

VALORACIONES DEL ALUMNADO UNIVERSITARIO SOBRE LAS PRUEBAS OBJETIVAS

Javier Gil Flores

jflores@us.es

Departamento MIDE. Universidad de Sevilla

C/ Camilo José Cela, s/n

41.018-Sevilla

RESUMEN

El estudio se centra en el uso de las pruebas objetivas para la evaluación del aprendizaje universitario. Tomando como referencia las ventajas e inconvenientes señaladas en la literatura, hemos recogido las valoraciones que el alumnado de diferentes titulaciones universitarias hace de este tipo de pruebas. Los resultados muestran una discreta valoración de las pruebas objetivas, que difiere según el área de enseñanza en la que nos situemos y es mejor entre los alumnos que entre las alumnas. El trabajo finaliza con una discusión de los resultados, que da lugar a plantear recomendaciones para mejorar la construcción y el uso de este instrumento de evaluación.

Palabras clave: *enseñanza universitaria, evaluación del aprendizaje, pruebas objetivas.*

ABSTRACT

This study focuses on the use of objective tests for learning assessment at University level. Taking into account the advantages and disadvantages reported in previous studies, we have collected some data related to the assessment that students from different degrees make of this kind of tests. Results show a moderate appraisal of objective tests, which differs depending on the subject and the sex of the students, with better results with female students. The paper finishes with a discussion of the results, which provides some recommendations to improve the construction and use of this evaluation tool.

Key words: *University teaching, learning assessment, objective tests*

INTRODUCCIÓN

La evaluación del rendimiento constituye una tarea de crucial importancia en el ámbito de los procesos de enseñanza-aprendizaje universitarios. Para el profesor, la evaluación del rendimiento provee retroalimentación sobre la eficacia de su actuación docente, permitiéndole identificar los resultados de su trabajo, determinar qué contenidos no han sido suficientemente adquiridos y, en consecuencia, tomar decisiones que podrían afectar al modo en que se enfoca la enseñanza. Desde la perspectiva del alumno, la evaluación del rendimiento supone una piedra de toque para valorar su progreso en el aprendizaje, pero sobre todo, representa el modo en que la institución universitaria comprueba su grado de competencia en las diferentes disciplinas académicas, de cara a otorgarle una titulación. La repercusión personal que conlleva el éxito o fracaso del alumno en una materia hace que la evaluación sea un foco de atención prioritario del alumno, y que en función de ésta se oriente en buena medida su trabajo.

Cuando la evaluación del aprendizaje tiene como fin valorar el nivel de logro alcanzado para otorgar una calificación académica, la objetividad en la asignación de tales calificaciones es una aspiración docente. Esta es una de las razones por las que el uso de las pruebas objetivas se ha extendido ampliamente en la enseñanza universitaria, tanto para evaluar el rendimiento alcanzado al cursar una materia como para la selección o la obtención de un grado. En cierto modo, el uso de las pruebas objetivas se ha convertido en uno de los procedimientos más populares para la evaluación del rendimiento universitario, presentando frente a los exámenes tradicionales contruidos mediante preguntas de ensayo toda una serie de ventajas que la profusa literatura sobre el tema se ha encargado de destacar (Heywood, 1989; Ory y Ryan, 1993; Haladyna, 1994; Smith, Brown y Race, 1996; Linn y Gronlund, 2000).

Una prueba objetiva puede constar de un número amplio de items, por lo que es posible cubrir con mayor exhaustividad los contenidos de aprendizaje objeto de evaluación, y lograr así una elevada validez curricular (ajuste del instrumento a la materia objeto de evaluación). En la práctica, posiblemente el principal atractivo de una prueba objetiva es su rapidez y facilidad para calificarla, que la hacen especialmente adecuada en situaciones en las que el número de estudiantes evaluados es alto. A diferencia de las pruebas de ensayo, en las que la amenaza de una valoración subjetiva está presente, para las preguntas incluidas en una prueba objetiva el profesor puede determinar inequívocamente si la respuesta del alumno es o no acertada. El uso de plantillas de respuesta permite la mecanización del proceso de corrección, con la posibilidad de conexión a programas informáticos capaces de analizar las características técnicas de la prueba y la idoneidad de los items. El cálculo de medidas de dificultad y discriminación para los items, o de los correspondientes parámetros según modelos de la Teoría de Respuesta al Ítem (TRI), ha permitido construir bancos de items con las indudables ventajas que ello representa de cara a la construcción de pruebas o la administración informatizada.

Un formato habitual para las preguntas que forman parte de una prueba objetiva es el ítem de opciones múltiples, que sitúa al alumno ante varias respuestas, para que elija entre ellas la respuesta correcta. El predominio de este formato de preguntas ha llevado incluso a identificar prueba objetiva y prueba de opciones múltiples (Heywood, 1989:285). Otros formatos de respuesta pueden ser utilizados, tales como los items de

verdadero-falso, de emparejamientos, de completar frases, de respuesta breve. La amplia utilización de las pruebas objetivas basadas en cuestiones de opciones múltiples ha hecho que la construcción y uso de este tipo de items centre la atención de numerosos estudios empíricos y reflexiones teóricas.

Si bien el uso de pruebas de opciones múltiples conlleva una serie de ventajas que hemos destacado anteriormente, también se han señalado límites a su utilización. Así, un problema estaría en la dificultad para la redacción de buenas preguntas, de tal manera que se suele proponer una buena respuesta correcta pero malas respuestas incorrectas o distractores (Gibbs, Habeshaw y Habeshaw, 1986).

Otra limitación apuntada se encuentra en el tipo de aprendizaje que permiten medir. Se ha afirmado que las pruebas objetivas se prestan a la evaluación de aprendizajes simples (definiciones, conceptos, recuerdo de datos, reconocimiento de hechos, etc.), sin ser adecuadas para evaluar aspectos reveladores de un conocimiento profundo, tales como capacidad de crítica, análisis, reflexión. En este sentido, Scouller (1997) encontró que la razón que lleva a los estudiantes a preferir las preguntas de opciones múltiples es precisamente su menor exigencia en cuanto a procesos cognitivos que deben ser desarrollados para responder. En contra de este tópico, la investigación empírica ha demostrado la similitud entre las competencias evaluadas a través de preguntas abiertas y preguntas de opciones múltiples, y los estudios sobre la construcción de este tipo de pruebas han subrayado la posibilidad de construir items variados en cuanto a demanda cognitiva y adecuados para la evaluación de distintos aprendizajes (Bennett, Rock y Wang, 1991; Stape, 1995; Martínez, 1999; Berezina y Berman, 2000). El debate sobre esta cuestión es interesante si tenemos en cuenta que en la práctica habitual, el alumno responde al estudio de una materia en función de los requerimientos inmediatos que la obtención de una evaluación positiva le exige (Gibbs, 1999). En cualquier caso, la mayoría de los manuales sobre evaluación del aprendizaje insisten en la necesidad de un enfoque múltiple, que descarte planteamientos de evaluación apoyados exclusivamente en este tipo de pruebas (Biggs, 2002).

