ANÁLISIS DEL FUNCIONAMIENTO DIFERENCIAL DE ÍTEMS DE RESPUESTA GRADUADA: UNA APROXIMACIÓN METODOLÓGICA

F. Javier Murillo Torrecilla
Centro de Investigación, Documentación y Evaluación (M.E.C.). Servicio de Investigación
María Castro Morera
Universidad Complutense de Madrid

En esta comunicación se aborda el estudio, desde un punto de vista metodológico y en base a la I.R.T., del funcionamiento diferencial de un tipo particular de ítems: los ítems de respuesta graduada. Así, se proponen tres procedimientos para calcular el «grado de sesgo» de ítems respuesta dicotómica, se presenta un modelo para estudiar ítems de respuesta graduada y, combinando ambas aproximaciones, se proponen tres métodos para calcular el grado de sesgo en ítems de respuesta graduada.

FUNCIONAMIENTO DIFERENCIAL DE LOS ÍTEMS

Intuitivamente se puede decir que un ítem funciona diferencialmente para dos grupos si dos sujetos de distintos grupos con el mismo nivel de habilidad tienen diferente rendimiento en ese ítem. Esta idea se puede traducir desde la perspectiva de la I.R.T. en la siguiente proposición: un ítem se comporta igual con dos grupos si las curvas características de ese ítem (C.C.I.) son idénticas para ambos grupos (Holland y Wainer, 1993). Dos definiciones equivalentes pueden ser:

1) Un ítem se comporta igual para dos grupos si el área comprendida entre las dos curvas características de los ítems es cero para todo entorno de cualquier nivel de habilidad (q) (Hambleton y Swaminathan, 1985).
2) Un ítem funciona igual para dos grupos si, para todo par de sujetos pertenecientes a distintos grupos y con el mismo nivel de habilidad, la probabilidad de acierto de ese ítem es idéntica (Pine, 1977; Lord, 1980; Hambleton y Swaminathan, 1985).

De la primera definición se deduce el **procedimiento de las áreas** para conocer el D.I.F., dado que el área comprendida entre las dos C.C.Is. se puede hallar calculando la integral de la diferencia de las funciones de probabilidad. Así, se establece el «grado de sesgo» a través de la siguiente fórmula:

$$Sesgo=\int_{-\infty}^{\infty}[P_i^1(\theta)-P_i^2(\theta)]\,d(\theta)$$

donde $P^1_i(q)$ y $P^2_i(q)$ son las funciones que definen las curvas del cada grupo.

Esta integral definida es fácil de calcular para el Modelo logístico de un parámetro. En ese caso se obtiene:

$$Sesgo=\int_{-\infty}^{\infty}[\frac{e^{(\theta-b)}}{1+e^{(\theta-b)}}-\frac{e^{(\theta-b')}}{1+e^{(\theta-b')}}]\,d(\theta)=b'-b$$

Este resultado, fácil y práctico, ofrece la posibilidad de calcular una aproximación al grado de sesgo estimando únicamente los parámetros *b* de cada grupo.

Para otro tipo de modelos, dada la dificultad que supone resolver esa integral, se puede realizar una aproximación numérica de la misma. Aunque la forma más habitual de hacerlo es mediante la fórmula de los rectángulos (Linn y cols., 1981), se obtiene una mejor **aproximación mediante la fórmula de Simpson**. Así:

$$Sesgo=\int_a^b P_i^*(\theta)\,d(\theta)\approx\frac{b-a}{6m}[y_0+y_{2m}+2(y_2+y_4+...+y_{2m-2})+4(y_1+y_3+...+y_{2m-1})$$

donde 2m es el número de intervalos, y_0, y_1,... y_{2m} son los valores de la función diferencia $P^*_i(q)$ en los puntos q_0, ..., q_{2m}, a es el máximo valor para el cual las C.C.Is. de cada grupo son simultáneamente cero y b es el mínimo valor donde las dos C.C.Is. toman el valor de uno.

De la segunda definición propuesta se extrae el procedimiento que se denominará **de las probabilidades empíricas**. Consiste en comparar las probabilidades empíricas de acierto para cada grupo de sujetos con niveles de habilidad similares. Así, se calibra el test para toda la población y se halla el nivel de habilidad para cada sujeto. En un segundo paso, se agrupan los niveles de habilidad en intervalos, (por ejemplo, 12 intervalos entre q=-3 y q=3) se hallan las probabilidades de acierto de cada intervalo de habilidad separadamente para cada grupo (y_i^j) y, con estas puntuaciones, se calcula el grado de sesgo mediante la siguiente fórmula:

$$Sesgo=\frac{b-a}{n}[(y_1^1-y_1^2)+(y_2^1-y_2^2)+...+(y_n^1-y_n^2)]=\frac{b-a}{n}\sum_{i=1}^{n}(y_i^1-y_i^2)$$

donde y_i^1 e y_i^2 son las probabilidades empíricas en cada intervalo i por los grupos 1 y 2.

