

Andlisis de la calidad en las pruebas objetivas en la enseanza universitaria de Fisioterapia

S. Hernandez Sanchez, J.V. Toledo Marhuenda, M.C. Lillo Navarro, S. Ramirez Gomez, EJ. Poveda Pagan, M.L. Bataller Richart

Area de Fisioterapia. Dpto. Patologia y Cirugia. Universidad Miguel Hernandez de Elche.

RESUMEN

Desde hace ya varios afios, la evaluation de la calidad en los diferentes campos de la enseanza universitaria constituye un signo de autocritica y progreso. La preocupacion por el desarrollo de un maximo nivel en la calidad docente ha adquirido una relevancia maxima a causa de la inminente llegada del proceso de convergencia europea.

El andlisis que efectuamos a continuation, sobre la calidad en las pruebas objetivas de Fisioterapia, responde a una serie de reflexiones entorno a las siguientes hipdtesis: ¿Se utilizan correctamente las pruebas objetivas en la evaluacdn del aprendizaje por parte de los alumnos de Fisioterapia? ¿Somos conscientes del sesgo que supone, para la evaluation del aprendizaje, el uso de pruebas objetivas como medio unico de evaluation?

El calculo de parametros tales como el fndice de dificultad de un item o su indice de discrimination nos permite obtener una informacdn más profunda acerca de la validez de los items en las pruebas objetivas.

Una vez analizada la prueba, es necesario un ejercicio de reflexion critica que posibilite la ejecucion de actuaciones posteriores de mejora en el proceso evaluativo.

Palabras clave: evaluation; calidad; pruebas objetivas; Fisioterapia.

ABSTRACT

For several years now, the evaluation of the quality in the different university learning fields is considered a sign of self-criticism and progress. The concern about the development of the maximum level in teaching quality has acquired the highest relevance due to the imminent arrival of the process of European convergence.

The analysis carried out below on the quality of the objective tests used for the university degree of Physiotherapy is the reply to a series of thoughts derived from the following hypothesis: Are these objective tests properly used for the assessment of the learning obtained by Physiotherapy students? Are we aware that the assessment of the knowledge by using objective tests as the only and unique tool of assessment could mean a way of slant?

The calculation of parameters such as the level of difficulty of an item or its discrimination degree allows us to get more concise information about the validity of the items in the objective tests.

Once analyzed the test, an exercise of critical thinking which makes it possible the execution of later improvement actions in the assessment process is necessary.

Key words: assessment, quality, objective tests, Physiotherapy.

OBJETIVOS

— **Mostrar la experiencia del Equipo de Evaluación y Mejora de la Calidad Docente del área de Fisioterapia de la Universidad Miguel Hernández (UMH) en la revisión crítica de las pruebas objetivas, utilizadas en la evaluación de conocimientos teóricos de las asignaturas de Fisioterapia.**

— **Recordar algunas estrategias e instrumentos para la valoración de la calidad de las pruebas objetivas, utilizadas en las evaluaciones de la parte teórica de asignaturas universitarias.**

— **Generar una reflexión entre el colectivo de profesionales docentes, implicados en la evaluación del aprendizaje de los alumnos de Fisioterapia, que utilizan las pruebas objetivas.**

INTRODUCCIÓN

Si realizamos una revisión de los programas de las asignaturas del Área de Fisioterapia de las diferentes universidades españolas publicados en internet, se puede comprobar que el uso de pruebas objetivas es una práctica frecuente y extendida en la evaluación del aprendizaje teórico de los estudiantes de Fisioterapia.

Según el profesor Mateo [1] «*las pruebas objetivas son un conjunto vertebrado de preguntas claras y precisas que demandan del alumnado una respuesta igualmente breve, generalmente limitada a una elección entre una serie de alternativas que se le proporcionan*».

Este autor, añade que el calificativo de «objetivas», se refiere al tratamiento de los datos y a las condiciones de aplicación, pero no porque sean más objetivas desde el punto de vista de una mejor valoración del aprendizaje del alumno.

Algunas ventajas del uso de las pruebas de elección única [2, 3, 4, 5, 6, 7, 8]

— **Este tipo de prueba suele ser bastante utilizada ya que permite evaluar a un gran número de alumnos de forma simultánea.**

— **Tiene un carácter objetivo en su presentación ya que solo existe una respuesta correcta. Se supone que todos los alumnos deben interpretar las preguntas de la misma forma, siempre y cuando estén correctamente elaboradas.**

— **Se puede utilizar una gran variedad de ítems para evaluar diferentes objetivos y abarcar distintos contenidos.**

— **Poseen una elevada fiabilidad. Con cinco respuestas alternativas a la pregunta, se reduce a un 20 % la oportunidad de «adivinar» la respuesta correcta.**

— **Permite una corrección rápida y sencilla, incluso automática. Se reduce la influencia de factores extraños y externos en la puntuación, como por ejemplo la falta de objetividad del profesor.**