Un inconveniente más se halla en que los estudiantes no tienen la oportunidad de explicar por qué eligen un distractor, posibilidad ésta que resultaría interesante de cara a conocer el modo de razonar de los alumnos y mejorar los items (Kottke, 2001). Tampoco los estudiantes reciben explicaciones sobre las razones por las que una opción es o no correcta, lo cual limita las posibilidades de un uso formativo de este tipo de instrumentos. Es más, se ha descrito el efecto negativo que las respuestas falsas tienen sobre el aprendizaje del alumno, que tenderá a considerarlas correctas si vuelve a encontrarlas en una segunda prueba (Toppino y Luipersbeck, 1993). La solución a este efecto no deseado podría estar en las estrategias favorecedoras de una retroalimentación al alumno, que pasarían por la simple revelación de las respuestas correctas, o el comentario y discusión de las diferentes opciones de respuesta una vez finalizada la prueba (Talmir, 1991; Epstein y otros, 2001; Epstein y Brosvic, 2002).

Estrategias del alumno a la hora de responder, tales como la adivinación ensayada cuando se desconoce la respuesta correcta o la respuesta al azar, han dado lugar al desarrollo de soluciones estadísticas adecuadas para la detección y/o corrección de tales problemas (Choppin, 1985; Burton y Miller, 1999; Charter, 2000; Belleza y Belleza, 2002). Tratando de atenuar el efecto del azar, se han adoptado estrategias basadas en

introducir más opciones y añadir una escala de confianza en la respuesta (De Pablos y otros, 1983). Una interesante vía para discriminar entre la respuesta al azar y la respuesta debida a un conocimiento parcial de la materia por parte del examinado ha sido aportada por Bush (2001), quien propone un nuevo formato de ítem en el que es posible seleccionar más de una respuesta cuando el examinado no tiene plena confianza sobre una opción. La justificación de la respuesta por parte del examinado (Dodd y Leal, 2002; Nield y Wintre, 2002) constituye otra alternativa para descartar la posibilidad de un acierto por azar, si bien en este caso nos estaríamos alejando del formato de opciones múltiples para convertir a este tipo de ítems en preguntas próximas a las cuestiones de respuesta breve.

Para hacer frente a la copia fraudulenta durante la realización del examen, el recurso habitual ha venido siendo la elaboración de diferentes versiones de una misma prueba, recurso que sigue utilizándose en trabajos recientes (Denyer y Hancock, 2002).

Una conducta del examinado que ha centrado la atención de diversos estudios es el cambio de la opción de respuesta inicialmente marcada, analizándose las razones y los efectos del mismo. Los resultados apuntan hacia un beneficio, en términos de puntuación final, para los alumnos que realizan tales cambios, salvo cuando la razón por la que se realizan es un intento de adivinación de la respuesta (Shatz y Best, 1987; Geiger, 1991; Schwarz y otros, 1991). Según el estudio de Friedman y Cook (1995), los efectos sobre la calificación conseguidos al realizar el cambio de respuesta no parecen relacionarse con el estilo cognitivo de los examinados (dependencia/independencia de campo; impulsividad/reflexividad).

Otra conducta habitual es la de escribir sobre la prueba, subrayando términos, marcando provisionalmente más de una respuesta para luego revisarlas y adoptar una decisión definitiva, o haciendo diagramas, esquemas, gráficos y anotaciones diversas. Se ha hallado una correlación significativa entre el número de marcas realizadas durante el examen y el resultado obtenido en el mismo (LoSchiavo y Shatz, 2002), revelándose una opinión favorable de los estudiantes sobre la posibilidad de escribir sobre la prueba y la contribución de esta práctica a un mayor rendimiento en la misma.

La investigación sobre pruebas objetivas y, en particular, pruebas constituidas por preguntas de opciones de respuesta múltiples ha considerado algunas variables personales en función de las cuales valorar el funcionamiento de tales instrumentos de evaluación. Los estudios que han tomado en consideración el sexo de los examinados (Murphy, 1982; Anderson, 1989; Garner y Engelhard, 1999; Beller y Gafni, 2000; Bielinski y Davison, 2001) apuntan diferencias en los resultados a favor de hombres o mujeres de manera inconsistente, por lo que resulta difícil llegar a conclusiones generales sobre el efecto del género en las puntuaciones obtenidas mediante este tipo de pruebas. En cuanto a la conducta de los examinados, se ha encontrado una mayor tendencia en los hombres a adivinar la respuesta correcta y en las mujeres a la omisión cuando la desconocen (Ben-Shakhar y Sinai, 1991).

La ansiedad ante los exámenes constituye otro de los rasgos abordados en el estudio de las pruebas objetivas. Se ha observado una mayor ansiedad entre mujeres que entre hombres (Pamphlett y Farnill, 1995) a la hora de afrontar una prueba de opciones múltiples, si bien ésta parece manifestarse en menor medida ante las pruebas de opciones múltiples que ante las pruebas de ensayo (Choi, 1998). Entre los diferentes formatos de

items característicos de una prueba objetiva, Shaha (1984) encontró que los items de emparejamiento provocan menos ansiedad que los de opciones múltiples.

A veces, se ha afirmado que determinados estudiantes poseen mayores destrezas que otros para enfrentarse a pruebas objetivas. Un reciente estudio con alumnos universitarios ha revelado que los resultados obtenidos en pruebas de opciones múltiples no están relacionados con la percepción que tienen los estudiantes sobre sus habilidades para responder a exámenes (Smith, 2002).

En definitiva, la mayor parte de los trabajos realizados sobre las pruebas objetivas analizan el modo en que éstas han de ser construidas, su funcionamiento cuando se utilizan para la evaluación de aprendizajes, las conductas o estrategias del alumnado a la hora de responder, o la incidencia de sus características personales en la respuesta a las mismas y en los resultados logrados. De tales trabajos se deducen argumentos en los que apoyar una valoración de las pruebas objetivas.

Haciendo frente a la escasez de estudios interesados por las opiniones de los estudiantes acerca de este instrumento de evaluación, en el presente trabajo trataremos de conocer el valor que los alumnos atribuyen a las pruebas objetivas utilizadas en la evaluación del rendimiento, sometiéndolo a su consideración diferentes aspectos que la literatura ha abordado al valorar este tipo de pruebas y que han sido recogidos a lo largo de estos párrafos introductorios.

