

ACTITUDES HACIA LOS TEST INFORMATIZADOS APLICADOS POR INTERNET CON FORMATO *RESPONDER HASTA ACERTAR*¹

ATTITUDES TOWARDS THE COMPUTERIZED TEST ADMINISTERED VIA INTERNET WITH ANSWER-UNTIL-CORRECT ITEM FORMAT

MENÉNDEZ, L.; HIERRO, P. Y MUÑIZ, J.
Universidad de Oviedo

Resumen

Internet y las nuevas tecnologías están jugando un papel cada vez más importante en la evaluación psicológica y educativa. En los últimos años estamos asistiendo a un creciente trasvase de pruebas tradicionales de lápiz y papel a tests informatizados, lo cual plantea distintos problemas, entre ellos el de la equivalencia entre ambos tipos de pruebas. Aunque ambas modalidades de aplicación podrían considerarse equivalentes, en la práctica los resultados obtenidos difieren tanto en aspectos objetivos, tales como las características técnicas de la red, o la seguridad, como subjetivos, por ejemplo la ansiedad ante el ordenador, las actitudes, la motivación, o el feedback recibido. El presente trabajo trata de investigar las actitudes de las personas evaluadas hacia las pruebas informatizadas aplicadas vía Internet con formato Responder Hasta Acertar. Para ello se construyó una escala de actitudes de nueve ítems tipo Likert, que se aplicó a 990 estudiantes (574 chicos y 416 chicas) de edades comprendidas entre 13 y 21 años, con media 16,5 y desviación típica de 2. Todos los estudiantes tenían experiencia en pruebas informatizadas aplicadas con formato responder hasta acertar. Los resultados obtenidos muestran que los estudiantes tienen una clara predilección por la modalidad de evaluación informatizada, frente a

Abstract

Psychological and Educational assessment are deeply influenced by the computer technology and the Internet. Currently more and more classic paper and pencil assessment instruments are transformed into computer format, which generates different technical problems, among them the equivalence of the results obtained from both formats. In practice, the results obtained from paper and pencil and computerized modalities differ in many aspects, both objective and subjective, due to variables such as anxiety towards computers, attitudes, motivation, or the feedback received. In this paper the attitudes of the students towards computerized tests administered via Internet, using the answer until correct format, will be investigated. An attitude scale of nine Likert type items was developed, each item ranging from zero to ten. A sample of 990 students was used (574 boys and 416 girls), with ages ranging from 13 to 21 years, mean of 16,5 years and a standard deviation of 2. The results show that the students have a very positive attitude towards computerized tests administered via Internet using the answer until correct format. The partial credit assessment, the objectivity of this type of test, and the feedback offered by the answer until correct format, seem to be the main reasons for this positive attitude of the students towards this type of tests. This

¹ Agradecimientos: Este trabajo ha sido financiado por el Ministerio de Educación y Ciencia de España (Ref. SEJ2005-08924) y por la FICYT del Principado de Asturias (Ref.: IB05-027).

las pruebas clásicas de lápiz y papel. Saber que su conocimiento parcial será tenido en cuenta, la objetividad que las pruebas informatizadas proporcionan, y el feedback inmediato proporcionado por el formato *Responder hasta acertar*, son los tres factores hacia los que muestran una actitud más positiva. Todo hace pensar que esta actitud tan favorable de los estudiantes hacia los tests computerizados va a ser de gran ayuda en el futuro a la hora de introducir y generalizar este tipo de evaluación en la escuela.

Palabras Clave

Actitudes, Tests informatizados, *Responder hasta acertar*, crédito parcial, Feedback.

Los avances informáticos e Internet están jugando un papel muy importante en los cambios que se están produciendo actualmente en la evaluación psicológica y educativa (Naglieri & Drasgow, 2004; Olea & Posonda, 1998). No obstante, no conviene olvidar que estos avances no están exentos de problemas de todo tipo, tanto psicométricos como éticos y deontológicos (Coyne & Bartram, 2006).

Aunque las pruebas de lápiz y papel clásicas y las informatizadas podrían considerarse, salvo en el soporte, equivalentes, en la práctica, los resultados obtenidos en cada modalidad de evaluación no son concluyentes y difieren en importantes aspectos que conviene tener presentes a la hora de establecer las consiguientes equivalencias. Así, por ejemplo, algunos autores encuentran correlaciones entre 0,74 y 0,93 cuando se comparan la administración de determinados tests psicológicos con ambas modalidades de evaluación (Bedeian & Feild, 2006; Mead, 2002; Stanton, 1998). Otros encuentran correlaciones entre los tests de lápiz y papel y las informatizadas sensiblemente inferiores (Harris, 2006; Ployhart, 2003).

Por lo que se refiere a Internet, los tests informatizados aplicados a través de este medio proporcionan grandes ventajas, como son la ve-

positive attitude of the students towards computerized tests is going to be an important factor in order to generalize computerized assessment in the schools.

