

ELECCIÓN DE LA DIFICULTAD DE LOS TESTS AUTOADAPTADOS: UN ESTUDIO PILOTO

by/por

[Article record](#)

[About authors](#)

[HTML format](#)

P. M. Hontagas (Pedro.M.Hontagas@uv.es)

J. Olea

V. Ponsoda

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

Abstract

This study compared several selection of difficulty strategies in self-adapted testing. An ascendant but non linear tendency is observed. The different strategies (flexible, failure intolerant, inflexible and failure tolerant) differed whit regard to ability, standard error, number of item passed and testing time. A moderated positive linear relation between ability, difficulty of previous item and difficulty level was obtained.

Resumen

Se estudia la conducta de elección de la dificultad en Tests Autoadaptados Informatizados (TADIs). Aparece una tendencia ascendente no lineal en los niveles de dificultad que se seleccionan a lo largo del test. Se observaron diferentes estrategias de elección (flexible, intolerante a la frustración, inflexible y tolerante a la frustración), que puede venir determinadas por los niveles de habilidad y ansiedad de los sujetos. Se obtuvieron entre ellas diferencias en habilidad estimada, precisión de las estimaciones, número de ítems acertados y tiempo invertido en el test. Existe una relación linal moderada entre las estimaciones de habilidad, la dificultad del ítem previo y el nivel de dificultad seleccionado.

Keywords

self-adapted testing

Descriptores

Tests Autoadaptados Informatizados

1. Introducción.

El avance de la informática y el desarrollo de la teoría de respuesta a los ítems (TRI) ha fomentado en las últimas décadas la aparición de nuevos procedimientos de medida distintos de los tests convencionales de lápiz y papel. Los primeros tests adaptativos en formato de lápiz y papel (Lord, 1971) y los test convencionales informatizados (Koson, Kitchen, Kochen y Stodolosky, 1970) han dado paso a los tests adaptativos informatizados (TAIs; Johnson y Mihal, 1973), entre los cuales, los denominados tests autoadaptados informatizados (TADIs; Rocklin y O'Donnell, 1987) constituyen una de sus variantes más recientes.

Un TAI consiste en un banco de ítems con propiedades conocidas (calibrados normalmente bajo algún modelo de la TRI) y un procedimiento para seleccionar los más apropiados para la habilidad de cada persona (un algoritmo informático que presenta ítems más difíciles o más fáciles según se acierten o fallen los anteriores). Su principal ventaja frente a los TC y TCI es la eficiencia: permite obtener medidas más precisas con el mismo número de ítems o conseguir la misma precisión con casi la mitad de elementos (Wainer, 1990; Olea y Ponsoda, 1996). Ahora bien, un TAI, además de adaptar la dificultad de los ítems a la habilidad de los sujetos, presenta una situación afectiva y motivacional semejante para todos, puesto que se

enfrentan a ítems con el mismo grado de desafío a su capacidad y obtienen un éxito similar (aproximadamente el 50% de aciertos). Sin embargo, las personas difieren en sus reacciones a la dificultad del test, bien por la orientación motivacional hacia el éxito o el fracaso (Atkinson y Litwin, 1960), la ansiedad que experimentan ante una situación de evaluación (Rocklin y Thompson, 1985) o el concepto que tienen de sí mismas (Vispoel, Rocklin y Wang, 1994). De modo que la uniformidad afectiva y motivacional generada por un TAI no siempre es una situación óptima de evaluación y la incidencia en el rendimiento de los factores señalados puede afectar a la validez de las medidas (Rocklin, O'Donnell y Holst, 1995). Recientemente, Rocklin y O'Donnell (1987) han sugerido modificar el procedimiento de selección de ítems de los TAI para paliar estos inconvenientes, dando lugar a los TADI. El cambio consiste en permitir al sujeto elegir la dificultad de los ítems para que pueda adaptar la situación de evaluación a su estado afectivo y motivacional. La finalidad es evitar el efecto de factores extraños sobre el rendimiento en el test, especialmente la ansiedad, y se diferencia básicamente de un TAI en la forma de selección de ítems: el propio sujeto vs. el algoritmo informático. En resumen, un TADI consiste en un banco de ítems dividido en niveles de dificultad (5 a 9) y un procedimiento de selección de ítem controlado por el sujeto. Normalmente, incluye la instrucción de escoger siempre ítems del nivel más difícil que el sujeto sea capaz de responder correctamente y suele ofrecer feedback después de cada respuesta.

La mayor parte de la investigación sobre TADI se han centrado en analizar sus ventajas e inconvenientes respecto a otros procedimientos informatizados. Los resultados de estos estudios indican que este tipo de tests disminuye la ansiedad, atenúa la relación entre ansiedad y rendimiento, da lugar a puntuaciones en el rasgo más elevadas en algunos casos, requiere un

tiempo de aplicación mayor y produce errores típicos de estimación más altos. En definitiva, un TADI ofrece medidas más válidas, aunque menos fiables y eficientes, siendo estas cualidades independientes de dar o no feedback de los resultados (Rocklin, 1994; Wise, 1994). Por otra parte, cabe destacar la poca atención que se ha prestado a la forma en que tiene lugar el proceso de selección de ítems y a las consecuencias que las diferencias individuales pueden tener para las propiedades psicométricas del TADI y el bienestar individual. Hay que tener presente que, al trasladar el control de la aplicación del test desde el programa informático a los sujetos, disminuyen las garantías de que éstos actúen cumpliendo las condiciones para que el procedimiento sea eficaz y se abre la posibilidad de que se utilicen estrategias diferentes, cuya existencia y consecuencias es necesario conocer. Respecto a la eficacia, un TADI requiere que los sujetos elijan ítems con una dificultad próxima a su capacidad; por el contrario, si escogen ítems muy fáciles o muy difíciles para ellos, su habilidad será estimada con poca precisión. Los resultados del único estudio disponible que aborda esta cuestión presentan altas correlaciones entre habilidad y dificultad (Johnson, Roos, Wise y Plake, 1991). En cuanto al proceso de selección de ítems, las personas suelen elegir ítems fáciles al principio del test y posteriormente ítems más difíciles, siguiendo así una tendencia creciente en dificultad conforme avanza el test (Rocklin y O'Donnell, 1987). Los principales factores propuestos como predictores de dichas elecciones son la ansiedad, la competencia percibida y el acierto o error en los ítems anteriores. La ansiedad y la competencia percibida están asociadas principalmente con la elección de la dificultad en los primeros ítems. Los sujetos con mayor ansiedad y menor competencia comienzan por ítems más fáciles que los demás. Esta asociación disminuye en las elecciones siguientes, donde el feedback recibido del resultado pasa a ser el aspecto más importante (Johnson et al., 1991).

