

Este documento se encontraba antiguamente en la dirección http://www2.uca.es/RELIEVE/V4N2_4.HTM. Por favor, a partir de ahora utilice la siguiente referencia: "Revuelta, J., Ponsoda, V. y Olea, J. (1998). Métodos para el control de las tasas de exposición en tests adaptativos informatizados. *RELIEVE*, vol. 4, n. 2. Consultado en http://www.uv.es/RELIEVE/v4n2/RELIEVEv4n2_4.htm en (poner fecha)."

Este artículo ha sido consultado en **009543** ocasiones desde la fecha de su publicación

Revista **EL**ectrónica de **I**nvestigación y **E**valuación **E**ducativa // 1998 // Volumen 4 // Número 2_4
ISSN 1134-4032 // D.L. SE-1138-94

MÉTODOS PARA EL CONTROL DE LAS TASAS DE EXPOSICIÓN EN TESTS ADAPTATIVOS INFORMATIZADOS

por

Javier Revuelta

Vicente Ponsoda

Julio Olea

Julio.Olea@uam.es

Universidad Autónoma de Madrid

RESUMEN

En este trabajo se presentan tres nuevos métodos de control de las tasas de exposición de los ítems en tests adaptativos informatizados. El primero de ellos añade un componente aleatorio al método de máxima información. En el segundo se controla directamente la tasa de cada ítem, y sólo se administran aquellos ítems cuya tasa sea menor de un cierto valor. En el tercer método es una mezcla de los dos anteriores.

Los tres métodos se ponen a prueba mediante simulación, junto con el método de máxima información y el de Sympon y Hetter. Se estudian los resultados en cuanto a precisión del test y control de la tasa. Finalmente, se discuten las ventajas de cada método y se sugieren algunas líneas de investigación.

ABSTRACT

Three new item exposure control methods for computerized adaptive tests are introduced. The first one is based on maximum information, and adds a random component to the information provided by each item. The second one directly controls the exposure of each item, and apply only the items with and exposure lower than some arbitrary value. The third method is a mixture of the other two.

The three methods, plus maximum information and the Sympon and Hetter method, are compared in a simulation study, with respect to test precision and exposure rate. Finally, we comment some future research lines.

1. INTRODUCCIÓN

Uno de los objetivos de los test adaptativos informatizados (TAIs) es conseguir una estimación precisa del nivel de habilidad de los sujetos evaluados con un número reducido de ítems. La forma de lograrlo

consiste en seleccionar para cada sujeto aquellos ítems que resultan más informativos sobre su nivel de habilidad, esto es lo que se conoce como método de máxima información (MI).

La cantidad de información que proporciona un ítem depende de sus parámetros y de los del sujeto evaluado, aunque puede afirmarse que, en general, si se usa el modelo logístico de tres parámetros un ítem será más informativo cuanto mayor sea su capacidad discriminativa (parámetro 'a'), más ceca este su nivel de dificultad (parámetro 'b') del nivel del sujeto (q) y menor sea la probabilidad de acierto al azar (c) (Hambleton y Swaminathan, 1985, págs. 104-105). Por lo tanto, los ítems que sean altamente discriminativos serán utilizados con mayor frecuencia en la aplicación de TAIs debido a la mayor información que proporcionan sobre θ (Revuelta y Ponsoda, 1996).

En la literatura sobre TAIs se denomina "exposición de un ítem" a la frecuencia con que dicho ítem se utiliza en repetidas administraciones de un test (Wainer, 1990). Algunos autores (Hulin, Drasgow y Parsons, 1983) han encontrado que la tasa de exposición es muy variable de unos ítems a otros. Es decir, mientras algunos ítems se utilizan en todos o casi todos los tests, otros permanecen infrautilizados.

La excesiva variabilidad en las tasas de exposición causa problemas de orden práctico en la aplicación de los tests. Los ítems más populares pueden ser conocidos por los sujetos antes de recibir el test, por ejemplo en un diseño test-retest. Esto puede causar una pérdida de las propiedades psicométricas de los ítems y, por tanto, de la validez del test.

En consecuencia, puede ser ventajoso reducir la tasa de exposición de los ítems más populares a costa de incluir en el test aquellos que son utilizados con menor frecuencia. Esto es lo que se denomina "control de la tasa de exposición".