Key Words

Attitudes, Computerized tests, Answer until correct, Partial credit, Feedback

locidad, el bajo coste en la recogida de datos, confortabilidad para el usuario, utilización de ítems complejos e interactivos, etc. Pero también se plantean problemas nuevos, como la seguridad, confidencialidad de la información, autenticidad, o la vulnerabilidad de las plataformas y servidores. También el propio usuario puede verse afectado por la complejidad técnica implicada. Desde el punto de vista del sujeto examinado, tres son, según Harris (2006) los factores principales que conviene tener presentes cuando intentamos sustituir las tradicionales pruebas de evaluación por sus equivalentes a través de Internet: a) percepción de objetividad o justicia por parte de los usuarios b) percepción de seguridad y confidencialidad y c) la confortabilidad y motivación. Respecto al primer punto, uno de los aspectos más valorados por los usuarios de los ordenadores y en general de las pruebas informatizadas es la imparcialidad y objetividad que de ellos se derivan. La estandarización de las pruebas, el formato sencillo y diáfano de los tests de elección múltiple y la corrección imparcial de sus resultados son los aspectos mejor considerados por los examinados. La privacidad inherente a los procedimientos automatizados de evaluación, junto con la confianza en la seguridad de sus protocolos de corrección, son dos aspectos igualmente valorados por los usuarios

de los tests informatizados. El tercer factor citado, la confortabilidad y motivación, juega un papel fundamental en la validez de la prueba. Respecto a este último apartado, conviene señalar que la ansiedad de ejecución que va asociada a la realización de gran parte de los tests psicológicos, especialmente los de rendimiento o aptitudes, suele incrementarse notablemente como consecuencia de la interacción del usuario con los instrumentos tecnológicos (Dutke & Stöber, 2001). Una de las razones de este incremento de la ansiedad se debe a la sobrecarga cognitiva que el usuario experimenta al tener que atender simultáneamente a los contenidos propios del test y a la interacción con el soporte informático a través de la cual se administra la prueba de evaluación. Los factores de personalidad, de género (Whitaker, 2004), culturales, de edad y de experiencia previa informática pueden estar relacionados con la ansiedad hacia estas pruebas produciendo impactos adversos en determinados grupos de usuarios. Los resultados obtenidos hasta el momento no son concluyentes y en algunos casos no coinciden los resultados observados. Así por ejemplo mientras que algunos autores observan impactos adversos para las mujeres en la utilización de las pruebas informatizadas frente a los hombres (Chua, Chen, & Wong, 1999), otros estudios por el contrario no encuentran diferencia alguna en la ansiedad en función del sexo (Whitaker, 2004).

Según Bartram (2006) hay tres formas fundamentales en la aplicación de los tests informatizados. El modo abierto, sin barreras de ningún tipo para la realización de las pruebas. Son pruebas a las que se puede acceder desde Internet sin restricción alguna. Modo controlado, cuando se utilizan controles de acceso por medio de un registro de datos con login y password. Modo supervisado, cuando para velar por el óptimo desarrollo de la prueba, alguien supervisa todo el proceso garantizando tanto la identidad del usuario como el cumplimiento de las normas establecidas de la prueba. Según Bartram (2006) las funciones más importantes de este supervisor serían: a) Autenticación de la identidad de los usuarios, b) Establecimiento del oportuno rapport a fin de reducir la ansiedad e incrementar la motivación de los usuarios, c) Proporcionar las instrucciones precisas para la realización de la prueba, d) Velar por el normal desarrollo de la prueba, solucionando los posibles problemas que pueden surgir, e) Validación de los resultados del test asegurándose de que los resultados obtenidos coinciden con los esperados en la prueba, f) Asegurarse de que los materiales utilizados son originales y no han sufrido manipulación externa. En las tablas 1 y 2 se pueden observar resumidas los aspectos positivos y negativos de la evaluación informatizada, tanto para el administrador del test como para las personas evaluadas.

Tabla 1. Aspectos positivos y negativos de la evaluación informatizada desde el punto de vista del administrador del test.

Aspectos positivos	Aspectos negativos
<p><i>Internet</i></p> <p>Administración remota. Seguridad de los datos. Posibilidad de compartir los bancos de ítems. Usuarios potencialmente ilimitados. Bajo costo en la administración del test.</p>	<p><i>Internet</i></p> <p>Limitaciones técnicas de la conexión (ancho de banda, Fire-wall, persistencia en la conexión, etc.). Limitaciones del navegador e interface gráfica Alto costo en la implementación del test.</p>
<p><i>Red Local</i></p> <p>Además de los anteriores... Eficiencia, rapidez en la corrección de las pruebas, fiabilidad. Posibilidad de administrar tests adaptados y autoadaptados. Posibilidades ilimitadas del formato de ítems (imágenes, videos, textos, etc.).</p>	<p><i>Red Local</i></p> <p>Además de los anteriores... Brecha digital. Funcionamiento diferencial del test (sexo, raza, cultura informática).</p>
<p><i>Responder hasta acertar</i></p> <p>Además de los anteriores... Evaluación del crédito parcial. Posibilidad de adaptar con poco esfuerzo los bancos de ítems a este formato y conseguir con ellos el crédito parcial. Produce aprendizaje en los usuarios por el feedback inmediato del formato. Reduce la ansiedad que producen las pruebas informatizadas, disminuyendo el error de medida.</p>	<p><i>Responder hasta acertar</i></p> <p>Además de los anteriores... Mayor tiempo de respuesta. Tendencia a demorar la respuesta.</p>

Tabla 2. Aspectos positivos y negativos de la evaluación informatizada desde el punto de vista del examinado.