En otra dirección, Rocklin (1989) ha analizado tres estrategias de selección de ítems, definidas en base al éxito o fracaso en los ítems anteriores: a) una estrategia flexible consiste en elegir el ítem siguiente más difícil si acierta y más fácil si falla, b) una estrategia de tolerancia ante el fracaso consiste en elegir un ítem con mayor dificultad si acierta y con el mismo nivel de dificultad si falla, y c) una estrategia de intolerancia al fracaso consiste en mantener el mismo nivel de dificultad si acierta y elegir un nivel más fácil si falla. Los resultados obtenidos con una muestra muy reducida ($n=29$) indican que la mayoría de los sujetos sigue una estrategia flexible, utilizando en menor medida las demás; sin embargo, en el estudio realizado posteriormente por Johnson et al. (1991) se afirma que en muy pocos casos se siguen estrictamente las estrategias anteriores.

Así pues, la literatura sobre el proceso de cumplimentación de un TADI pone de relieve el reducido número de estudios existente (Rocklin, 1989; Johnson et al. (1991), el pequeño tamaño de la muestra utilizada, las limitaciones de las estrategias propuestas y el insuficiente conocimiento de las consecuencias que conlleva seguir cada una. En este sentido, el presente trabajo propone analizar la conducta de elección de la dificultad en un TADI de conocimientos de inglés con una muestra más amplia, incluye un repertorio mayor de estrategias y analiza las diferencias entre estrategias en diversas variables relativas al test y al bienestar individual. Concretamente, nuestro estudio pretende, en primer lugar, determinar el grado de ajuste entre la habilidad de los sujetos y la dificultad de los ítems elegidos; en segundo lugar, caracterizar el perfil de elecciones de la dificultad y clarificar la influencia de diversos factores en este proceso (sexo, ansiedad, habilidad, éxito-fracaso, dificultad previa); y finalmente, conocer las estrategias que los sujetos utilizan para elegir la dificultad y sus consecuencias personales (ansiedad) y psicométricas (habili-

dad estimada, error típico de estimación, tasa de aciertos, tiempo de respuesta, dificultad de los ítems seleccionados).

2. Método

Muestra

En este trabajo participan 111 jóvenes estudiantes de C.O.U., pertenecientes a varios centros de la comunidad de Madrid, de los cuales el 41.8% son varones y el 58.2% son mujeres. La edad de los participantes se encuentra entre los 17 y los 19 años.

Instrumentos de medida.

La **ansiedad** situacional es medida con la dimensión Ansiedad Estado del STAI (Spielberger, Gorsuch y Lushene, 1970), que consta de 20 ítems tipo Likert con 4 alternativas de respuesta (0=nada, 1=poco, 2=bastante y 3=mucho). Ha sido dividida en dos partes equivalentes, con las que se ha construido una versión informatizada, para evaluar la ansiedad antes (pretest) y después (postest) de la aplicación del test autoadaptado (las propiedades psicométricas de ambas mitades y su equivalencia se estudiaron en un trabajo previo: Hontagas, Canela y Agustín, 1996). La consistencia interna obtenido es 0.69 y 0.71 (coeficiente alfa). Las puntuaciones totales están expresadas en la misma escala que los ítems.

El **test autoadaptado** mide conocimientos de vocabulario inglés con un banco de 221 ítems de 5 alternativas de respuesta, calibrado con el modelo de TRI de tres parámetros (ver sus propiedades psicométricas en Ponsoda, Olea y Revuelta, 1994; Olea, Ponsoda, Revuelta y Belchi, 1996, Ponsoda, Wise, Olea y Revuelta, 1997). Los ítems del banco son agrupados en 5 niveles de dificultad, según el valor de su parámetro b : 1) Muy fácil ($b < -1.8$), 2) Fácil ($-1.8 \leq b < -0.6$), 3) Dificultad media ($-0.6 \leq b < 0.6$), 4) Difícil ($0.6 \leq b < 1.8$) y 5) Muy difícil ($b > 1.8$). El algoritmo del programa informático que controla la aplicación del test utiliza como punto de arranque un ítem central elegido aleato-

riamente en el intervalo de dificultad -1 y 1, a continuación, presenta el ítem que tiene un mayor poder discriminativo en cada categoría de dificultad y, finalmente, termina al administrar una longitud fija de 20 ítems. Asimismo, el programa calcula y registra la habilidad estimada de cada sujeto y su error típico, el número de aciertos, el tiempo de respuesta y la dificultad de los ítems presentados. La habilidad es calculada por el método de máxima verosimilitud y se expresa en puntuaciones z con media 0 y desviación típica 1. El tiempo de respuesta mide los segundos transcurridos desde la presentación del ítem en la pantalla del ordenador hasta que se contesta. Y la dificultad de los ítems es considerada tanto objetiva como subjetivamente. La dificultad objetiva es el valor del parámetro b del modelo de TRI utilizado y está expresada en la escala z con media 0 y desviación típica 1. La dificultad subjetiva se refiere al nivel elegido por el sujeto entre las categorías que se le ofrecen y se encuentra en el rango 1 a 5.

Procedimiento.

Las sesiones fueron realizadas en una sala con 6 ordenadores, como si se tratara de un examen adicional de la asignatura de inglés. Los sujetos fueron distribuidos aleatoriamente en los ordenadores disponibles y recibieron una explicación del examinador sobre las condiciones generales de la prueba. Antes de comenzar, se presenta en la pantalla las siguientes instrucciones: “Tu tarea principal será responder a un test de vocabulario de inglés. Antes y después del test de inglés, tienes que contestar a una pocas preguntas sobre cómo te sientes en ese momento. Es muy importante que tus respuestas sean sinceras. Nadie obtendrá información de ellas. Sólo tus resultados en el test de inglés será entregados a tu profesor”. “En el test de inglés tienes que elegir la dificultad de cada ítem. Antes de cada ítem, el ordenador te preguntará por la dificultad que quieres para el ítem siguiente. Tu respuesta tiene que ser un número del 1

(muy fácil) al 5 (muy difícil). Intenta elegir los ítems más difíciles que creas que puede contestar correctamente. Intenta evitar los ítems muy fáciles o muy difíciles para tu nivel de inglés. Tu puntuación final dependerá de los ítems que aciertes y su dificultad”. A continuación de las instrucciones, se aplican todos los tests controlados por el ordenador. En primer lugar, se presenta el inventario de ansiedad pretest. En segundo lugar, se administra el test de inglés, compuesto por una parte de prueba con 4 ítems y otra de aplicación efectiva de 20 ítems. Se concede un tiempo límite de 15 segundos para contestar cada ítem y no se ofrece feedback del resultado. En tercer lugar, se presenta la forma correspondiente a la ansiedad posttest. Y, por último, se solicitan los datos personales.