El objetivo que se busca con un método de control es lograr la mayor variedad posible en los ítems que se administran en diferentes tests. En primer lugar es necesario reducir la tasa máxima de exposición, en comparación con la obtenida con MI. En segundo lugar, es conveniente evitar que haya ítems infrautilizados. Finalmente, es deseable lograr estos objetivos no sólo en el total de tests administrados, sino también en el total de tests aplicados a cada uno de los grupos de habilidad de los sujetos evaluados. De esta forma se asegura que si un mismo sujeto es evaluado varias veces con el mismo test, y su nivel de rasgo no sufre grandes cambios, los ítems que recibe son distintos.

El control de la exposición implica que en algunas ocasiones no se emplean los ítems más informativos del banco, con la consecuencia de que cabe esperar una reducción en la precisión en comparación con MI.

En la literatura aparecen descritos varios métodos de control de la exposición, los cuales pueden clasificarse en dos grupos (Stocking, 1993):

1- Indirectos. Son los primeros que se desarrollaron. Básicamente añaden un componente aleatorio al método MI. Entre ellos puede citarse:

1- McBride y Martin (1983). El primer ítem del test se escoge al azar entre los 5 más informativos, el segundo entre los 4, y así sucesivamente hasta el quinto y siguientes, en que se escoge el más informativo.

2- Randlesque. (Kingsbury y Zara, 1989). Durante todo el test, el ítem que se va a aplicar se selecciona aleatoriamente entre los 5 más informativos.

Otros métodos indirectos pueden consultarse en: Lewis, Subhiyah y Morrison (1995) , y Morrison, Subhiyah y Nungester (1995).

2- Directos. También emplean un componente aleatorio, pero su característica esencial es que incluyen parámetros que controlan de forma explícita la tasa de exposición.

El método directo originario fue propuesto por Sympon y Hetter (1985), y consiste en asignar a cada ítem un parámetro 'k', entre 0 y 1, que controla su tasa de exposición. Durante el test, cada vez que se selecciona un ítem se genera un valor aleatorio uniforme entre 0 y 1. Si

dicho valor es menor que el parámetro se administra el ítem. En caso contrario el ítem queda inhabilitado para el test actual y se selecciona otro ítem. Algunas extensiones de este método han sido propuestas por Stocking (1993), Stocking y Lewis (1995) y Davey y Parshall (1995).

De todos los métodos descritos, únicamente el de Davey y Parshall (1995) controla las tasas de exposición condicionadas al nivel de habilidad del sujeto. En todos los demás el control se realiza sobre el total de tests administrados.

2. MÉTODOS PROPUESTOS

A continuación se describen tres nuevos métodos de control de las tasas de exposición: uno directo, otro indirecto y una mezcla de ambos. En los tres métodos el control es incondicional (sobre el total de tests administrados) y consisten en asignar un peso "w" a cada ítem "i" en cada selección. El ítem de mayor peso se selecciona y administra.

1- Progresivo (PR). Los primeros ítems del test se seleccionan de forma casi aleatoria, y en las siguientes selecciones se concede cada vez más importancia a la información proporcionada por cada ítem (Revuelta, 1995. Revuelta y Ponsoda, 1996). Los pesos se calculan mediante la fórmula:

$$w_i = (1-s) R_i + s I_i(\theta)$$

Donde "s" es igual al número de ítems administrados dividido por la longitud máxima del test. R es un número aleatorio entre 0 y el valor máximo de la información en el banco de ítems, e $I(\theta)$ es la información aportada por el ítem para el nivel estimado de habilidad.

Este método puede considerarse un refinamiento de los de McBride y Martin (1983) que incluye un componente aleatorio en los primeros ítems, y el de Kingsbury y Zara (1989), que incluye dicho componente durante todo el test.

2- Restringido (Rk). Los ítems se seleccionan por MI, sin tomar en consideración aquellos cuya tasa de exposición en los test anteriores supere el valor 'k' (Revuelta, 1995. Revuelta y Ponsoda, 1996). El peso de cada ítem es:

$$w_i = \begin{cases} I_i(\theta), & \text{si } a_i / t \leq k \\ 0, & \text{en caso contrario} \end{cases}$$

Donde 'a' es el número de veces que se ha utilizado el ítem en los 't' tests anteriores. Con este método cada uno de los ítems está inhabilitado en un k% de los tests. Por lo tanto, para conseguir que el banco del

ítems no se agote en ningún test, el valor de 'k' debe ser superior a la longitud del test dividida por el tamaño del banco.

El método Rk es una simplificación de Sympton y Hetter (1985), con la cual se pretende evitar la complejidad que supone asignar los parámetros 'k' de este método. Este proceso se realiza mediante simulación, y los parámetros resultantes son dependientes de las condiciones del test en dichas simulaciones: longitud, distribución de θ , etc.