Aspectos positivos	Aspectos negativos
<p><i>Internet y Red Local</i></p> <p>Efecto de halo e interés por las TICS. Motivación. Rapidez en la ejecución del test. Posibilidad de corrección inmediata del test. Evaluación amplia del constructo. Imparcialidad en la corrección del test. Evita la expresión escrita. Expectativas de obtener puntuaciones extras (<i>Cheating-Gaming system</i>, adivinación).</p>	<p><i>Internet y Red Local</i></p> <p>Problemas técnicos (Ancho de banda limitado, restricciones en el interface gráfico —en el caso de Internet—, etc.). Escasa familiaridad con la tecnología informática. Sobrecarga cognitiva. Ansiedad. Distracción con los elementos del interface gráfico. Desconfianza sobre la eficiencia de esta modalidad de evaluación. Problemas de comprensión de los enunciados del ítem.</p>
<p><i>Responder hasta acertar</i></p> <p>Además de los anteriores... Feedback inmediato y/o diferido. Aprendizaje asociado al test. Posibilidad del crédito parcial.</p>	<p><i>Responder hasta acertar</i></p> <p>Además de los anteriores... No se permite la revisión de ítems. Desmotivación ante los errores cometidos.</p>

El formato *Responder hasta acertar*

De entre los muchos formatos de elección múltiple utilizados para administrar pruebas informatizadas el de *Responder hasta acertar* (RHA) tiene una larga tradición. Bajo este formato al examinado se le presenta el enunciado del ítem y se le muestran las opciones alternativas o distractores a fin de que vaya eliminando uno a uno hasta alcanzar la respuesta correcta. Tras cada eliminación del distractor realizada, se le proporciona información inmediata (feedback) acerca del resultado de su opción (acierto o error). Este método ha sido utilizado tradicionalmente con pruebas de papel y lápiz utilizando formularios impresos con el sistema tradicional de «rasca», aunque su impulso decisivo se produce a partir del uso de los ordenadores en el ámbito educativo. Sus poderosos efectos en el reforzamiento inmediato del aprendizaje ya han sido señalados en la década de los 50 por Pressey (1950). El inmediato feedback proporcionado tras cada elección realizada, la posibilidad de la estimación del crédito parcial (Hutchinson, 2001; Wilcox, 1981, 1982, 1983), y la acogida favorable que los estudiantes presentan hacia este tipo de formato, es considerado por algunos autores como uno de los mejores para aplicaciones de carácter didáctico (Ben-Simon, Budescu, & Nevo, 1997; DiBattista, Mitterer, & Grosse, 2004; Hanna, 1975; Hirose, 2000; Kane & Moloney, 1978).

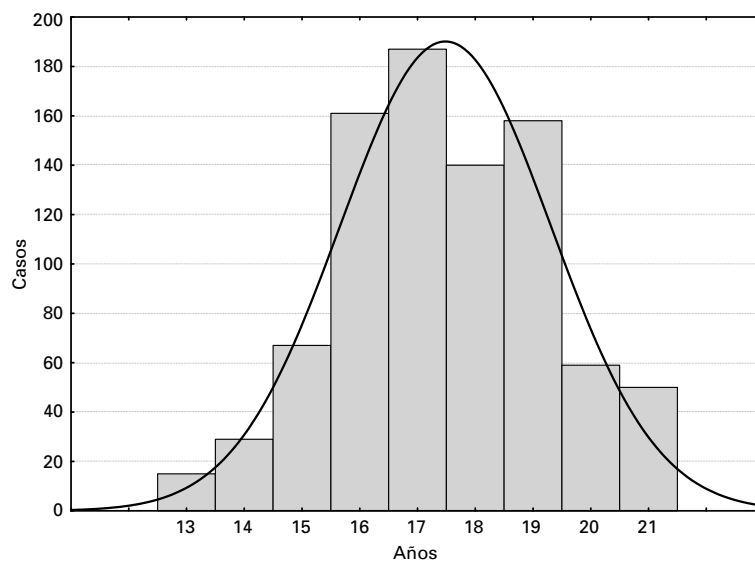
El objetivo de nuestro trabajo será precisamente estudiar el grado de satisfacción que los estudiantes muestran hacia los tests informatizados, y especialmente hacia aquellas pruebas de elección múltiple, administradas a través de Internet, con el formato de respuesta de Responder hasta acertar. Se trata de saber si los estudiantes prefieren la evaluación informatizada a la clásica de papel y lápiz, si la encuentran clara y objetiva. Se prestará especial atención al formato responder hasta acertar, de aplicación muy útil en contextos educativos, dado que proporciona feedback inmediato a los estudiantes tras su respuesta. Nos guía la hipótesis de que las actitudes de los estudiantes hacia estas pruebas informatizadas deberían de ser mucho más positivas que las relativas tanto a las pruebas de lápiz y papel, como a los formatos clásicos de acierto-error, en los que no se recibe retroalimentación inmediata sobre las respuestas.

Método

Participantes

Participaron un total de 990 estudiantes asturianos de ambos sexos (574 hombres y 416 mujeres) de enseñanza secundaria, bachillerato y ciclos formativos, de edades comprendidas entre los 13 y 21 años, con media 16,5 y desviación típica 2. En el gráfico 1 aparece la distribución de la muestra por edades.