PROBLEMA Y OBJETIVOS

De acuerdo con la idea apuntada en el apartado anterior, el problema abordado en nuestro estudio es la valoración que hacen los estudiantes universitarios sobre las pruebas objetivas como instrumento para la evaluación del aprendizaje. El problema que planteamos se concreta en al menos tres objetivos, que tratamos de cubrir:

- Determinar la valoración que el alumnado universitario hace de las pruebas objetivas en general, y particularmente de los diferentes aspectos que se vienen manejando en la literatura generada acerca de este tópico, tales como la formulación de las preguntas, su contenido, el tipo de aprendizaje que permiten medir, las estrategias de respuesta adoptadas por los evaluados, el papel del azar, la utilización de este tipo de pruebas, el modo en que son calificadas, o las posibilidades de retroalimentación a partir de los resultados.
- Analizar la existencia de diferentes valoraciones en función del contexto académico en el que se ubica el alumno, de tal manera que podamos comprobar si se mantienen o no diferentes puntos de vista sobre las pruebas objetivas entre el alumnado de titulaciones adscritas a áreas de enseñanza universitaria diversas.
- Estudiar la incidencia de la variable sexo en las valoraciones que el alumnado hace acerca de las pruebas objetivas, determinando si existe una valoración diferenciada entre ambos géneros.

Conocer las valoraciones de los alumnos acerca de un elemento clave en la evaluación del rendimiento universitario, como son las pruebas objetivas, permite contrastar a partir de sus destinatarios las ventajas e inconvenientes que se hacen de este instrumento en la

literatura especializada. De las principales críticas suscritas por los encuestados, podrían derivarse los elementos fundamentales a los que debería prestarse atención si pretendemos mejorar técnicamente la construcción y uso de este tipo de pruebas al tiempo que incrementamos la aceptación de las mismas por parte de los estudiantes.

METODOLOGÍA

El estudio se ha apoyado en una metodología de encuesta, llevada a cabo mediante la administración de un cuestionario escrito. La adopción de este método de trabajo responde a la intención de obtener las valoraciones de un colectivo amplio de sujetos, con un coste relativamente bajo en tiempo y esfuerzo. Partimos de una concepción previa acerca de qué aspectos van a ser valorados, los cuales surgen de la revisión de la literatura y se reflejan en el instrumento empleado para recoger las valoraciones de los estudiantes.

Definiendo como población objeto de estudio al colectivo de estudiantes que cursan los últimos años de las respectivas titulaciones impartidas en los centros de la Universidad de Sevilla, hemos pretendido recoger las valoraciones de una muestra amplia de sujetos, en la que se vieran representados estudiantes que cursan titulaciones pertenecientes a diversas áreas de enseñanza universitaria. Un sondeo previo entre las Delegaciones de Alumnos de los respectivos centros de la Universidad de Sevilla llevó a descartar la participación de alumnado que cursa titulaciones del área de Artes y Humanidades, dada la baja utilización que de las pruebas objetivas se hace en tales carreras. Se seleccionaron en consecuencia algunas titulaciones de las áreas de Ciencias Exactas y Naturales (Biología), Ciencias Sociales y Jurídicas (Psicología, Relaciones Laborales), Ciencias de la Salud (Farmacia, Medicina, Odontología) e Ingeniería y Tecnología (Ingeniero Industrial, Ingeniero de Telecomunicaciones).

Los sujetos invitados a participar en el estudio habrían de contar con una cierta experiencia como alumnos examinados mediante pruebas objetivas para la evaluación del rendimiento. Tratando de garantizar este criterio, se consideraron titulaciones donde existe una práctica habitual de utilización de pruebas objetivas, y tuvimos en cuenta únicamente los grupos de segundo ciclo. Una vez seleccionados los sujetos, para comprobar que la muestra satisface este criterio se preguntó a los encuestados por el nivel de uso de las pruebas objetivas a lo largo de los años cursados en su carrera. De acuerdo con los datos recogidos, para el 34,6% de los participantes las pruebas objetivas son el procedimiento más utilizado en la evaluación de las materias que cursa; para el 58,3% este tipo de pruebas se encuentra entre los procedimientos habituales; y sólo un 7,2% los tiene por instrumentos raramente utilizados.

A los criterios de diversidad y experiencia, se sumó el de accesibilidad a los grupos, que se ha visto facilitada por la aceptación de los profesores para que el instrumento de recogida de datos pudiera ser administrado en alguna de sus clases. Finalmente, se recogieron valoraciones de un total de 762 estudiantes, de los cuales el 41,3% eran hombres y el 58,7% mujeres. La distribución de la muestra por titulaciones queda recogida en la tabla 1.

TABLA 1
DISTRIBUCIÓN DE LOS SUJETOS ENCUESTADOS POR ÁREAS Y TITULACIONES

Área	n	%	Titulación	n	%
<i>Ciencias Exactas y Naturales</i>	112	14,7	· Biología	112	14,7
<i>Ciencias Sociales y Jurídicas</i>	259	34,0	· Psicología	119	15,6
			· Relaciones Laborales	140	18,4
<i>Ciencias de la Salud</i>	229	30,1	· Farmacia	83	10,9
			· Medicina	72	9,4
			· Odontología	74	9,7
<i>Ingeniería y Tecnología</i>	162	21,3	· Ingeniero Industrial	30	3,9
			· Ingeniero de Telecom..	132	17,3
Total	762	100,0	Total	762	100,0

Para obtener las valoraciones de los estudiantes integrados en la muestra, se construyó la *Escala de Valoración de las Pruebas Objetivas por Estudiantes Universitarios*, que consta de un total de 14 enunciados valorativos, de los cuales la mitad encierra una valoración positiva y la otra mitad una valoración negativa. En relación a cada ítem, se pide al encuestado que se posicione en una escala que va de 1 a 7, teniendo en cuenta el grado en que se identifica con la valoración contenida en el ítem. En el instrumento se incluye además un ítem final, recogiendo la valoración global del alumno sobre las pruebas objetivas, y una pregunta acerca de la frecuencia de uso de este tipo de pruebas en el contexto universitario en el que se desenvuelve. Para facilitar la comprensión por parte de los encuestados, las alusiones a pruebas objetivas se hicieron en términos de «examen tipo test», dado que esta denominación es la utilizada por la mayor parte del alumnado.

La recogida de datos se llevó a cabo durante el segundo cuatrimestre del curso 2002/03, realizando una administración colectiva del instrumento por medio de encuestadores desplazados a los respectivos grupos de clase.