3- Progresivo Restringido (Prk). En estudios anteriores (Revuelta, J, 1995. Revuelta, J. y Ponsoda, V. 1996) hemos encontrado que al utilizar el método PR la tasa máxima de exposición alcanza valores excesivamente altos, mientras que el número de ítems sin utilizar es reducido. Lo contrario sucede al utilizar el método Rk: la tasa máxima es óptima y el número de ítems sin utilizar es demasiado alto. Por esta razón, se propone un método combinado PR-Rk, el cual asigna un peso a cada ítem en la misma forma en que lo hace el Progresivo, pero además se controla la tasa máxima de exposición en la misma forma que con el método Rk. La fórmula del peso para cada ítem 'i' es:

$$w_i = \begin{cases} (1-s) R_i + s I_i(\theta), & \text{si } a_i / t \leq k \\ 0, & \text{en caso contrario} \end{cases}$$

Al igual que en los métodos anteriores, se selecciona y administra el ítem de mayor peso.

El objetivo de este trabajo es comparar los tres métodos propuestos (PR, Rk y PRk) con MI y SH en cuanto a la precisión del test y al control de la tasa. Respecto a la precisión, es necesario comparar el promedio de la diferencia entre θ verdadera y estimada con cada uno de los métodos, y también la diferencia en precisión entre diferentes sujetos con cada método. La razón es que con los métodos Rk y PRk el tamaño del banco de ítems disponibles no es el mismo para todos los sujetos. Por este motivo, es posible que las diferencias en precisión entre diferentes sujetos sean mayores que con los demás métodos.

En cuanto a la tasa de exposición. Es necesario comparar los valores de la tasa máxima, la cantidad de ítems infrautilizados y las diferencias en tasa entre los ítems.

A continuación se describe un estudio de simulación en que se comparan estos resultados con cada método, finalmente, las conclusiones.

3. MÉTODO

Condiciones. Se compararon los métodos MI, PR, Rk, PRk, SH y control (CO). En este último la selección de ítems se realizó de forma aleatoria. En los métodos Rk y PRk se empleó una tasa máxima del 40%, por esta razón el método se denominan en lo sucesivo R40 y PR40. La misma tasa se empleó en la estimación de los parámetros de exposición de SH.

Los seis métodos se compararon en tests de 20, 40 y 60 ítems, tomados de un banco de 500 en el cual se empleó el modelo logístico de 3 parámetros (3pl). Combinando 6 métodos por 3 longitudes resulta un total

de 18 condiciones.

Procedimiento. En cada condición se repitió cinco veces la aplicación de un TAI a los mismos 2000 sujetos simulados, cuya habilidad se distribuía $N(0, 1)$.

No se utilizó un banco de ítems reales, sino que este se generó específicamente para esta simulación. Cada ítem consta únicamente de los tres parámetros: 'a', 'b' y 'c' característicos del modelo 3pl, del parámetro 'k' propio del método SH (el cual se estimó independientemente para cada una de las 18 condiciones), y de dos contadores: 'nt' que indica el número de tests que se han aplicado, y 'na' que indica el número de veces que el ítem ha sido administrado. A partir de los valores 'na' y 'nt' se realiza el control de la exposición en los métodos R40 y PR40.

Los parámetros 'a', 'b' y 'c' se generaron a partir de las siguientes distribuciones (Baker, 1992, págs. 203-207): 'a': Lognormal (-0.25, 0.5), 'b': $N(0, 1)$ y 'c': Beta (5, 17). Esto produjo en un parámetro 'a' con media 0.93, desviación típica 0.41, mínimo 0.41 y máximo 2.37. Los mismos estadísticos correspondientes a 'b' son: 0.07, 0.95, -2.64 y 2.55. Los estadísticos de 'c' son: 0.23, 0.09, 0.029 y 0.51.

Variables analizadas. Se calculó la media en las cinco repeticiones de cada condición de las siguientes variables: 1- Sesgo: nivel verdadero del sujeto menos nivel estimado. 2- RMSD: raíz cuadrada de la media de las diferencias al cuadrado entre habilidad real y estimada, es un indicador de la precisión media del test. 3- RDSD: raíz cuadrada de la desviación típica de las diferencias al cuadrado, es un indicador de las diferencias en precisión entre sujetos. 4- Coeficiente de variación de la tasa de exposición. 5- Valor máximo de la tasa. 6- Valor mínimo y 7- Número de ítems sin usar.