— **Facilita el análisis estadístico de los datos obtenidos en la prueba.**

Inconvenientes que pueden surgir respecto al uso de estas pruebas

[2, 3, 5, 7, 8, 17]

— **Utilizando solamente instrumentos de este tipo, la evaluación se centra sobre todo en el ámbito cognoscitivo. No obstante, es difícil evaluar otros componentes del aprendizaje como los actitudinales o los objetivos de procedimiento. Por ejemplo, el uso de estas pruebas no es adecuado si queremos evaluar la capacidad organizativa o la creatividad.**

— **La elaboración de las preguntas test, y de la prueba en general, requiere bastante tiempo y dedicación del profesor.**

— Una construcción correcta de la prueba debe partir de los objetivos programados, trabajados en clase durante un periodo de tiempo determinado, y debe establecerse en base al contenido correspondiente. No se debe caer en el error de confeccionar items para valorar objetivos que no estaban programados o para evaluar una materia que se ha considerado adicional o secundaria durante el desarrollo del curso.

— Puede ocurrir que este tipo de prueba favorezca la memorización en detrimento de otras capacidades como la comprensión, aunque existen diferentes opiniones al respecto.

— Algunos autores mantienen que este tipo de pruebas refuerza el pensamiento selectivo y parcial, en detrimento del pensamiento global, crítico y constructivo.

ANÁLISIS DE LA CALIDAD DE UN TEST

El profesor Mateo [1] afirma que, «En los procesos de evaluación pueden aparecer numerosas fuentes de error, y el propio instrumento de valoración puede ser una de ellas si no se elabora cuidadosamente». Mediante el análisis de la calidad de una prueba de evaluación se puede conocer el grado de validez y fiabilidad, y, por tanto, su pertinencia para valorar la consecución de objetivos de aprendizaje propuestos inicialmente.

Este proceso es el que se establece en el Equipo de Evaluación y Mejora de la Calidad Docente del Área de Fisioterapia de la UMH, tras la realización de las pruebas de evaluación teóricas. Varias asignaturas utilizan estas pruebas, junto a otras, para la evaluación de los alumnos. Estas, generalmente, consisten de sesenta preguntas tipo test (elección única) y cuatro cuestiones de respuesta corta abierta.

Nos ocuparemos, en este artículo, de la valoración crítica de las preguntas test que aparecen en dichas pruebas.

En el examen test, los alumnos reflejan sus respuestas en plantillas especiales para una posterior corrección automática en la Unidad de Asistencia y Asesoramiento Académico del centro docente. Una vez finalizado este proceso, y conocidos los resultados de la evaluación, se procede a valorar la calidad del test mediante un proceso similar al propuesto por Tenbrink [4]. Este autor, propone los siguientes apartados para la valoración de la calidad de una prueba de test:

1. Observación de indicadores
2. Análisis de los items de la prueba
3. Análisis de fiabilidad. Es un indicador de la consistencia interna. Su cálculo se efectúa con fórmulas matemáticas, como por ejemplo la KR-20 o KR-21 [6]
4. Estudio de la validez. Se emplea para conocer «El grado de precisión con el que la prueba utilizada mide verdaderamente aquello para lo que fue diseñada como instrumento de evaluación» [6],

A continuación, desarrollaremos los dos primeros puntos de la propuesta de Tenbrink, por considerar que son instrumentos de fácil manejo para el profesor y que le reportan información rápida y valiosa sobre la calidad de los items de la prueba.

Observación de indicadores

Muestra indicios de que la prueba podría ser errónea en alguno de sus componentes. Estos datos iniciales deben servir para centrar la atención del profesor y para profundizar, si fuera necesario, en un análisis más detallado. Se puede considerar un «primer vistazo»

TABLA 1.

<i>Errores comunes</i>	<i>Indicadores</i>
a. Entasis en un tema inadecuado b. ítems ambiguos c. Valoración de objetivos que no están en el programa	<ul style="list-style-type: none"> - índice de discriminación bajo (explicado más adelante) - Los alumnos se quejan: «ha preguntado cosas que no dimos en clase o no estudiamos» «a esto no le dio importancia» <ul style="list-style-type: none"> - "Si se interpreta de este modo, podría ser la opción A, en lugar de la que Ud. dice» - Desequilibrio de cantidad de ítems para un tema o concepto concreto
a. Preguntas difíciles de leer	<ul style="list-style-type: none"> - Enunciados muy extensos o confusos - Los alumnos hacen muchas preguntas sobre ese ítem durante el examen

para obtener las primeras impresiones sobre la calidad de la prueba (tabla 1).

Analizar los ítems de la prueba

Nos detendremos en este apartado, por su aplicación práctica y su utilidad en la mejora de la docencia.

El análisis individual de cada pregunta del examen test tiene una justificación básica y lógica. Si los componentes individuales de la prueba son de calidad, y a continuación se detallara este término, la prueba en general también se puede considerar buena.