Estrategias de selección de ítems.

El criterio utilizado para definir las estrategias es la conexión entre la respuesta del ítem anterior y la dificultad del ítem siguiente; es decir, la elección depende de la contingencia de resultados (acierto vs. error). Las estrategias consideradas son las tres propuestas por Rocklin (1989), denominadas flexible, tolerante a la frustración e intolerante a la frustración (ver introducción), y tres adicionales, denominadas inflexible, incongruente e indefinida. La estrategia inflexible consiste en elegir un nivel de dificultad inicial y mantenerlo siempre durante toda la prueba. Es una estrategia no contingente con los resultados de las respuestas. La estrategia indefinida consiste en cumplir en el mismo grado los requisitos de varias estrategias anteriores. Se trata de una estrategia que describe la indeterminación del comportamiento del sujeto. Y la estrategia incongruente consiste en elegir el siguiente ítem más fácil, cuando se acierta, o más difícil, cuando se falla. En este caso se contemplan conductas poco razonables o el desinterés por la aplicación del test.

Análisis de datos.

Los datos proceden de los estudios experimentales de Ponsoda, Olea, Revuelta y Wise

(1995) y Ponsoda, Wise, Olea y Revuelta (1997). Los análisis efectuados para determinar su equivalencia indican la ausencia de diferencias significativas en las variables consideradas (tabla 1), salvo ansiedad pretest ($p=.01$). Dado

que el procedimiento utilizado es el mismo en los dos estudios y que los sujetos presentan características muy semejantes, el presente trabajo se centra en la muestra global.

Tabla 1.- Estadísticos descriptivos de las características del test.

Variables	Estudio 1		Estudio 2		Muestra global	
	Media	D. Típica	Media	D. Típica	Media	D. Típica
Ansiedad pre-test	1.08	.41	.89	.36	1.00	.40
Ansiedad post-test	.93	.37	.83	.43	.88	.40
Habilidad estimada	.34	.83	.37	.88	.35	.85
Error de estimación	.28	.04	.28	.05	.28	.04
Número de aciertos	69.18	12.02	68.85	14.45	69.04	13.08
Tiempo de respuesta	7.14	1.66	7.57	1.67	7.33	1.67
Dificultad subjetiva	3.08	.56	3.08	.66	3.08	.61
Dificultad objetiva	.03	.66	.04	.75	.03	.70
N	61		48*		109*	

* 2 casos eliminados

El grado de correspondencia entre habilidad y dificultad es explorado mediante la correlación y la raíz de la media de diferencias cuadráticas de ambas variables a lo largo de la secuencia de items presentados. Se incluye solo la dificultad subjetiva por ser la faceta que está bajo el control del sujeto (no obstante, los resultados son similares a la dificultad objetiva: parámetro b y $zeta$). Las estimaciones de la habilidad son transformadas con los mismos criterios empleados para dividir el banco de items en 5 niveles de dificultad (ver apartado instrumentos de medida).

La descripción del perfil de dificultad de las elecciones es efectuada considerando tres momentos en la aplicación del test: el comienzo (ítem 1), la parte media (ítem 11) y el final (ítem 20) (Johnson et.al., 1991). Asimismo, la dificultad del TADI es analizada globalmente mediante una regresión lineal para cada sujeto de la posición serial de los items sobre la dificultad elegida. La interceptal describe el nivel de la dificultad inicial y la pendiente la tasa de cambio (Rocklin y O'Donnell, 1987). También,

se consideran la posibilidad de una relación no lineal. A continuación, un conjunto de análisis de regresión jerárquica explora el poder predictivo de diversas variables introducidas secuencialmente en cuatro bloques:

$$\text{Bloque 1 : } D_i = S + A + H$$

$$\text{Bloque 2 : } = R_{i-1} + D_{i-1} + R_{i-1} \times D_{i-1}$$

$$\text{Bloque 3 : } = R_{i-2} + D_{i-2} + R_{i-2} \times D_{i-2}$$

$$\text{Bloque 4 : } = R_{i-3} + D_{i-3} + R_{i-3} \times D_{i-3}$$

En el primer bloque se introducen el sexo (S), la ansiedad pretest (A) y la habilidad estimada (H). Y en los siguientes bloques se añade la respuesta (R), la dificultad (D) y la interacción de ambas ($R \times D$) de los tres items anteriores (salvo en los items 1, 2 y 3). Se trata de determinar los factores que predicen la elección inicial y cómo varía el poder predictivo conforme avanza la prueba.

Las estrategias de selección de items seguida por cada sujeto analizada del siguiente modo: a) se elaboran vectores con las elecciones que tendrían que haber realizado en el caso de utilizar cada una de las estrategias teóricas; b) se

comparan las elecciones reales con las elecciones teóricas de cada estrategia mediante la raíz de la media de errores cuadráticos (RMEC) y c) se clasifica a los sujetos en la estrategia donde obtengan el mejor ajuste. Por último, son realizados análisis diferenciales por habilidad estimada, error típico de estimación, número de aciertos, tiempo de respuesta, dificultad objetiva y subjetiva y ansiedad pretest, ansiedad posttest y cambio en ansiedad.

3. Resultados

Descripción de las características del test.

Los resultados de la muestra global en el test se describen en la tabla 1. En primer lugar, se observa que la ansiedad antes y después de la aplicación refleja una situación de 'poca' ansiedad, según las etiquetas verbales de la escala de los ítems (media pretest= 1 y media posttest= 0.88). El nivel de ansiedad inicial sufre una reducción estadísticamente significativa al finalizar la aplicación (posttest-pretest= -.11; $p=.001$), poniendo de manifiesto el efecto positivo que los tests autoadaptados tienen sobre el estado afectivo y motivacional de los sujetos. En segundo lugar, la habilidad estimada es ligeramente superior a la media de la escala zeta y su distribución (media= 0.35 y d. típica= 0.85) es muy similar a la muestra de calibración (media=0.14, d.típica=0.72; Olea, Ponsoda, Revuelta y Belchi, 1996). En tercer lugar, el error típico de estimación es 0.28 y equivale a un coeficiente de fiabilidad de 0.92, en el ámbito de la teoría clásica de test (Thissen, 1990). Por tanto, la precisión media de las estimaciones es bastante alta. El rango de los errores se encuentra entre .23 y .49, salvo dos sujetos con valores muy elevados de 2.98 y 3.07, que han acertado todos los ítems, al ser muy fáciles para su capacidad. Esta circunstancia podría ser una limitación del banco de ítems por carecer elementos