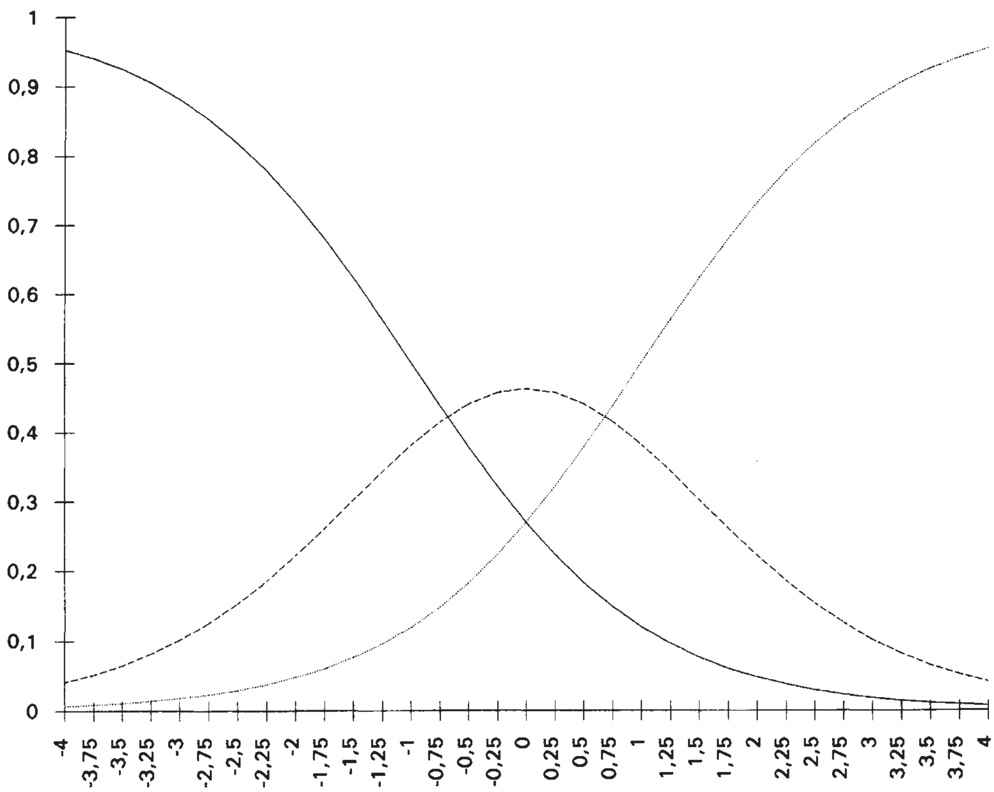
MODELO PARA ÍTEMS DE RESPUESTA GRADUADA

Los modelos tradicionales de I.R.T. están elaborados para ítems respuesta dicotómicas. Sin embargo, existen muchas pruebas donde es posible encontrar ítems con otras alternativas; por ejemplo, con las alternativas de correcto, parcialmente correcto e incorrecto. Para esos ítems, llamados de respuesta graduada, Samejima (1969) propuso el denominado **modelo diferencial**. Según el mismo, cada ítem está definido por tantas curvas como posibles respuestas del mismo. Cada una de las curvas está definida por la siguiente fórmula:

$$P_i(\theta)=P_i^*(\theta)-P_{(i+1)}^*(\theta)$$

para *i=1, ..., n-1*, donde *P*(q)* es la regresión de la puntuación de un ítem dicotómico sobre la habilidad, cuando todas las categorías de respuesta menores que *i* son puntuadas como 0, y aquellas mayores o iguales a *i* se puntúan 1. Esta propuesta sirve para cualquier modelo de respuesta al ítem. En la siguiente gráfica se pueden observar las curvas características de los ítems que representan cada una de las tres posibles respuestas de este ítem:

CCIs de un Item con el Modelo de Respuesta Graduada



Para el Modelo Logístico de dos parámetros las curvas características para ítems de respuesta graduada están definidas por:

$$P_0(\theta) = \frac{1}{1+e^{a_1(\theta-b_1)}}$$

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

para i = 1, ..., n-1.

$$P_n(\theta) = \frac{e^{a_n(\theta-b_n)}}{1+e^{a_n(\theta-b_n)}}$$

FUNCIONAMIENTO DIFERENCIAL DE ÍTEMS DE RESPUESTA GRADUADA

En este tercer bloque se aplicarán los tres procedimientos de detección del D.I.F.: de las áreas, de la aproximación numérica a través de la fórmula de Simpson y de las probabilidades empíricas al modelo de ítems de respuesta graduada descrito.

Dado que los ítems de respuesta graduada no tienen una sola curva característica sino tantas como posibles respuestas del ítem (n), los procedimientos de detección de D.I.F. deberán aplicarse a cada una de las curvas que definen el ítem, o, con más exactitud, a n-1 de las curvas que lo definen. Ello tiene un sentido lógico, dado que puede darse el caso de un ítem que se comporte de manera diferente para una o varias de las alternativas pero no para el ítem tomado conjuntamente.

