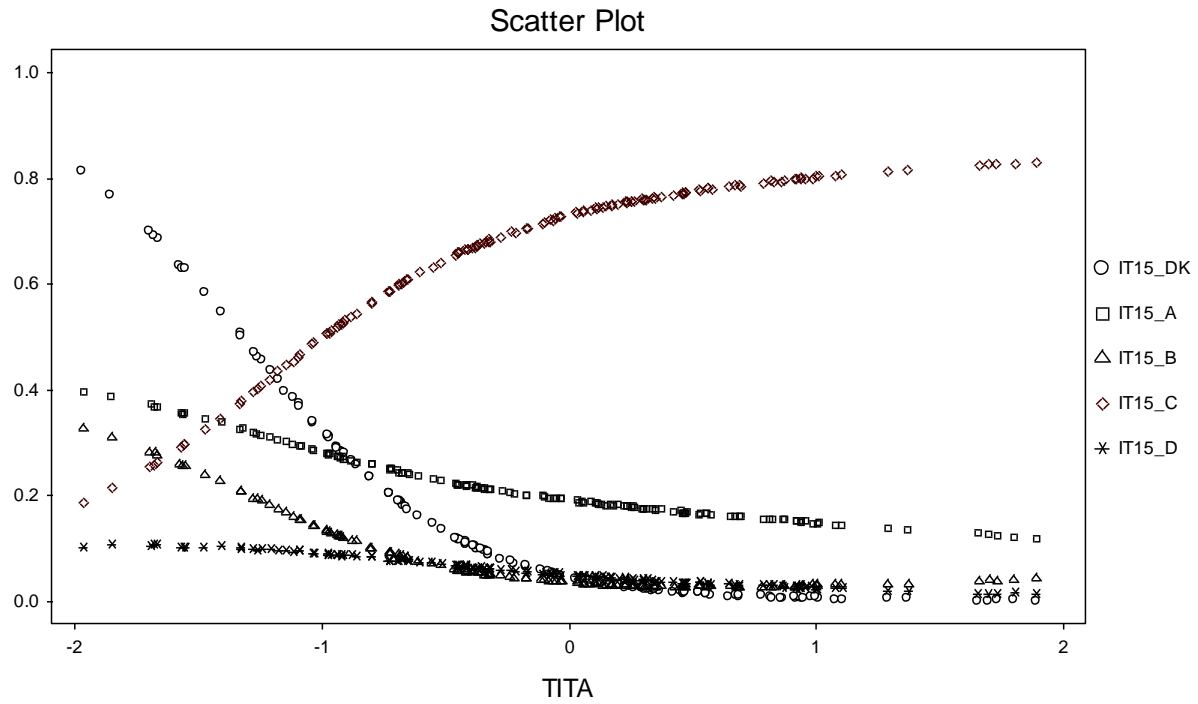
Figura 1. Curvas de respuesta para las opciones del ítem 1. La opción correcta es d).



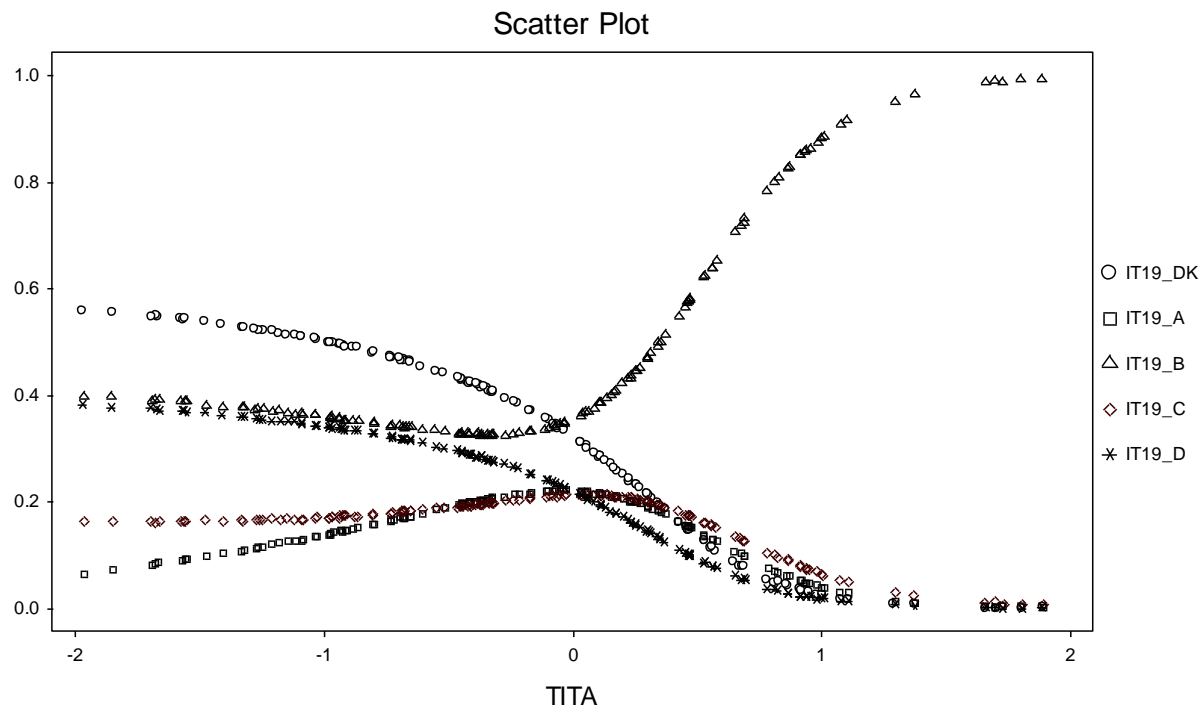
**Figura 2.** Curvas de respuesta para las opciones del ítem 4. La opción correcta es d).



**Figura 3.** Curvas de respuesta para las opciones del ítem 7. La opción correcta es b).



**Figura 4.** Curvas de respuesta para las opciones del ítem 15. La opción correcta es c).



**Figura 5.** Curvas de respuesta para las opciones del ítem 19. La opción correcta es b).