más difíciles para medir a sujetos con alta habilidad. Sin embargo, los ítems elegidos tienen dificultad media y el banco dispone de ítems de mayor dificultad. Así pues, queda de relieve la principal limitación de los tests autoadaptados: si la dificultad de los ítems no se ajusta a la capacidad del sujeto, las estimaciones son imprecisas y carentes de utilidad. En cuarto lugar, la tasa de aciertos se aproxima al 69%, cantidad superior a la obtenida normalmente en un test adaptativo, donde se suele acertar y fallar el 50% aproximadamente (en el modelo de 3 parámetros es algo más debido al parámetro c). En quinto lugar, el tiempo de respuesta presenta un promedio de 7,33 segundos. El límite máximo para dar la respuesta (15 segundos) es alcanzado en pocos casos, no siendo un factor de presión generador de ansiedad. Por último, la dificultad de los ítems en el conjunto del test tiene un valor medio, tanto en relación a la dificultad objetiva (0.03) como a la subjetiva (3.08).

Ajuste entre la dificultad elegida y la habilidad estimada.

La correlación media entre dificultad y habilidad es 0.65, con valores muy similares en toda la secuencia (de 0.53 a 0.68, tabla 2). Estos resultados son inferiores a los descritos por Johnson et al. (1991) para los 11 primeros ítems (ítem 1= 0.68, ítem 11= 0.83). La raíz de la media de diferencias cuadráticas (RMDC) presenta una media de 0.50, con un rango entre 0.41 y 0.71 (tabla 4). Las discrepancias mayores se producen en los tres primeros ítems, estando la mayoría de los restantes por debajo de la media (gráfico 2). Teniendo en cuenta que el índice RMDC puede estar entre 0 (ajuste perfecto) y 4 (desajuste perfecto), un valor medio de 0.5 puede ser interpretado como un grado de correspondencia razonable.

TABLA 2. Ajuste dificultad y habilidad

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Correlación	.57	.66	.53	.67	.59	.60	.60	.61	.54	.68	.66	.58	.66	.64	.56	.61	.60	.65	.61	
Raíz media error cuadrático	.72	.67	.61	.49	.50	.44	.48	.42	.51	.42	.47	.50	.41	.48	.51	.46	.49	.41	.53	.50

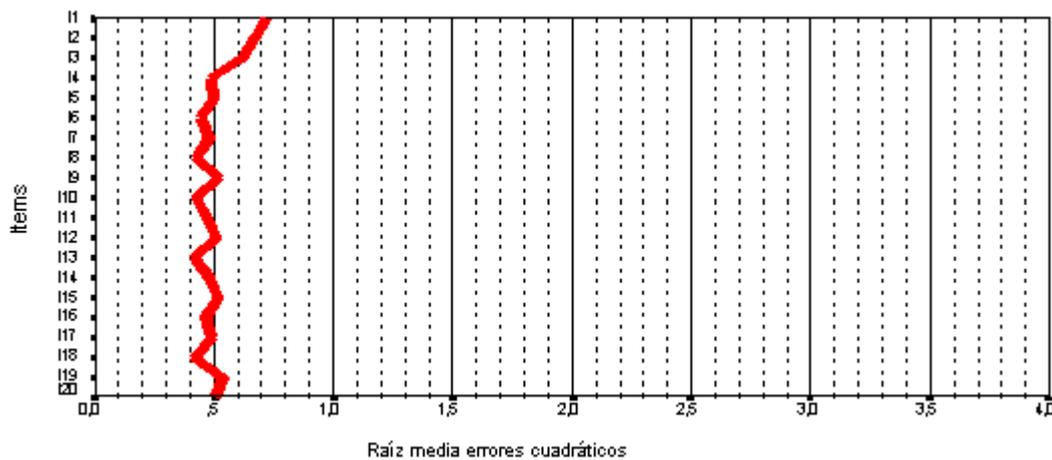


GRAFICO 1.- Perfil de ajuste dificultad-habilidad (RMEC)

Dificultad inicial y tendencia durante de la prueba.

La elección de la dificultad en los ítems 1, 11 y 20 permite apreciar claramente que la mayoría de los sujetos prefieren comenzar el test con un ítem de dificultad media o baja (tabla 3): el 86.3% eligen los niveles 1 y 2. La distribución

de frecuencias en los 5 niveles de dificultad presenta una ligera tendencia creciente: hay una disminución en los niveles más fáciles (1 y 2) y un aumento en los más difíciles (5). A la mismas consideraciones conducen las medias de los ítems. Estos resultados están en la línea de los obtenidos por Johnson et.al. (1991).

	1	2	3	4	5	Media	D. típica
Item 1	5.5	32.1	54.2	6.4	1.8	2.67	.76
Item 11	2.8	20.2	49.5	22.9	4.6	3.06	.85
Item 20	1.0	18.3	56.0	17.4	7.3	3.11	.82

TABLA 3. Elección de la dificultad en los ítems 1, 11 y 20

En segundo lugar, la regresión lineal de la posición serial sobre la dificultad es estadísticamente significativa, a nivel individual, sólo en el 22.5 % de los casos (3, $p < .001$; 6, $p < .01$ y 16 y $p < .05$), de los cuales 20 tienen pendiente positiva y 5 pendiente negativa. A nivel agregado, la pendiente media es nula (media= 0.06, d.típica= 0.34), $t = 0.02$, $p > .05$) y la interceptal se sitúa en el punto medio (media= 2.98,

d.típica= 0.67). Es decir, se elige inicialmente un nivel de dificultad moderado y se permanece en torno a él durante el resto del test. Los resultados no significativos a nivel agregado y los escasos significativos a nivel individual no permiten afirmar que las personas actúen siguiendo un patrón lineal de dificultad creciente, tal como señalan algunos autores (Rocklin, O'Donnell y Holst, 1995).

Si atendemos al perfil de dificultad media de los ítems en el conjunto de la muestra (gráfico 1), la tendencia no es completamente nula. Existen momentos diferenciados en la aplicación del test que sugieren una relación no lineal. Al comienzo hay un predominio de ítems fáciles, incrementándose su dificultad hasta llegar al ítem 5. Posteriormente, la dificultad desciende ligeramente y se mantiene relativamente constante con pequeñas oscilaciones. Al explorar el ajuste de diversas funciones no lineales, se aprecia una fuerte relación entre dificultad y orden de los ítems (gráfico 2 y tabla 4).