Figura 1. Distribución de la muestra por edades.



Instrumentos

Se construyó una escala con el fin de explorar las actitudes de los estudiantes hacia la evaluación informatizada utilizando el formato *Responder hasta acertar*. La escala (Tabla 3) consta de nueve ítems tipo Likert, que van desde cero hasta diez.

Aunque los estudios realizados sobre el número óptimo de categorías que deben de tener las escalas de tipo Likert apuntan de cuatro a siete categorías (Cañadas Osinski & Sánchez Bruno, 1998; García Cueto, Muñiz, & Lozano, 2003; Hernández Baeza, Muñiz, & García Cueto, 2000; Morales, Urra, & Blanco, 2003; Pérez Santamaría, Rodríguez Testal, & Romero de Lora, 2002), aquí se han utilizado inicialmente 11 categorías (0: Totalmente en desacuerdo; 10: Totalmente de acuerdo). Dos son las razones fundamentales por las que nos hemos inclinado por ese número de categorías. Por un lado la familiaridad que los estudiantes tienen con este tipo de escala utilizada para las calificaciones escolares (de cero a diez puntos). Por otro, a fin de poder posteriormente colapsar las categorías e investigar el número óptimo de categorías que deberían utilizarse, tanto desde una óptica clásica como su uso con modelos de Teoría de Respuesta a los Ítems.

Procedimiento

Los 990 estudiantes de la muestra participaron en una evaluación informatizada sobre contenidos del currículo de Historia de secundaria, utilizando un formato de elección múltiple en la modalidad *Responder hasta acertar*. Los 25 ítems utilizados eran de elección múltiple con cinco alternativas, una de las cuales era la correcta, aplicándose a través de Internet. Tras la aplicación de esta prueba se aplicó a todos los estudiantes la escala de actitudes hacia las pruebas informatizadas con formato responder hasta acertar. La aplicación de la escala se hizo también vía Internet en grupos de 15 estudiantes. La aplicación fue supervisada en todo momento por un profesor que vela por el desarrollo normal del procedimiento.

Análisis de datos

La fiabilidad, la validez y el análisis de los ítems de la escala se analizaron tanto desde la óptica clásica como mediante los modelos de Teoría de Respuesta a los Ítems. Se utilizó el modelo de crédito parcial (Masters, 1982). Los datos se analizaron con los programas SPSS versión 12.0, Multilog versión 7.0 y Winsteps versión 3.63.0 (Linacre, 2006).

Tabla 3. Escala utilizada para evaluar la satisfacción de los estudiantes hacia las pruebas informatizadas con formato *Responder hasta acertar*

- | |
|---|
| <ol style="list-style-type: none"> 1. Recibir información inmediata sobre los aciertos y los errores en los ítems resulta agradable. 2. La información inmediata sobre los aciertos y los errores facilita el aprendizaje de la materia. 3. Recibir información inmediata sobre los aciertos y los errores me proporciona una mayor tranquilidad. 4. Este tipo de pruebas generan menos ansiedad que las habituales de papel y lápiz. 5. Este tipo de pruebas permite evaluar de forma objetiva lo que se ha estudiado. 6. En general prefiero este tipo de prueba a las clásicas que no dan información inmediata sobre los aciertos y los errores. 7. Si en la prueba de la PAU (Selectividad) me dejasen elegir, preferiría este tipo de prueba a las que se hacen en la actualidad. 8. Prefiero utilizar el ordenador para hacer los exámenes que hacerlos a mano en papel. 9. Si me dan a elegir, prefiero examinarme de una materia con una prueba tipo test a hacerlo mediante preguntas a desarrollar. |
|---|

Resultados

Como se puede observar en la Tabla 4, las puntuaciones medias de los ítems en la escala son elevadas, por encima de los 7 puntos, lo que indica que existe una actitud muy positiva de los estudiantes hacia las pruebas informatizadas

con el formato *Responder hasta acertar*. El coeficiente alfa de la escala fue de 0,75, y las correlaciones ítem-test son todas superiores a 0,30, lo que nos indica una aceptable discriminación de los ítems que conforman la escala. Las correlaciones entre los ítems (Tabla 5) también resultaron elevadas.

Tabla 4. Estadísticos descriptivos de la escala.

ÍTEMS de la escala	Hombres		Mujeres		Total		Correlación ítem-total corregida	Coeficiente Alfa eliminado el ítem
	Media	Desviación Típica	Media	Desviación Típica	Media	Desviación Típica		
1	7,95	2,41	7,86	2,27	7,91	2,35	0,35	0,74
2	7,69	2,42	7,85	2,23	7,76	2,34	0,38	0,74
3	7,09	2,59	7,04	2,42	7,07	2,51	0,33	0,75
4	8,17	2,53	8,05	2,59	8,12	2,55	0,47	0,73
5	7,02	2,60	7,09	2,31	7,05	2,48	0,47	0,73
6	7,35	2,83	7,46	2,72	7,40	2,78	0,43	0,73
7	7,61	2,71	7,5	2,80	7,57	2,74	0,54	0,71
8	7,78	2,71	7,51	2,71	7,66	2,70	0,53	0,71
9	7,85	2,70	7,65	2,78	7,77	2,73	0,39	0,74

Tabla 5. Matriz de correlaciones entre los ítems de la escala.