Así, si eliminamos la curva $P_0(q)$, y aplicando el **procedimiento de las áreas** habría que calcular:

$$Sesgo^{[j]} = \int_{-\infty}^{\infty} [P_i^{[j](1)}(\theta) - P_i^{[j](2)}(\theta)]\, d(\theta)$$

para $j = 1, \ldots, n$.

Sustituyendo:

$$Sesgo^{[j]} = \int_{-\infty}^{\infty} [(P_i^{*[j](1)}(\theta) - P_i^{*[j+1](1)}(\theta)) - (P_i^{*[j](2)}(\theta) - P_i^{*[j+1](2)}(\theta))]\, d(\theta)$$

para $j = 1, \ldots, n-1$, y

$$Sesgo^{[n]} = \int_{-\infty}^{\infty} [P_i^{*[n](1)}(\theta) - P_i^{*[n](2)}(\theta)]\, d(\theta)$$

Estas fórmulas son generales para cualquier modelo de I.R.T.

Si se utiliza el modelo de Rasch, la integral tiene fácil solución:

$$Sesgo^{[j]} = (b_i^{*[j](1)} - b_i^{*[j+1](1)}) - (b_i^{*[j](2)} - b_i^{*[j+1](2)} \quad para\ j=1, \ldots, n-1$$

$$Sesgo^{[n]} = b_i^{*[n](1)} - b_i^{*[n](2)}$$

El otro procedimiento a analizar es el que se ha denominado **procedimiento de las áreas a través de la fórmula de Simpson.** Según el mismo, el sesgo para cada una de las curvas que definen el ítem se calcula de la siguiente manera:

$$Sesgo^{[j]} \approx \frac{b-a}{6m} [y_0^{[j]} + y_{2m}^{[j]} + 2(y_2^{[j]} + y_4^{[j]} + \ldots + y_{2m-2}^{[j]}) + 4(y_1^{[j]} + y_3^{[j]} + \ldots + y_{2m-1}^{[j]})]$$

donde [j] corresponde a cada una de las curvas diferenciales y 2m es el número de intervalos.

Por último, el **procedimiento de las probabilidades empíricas** se resuelve mediante la siguiente fórmula:

$$Sesgo^{[j]} \approx \frac{b-a}{n} \sum_{i=1}^{n} (y_i^{[j]1} - y_i^{[j]2})$$

donde [j] corresponde a cada una de las curvas diferenciales.

CONCLUSIONES

Las aportaciones de la comunicación presentada se pueden resumir en los siguientes puntos:

1. Se han propuesto tres **procedimientos para el cálculo del funcionamiento diferencial de los ítems**.

a) En primer lugar, mediante el procedimiento de las áreas, se ha buscado una solución para el Modelo de Rasch. Esta solución está caracterizada por la extrema sencillez que supone su puesta en marcha, puesto que la estimación del «grado de sesgo» depende exclusivamente de los parámetros de dificultad, sin embargo, tiene el inconveniente de perder mucha información si el modelo no se ajusta a los datos.
b) En segundo lugar, se ha realizado una mejor estimación del área comprendida entre las curvas características de los ítems mediante la fórmula de Simpson. Con este cálculo se ofrece una aproximación bastante exacta al «grado de sesgo», aunque sigue teniendo el inconveniente de partir de las probabilidades estimadas a través del modelo.
c) Por último, se ha propuesto el procedimiento denominado de las probabilidades empíricas, que tiene la ventaja de trabajar con las puntuaciones directas, por lo que aporta información muy ajustada de la realidad. Sin embargo, tiene como inconveniente que es necesario un importante número de sujetos para poder utilizar la propuesta.

2. Se ha desarrollado el **modelo diferencial** de Samejima (1969) para ítems de respuesta graduada.
3. Se han aplicado los tres procedimientos de detección del D.I.F. al modelo de ítems de respuesta graduada, obteniendose tres procedimientos novedosos para estudiar el **funcionamiento diferencial de los ítems de respuesta graduada** de una manera sencilla y práctica.

REFERENCIAS BIBLIOGRÁFICAS

HAMBLETON, R. K. y SWAMINATHAN, H. (1985): *Item response theory. Principles and applications.* Boston: Kluwer-Nijhoff Pub.

HOLLAND, P. W. y WAINER, H. (1993): *Differential ítem functioning.* Hillsdale, N.J.: LEA.

LINN, R. L., LEVINE, M. V., HASTINGS, C. N. y WARDROP, J. L. (1981): An investigation of item bias ina a test of reading comprehension. *Applied Psychological Measurement,* 5, 159-173.

LORD, F. M. (1980): *Applications of ítem response theory to practival testing problems.* Hillsdale, N.J.: LEA.

PINE, S. M. (1977): Applications of item response theory to the problem of test bias. En D. J. WEISS (ed.): *Applications of computerize adaptive testing.* Research Report 77-1. Minneapolis, University of Minnesota, Psychometric Methods Program, Departament of Psychology.

SAMEJIMA, R. (1969): Estimation of latent ability using a response pattern of gradded scores. *Psychometric* Monograph, 17. Psychometric Society.