4. RESULTADOS

La precisión depende fundamentalmente de la longitud del test. Con independencia del método, al aumentar la longitud aumenta también la precisión.

Con todos los métodos se obtuvo un sesgo negativo, indicando que la θ estimada es mayor que la verdadera, aunque las diferencias entre métodos no son significativas. En cuanto al RMSD, la mejor precisión se obtuvo con MI y la peor con CO. El valor de F es 57.51 ($p < 0.00...$) aunque únicamente son significativas las diferencias entre CO y el resto de los métodos. Lo mismo puede decirse de las diferencias en precisión entre los sujetos con cada método, medidas por el RDSD. En este caso el valor de F es 47.21 ($p < 0.00...$), aunque sólo son significativas las diferencias entre CO y el resto de los métodos (ver tabla 1).

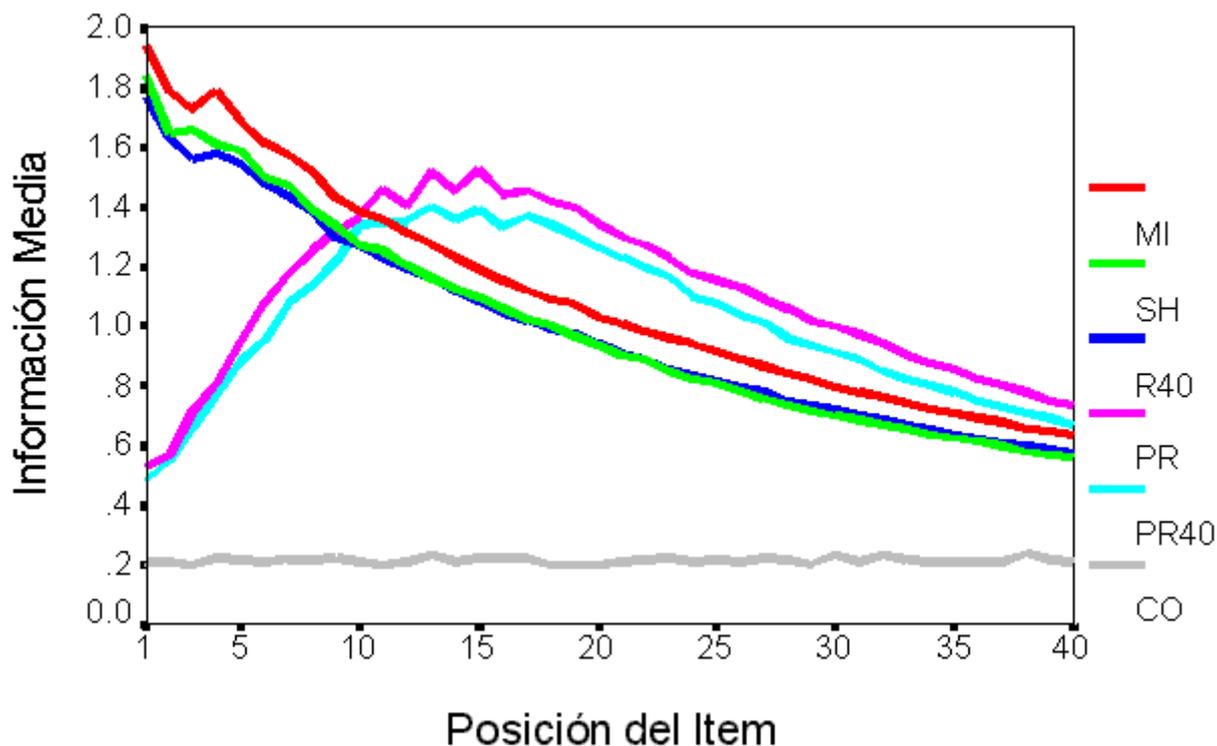
Tabla 1

Precisión del test por longitud y método de selección de ítems

VD	Método	Longitud		
		20	40	60
Sesgo	MI	-0.008	-0.004	-0.001
	PR	-0.014	-0.008	-0.005
	R40	-0.009	-0.007	-0.005
	PR40	-0.014	-0.008	-0.001
	SH	-0.013	-0.004	-0.002
	CO	-0.012	-0.010	-0.005
RMSD	MI	0.212	0.165	0.145
	PR	0.218	0.165	0.146
	R40	0.217	0.167	0.153
	PR40	0.217	0.168	0.153
	SH	0.217	0.166	0.151
	CO	0.631	0.415	0.337
RDSD	MI	0.324	0.228	0.201
	PR	0.360	0.230	0.206
	R40	0.376	0.230	0.210
	PR40	0.320	0.237	0.218
	SH	0.359	0.232	0.205
	CO	0.887	0.606	0.506

La función de información a lo largo del test muestra algunas diferencias entre los métodos. Esta consiste en la información media aportada por los items en la primera posición de los 2000 tests, respecto a la habilidad estimada final. La media de los items en la segunda posición, etc.