	1	2	3	4	5	6	7	8
1	,30							
2	,38	,33						
3	,20	,20	,17					
4	,24	,25	,26	,32				
5	,18	,19	,16	,30	,29			
6	,16	,18	,11	,37	,32	,35		
7	,14	,20	,12	,35	,28	,32	,52	
8	,05	,14	,09	,23	,21	,22	,39	,45

Tanto el valor significativo (0,001) del test de esfericidad de Bartlett hallado, como el alto valor de la adecuación muestral (KMO = 0,823), permiten llevar a cabo una análisis factorial de los ítems de la escala. En la Tabla 6 aparece la varianza explicada por los factores obtenidos. Como se puede ver en la Tabla 7, dos factores explican el 50% de la varianza total. No se puede hablar de una escala estrictamente unidimensional, ni ése fue el objetivo al construirla.

El primer factor (ítems 6, 7, 8, 9) se refiere a la preferencia de los estudiantes hacia las pruebas informatizadas frente a las clásicas de papel y lápiz, mientras que el segundo (ítems 1, 2, 3) alude a la importancia del feedback en el formato responder hasta acertar. La correlación entre ambos factores fue de 0,30.

Se aplicó el modelo de crédito parcial de Masters (1982) a los ítems de la escala. El modelo se ajusta razonablemente a los datos, si bien hay cuatro ítems (1, 3, 7, 8) cuyo ajuste es peor, como se puede ver en la Tabla 8. Por otro lado, tanto la observación de las funciones de información de las categorías de respuesta de la escala, como los análisis de bondad de ajuste realizados colapsando las 11 alternativas de respuesta inicial a 3 y a 5 categorías, nos permite concluir que, con 5 categorías de respuesta, la escala funcionaría con la misma eficiencia que con las 11 categorías originales. En el Gráfico 2 se puede observar la Función de Información de la escala y el error de medida. Como se puede ver, la escala resulta más precisa para niveles altos del constructo evaluado.

Tabla 6. Varianza explicada por los factores de la escala (método de extracción: Componentes principales)

Componente	Valores Propios	% de varianza explicada	% de varianza acumulado
1	3,06	34,0	34,0
2	1,39	15,5	49,5
3	0,83	9,2	58,7
4	0,71	7,9	66,6
5	0,70	7,7	74,3
6	0,68	7,5	81,8
7	0,62	6,9	88,7
8	0,56	6,2	94,9
9	0,46	5,1	100,0

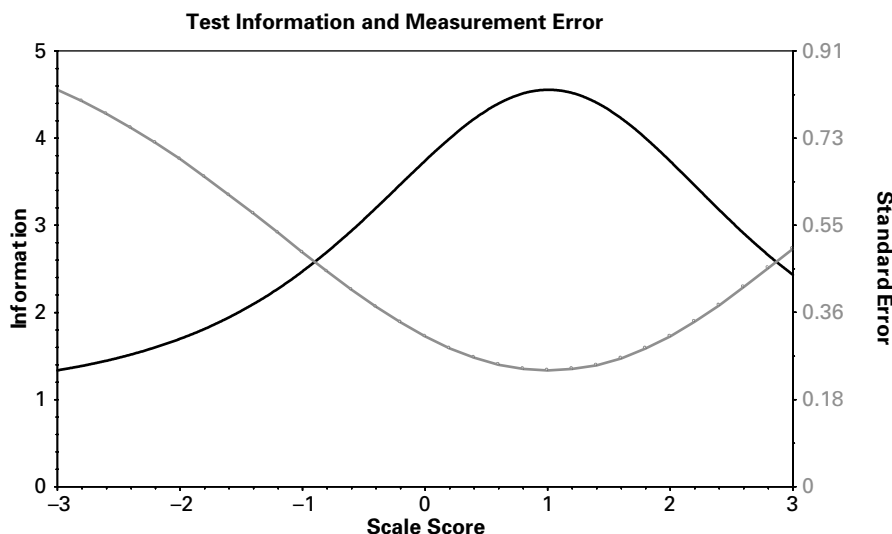
Tabla 7. Análisis factorial de los ítems de la escala (Método de extracción de componentes principales, con criterio de Kaiser y rotación oblicua).

Ítems	Factores	
	1	2
8	0,786	
7	0,780	
9	0,673	
4	0,608	
6	0,574	
5	0,516	0,502
3		0,759
1		0,745
2		0,666

Tabla 8. Ajuste de los 9 ítems de la escala de satisfacción al modelo de Crédito Parcial de Masters (1982). (PUNT: Puntuación de dificultad en logits; Err. Est.: Error estándar de la puntuación del ítem. INFIT: Ajuste interno OUTFIT: Ajuste externo; MNSQ y ZSTD= medias cuadráticas de los residuales no estandarizados y estandarizados).

ÍTEMS	Crédito Parcial		INFIT		OUTFIT	
	PUNT.	Err. Est.	MNSQ	ZSTD	MNSQ	ZSTD
1	-0,08	0,02	1,11	2,1	1,17	2,8
2	-0,11	0,02	1,07	1,4	1,07	1,2
3	0,04	0,02	1,18	3,7	1,20	3,9
4	-0,06	0,02	0,93	-1,3	0,92	-1,1
5	0,11	0,02	0,97	-0,6	0,99	-0,2
6	0,06	0,01	1,01	0,3	1,04	0,7
7	0,05	0,01	0,86	-3,0	0,83	-2,9
8	-0,01	0,01	0,86	-3,1	0,81	-3,1
9	0,00	0,01	1,07	1,4	1,11	1,7

Figura 2. Función de información y error de medida de la escala.