El análisis de los datos se ha apoyado en técnicas estadísticas. Concretamente, se ha llevado a cabo un estudio descriptivo de las respuestas ofrecidas por los encuestados, para las cuales se han calculado la media y la desviación típica. Con el fin de examinar las diferencias de valoración en función del área de enseñanza, se recurrió al análisis de la varianza de un factor, y en el caso de las diferencias en función del género se aplicó la prueba t de Student, previa comprobación del supuesto de igualdad de varianzas mediante la prueba de Levene.

RESULTADOS

La valoración de los estudiantes

Atendiendo a las respuestas de los encuestados al ítem final mediante el que se recogió una valoración general sobre las pruebas objetivas (*En general, mi valoración*

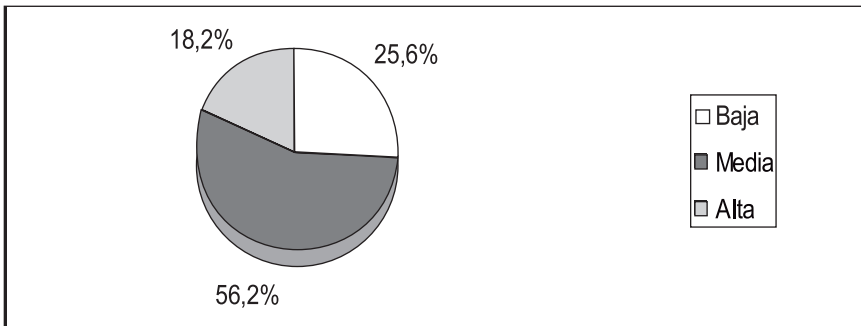


FIGURA 1

Ciclograma para la valoración general de los alumnos sobre las pruebas objetivas

TABLA 2
 MEDIAS Y DESVIACIONES TÍPICAS PARA LOS ITEMS
 DE LA ESCALA DE VALORACIÓN

	n	Media	Desv. típ.
1. Los exámenes tipo test son preferibles a otros procedimientos cuando se evalúa a grupos muy numerosos.	758	4,34	1,822
2. Al realizar un examen tipo test la calificación es conocida por el alumno rápidamente.	760	4,56	1,905
3. La calificación en un examen tipo test depende de la subjetividad del profesor.	759	2,41	1,843
4. Los exámenes tipo test cubren ampliamente todos los temas objeto de evaluación.	760	4,77	1,740
5. Los exámenes tipo test miden sobre todo la capacidad para aprender de memoria datos, hechos, definiciones, etc.	762	4,33	1,826
6. Los resultados de un examen tipo test reflejan el nivel de aprendizaje que ha logrado el alumno.	759	2,91	1,583
7. El modo en que deben indicarse las respuestas a un examen tipo test queda claramente explicado en las instrucciones.	761	5,55	1,537
8. Las preguntas en exámenes tipo test resultan ambiguas.	761	5,30	1,505
9. Los exámenes tipo test se responden con menos esfuerzo que otro tipo de exámenes.	754	4,38	1,913
10. Los exámenes tipo test dan la posibilidad de aprender a partir de los propios errores, tomando conciencia de por qué una respuesta dada resulta incorrecta.	758	3,71	1,769
11. Los exámenes tipo test favorecen a un determinado tipo de alumnos.	759	4,79	1,738
12. En un examen tipo test es posible adivinar las respuestas correctas cuando no se ha estudiado la materia.	755	3,42	1,735
13. El éxito de un examen tipo test depende del azar.	759	3,13	1,681
14. En un examen tipo test es fácil copiar.	750	4,09	1,849

sobre los exámenes tipo test es...), podemos considerar que los estudiantes universitarios no muestran un especial entusiasmo por este tipo de instrumentos. De acuerdo con los resultados del análisis descriptivo, un 25,2% de los alumnos expresaron una valoración baja sobre los exámenes tipo test, el 56,2% hicieron una valoración media, y emitieron una valoración alta sólo el 18,2%. En el ciclograma de la figura 1 queda reflejada esta distribución.

El examen del comportamiento de cada uno de los restantes ítems nos permite concretar con mayor detalle el sentido de esta valoración general. En la tabla 2 recogemos las medias y desviaciones típicas obtenidas respectivamente para los ítems de la Escala.

Atendiendo a los resultados mostrados en la tabla 2, podemos afirmar que entre los enunciados que cuentan con un mayor respaldo por parte de los estudiantes aparecen tanto valoraciones positivas como negativas. Si atendemos a las cinco puntuaciones más altas, encontramos que para los alumnos las pruebas objetivas se caracterizan por la claridad de las instrucciones ofrecidas acerca de cómo responder (media de 5,55 en ítem 7), la amplia cobertura que consiguen sobre la materia objeto de examen (4,77, en ítem 4) o la rapidez con que son conocidas las calificaciones obtenidas en la prueba (4,56, en ítem 2). Por el contrario, las críticas más respaldadas apuntan hacia la ambigüedad de las preguntas formuladas (media de 5,30 en ítem 8) y el hecho de que las pruebas objetivas favorecen a un determinado tipo de alumnos (4,79 en ítem 11).

De forma complementaria, el examen de las puntuaciones más bajas refleja qué aspectos cuentan con poco apoyo en el alumnado a la hora de ser atribuidos a las pruebas objetivas. Según este análisis, han sido escasamente refrendadas posibles críticas como considerar que el resultado en este tipo de pruebas esté influido por la subjetividad del profesor (media de 2,41 en ítem 3) o dependa del azar (3,13 en ítem 13). No obstante, tampoco ha sido respaldada una cualidad deseable en cualquier instrumento de evaluación, como es la validez, que supondría una correspondencia entre los resultados obtenidos en una prueba objetiva y el nivel de aprendizaje logrado por el alumno (2,91 en ítem 6).

Valoración según áreas de enseñanza

La diferenciación en las valoraciones emitidas por el alumnado de las cuatro áreas de enseñanza representadas en la muestra afectan a una buena parte de los elementos considerados. Así lo indican los resultados del análisis de la varianza que hemos llevado a cabo tomando como variables dependientes cada una de los ítems de la Escala y como factor generador de los grupos la variable área de enseñanza. En la tabla 3 se incluyen las sumas de cuadrados, grados de libertad, medias cuadráticas, estadísticos F y probabilidades asociadas a éstos para los ítems de la Escala en los que hemos hallado diferencias significativas entre grupos, tomando un nivel de significación $\alpha=0,05$.

Para explorar el sentido de las diferencias, hemos recogido en la tabla 4 los valores que alcanza la puntuación media de los ítems en cada una de las áreas de enseñanza. De nuevo incluimos en la tabla únicamente los ítems que merecieron valoraciones significativamente diferentes por parte del alumnado de cada área.