Las ecuaciones que mejor describen la tendencia son la función cúbica (86% varianza explicada: aumento rápido inicial y estabilización posterior, con una leve inflexión que marca un sucesivo descenso y ascenso) y la función inversa (69% varianza explicada: incremento muy acelerado en los primeros ítems e incremento muy lento en el resto). También cabe señalar que las ecuaciones no lineales predicen mejor de la dificultad inicial: las ordenadas en el origen se encuentran muy próximas a su valor empírico (media=2.67).

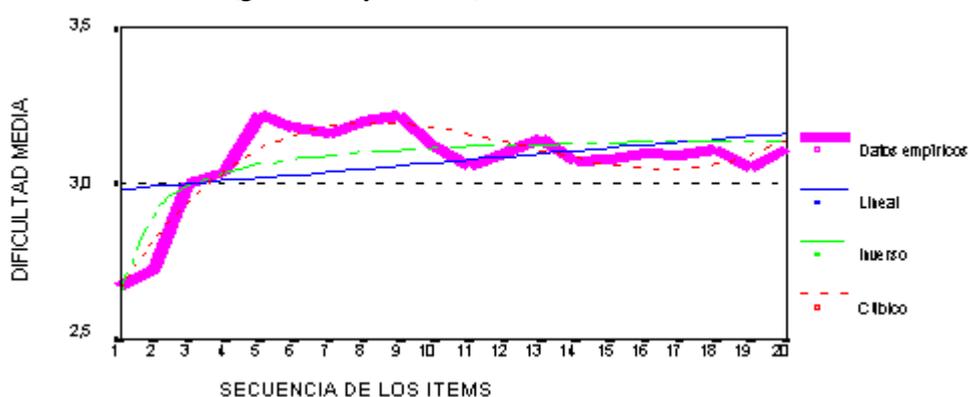


GRAFICO 2. Dificultad media de los ítems.

TABLA 4. Ecuaciones de regresión lineal y no lineal.

	R2	p	b0	b1	b2	b3
Función lineal	.159	.082	2.972	.010		
Función cuadrática	.563	.001	2.743	.072	-.003	
Función cúbica	.863	.000	2.471	.211	-.019	.0005
Función inversa	.687	.000	3.168	-.528		
Función logarítmica	.442	.001	2.826	.116		
Función potencial	.457	.001	2.818	.040		

Factores que predicen la elección de la dificultad de los ítems.

En el análisis de regresión jerárquico, las variables que predicen la elección de la dificultad son la capacidad del sujeto y la dificultad elegida en el ítem anterior (tabla 5, superior). El sentido de la predicción indica que las personas con mayor habilidad eligen ítems más difíciles y viceversa. Asimismo, la elección de un ítem previo fácil predice que la elección del siguiente

ítem también será fácil; sin embargo, al considerar la influencia de ítems más alejados del actual, la dificultad deja de tener poder predictivo. En algunas ocasiones la respuesta previa es significativa, mostrando débilmente que los aciertos conducen a ítems más difíciles y los fallos a ítems más fáciles. Por el contrario, la interacción entre la dificultad y la respuesta al ítem previo, el sexo y la ansiedad pretest no parecen predecir la elección de la dificultad

más allá de la capitalización del azar. Adicionalmente, hemos realizado un análisis de regresión con las variables que han mostrado algún poder predictivo (habilidad del sujeto, dificultad del ítem previo y respuesta dada al ítem

anterior) para simplificar los resultados. Los datos obtenidos son muy similares a los comentados, explicando entre el 33 y el 61% de la varianza.

TABLA 5.- Análisis de regresión jerárquica y simultánea.

PREDICTORES	CRITERIOS: Dificultad ITEM i																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
R. JERARQUICA																				
Sexo					+		+					+								(+)
Ansiedad													-							(-)
Habilidad	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+
R ²	.33	.50	.37	.50	.44	.46	.39	.50	.36	.53	.49	.46	.59	.43	.41	.42	.45	.50	.40	.43
ITEM i-1																				
Dificultad			+	+	+	+		+		+		+	+	+		+	+	+	+	+
Respuesta				+	+		+					(+)	(+)			(+)		+	+	+
Interacción			+																	
+ R ²	-	.08	.08	.07	.09	.07	ns	.09	.10	ns	.11	.10	.04	ns	.13	.15	.08	.13	.12	.11
ITEM i-2																				
Dificultad				+	+			+						+					+	+
Respuesta																				
Interacción																				
+ R ²	-	-	.06	.02	ns	ns	.06	ns	ns	ns	ns	ns	ns	.04	ns	ns	ns	.04	.08	ns
ITEM i-3																				
Dificultad					+															+
Respuesta																				
Interacción																				
+ R ²	-	-	-	ns	.07	ns	ns	ns	ns	ns	ns	ns	ns	.04	ns	ns	ns	ns	.06	ns
Total R ²	.33	.58	.51	.59	.60	.53	.45	.59	.46	.53	.60	.56	.71	.44	.54	.57	.53	.67	.66	.54
R. MULTIPLE																				
Habilidad	.55	.51	.24	.40	.22	.33	.49	.41(.22)	.58		.35	.30	.54	.48	.27	.28	.37	.36	.23	.33
Dificultad	-	.26	.50	.44	.55	.40	.42	.48(.21)			.49	.49	.28(.22)	.51		.53	.40	.50	.50	.45
Respuesta	-		.20	.19	.17	.21							.16							.16
R ²	.33	.53	.43	.57	.50	.50	.41	.58	.43	.52	.59	.51	.58	.43	.47	.56	.50	.61	.51	.50

Los coeficientes son significativos a nivel .05, salvo los que están paréntesis, que los son entre .05 y .07.

Estrategias de selección de ítems.

Globalmente, la distribución de elecciones de dificultad (tabla 6) indica que el comportamiento más frecuente es elegir el mismo nivel (61.5%). A pesar de esta tendencia a no cambiar, existen cambios en el 38.5% de las decisiones y parecen responder a una estrategia flexible: el aumento de la dificultad ocurre más

después de un acierto (26.1% de aciertos frente a 7.5% de fallos) y la disminución después de un error (41.3% de fallos frente a 7.9% de aciertos). Los resultados son semejantes a los obtenidos por Johnson et al. (1991).

Tabla 6.- Porcentaje de cambios de nivel de dificultad en función de los resultados al ítem anterior.