En líneas generales, la mayor parte de los estudiantes que han sido consultados han valorado muy positivamente los tests informatizados como instrumentos para evaluar sus conocimientos (Tabla 9). Así por ejemplo un 82% de ellos manifiestan estar de acuerdo o totalmente de acuerdo (42%) con el enunciado del ítem 1: *Recibir información inmediata sobre los aciertos y los errores en los ítems resulta agradable*. De igual modo, aproximadamente el 80% de los estudiantes, coinciden al señalar que la retroalimentación de sus aciertos y errores incide favorablemente en su proceso de aprendizaje (ítem 2), atribuyéndole un 67% de ellos un papel *muy im-*

portante en la reducción de la ansiedad cuando se enfrentan a los tests de evaluación (ítem 3).

Estas pruebas informatizadas les proporcionan, según ellos, mayor tranquilidad que las pruebas clásicas de lápiz y papel. Así pues, un 81% de los estudiantes afirman afrontar este tipo de pruebas con mayor tranquilidad al proporcionárseles información sobre sus aciertos y errores (ítem 4). Respecto a la validez aparente que, para los estudiantes tienen estos procedimientos informatizados de evaluación, la mayoría de ellos (71%) coinciden en considerarlos como pruebas objetivas y válidas para este cometido (ítem 5).

Tabla 9. Porcentajes de las respuestas de los estudiantes a los ítems presentados.

Categoría	1	2	3	4	5	6	7	8	9
0: Total desacuerdo	2%	1%	2%	3%	3%	4%	4%	3%	2%
1	1%	1%	1%	1%	1%	1%	0%	1%	1%
2	0%	0%	1%	1%	1%	1%	1%	0%	0%
3	2%	3%	3%	3%	4%	4%	4%	4%	2%
4	1%	1%	3%	1%	2%	3%	1%	2%	1%
5	13%	13%	23%	11%	18%	15%	16%	17%	13%
6	4%	4%	6%	4%	7%	4%	5%	4%	4%
7	21%	21%	18%	12%	21%	15%	13%	10%	21%
8	7%	8%	7%	6%	10%	9%	7%	6%	7%
9	8%	9%	8%	6%	8%	7%	7%	6%	8%
10: Total acuerdo	42%	38%	28%	53%	24%	38%	42%	45%	42%

Las respuestas dadas a los ítems 6, 7, 8 y 9 no dejan la menor duda de la clara preferencia de los estudiantes hacia este tipo de procedimientos informatizados de evaluación. Así por ejemplo, un 72% de ellos afirman preferir realizar pruebas que proporcionen feedback sobre aciertos y errores que aquellas otras pruebas que carecen de esta posibilidad (ítem 6). Un 74% de ellos optarían en la prueba de acceso a la universidad por pruebas de este tipo frente a las clásicas de lápiz y papel (ítem 7) y casi la mitad de ellos (48%) están completamente de acuerdo en sustituir las clásicas pruebas de desarrollo escrito por estas de elección múltiple (ítem 9).

Por lo que se refiere a la distribución de estas actitudes en relación a los niveles educativos (Tabla 10), no se observan grandes diferencias entre ellos respecto a cómo valoran el feedback proporcionado en este tipo de pruebas de *Responder hasta acertar*. Aunque son los estudiantes de Bachillerato y de ciclos formativos los que manifiestan una valoración más positiva acerca de la importancia que el feedback tiene en el aprendizaje, estas diferencias no alcanzan, sin embargo, la significación estadística en ninguno de los tres primeros ítems de la escala. Son, sin embargo, estos mismos alumnos de bachillerato y ciclos formativos los que menos convencidos están respecto a la reducción de la ansiedad que las pruebas informatizadas presentan con respecto a las prueba de lápiz y papel (ítem 4) ($\chi^2_6 = 14,6$; $p = 0,023$) y los que más críticos parecen ser respecto a la validez de las pruebas informatizadas para valorar de forma objetiva sus conocimientos (ítem 5) ($\chi^2_6 = 15,6$; $p = 0,016$).

En el mismo sentido se manifiestan los estudiantes de bachillerato frente a los de secundaria cuando optan por las pruebas informatizadas o por las pruebas clásicas de lápiz y papel. Así, por ejemplo son los estudiantes de bachillerato los que menos interés muestran hacia este tipo de evaluación ($\chi^2_6 = 13,2$; $p = 0,04$).