TABLA 3
ANÁLISIS DE LA VARIANZA PARA LOS ITEMS DE LA ESCALA
TOMANDO COMO FACTOR EL ÁREA DE ENSEÑANZA

	Fuente de variación	Suma de cuadrados	g.l.	Media cuadrática	F	Sig.
1. Los exámenes tipo test son preferibles a otros procedimientos cuando se evalúa a grupos muy numerosos.	Inter-grupos	159,940	3	53,313	17,079	,000
	Intra-grupos	2353,601	754	3,121		
2. Al realizar un examen tipo test la calificación es conocida por el alumno rápidamente.	Inter-grupos	136,802	3	45,601	13,175	,000
	Intra-grupos	2616,651	756	3,461		
3. La calificación en un examen tipo test depende de la subjetividad del profesor.	Inter-grupos	29,636	3	9,879	2,930	,033
	Intra-grupos	2545,379	755	3,371		
4. Los exámenes tipo test cubren ampliamente todos los temas objeto de evaluación.	Inter-grupos	92,690	3	30,897	10,595	,000
	Intra-grupos	2204,552	756	2,916		
5. Los exámenes tipo test miden sobre todo la capacidad para aprender de memoria datos, hechos, definiciones, etc.	Inter-grupos	54,539	3	18,180	5,552	,001
	Intra-grupos	2482,122	758	3,275		
7. El modo en que deben indicarse las respuestas a un examen tipo test queda claramente explicado en las instrucciones.	Inter-grupos	58,923	3	19,641	8,558	,000
	Intra-grupos	1737,276	757	2,295		
12. En un examen tipo test es posible adivinar las respuestas correctas cuando no se ha estudiado la materia.	Inter-grupos	25,568	3	8,523	2,852	,036
	Intra-grupos	2243,841	751	2,988		
14. En un examen tipo test es fácil copiar.	Inter-grupos	173,935	3	57,978	18,123	,000
	Intra-grupos	2386,604	746	3,199		

Como puede observarse a partir de los resultados mostrados en la tabla 4, las puntuaciones medias más altas se han registrado entre el alumnado del área de Ciencias Exactas y Naturales, que alcanza los máximos para cinco de los ocho items: los números 1, 3, 4, 5 y 7. De acuerdo con el contenido de tales items, este colectivo es el que

asume en mayor grado las críticas de que los exámenes tipo test miden sobre todo un aprendizaje memorístico (media de 4,82) y las calificaciones obtenidas dependen de la subjetividad del profesor (2,68). No obstante, es en este grupo donde se expresan con mayor fuerza las preferencias por los exámenes tipo test frente a otros procedimientos cuando se evalúa a grupos numerosos (5,07), se valora la cobertura lograda sobre la materia objeto de evaluación (5,15), y se consideran claramente explicadas las instrucciones que se dan para responder (6,00). En estos tres aspectos, las puntuaciones medias más bajas correspondieron al alumnado del área de Ingeniería y Tecnología, con valores de 3,59; 4,23 y 5,12 respectivamente. El alumnado de Ingeniería y Tecnología registra además las puntuaciones más bajas en los ítems 12 y 14 (medias respectivas de 3,15 y 3,39), que suponen el menor respaldo encontrado para críticas como las posibilidades de adivinación de respuestas en este tipo de exámenes y la facilidad para la copia fraudulenta.

TABLA 4
 MEDIAS ALCANZADAS EN CADA ÁREA DE ENSEÑANZA PARA
 LOS ÍTEMS DE LA ESCALA QUE PRESENTAN DIFERENCIAS SIGNIFICATIVAS

	Exactas y Naturales	Sociales y Jurídicas	Ciencias de la Salud	Ingeniería y Tecnología
1. Los exámenes tipo test son preferibles a otros procedimientos cuando se evalúa a grupos muy numerosos.	5.07	4.53	4.29	3.59
2. Al realizar un examen tipo test la calificación es conocida por el alumno rápidamente.	4.75	5.08	4.14	4.19
3. La calificación en un examen tipo test depende de la subjetividad del profesor.	2.68	2.39	2.15	2.60
4. Los exámenes tipo test cubren ampliamente todos los temas objeto de evaluación.	5.15	4.65	5.10	4.23
5. Los exámenes tipo test miden sobre todo la capacidad para aprender de memoria datos, hechos, definiciones, etc.	4.82	4.02	4.40	4.40
7. El modo en que deben indicarse las respuestas a un examen tipo test queda claramente explicado en las instrucciones.	6.00	5.69	5.47	5.12
12. En un examen tipo test es posible adivinar las respuestas correctas cuando no se ha estudiado la materia.	3.21	3.52	3.59	3.15
14. En un examen tipo test es fácil copiar.	3.54	4.54	4.33	3.39

Los ítems 2 y 14 obtienen su mayor puntuación entre los alumnos del área de Ciencias Sociales y Jurídicas, por lo que es este colectivo el que en mayor medida considera que las calificaciones de pruebas objetivas se conocen con rapidez (5,08) y que resulta fácil copiar durante la realización de este tipo de exámenes (4,54). Resulta característico de este grupo el mínimo respaldo a la crítica de que las pruebas objetivas miden un aprendizaje memorístico. Y entre el alumnado de Ciencias de la Salud destaca la baja adhesión a la crítica de subjetividad del profesor al calificar este tipo de pruebas, ítem para el que se ha registrado la menor puntuación media en todos los grupos (2,15).

Incidencia de la variable sexo en la valoración

Al aplicar la prueba t de contraste de medias para comparar las valoraciones que realizan estudiantes de uno y otro sexo, se han encontrado diferencias significativas ($\alpha=0.05$) en un total de seis ítems. En la tabla 5 se recogen los resultados de la aplicación, incluyendo únicamente los seis ítems para los que se puede rechazar la hipótesis nula de igualdad de medias entre valoraciones de alumnos y alumnas.

TABLA 5
PRUEBA T PARA LA COMPARACIÓN DE LAS VALORACIONES
EN FUNCIÓN DEL SEXO

	t	g.l.	Sig.
6. Los resultados de un examen tipo test reflejan el nivel de aprendizaje que ha logrado el alumno.	2,182	743	,029
7. El modo en que deben indicarse las respuestas a un examen tipo test queda claramente explicado en las instrucciones.	-2,698	745	,007
8. Las preguntas en exámenes tipo test resultan ambiguas.	-2,256	745	,024
9. Los exámenes tipo test se responden con menos esfuerzo que otro tipo de exámenes.	2,570	739	,010
13. El éxito de un examen tipo test depende del azar.	-2,836	743	,005
14. En un examen tipo test es fácil copiar.	-2,521	735	,012

Las puntuaciones medias alcanzadas por los y las estudiantes al valorar las pruebas objetivas han sido recogidas en la tabla 6, donde se incluyen sólo los ítems en los que se detectaron diferencias significativas para un nivel de significación $\alpha=0,05$.