	RESPUESTAS		
DIFICULTAD	Aciertos	Fallos	Total
Aumento	26.1	7.5	20.4
Igual	66.0	51.2	61.5
Disminución	7.9	41.3	18.1
Total	69.5	30.5	100.0

Atendiendo a estrategias específicas, los sujetos siguen en su mayoría una estrategia flexible (60.4 %), un grupo menos numeroso utiliza la estrategia intolerante a la frustración (25.2 %) y prácticamente el resto emplea las estrategias inflexible (8.1 %) o tolerante a la frustración (3.6 %). Ningún caso presenta un patrón incongruente y sólo uno es caracterizado como indefinido. También se ha detectado otra estrategia no considerada previamente que puede calificarse como una conducta inadecuada para un TADI. Se trata de un comportamiento muy conservador seguido por 2 sujetos que eligen ítems extremadamente fáciles para su capacidad (una estrategia opuesta de alto riesgo, consistente en elegir ítems excesivamente difíciles, no se da). Los análisis diferenciales han sido realizados con las estrategias que tienen suficiente número de sujetos: flexible (F), tolerante a la frustración (T), inflexible (IF) e intolerante a la frustración (IT). El grado de ajuste entre elecciones teóricas y observadas es similar en todas ellas ($F=.82$, $p=.44$, tabla 7), a excepción de la estrategia inflexible cuyo ajuste es perfecto.

Las estrategias mencionadas presentan diferencias en todas las variables referidas al test: habilidad estimada, error típico de estimación, dificultad de los ítems, porcentaje de aciertos y tiempo de respuesta (tabla 7). Los sujetos de mayor habilidad se comportan de manera flexible mientras que los de menor habilidad lo ha-

cen de manera inflexible e intolerante ($F=7.26$, $p=.00$). La precisión de las estimaciones es significativamente menor en las estrategias tolerante e intolerante en comparación con la estrategia inflexible ($F=4.47$, $p=.02$), debido principalmente a que la diferencia entre la habilidad estimada para los sujetos (θ) y la dificultad de los ítems presentados (b) es mucho mayor ($F=13.58$, $p=.00$). Los sujetos de las estrategias flexible y tolerante eligen y contestan ítems más difíciles que los de las estrategias inflexible e intolerante (subjetiva, $F=23.31$, $p=.00$; objetiva, $F=23.59$, $p=.00$). Y las personas que siguen la estrategia tolerante obtienen menor porcentaje de aciertos ($F=9.34$, $p=.03$) y emplean mayor tiempo de respuesta ($F=3.63$, $p=.00$) que las que siguen otras estrategias. En cuanto al bienestar de los sujetos (tabla 8), los resultados indican que no hay diferencias en ansiedad entre las estrategias ($F=.84$, $p=.84$). No obstante, la ansiedad inicial disminuye con el transcurso de la prueba ($F=5.00$, $p=.03$), siendo diferente el cambio en función de la estrategia utilizada ($F=3.75$, $p=.01$). La reducción de la ansiedad tiene lugar en las estrategias intolerante e inflexible, mientras que en las estrategias flexible y tolerante la ansiedad permanece relativamente constante (gráfico 3).

Tabla 7.- Análisis diferencial de las propiedades del test según las estrategias de selección de items.

	1 FLEXIBLE n=67		2 n=4		3 TOLERANTE n=9		4 n=28		INFLEXIBLE		INTOLERANTE		ANOVA		F	p	Post.
	Media	D.T.	Media	D.T.	Media	D.T.	Media	D.T.	Media	D.T.	Media	D.T.	Media	D.T.			
Ajuste		1.27	(.28)	1.12	(.35)	0	(0)	1.31	(.28)	.82	.44	*					
Habilidad		.62	(.76)	.21	(.60)	-.16	(.88)	-.12	(.84)	7.26	.00	1- 3,4					
Error típico		.27	(.05)	.32	(.05)	.25	(.01)	.29	(.04)	4.47	.02	3- 2,4					
Theta-b		.31	(.39)	-.60	(.34)	.30	(.21)	.48	(.40)	13.58	.00	2- 1,3,4					
Dif. subjetiva	3.32		(.44)	3.79	(.54)	2.63	(.70)	2.53	(.47)	23.31	.00	1,2- 3,4					
Dif. objetiva	.31		(.50)	.81	(.63)	-.48	(.78)	-.60	(.56)	23.59	.00	1,2- 3,4					
% Aciertos		69.03	(12.53)	40.00	(7.07)	69.44	(9.17)	73.39	(11.06)	9.34	.00	2- 1,3,4					
Tiempo resp.		7.11	(1.76)	9.81	(.55)	7.27	(1.36)	7.48	(1.39)	3.63	.02	2- 1,3,4					

N=108, * Excluida estrategia flexible

Tabla 8.- Análisis diferencial de la ansiedad según estrategias de selección de items y momento de aplicación del test.

Momento de aplicación	Estrategia de selección de items								Total			
	FLEXIBLE		TOLERANTE		INFLEXIBLE		INTOLERANTE		Media	D.T.		
	Media	D.T.	Media	D.T.	Media	D.T.	Media	D.T.				
PRETEST	.96	(.42)	.80	(.29)	1.10	(.35)	1.11	(.36)	1.00	(.40)		
POSTEST	.92	(.44)	.83	(.05)	.90	(.28)	.82	(.36)	.89	(.40)		
Total	.94	(.43)	.81	(.20)	1.00	(.32)	.96	(.38)				
ANOVA	ESTRATEGIA		F= .84 , p= .84		MOMENTO		F= 5.00 , p= .03		Interacción		F= 3.75 , p= .01	

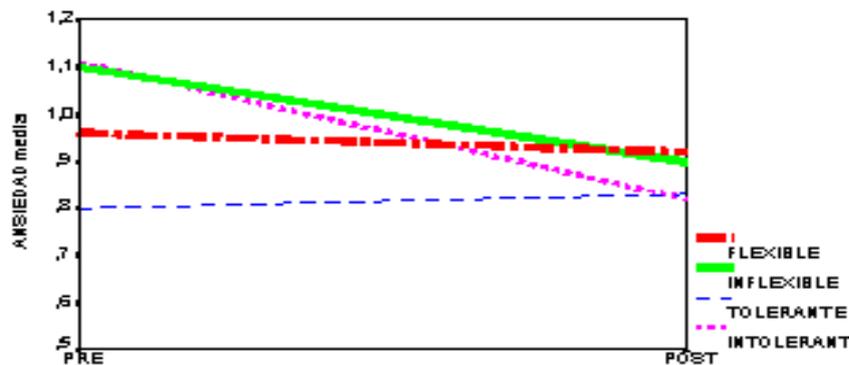


Gráfico 3.- Medias de ansiedad en estrategias de selección de ítems y momento de aplicación del test.