Discusión y Conclusiones

Los resultados muestran una disposición muy positiva de los estudiantes a la inclusión de las nuevas tecnologías de la información y comunicación (TIC) en sus actividades académicas. Cuando éstas son oportunas pueden facilitar los procesos de aprendizaje haciendo más agradable la tarea de aprender. La razón de ello se debe, muchas veces, a la posibilidad de retroalimentación que conlleva la automatización de la información, lo que actúa como un elemento muy poderoso en la motivación de estos (Shute, 2007). Sin embargo, esta visión favorable hacia las pruebas informatizadas no es constante en todos los niveles educativos. Parece existir una clara diferencia entre las opiniones de los estudiantes de secundaria y la de los estudiantes de bachillerato respecto a la valoración que ellos hacen respecto al uso de las pruebas informatizadas sobre las clásicas de lápiz y papel. Aunque en líneas generales existe un claro consenso en valorarlas positivamente, son los estudiantes de bachillerato los que más críticamente se muestran con esta tecnología. Las diferencias entre medias son significativas solo para el ítem 5 ($F = 13,3$; $p = 0,001$) y para el ítem 9 ($F = 21,3$; $p = 0,001$) en donde los

Tabla 10. Estadísticos descriptivos de la escala según el nivel académico (ESO = 683, Bachillerato = 307, Total = 990).

ÍTEMS	ESO		BACHILLERATO		TOTAL	
	Media	Desv. típ.	Media	Desv. típ.	Media	Desv. típ.
1	7,83	2,40	8,09	2,22	7,91	2,35
2	7,76	2,40	7,77	2,20	7,76	2,34
3	7,12	2,58	6,95	2,37	7,07	2,52
4	8,16	2,66	8,05	2,31	8,12	2,56
5	7,24*	2,45	6,62*	2,51	7,05	2,48
6	7,41	2,84	7,37	2,64	7,40	2,78
7	7,71	2,68	7,24	2,87	7,57	2,75
8	7,84	2,70	7,27	2,70	7,66	2,71
9	8,03*	2,60	7,18*	2,93	7,77	2,73

* Diferencias significativas 0,001.

estudiantes de bachillerato no están tan seguros como los de la ESO respecto a esos enunciados. La causa de esta diferencia la atribuimos a la mayor experiencia que estos estudiantes tienen con respecto a la administración de este tipo de pruebas frente al componente de novedad y expectación que para los estudiantes de secundaria posee todo lo relacionado con las TICS. A diferencia de otros modelos de formatos, donde la solución a las preguntas planteadas se proporciona al final de la prueba, esta retroalimentación inmediata de las respuestas del formato *Responder hasta acertar* (RHA) junto con la posibilidad de tomar en consideración el conocimiento parcial, parece disminuir la amenaza de la prueba de evaluación al proporcionar una percepción de control situacional sobre la tarea a realizar. Como consecuencia de ello se produce una estimulación más intensa que con las pruebas de retroalimentación diferida, propiciando con ello, adicionalmente, el aprendizaje de la materia (Whitelock, 2006). Aunque este formato de RHA es preferido a los otros alternativos, hay que tener presente, sin embargo, el objetivo de los tests a la hora de valorar estas actitudes. Cuando estas pruebas están planteadas como meros procedimientos para la auto-evaluación y el aprendizaje de una materia en particular, la disposición del alumno es diferente a cuando se le plantean como pruebas de evaluación de conocimientos con implicaciones en sus calificaciones escolares. En este segundo caso, los estudiantes que, tras los primeros intentos comienzan a constatar fracasos en su intento de responder correctamente al primer intento, comienzan a mostrar desaliento mucho antes que con aquellas otras pruebas en donde los resultados se proporcionan al final del ejercicio (Whitelock, 2006). Con éstas últimas el estado motivacional suele persistir más tiempo al mantener el alumno la esperanza de obtener buenos resultados hasta el final del ejercicio. A fin de solventar este tipo de in-

convenientes, se han propuesto variantes del formato RHA, como por ejemplo la eliminación de los distractores incorrectos con retroalimentación hasta alcanzar la respuesta correcta (Ben-Simon, Budescu, & Nevo, 1997).

Otro aspecto a reseñar es que son los mejores estudiantes los que manifiestan un mayor interés hacia estas pruebas con retroalimentación inmediata, sin importarles mucho el formato de la prueba ya sea ésta de elección múltiple o de respuesta construida. Así podrían interpretarse las correlaciones positivas encontradas entre el ítem 1 y 3 y la puntuación total obtenida en el test de rendimiento académico de Historia ($r = 0.13$; $p < 0.05$) y el ítem 9 ($r = -0.16$; $p < 0.05$). Por otro lado, son los alumnos con menor rendimiento académico los que más se inclinan por este tipo de pruebas de elección múltiple, considerándolas más atractivas que las de respuesta construida en las que prevalece la expresión escrita, tal vez al considerarlas como una medio de eludir la precisión en el lenguaje. En este sentido entendemos la correlación negativa y significativa que encontramos entre puntuación del test de rendimiento e ítem 9 de la escala.

Independientemente del efecto que sobre el aprendizaje tienen estos formatos de respuesta, no hay duda de que el formato RHA se presenta como una opción muy interesante para procesos de auto-evaluación y de aprendizaje, puesto que: a) permite evaluar el crédito parcial del conocimiento, b) produce feedback inmediato, c) produce gran satisfacción en los estudiantes, y d) permite adaptar sin problemas bancos de ítems puntuados dicotómicamente a puntuaciones politómicas, con el ahorro consiguiente de ítems que esto supone en los tests adaptativos informatizados. Los aspectos favorables que los tests adaptativos presentan respecto a la confortabilidad del usuario se sumarían a los, ahora, aportados por este formato.