TABLA 6
 MEDIAS ALCANZADAS POR ALUMNOS Y ALUMNAS
 EN LOS ITEMS DE LA ESCALA QUE PRESENTAN DIFERENCIAS SIGNIFICATIVAS

	Hombre	Mujer
6. Los resultados de un examen tipo test reflejan el nivel de aprendizaje que ha logrado el alumno.	3.06	2.80
7. El modo en que deben indicarse las respuestas a un examen tipo test queda claramente explicado en las instrucciones.	5.38	5.69
8. Las preguntas en exámenes tipo test resultan ambiguas.	5.16	5.41
9. Los exámenes tipo test se responden con menos esfuerzo que otro tipo de exámenes.	4.60	4.23
13. El éxito de un examen tipo test depende del azar.	2.93	3.28
14. En un examen tipo test es fácil copiar.	3.89	4.23

De acuerdo con los valores recogidos en la tabla 6, las alumnas valoran de manera menos favorable que los alumnos las pruebas objetivas, con la única excepción de la claridad que perciben en las instrucciones de respuesta (media de 5,38 en hombres frente a 5,69 en mujeres). Las alumnas han respaldado en mayor medida que los alumnos críticas como la ambigüedad en la formulación de las preguntas (media de 5,16 en hombres frente a 5,41 en mujeres), la influencia del azar en el éxito logrado en este tipo de pruebas (2,93 frente a 3,28), o la facilidad para copiar (3,89 frente a 4,23). Y ante enunciados que encierran una valoración positiva de los exámenes tipo test, las alumnas han expresado un menor respaldo. Esto ocurre al valorar el grado en que los resultados de estos exámenes reflejan el nivel de aprendizaje del alumno (media de 3,06 en hombres frente a 2,80 en mujeres). Las mujeres apoyan en menor medida que los exámenes tipo test se respondan con menor esfuerzo que otro tipo de exámenes (media de 4,60 en hombres frente a 4,23 en mujeres).

DISCUSIÓN Y CONCLUSIONES

Una primera conclusión a la que podemos llegar, a la vista de los resultados obtenidos, es que las valoraciones realizadas por el alumnado universitario participante en este estudio no revelan precisamente entusiasmo por las pruebas objetivas. Las valoraciones más positivas se han centrado en aspectos relacionados con la calificación de las pruebas (no dependientes de la subjetividad del profesor ni del azar, y conocidas en breve plazo de tiempo), con las instrucciones ofrecidas al alumno (resultan claras) y con el contenido de las pruebas (cobertura sobre la materia objeto de evaluación), mientras que las más negativas se dirigieron a la formulación de las cuestiones (resultan ambiguas), su exactitud al evaluar (no reflejan el nivel de aprendizaje logrado por el alumno) y el hecho de que requieren ciertas habilidades para responder (favorecen

a determinado tipo de alumnos). En los párrafos que siguen trataremos de valorar y contextualizar estos resultados.

Las consecuencias de la evaluación para el alumnado podrían ser una de las causas por las que este aspecto de la vida académica no goza de excesiva popularidad. De hecho, someterse a evaluación provoca incomodidad en el alumnado, encontrándose la ansiedad ante los exámenes entre los problemas más habituales en la población universitaria (Valero, 1999). Como consecuencia de ello, cabía esperar que las valoraciones del alumnado sobre los procedimientos de evaluación tendieran a ser poco generosas. Asumiendo este efecto, pensamos sin embargo que a las pruebas objetivas no corresponde la peor parte; en un reciente trabajo en el que se han estudiado las preferencias de los estudiantes sobre diferentes tipos de exámenes, las pruebas objetivas se han situado por encima de las pruebas de ensayo (Nield y Wintre, 2002), y se ha comprobado que éstas última generan más ansiedad en el alumno que las pruebas objetivas (Choi, 1998).

Si hay algún rasgo que define a las pruebas objetivas es precisamente la objetividad en la asignación de una calificación. Desde la perspectiva de los alumnos, esta característica fundamental de las pruebas objetivas es una de las que se han reconocido y valorado positivamente. Para el alumnado, los resultados obtenidos en una prueba objetiva están libres de la subjetividad del profesor. Con ello, las pruebas objetivas estarían garantizando el derecho de los alumnos a una evaluación justa, que mida a todos los alumnos con el mismo rasero. En palabras de Contreras (1995:35), los alumnos suelen mostrarse muy críticos ante situaciones de evaluación manifiestamente injustas, ocasionadas por decisiones arbitrarias, discrecionales de los profesores, bien porque no responden a un criterio fijado o simplemente porque no existe equidad. La asignación objetiva de puntuaciones que se realiza al calificar una prueba objetiva permite salir al paso de estos problemas.

La posibilidad de una respuesta correcta por azar está presente para formatos de preguntas que suelen emplearse en las pruebas objetivas, con especial incidencia en los de elección de respuesta. En contra de lo que cabría esperar, las críticas en este sentido no han sido respaldadas por el alumnado. Si bien es cierto que el azar puede influir en los resultados, también lo es que la práctica totalidad del profesorado recurre a procedimientos que tratan de amortiguar el efecto del azar, generalmente mediante la introducción de una corrección en la puntuación final por los errores cometidos. Este tema ha centrado un debate interesante y ha dado lugar a alternativas diversas que hemos comentado en este trabajo (De Pablos y otros, 1983; Charter, 2000; Dodd y Leal, 2002; Nield y Wintre, 2002). No obstante, la preocupación sobre el tema permanece en el plano técnico más que en el de los usuarios, quienes no han destacado especialmente tal problema.

Los resultados de una prueba objetiva cuentan con una ventaja sobre las pruebas de ensayo en la inmediatez con que pueden ser conocidos por los alumnos, los cuales pueden incluso comprobar las plantillas de corrección en el momento de finalizar la prueba. Este valor no ha pasado desapercibido para los alumnos, que lo han señalado como una de sus fortalezas. Sin embargo, los alumnos no han prestado tanta atención al hecho de que el resultado que se conoce es únicamente la calificación o el número de aciertos y errores, sin posibilidad de que estos valores numéricos puedan retroalimentar

el aprendizaje de los estudiantes propiciando una función formativa de la evaluación. Para que el alumno pueda aprender a partir de los errores cometidos en una prueba, resulta imprescindible algún tipo de explicación o justificación sobre las razones por las que una respuesta concreta es considerada correcta o incorrecta.