4. Discusión

Atendiendo al primer objetivo del estudio, podemos concluir que existe una correspondencia 'moderadamente alta' entre habilidad y dificultad. Los individuos más competentes eligen ítems más difíciles y los menos competentes ítems más fáciles y lo hacen de manera consistente a lo largo del test. Por tanto, actúan siguiendo las instrucciones y parece que son conscientes de la dificultad que deben tener los ítems apropiados para ellos. En la fase inicial del test se produce la correspondencia más débil, lo que sugiere que los primeros ítems pueden servir para explorar el significado de los niveles de dificultad y conocer lo que son capaces de hacer. Estas circunstancias remarcan la importancia de las instrucciones y la fase de práctica previa a la aplicación del test y la necesidad de mejorarlas. En este sentido, la eficacia de un TADI podría ser mejorada incluyendo en las instrucciones ejemplos que ilustren claramente el significado de los niveles de dificultad, presentando ítems de práctica de cada nivel y utilizando los resultados para hacer una estimación inicial de la habilidad que sirva para sugerir al sujeto el nivel de dificultad inicial más apropiado para él.

En cuanto al segundo objetivo, el perfil de dificultad del test se caracteriza por comenzar con ítems relativamente fáciles y continuar con ítems más difíciles. El incremento de la dificultad sigue una tendencia no lineal definida

por un rápido aumento en los primeros ítems y la estabilización en torno a niveles de dificultad media. Los principales predictores de la dificultad han resultado ser la habilidad del sujeto y la dificultad del ítem anterior. La importancia de la habilidad evidencia el comportamiento adaptativo de los sujetos en el TADI y la estrecha correspondencia existente entre habilidad y dificultad antes mencionada. La influencia de la dificultad previa indica que cuando se consideran secuencias de sólo dos ítems domina un patrón de conducta caracterizado por elegir el mismo nivel que en el ítem anterior. De lo que se desprende que para explicar los cambios habría que plantear hipótesis que consideren secuencias de ítem más amplias, pues, además, es razonable pensar que diferentes personas necesiten distinto número de aciertos o errores consecutivos antes de decidir probar con un nuevo nivel. Por otra parte, hemos obtenido que la dificultad no depende de la respuesta previa, es decir, la decisión de aumentar o disminuir la dificultad no es 'linealmente' contingente con los aciertos y fallos anteriores. Una causa plausible para ello es que si las personas reaccionan de diferente modo al éxito y al fracaso, pueden utilizar diferentes estrategias de afrontamiento del test que se compensan cuando son consideradas conjuntamente. Así pues, en nuestra opinión, un tratamiento más fructífero sería centrarse en los distintos estilos o estrategias de respuesta y analizarlas separadamente. Por último, nuestros resultados sobre la ansiedad pre-

test no coinciden con la literatura (Rocklin y O'Donnell, 1987; Johnson, Ross, Wise y Plake, 1991), pues ni siquiera predicen la elección de la dificultad inicial. Esta discrepancia se debe al tratamiento univariado empleado en los estudios precedentes frente al análisis multivariado de nuestro estudio. Desde un punto de vista univariado, también nosotros encontramos una relación negativa entre ansiedad y dificultad inicial ($r=-.29$, $p=.01$); sin embargo, como ambas variables están relacionadas con la habilidad (ansiedad-habilidad, $r=-.20$, $p=.05$; dificultad-habilidad, $r=.77$, $p=001$), la asociación desaparece al parcializar los efectos. En definitiva, los estudios anteriores no tienen en cuenta que si los sujetos tienen la misma habilidad, la ansiedad y la dificultad son independientes.

Y respecto al tercer objetivo, los resultados permiten afirmar que el proceso de cumplimentación de un TADI no es uniforme. Los sujetos utilizan diferentes estrategias de selección de ítems, que pueden estar determinadas por características cognitivas (habilidad) y afectivas (ansiedad) y que, a su vez, pueden influir en la calidad de las medidas en el test (precisión). Por un lado, la habilidad, en cuanto característica previa y relativamente estable, y la ansiedad, como factor situacional de la evaluación, pueden predisponer o determinar a los sujetos a emplear estrategias concretas. Las diferencias en habilidad no pueden ser debidas a la imprecisión de las medidas ni al uso de estrategias concretas, pues las estrategias flexible e intolerante, que comprenden el 85.6 % de la muestra, difieren en habilidad ($F=.62$, $IT=-.12$), siendo estimada ésta con una precisión equivalente ($F=.27$, $IT=.29$). Las personas con menor ansiedad y mayor habilidad adoptan las estrategias más desafiantes (flexible y tolerante), mientras que aquéllas con mayor ansiedad y menor habilidad adoptan las estrategias más cautas (inflexible e intolerante). Estas últimas son las que obtienen los beneficios psicológicos del procedimiento autoadaptado, puesto que

reducen significativamente su estado de ansiedad. Por otra parte, desde un punto de vista psicométrico el aspecto más destacable es que las estrategias conllevan diferencias en la precisión de las estimaciones. En este mismo ámbito, también pueden resultar de interés las siguientes observaciones. En primer lugar, hay evidencia de que ocurren conductas inadmisibles en un TADI (v.gr., las propias de estrategias radicalmente conservadoras o arriesgadas que conducen a acertar o fallar todos los ítems por ser excesivamente fáciles o difíciles para el sujeto). Aunque su aparición es reducida (1.8 %), no deberían ocurrir en absoluto y un TADI eficiente ha de controlar este evento (v.gr., establecer límites en el error de estimación tras un cierto número de ítems, recordar al sujeto el comportamiento correcto cuando se sobrepasan y, en su caso, controlar la selección de ítems de forma encubierta). En segundo lugar, los estudios realizados hasta el momento definen las estrategias de selección de ítem a partir de un criterio simple y desde una perspectiva racional 'a priori'. Este modo de proceder puede ser mejorado mediante la utilización de criterios más elaborados (v.gr, secuencias de aciertos y fallos diferentes), la fundamentación de las estrategias en teorías psicológicas (v.gr., motivación de logro, toma de decisiones) o el uso previo de métodos exploratorios cualitativos (v.gr., estudio de casos). En tercer lugar, la existencia de varios tipos de TADI, según las estrategias de selección de ítems utilizadas, sugiere que su comparación separada con los TAI y con tests convencionales informatizados podría ayudar a clarificar algunos resultados inconsistentes obtenidos en estas investigaciones.