Referencias

- Bartram, D. (2006). The internationalization of testing and new models of test delivery on the Internet. *International Journal of Testing*, 6(2), 121-131.
- Bedeian, A. G., & Feild, H. S. (2006). The Measurement Equivalence of Web-Based and Paper-and-Pencil Measures of Transformational Leadership. *Organizational Research Methods*, 9(3), 339-368.
- Ben-Simon, A., Budescu, D., & Nevo, B. (1997). A comparative Study of Measures of partial Knowledge in Multiple-Choice Tests. *Applied Psychological Measurement*, 21(1), 65-88.

- Cañadas Osinski, I., & Sánchez Bruno, A. (1998). Categorías de respuesta en escalas tipo Likert. *Psicothema*, 10(3), 623-631.
- Coyne, I., & Bartram, D. (2006). Introduction to the special issue on the ITC guidelines on computer-based and Internet-delivered testing. *International Journal of Testing*, 6(2), 115-119.
- Chua, S. L., Chen, D., & Wong, F. L. (1999). Computer anxiety and its correlates: a meta-analysis. *Computers in Human Behavior* 15(5), 609-623.
- DiBattista, D., Mitterer, J. O., & Grosse, L. (2004). Acceptance by undergraduates of the immediate feedback assessment technique for multiple-choice testing. *Teaching in higher Education*, 9(1), 17-28.
- Dutke, S., & Stöber, J. (2001). Test anxiety, working memory, and cognitive performance: Supportive effects of sequential demands. In P. Press (Ed.), *Cognition & Emotion* (Vol. 15, pp. 381-389).
- García Cueto, E., Muñoz, J., & Lozano, L. M. (2003). Efecto de la reducción de alternativas en las escalas tipo Likert desde la perspectiva de la TRI. Paper presented at the IX Conferencia Española de Biometría, La Coruña, 28-30 de mayo de 2003.
- Hanna, G. S. (1975). Incremental reability and validity of multiple choice tests with an answer-until-correct procedure. *Journal of Educational Measurement*, 12, 175-178.
- Harris, M. M. (2006). Internet testing: the examinee perspective. In B. Hambleton (Ed.), *Computer based testing and the Internet: Issues and advances* (pp. 114-133): Jhon Wiley & sons, Ltd.
- Hernández Baeza, A., Muñoz, J., & García Cueto, E. (2000). Comportamiento del modelo de respuesta graduada en función del número de categorías de la escala. *Psicothema*, 12(2), 288-291.
- Hirose, I. (2000). Answer-Until-Correct Item Response Model with Restricted Number of Responses. *Japan journal of educational technology*, 24(1), 53-62.
- Hutchinson, T. P. (2001). Partial Knowledge and Answer-Until-Correct Tasks in Birds and Humans. *Biometrics*, 57(4), 1251-1252.
- Kane, M., & Moloney, J. (1978). The Effect of Guessing on Item Reliability under Answer-Until-Correct Scoring. *Applied Psychological Measurement*, 2(1), 41-49.
- Linacre, J. M. (2006). *Winsteps Rasch measurements computer program* Chicago: Winsteps.com.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Mead, A. D., Coussons-Read, M. (2002). The equivalence of paper- and Web-based versions of the 16PF questionnaire. In F. L. Oswald, Stanton, J.M. (Ed.), *Being Virtually Hired: Implications of Web Testing for Personnel Selection*. Toronto: 17th Annual Conference of the Society for Industrial and Organizational Psychology.
- Morales, V. P., Urora, S. B., & Blanco, B. A. (2003). *Construcción de escalas de actitudes tipo Likert*. Madrid: La Muralla.
- Naglieri, J., & Drasgow, F. (2004). Psychological Testing on the Internet. New problems, Old Issues. *American Psychologist*, 59(3), 150-162.
- Olea, J., & Posonda, V. (1998). Evaluación informatizada en contextos de aprendizaje. In Vizcarro & León (Eds.), *Nuevas tecnologías para el aprendizaje*. Madrid.
- Pérez Santamaría, F., Rodríguez Testal, J., & Romero de Loera, B. (2002). Preferencias por formatos de respuesta en cuestionarios para encuestas. *Metodología de Encuestas*, 4(1).
- Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. F. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology*, 56, 733-752.
- Pressey, S. L. (1950). Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *The Journal of Psychology*(29), 419-447.
- Shute, V. (2007). Focus on Formative Feedback. *ETS, Princeton, NJ*.
- Stanton, J. M. (1998). An empirical assessment of data collection using the internet. *Personnel Psychology*, 51(3), 709-725.
- Whitaker, B. (2004). Internet-based attitude assessment: does gender affect measurement equivalence? *Computers in Human Behavior* 23(3), 1183-1194.
- Whitlock, D. (2006). Electronic Assessment: Marketing, Monitoring and Mediating Learning. *International Journal of Learning Technology (IJLT)*, 2(2).
- Wilcox, R. (1981). Solving Measurement Problems with an Answer-Until-Correct Scoring Procedure. *Applied Psychological Measurement*, 5(3), 399-414.
- Wilcox, R. (1982). Some new results on an answer-Until-Correct Scoring procedure. *Journal of Educational Measurement*, 19(1), 67-74.
- Wilcox, R. (1983). How do examinees behave when taking multiple-choice tests? *Applied Psychological Measurement*, 7(2), 239-240.