De cara a propiciar una mejora en el uso de las pruebas objetivas, convendría tener en cuenta las valoraciones negativas que han apoyado los encuestados. Una primera cuestión tiene que ver con la redacción de la prueba. A la claridad en las instrucciones, motivada posiblemente por la experiencia continuada en este tipo de pruebas, se contraponen la ambigüedad en la formulación de los ítems. La redacción de buenos ítems para una prueba objetiva es tarea que entraña dificultad, de tal manera que no siempre las cuestiones que aparecen formando parte de este tipo de pruebas resultan correctas desde un punto de vista técnico. Una recomendación obligada sería, por tanto, la de prestar mayor atención por parte de los docentes a la construcción de ítems, respetando los criterios que se han ofrecido en la literatura al respecto. En el caso concreto de los ítems de opciones múltiples, es interesante por su actualidad, la recopilación hecha por Haladyna y otros (2002) a partir de la revisión de los manuales y trabajos dedicados al tema durante las últimas décadas.

Una segunda debilidad se centra en la validez de las pruebas objetivas. Si bien se reconoce que las pruebas objetivas permiten cubrir todos los temas objeto de evaluación, aspecto propiciado por la posibilidad de incluir un elevado número de cuestiones dentro de una prueba, también se critica la falta de correspondencia entre los resultados que se obtienen y el nivel de aprendizaje de los alumnos. Tal objeción se ha encontrado en estudios anteriores, y es común a la generalidad de los procedimientos de evaluación. Así, entre las conclusiones a las que se llegaba al analizar las valoraciones que los estudiantes hacen del rendimiento docente en la Universidad de Sevilla (García y otros, 1998), se incluía la falta de concordancia entre las calificaciones obtenidas y la opinión que el propio alumno posee sobre su aprendizaje, dado que el ítem donde se recogía esta idea era uno de los peor valorados en las encuestas para la evaluación del profesorado universitario.

Relacionado con lo anterior estaría la crítica de que las pruebas objetivas favorecen a determinado tipo de alumnos. En un estudio sobre el modo en que los estudiantes ven los exámenes que realizan, Van Etten y otros (1997) encontraron que los alumnos reconocen estudiar los exámenes siguiendo estrategias diferenciadas según el tipo de prueba que se les va a plantear. Asumiendo que distintas formas de evaluación requieren diferentes cualidades o habilidades, podría aceptarse también que determinados alumnos se adaptan mejor que otros a responder a las preguntas de una prueba objetiva. Para afrontar esta crítica suscrita por los encuestados en nuestro estudio, sería recomendable el diseño de sistemas de evaluación multimétodo en los que las pruebas objetivas no sean el único procedimiento usado para evaluar el aprendizaje.

Concluyendo sobre el segundo de los objetivos planteados en este estudio, podemos afirmar que las valoraciones acerca de las pruebas objetivas varían según las áreas de enseñanza, de tal manera que para determinados aspectos llegan a identificarse posiciones del alumnado claramente diferenciadas. Podemos ver este resultado como una manifestación más de la existencia de diferentes culturas universitarias, entendidas como el conjunto de patrones de comportamiento, reglas establecidas, formas de pensa-

miento, formas de relacionarse y organizarse, valores compartidos por los miembros de un colectivo, que configuran contextos singulares en los que se desarrollan los procesos de enseñanza-aprendizaje universitarios. En definitiva, la vida universitaria difiere de unos ámbitos académicos a otros, de tal forma que las posibilidades de uso y el valor asignado a las pruebas objetivas no necesariamente coinciden.

Finalmente, la diferente valoración de las pruebas objetivas inducida por el género ha venido a señalar una peor opinión de las alumnas respecto a este tipo de instrumentos de evaluación. No obstante, este resultado debe ser acogido con cautela, dada la inestabilidad de las diferencias entre géneros que suelen encontrarse en la literatura cuando se revisan diferentes estudios. Ya apuntábamos en la introducción de este trabajo las evidencias inconsistentes encontradas respecto a la obtención de mejores resultados entre alumnos de uno u otro sexo. A éstas podríamos añadir la falta de correspondencia entre los resultados del estudio de Ben-Shakhar y Sinai (1991), los cuales señalan en los hombres un mayor uso de estrategias de adivinación cuando se desconoce la respuesta correcta, y algunos hallazgos del presente trabajo, en el que hombres y mujeres no se han diferenciado por su forma de valorar las posibilidades de adivinación.

En el trasfondo del trabajo que hemos presentado se encuentra la idea de que los alumnos, como partícipes de los procesos de enseñanza-aprendizaje que se desarrollan en la universidad, poseen una percepción sobre los mismos derivada de su propia experiencia. Sus valoraciones acerca de las pruebas objetivas, recogidas aquí, nos indican cuáles de los inconvenientes atribuidos a este instrumento de evaluación son confirmados desde el punto de vista de los sujetos cuyo aprendizaje evaluamos. En consecuencia, en la medida en que se actúe para contrarrestar estos inconvenientes, siguiendo líneas como las apuntadas en párrafos anteriores, estaremos contribuyendo a que el procedimiento de evaluación sea mejor valorado y aceptado por los estudiantes. No obstante, de los resultados de este estudio se deduce también que la experiencia de evaluación mediante pruebas objetivas puede variar de unos contextos académicos a otros, de tal manera que las expectativas del alumnado sobre este tipo de instrumentos no son únicas.

REFERENCIAS BIBLIOGRÁFICAS

- Anderson, J. (1989). Sex-related differences on objective tests among undergraduates, en *Educational Studies in Mathematics*, 20 (2), 165-177.
- Beller, M. y Gafni, N. (2000). Can item format / multiple choice vs. open-ended) account for gender differences in mathematics achievement, en *Sex roles*, 42 (1-2), 1-21.
- Belleza, F.S. y Belleza, S.F. (2002). Detection of cheating on múltiple-choide tests by using error-similarity analysis, en Griggs, R.A. (Ed.) *Handbook for teaching introductory psychology, Vol.3*. Mahwah: Lawrence Erlbaum Associates, pp. 160-164.
- Ben-Shakhar, G. y Sinai, Y. (1991). Gender differences in multiple-choice tests: the role of differential guessing tendencies, en *Journal of Educational Measurement*, 28 (1), 23-35.
- Bennet, R.E.; Rock, D.A. y Wang, M. (1991). Equivalence of free-response and multiple-choice items, en *Journal of Educational Measurement*, 28 (1), 77-92.