Por último, queremos señalar varias limitaciones de nuestro trabajo que habrían de superar posteriores estudios. La más destacada es la 'ausencia de feedback' del resultado tras la respuesta a cada ítem. La relación de contingencia entre el resultado en el ítem anterior y la elección de la dificultad del ítem siguiente, en

el que se basa la definición de las estrategias de selección de ítems, descansa en el supuesto de que los sujetos perciben en mayor o menor grado la adecuación de sus respuestas; es decir, saben si han acertado o fallado un ítem por su familiaridad con el enunciado y sus alternativas. Evidentemente, existe una clara diferencia entre feedback objetivo y juicio subjetivo; no obstante, asumimos que cuando se trata de conocimientos sobre vocabulario inglés las diferencias no son muy marcadas. A pesar de todo, los resultados necesitan ser replicados incluyendo explícitamente feedback objetivo. La segunda limitación es el 'reducido número de niveles de dificultad' considerados. Utilizar solamente cinco niveles de dificultad da lugar a efectos de techo y suelo cuando los sujetos no tienen más categorías por encima o por debajo para elegir, quedando así afectados los índices de ajuste de las estrategias y la clasificación de los sujetos. Sería conveniente ampliar su número (v.gr., de 7 a 9 niveles) o diseñar métodos alternativos a dividir el banco de ítems en categorías fijas (v.gr., un método variable en el que los ítems más o menos difíciles según la habilidad estimada para el sujeto en cada momento). Asimismo, otra limitación importante es el 'desconocimiento de la distribución muestral del índice de ajuste', que no nos permite hacer hipótesis estadísticas acerca de los valores obtenidos. Clasificar a los sujetos en la estrategia donde obtienen el mejor ajuste es un criterio insuficiente, pues hay que tener presente que el 'mejor' ajuste no equivale a 'buen' ajuste. En definitiva, a pesar de las limitaciones señaladas, el presente estudio clarifica algunos aspectos de la conducta de las personas que responden a un TADI, ofrece sugerencias para mejorar este tipo de pruebas y abre diversas vías para futuras investigaciones.

Referencias

- Atkinson, J.W. y Litwin, G.H. (1970). Achievement motive and test anxiety conceived as motive to approach success and to avoid failure. *Journal of Abnormal and Social Psychology*, 60, 52-63.
- Hontagas, P., Canela, A. y Agustín, C. (1996). *Medida de la ansiedad desde la TRI: estudio del cuestionario de ansiedad-estado-rasgo*. Comunicación presentada al I Congreso de la Sociedad Española para el estudio de la Ansiedad y el Estrés, Benidorm.
- Johnson, D.F. y Mihal, W.L. (1973). Performance of blacks and whites in computerized versus manual testing environments. *American Psychologist*, August, 694-699.
- Johnson, P.L., Roos, L.L., Wise, S.L. y Plake, B.S. (1991). Correlates of examinee item choice behavior in self-adapted testing. *Mid-Western Educational Researcher*, 4 (3), 25-28.
- Koson, D.; Kitchen, C.; Kochen, M y Stodolosky, D. (1970). Psychological testing by computer: effect on response bias. *Educational and Psychological Measurement*, 30, 803-810.
- Lord, F.M. (1971). The self-scoring flexilevel test. *Journal of Educational Measurement*, 8, 147-151.
- Olea, J. y Ponsoda, V. (1996). Tests Adaptativos Informatizados. En J. Muñiz (Coor.). *Psicometría*. Madrid: Universitas.
- Olea, J., Ponsoda, V., Revuelta, J. y Belchí, J. (1996). Propiedades psicométricas de un tests adaptativo informatizado de vocabulario inglés. *Estudios de Psicología*, 55, 61-73.
- Ponsoda, V. Olea, J. y Revuelta, J. (1994). ADTEST: A computer-adaptative test based on the maximun information principle. *Educational and Psychological Measurement*, 54, 680- 686.
- Ponsoda, V., Olea, J., Revuelta, J. y Wise, S.L. (1995). *Tests adaptativos y autoadaptados informatizados: efectos en la ansiedad y en la precisión de las estimaciones*. Comunicación presentada en el IV Congreso de Metodología de las Ciencias Humanas y Sociales, La Manga (Murcia).
- Ponsoda, V., Wise, S.L., Olea, J. y Revuelta, J. (1997). An investigation of self-adapted test-

- ing in a Spanish high school population. *Educational and Psychological Measurement*, 57, 2, 210-221.
- Rocklin, T.R., O'Donnell, A.M. y Holst, P.M. (1995). Effects and underlying mechanisms of self-adapted testing. *Journal of Educational Psychology*, 87, 103-116.
- Rocklin, T.R. (1989). *Individual differences in item selection in self-adapted testing*. Annual meeting of the American Educational Research Association, San Francisco.
- Rocklin, T.R. (1994). Self-adapted testing. *Applied Measurement in Education*, 7, 3-14.
- Rocklin, T.R. y O'Donnell, (1987). Self-adapted testing: a performance-improving variant of computerized adaptative testing. *Journal of Educational Psychology*, 79, 315-319.
- Rocklin, T.R. y Thompson, J.M. (1985). Interactive effects of test anxiety, test difficulty, and feedback. *Journal of Educational Psychology*, 77, 368-372.
- Rocklin, T.R., O'Donnell, y Holst, P.M. (1995). Effects and underlying mechanisms of self-adapted testing. *Journal of Educational Psychology*, 87, 103-116.
- Vispoel, Rocklin y Wang (1994). Individual differences and test administration procedures: a comparison of fixed-item, adaptive, and self-adapted testing. *Applied Measurement in Education*, 7, 53-79.
- Wise, S.L. (1994). Understanding self-adapted testing: the perceived control hypothesis. *Applied Measurement in Education*, 7, 15-24.

AGRADECIMIENTOS

Esta investigación ha sido subvencionada parcialmente por la DGICYT (proyecto PS95-0046)

ABOUT THE AUTHORS / SOBRE LOS AUTORES

P. M. Hontagas (Pedro.M.Hontagas@uv.es). Universitat de València. **J. Olea**. Universidad Autónoma de Madrid. **V. Ponsoda**. Universidad Autónoma de Madrid. [Buscar otros artículos de estos autores en Google](#)

[Académico / Find other articles by those authors in Scholar Google](#)



RELIEVE

Revista ELectrónica de Investigación y EValuación Educativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).