Se calculó dicha función de información en la primera repetición de la condición de 40 items. El resultado indica que todos los métodos proporcionan los items mas informativos al principio del test, exceptuando PR y PRk con los cuales el máximo se alcanza entre los items 10 y 20. Con el método CO la información es una línea horizontal a lo largo de todo el test (ver gráfica 1).



En cuanto a la tasa de exposición, los resultados del método MI no son adecuados. El máximo está en el 100% debido a que la habilidad inicial es la misma para todos los sujetos, y por tanto el primer ítem del test también. Además, un gran porcentaje del banco permanece sin utilizar y las diferencias entre los ítems son grandes.

Los métodos R40, PR40 y SH tienen como ventaja que disminuyen la tasa máxima hasta el valor deseado, mientras que el número de ítems sin utilizar es similar al encontrado con MI. En cambio, con PR la tasa máxima es excesiva en tests largos debido a que no se ejerce control sobre. La ventaja de los métodos PR y PR40 es que disminuyen claramente el número de ítems sin utilizar, aunque ella (ver tabla 2).

Tabla 2

Control de la exposición por longitud del test y método de selección de ítems

VD	Método	Longitud		
		20	40	60
CV	MI	255.84	179.01	147.30
	PR	223.84	166.20	139.54
	R40	234.28	167.13	129.85
	PR40	220.91	160.76	126.03
	SH	234.04	168.32	132.35
	CO	11.17	7.67	5.98
Tasa mínima	MI	0.0	0.0	0.0
	PR	0.0	0.0	0.1
	R40	0.0	0.0	0.0
	PR40	0.0	0.0	0.1
	SH	0.0	0.0	0.0
	CO	2.7	6.4	9.6
Tasa máxima	MI	100.0	100.0	100.0
	PR	48.3	55.7	62.0
	R40	40.0	40.0	40.0
	PR40	40.0	40.0	40.0
	SH	41.7	42.6	42.0
	CO	2.7	10.1	14.3
% sin usar	MI	74.6	58.4	45.3
	PR	3.6	10.9	0.0
	R40	73.4	53.7	37.6
	PR40	10.1	0.1	0.0
	SH	74.0	57.4	42.7
	CO	0.0	0.0	0.0

5. CONCLUSIONES

Los resultados de este estudio no son concluyentes debido a que cada método muestra ventajas e inconvenientes. Las diferencias en precisión entre los métodos no son significativas, aunque el mejor valor lo proporcionó el método MI. En cambio, existen claras diferencias en el control de la tasa.

El principal objetivo del control: una tasa máxima no excesivamente alta, sólo se consigue con los métodos Rk, PRk y SH. No obstante, Rk y SH no proporcionan ventajas respecto al segundo objetivo: evitar la infrutilización de una parte del banco. Esto solo se consigue con los método PR y PRk.

Los métodos de control tienen cada uno sus propios inconvenientes. SH es complicado de implementar y susceptible a cambios en las condiciones del test, como cambios en la longitud o simplemente en la distribución de habilidad de los sujetos evaluados. La razón está en que los parámetros 'k' empleados en este método dependen de estas condiciones. Resultaría interesante comprobar hasta que punto los parámetros del

método SH son robustos a cambios en las condiciones del test, es decir, como varían las tasas de exposición al cambiar las condiciones del test manteniendo los mismos parámetros 'k'.

Rk resuelve algunos de los problemas de SH, aunque produce mayores diferencias de precisión entre los sujetos. En cuanto al progresivo, el inconveniente está en que no se ejerce control sobre la tasa máxima, y puede resultar excesiva.

En conclusión, el método combinado Prk parece la mejor solución de compromiso por sus resultados en precisión, tasa mínima y número de ítems sin utilizar. No obstante, debido a la gran cantidad de restricciones que este método impone en el criterio de máxima información, es posible que en futuros estudios aparezcan diferencias significativas en precisión al compararlo con MI, las cuales podrían reducirse aplicando tests más largos con PRk que con MI.