- Berezina y Berman (2000). Proof reading and multiple choice tests, en *International Journal of Mathematical Education in Science & Technology*, 31 (4), 613-619.
- Bielinski, J. y Davison, M.L. (2001). A sex difference by item difficulty interaction in multiple choice mathematics items administered to national probability samples, en *Journal of Educational Measurement*, 38 (1), 51-77.
- Biggs, J. (2002). *Teaching for quality learning at university*. Londres: The Society for Research into Higher Education & Open University.
- D.J. Burton, R.F. y Miller, D.J. (1999). Statistical modelling of multiple-choice and true/false tests: ways of considering, and of reducing, the uncertainties attributable to guessing, en *Assessment and Evaluation in Higher Education*, 24 (4), 399-411.
- Bush, M. (2001). A multiple choice test that rewards partial knowledge, en *Journal of Further and Higher Education*, 25 (2), 157-163.
- Charter, R.A. (2000). Determining random responding to objective tests, en *Journal of Psychoeducational Assessment*, 18 (4), 308-315.
- Choi, N. (1998). The effects of test format and locus of control on test anxiety, en *Journal of College Student Development*, 39 (6), 616-620.
- Choppin, B. (1985). The correction of guessing for objective tests, en *Evaluation in Education*, 9 (1), 71-80.
- Contreras, E. (1995). Derechos de los alumnos universitarios ante la evaluación de sus aprendizajes, en *Bordón*, 47 (1), 31-40.
- De Pablos, J. y otros (1983). La evaluación del alumno en la universidad: el proyecto CERT, en *Revista de Enseñanza Universitaria*, 6, 49-74.
- Denier, G. y Hancock, D. (2002). Graded multiple choice questions: rewarding understanding and preventing plagiarism, en *Journal of Chemical Education*, 79 (8), 961-964.
- Dood, D.K. y Leal, L. (2002). Answer justification: removing the «trick» from multiple choice questions, en Griggs, R.A. (Ed). *Handbook for teaching introductory psychology*, Vol. 3. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 99-100.
- Epstein, M.L. y otros (2001). Immediate feedback during academic testing, en *Psychological Reports*, 88 (3), 889-894.
- Epstein, M.L. y Brosvic, G.M. (2002). Immediate feedback assessment technique: multiple choice test that «behaves» like an essay examination, en *Psychological Reports*, 90 (1), 226.
- Friedman, S.J. y Cook, G.L. (1995). Is an examinee's cognitive style related to the impact of answer changing on multiple-choice tests?, en *Journal of Experimental Education*, 63 (3), 199-213.
- García, E. y otros (1998). *Estudio sobre la valoración del rendimiento docente en la Universidad de Sevilla*. Informe inédito encargado por la Comisión de Docencia de la Universidad de Sevilla.
- Garner, M. y Engelhard, G. (1999). Gender differences in performance on multiple-choice and constructed response, en *Applied Measurement in Education*, 12 (1), 29-51.
- Geiger, M.A. (1991). Changing multiple-choice answers: a validation and extension, en *College Student Journal*, 25 (2), 181-186.
- Gibbs, G. (1999). Using assessment strategically to change the way students learn, en Brown, S. y Glasner, A. (Eds.) *Assessment matters in Higher Education*. Londres: The Society for Research into Higher Education & Open University Press, 41-53.

- Gibbs, G.; Habeshaw, S. y Habeshaw, T. (1986). *53 interesting ways to asses your students*. Bristol: Technical and Educational Services.
- Haladyna, T.M.(1994). *Developing and validating multiple-choice test items*. Hillsdale: Lawrence Erlbaum Associates.
- Haladyna, T.M.; Downing, S.M. Y Rodríguez, M.C. (2002). A review of múltiple-choice item-writing guidelines for classroom assessment, en *Applied Measurement in Education*, 15 (3), 309-334.
- Heywood, J. (1989). *Assessment in Higher Education*. Chicester: John Wiley & Sons.
- Linn, R.L. y Gronlund, N. (2000). *Measurement and assessment in teaching*. Columbus: Merrill.
- LoSchiavo, F.M. y Shatz, M.A. (2002). Students' reasons for writing on multiple-choice examinations, en *Teaching of Psychology*, 29 (2), 138-140.
- Kottke, J.L. (2001). Students' reactions to written test item rebuttals, en *Journal of Instructional Psychology*, 28 (4), 256-258.
- Martínez, M. (1999). Cognition and the question of test item format, en *Educational Psychologist*, 34, 207-218.
- Murphy, R.J. (1982). Sex differences in objective test performance, en *British Journal of Educational Psychology*, 52 (2), 213-219.
- Nield, A.F. y Wintre, M.G. (2002). Multiple choice questions with an option to comment: student attitudes and use, en Griggs, R.A. (Ed.) *Handbook for teaching introductory psychology, Vol.3*. Mahwah: Lawrence Erlbaum Associates, pp. 95-99.
- Ory, J.C. y Ryan, K.E. (1993). *Tips for improving testing and grading*. Newbury Park: Sage Publications.
- Pamphlett, R. y Farnill, D. (1995). Effect of anxiety on performance in multiple choice examination, en *Medical Education*, 29 (4), 297-302.
- Schwarz, S.P. y otros (1991). Reasons for changing answers: an evaluation using personal interviews, en *Journal of Educational Measurement*, 28 (2), 163-171.
- Scouller, K. (1997). Students' perceptions of three assessment methods: asignment essay, multiple choice question examination, short answer examination. Comunicación presentada a *Higher Education Research and Development Society of Australasia*, Adelaide, 9-12 julio.
- Shaha, S.H. (1984). Matching tests: reduced anxiety and increased test-effectiveness, en *Educational and Psychological Measurement*, 44 (4), 468-481.
- Shatz, M.A. y Best, J.B. (1987). Students' reasons for changing answers on objective tests, en *Teaching of Psychology*, 14 (4), 241-242.
- Smith, L. (2002). The effects of confidence and perception of test-taking skills on performance, en *North American Journal of Psychology*, 4 (1), 37-50.
- Smith, B.; Brown, S. y Race, Ph. (1996). *500 tips on assessment*. Londres: Kogan Page Ltd.
- Stape, Ch.J. (1995). Techniques for developing higher-level objective test questions, en *Performance and Instruction*, 34 (3), 31-34.
- Talmir, P. (1991). Múltiple choice intes: how to gain the most out of them, en *Biochemical Education*, 19 (4), 188-191.
- Toppino, T.C. y Luipersbeck, S.M. (1993). Generality of the negative sugestion effect in objective tests, en *Journal of Educational Research*, 86 (6), 357-362.

- Valero, L. (1999). Evaluación de ansiedad ante exámenes, en *Anales de Psicología*, 15 (2), 223-231.
- Van Etten, S. y otros (1997). College students' beliefs about exam preparation, en *Contemporary Educational Psychology*, 22 (2), 192-212.

Fecha de recepción: 17 de octubre de 2004.

Fecha de aceptación: 25 de mayo de 2005.