Una extensión de los métodos Rk y PRk consiste en calcular la tasa de exposición condicionada al valor de la habilidad estimada, no sólo en el total de tests administrados. De esta forma no solamente se lograría controlar la tasa máxima en el total de tests administrados, sino también evitar que sujetos de habilidades similares reciban los mismos ítems.

Finalmente, todos los métodos propuestos deben ponerse a prueba en condiciones más realistas y con restricciones no psicométricas en la selección de ítems. Por ejemplo, utilizando tests estructurados. En los cuales el banco se divide en diversas áreas de contenido y es necesario incluir en el test una determinada cantidad de ítems de cada área.

6. REFERENCIAS

- Baker, F. B. 1992. *Item Response Theory, Parameter Estimation Techniques*. Marcel Dekker. New York.
- Davey, T., & Parshall, C.G. 1995. New algorithms for item selection and exposure control with computerized adaptive testing. *Annual meeting of the American Educational Research Association*. San Francisco.
- Hambleton, R.K., & Swaminathan, H. 1985. *Item Response Theory. Principles and Applications*. Boston: Kluwer-Nijhoff Pub.
- Hulin, C.L., Drasgow, F., & Parsons, C.K. 1983. *Item Response Theory: Applications to Psychological Measurement*. Homewood, IL.: Dow Jones-Irwin.
- Kingsbury, G.G., & Zara, A.R. 1984. Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*. Nº 2. Págs. 359-375.
- Lewis, M.J., Subhiyah, R.G., & Morrison, C.A. 1995. A comparison of classification agreement between adaptive and full-length tests under the 1-PL and 2-PL models. *Annual meeting of the American Educational Research Association*. San Francisco.
- McBride, J.R., & Martin, J.T. 1983. Reliability and validity of adaptive ability tests in a military setting. In D.J. Weiss (Ed.) *New Horizons in Testing*. Págs. 223-236. New York: Academic Press.
- Morrison, C., Subhiyah, R., & Nungester, R. 1995. Item exposure rates for unconstrained and content-balanced computerized adaptive tests. *Annual Meeting of the American Educational Research Association*. San Francisco.
- Revuelta, J. 1995. *El control de la exposición de los ítems en tests adaptativos informatizados*. Memoria de licenciatura. Universidad Autónoma de Madrid.
- Revuelta, J., y Ponsoda, V. 1996. *Metodos sencillos para el control de las tasas de exposición en tests adaptativos informatizados*. *Psicológica*, 17, 161-172.
- Stocking, M. 1993. *Controlling item exposure rates in a realistic adaptive testing paradigm*. Technical Report (RR 93-2). Princeton, NJ: Educational Testing Service.

Stocking, M., & Lewis, C. 1995. *A New Method of Controlling Item Exposure in Computerized Adaptive Testing. Technical Report (RR-95-25)*. Princeton, NJ: Educational Testing Service.

Sympson, J.B., & Hetter, R.D. 1985. Controlling item exposure rates in computerized adaptive testing. *Proceedings of the 27th Annual Meeting of the Military Testing Association*. San Diego, CA: Navy Personnel Research and Development Center.

Wainer, H. 1990. *Computerized adaptive testing: a primer*. Hillsdale, New Jersey: LEA.

TO KNOW MORE / PARA SABER MÁS

This article has been cited in... / Este artículo ha sido citado en...

- * Hontangas, Pedro; Ponsoda, Vicente; Olea, Julio; Abad, Francisco José (2000). Los test adaptativos informatizados en la frontera del siglo xxi: una revisión. *Metodología de las Ciencias del Comportamiento*, 2, 2: 183-216 [Párrafo añadido el 10 de Abril de 2007/ Paragraph added April 10th 2007].
- * Olea, Julio; Ponsoda, Vicente; Felipe, Luordes; Carretie, Luis (1998). Estrategias de evaluación informatizada y ansiedad: un estudio comparativo en una muestra de alto riesgo. *Ansiedad y Estrés*, 4, 1: 71-79 [Párrafo añadido el 10 de Abril de 2007/ Paragraph added April 10th 2007].
- * Olea Díaz, Julio; Revuelta, J.; Ximénez, M. C.; Abad, Francisco José (2000). Efectos psicométricos y psicológicos de la revisión de respuestas en tests fijos y adaptativos informatizados. *Psicológica*, 21, 1/2: 157-173. [Párrafo añadido el 10 de Abril de 2007/ Paragraph added April 10th 2007]



[Volver al índice general](#)



[Volver al índice del Vol. 4 N° 2](#)