La información necesaria para efectuar la valoración individual de los ítems se obtienen del análisis informático de los mismos mediante el programa EVALEC® que se lleva a cabo en la unidad de asesoramiento y apoyo a la docencia. Los datos que se obtienen con este análisis son:

- Índice de dificultad de un ítem
- Índice de discriminación
- Número respuestas correctas e incorrectas a una pregunta, así como las no contestadas

- Número de respuestas de los alumnos a la opción de cada pregunta
- Índices de fiabilidad KR-20 y KR-21.
- Media de la nota
- Media de respuestas correctas, incorrectas y no contestadas. Este último parámetro es el utilizado para calificar la dificultad del examen, en función de los siguientes intervalos:

- 0 a 49% de respuestas correctas (muy difícil)
- 50 a 69% de respuestas correctas (relativamente difícil)
- 70 a 89% de respuestas correctas (relativamente sencillo)
- 90 a 100% de respuestas correctas (muy fácil).

El análisis de un ítem en una prueba test nos permite conocer si el alumno ha adquirido los conocimientos planteados, y en qué medida, para conseguir un determinado objetivo de aprendizaje, establecido previamente en el programa docente y explicitado al inicio de la clase.

A continuación, se exponen diferentes parámetros utilizados para el estudio de la calidad de los ítems de un examen.

ÍNDICE DE DIFICULTAD de un ítem

[2, 3, 4, 5].

Indica la proporción de sujetos que respondieron correctamente un ítem. Un índice menor de dificultad en un ítem, se corresponde con un elevado número de alumnos que respondieron correctamente a la pregunta analizada. En el Área de Fisioterapia, la obtención de estos valores es automática. En cualquier caso, el procedimiento para su cálculo puede ser el siguiente:

1. Ordenar las pruebas corregidas en función de la puntuación obtenida. La ordenación será decreciente.

2. Seleccionar un porcentaje representativo de las pruebas mejor calificadas y uno similar de las de peor nota. Según los autores [10, 11, 12], oscila entre el 33% y el 25%. Kelly, [13] demostró que la utilización del 27% de las pruebas mejor y peor calificadas, se corresponde con el porcentaje más sensible y estable. De este modo, aplicaremos este porcentaje a la muestra, y se prescindirá del resto de pruebas para efectuar el estudio. Con estos dos grupos, superior e inferior, se trabajará para el resto de las operaciones. El uso de grupos extremos como el 10% o el 15% del total de las pruebas reducirá la fiabilidad de los resultados a causa del pequeño tamaño de la muestra (Anastas! 1980). Sin embargo, Guilbert [14] establece que se puede trabajar más cómodamente con un 25% o 33%, y evitar una cifra inabordable como el 27%.

1. La fórmula matemática que se aplica para obtener el valor numérico es la siguiente:

$Df = AS + AI/2n =$ Aciertos de un ítem/nº sujetos que contestan

Siendo:

AS: número de aciertos del grupo superior

AI: número de aciertos del grupo inferior

n: alumnos de cada subgrupo

Del resultado obtenido podemos interpretar que si:

$Df = 0$ Nadie responde correctamente el ítem

$Df = 1$ Todos responden correctamente

$Df =$ Entre 0 y 1. Su interpretación sería (tabla 2):

CONCLUSIONES A LAS QUE PODEMOS LLEGAR TRAS EL ESTUDIO DE LA DIFICULTAD DE UN ÍTEM

a. Cuando encontramos un ítem con un elevado número de aciertos, sobre el total de alumnos (interpretado como fácil o muy fácil), podemos pensar que:

— Los distractores planteados son inadecuados y hacen que, por eliminación, la pregunta sea realmente sencilla.

— Se están valorando objetivos esenciales o fundamentales, que se han trabajado mucho en el aula y con actividades diferentes. Esta es la razón por la que prácticamente to-

TABLA 2.

Valor Df	Interpretation
0.81 - 1:	Ítem muy fácil
0.61 - 0.80:	Ítem fácil
0.41 - 0.60:	Ítem de dificultad media
0.21 - 0.40:	Ítem difícil
0-0.21:	Ítem muy difícil

(Ebel y Frisbie 1986) [15]

dos los alumnos la contestan correctamente. Quizas, este hecho, que a priori prejuzga la bondad de un profesor o la sencillez de un item, sea su verdadero proposito: comprobar si los alumnos han adquirido ese objetivo esencial. Lo extrafio, seria obtener estos resultados sin que ese item correspondiera a un objetivo basico. De ser asi, esta situation deberia analizarse de forma minuciosa, con el proposito de observar las causas que han propiciado este resultado. Por ejemplo, errores de redaccion del item u otras causas que hayan influido en esos resultados.

b. Cuando observamos que un item es interpretado como dificil o muy dificil, se considera que, en esa cuestion, la mayoría de los alumnos no ha contestado correctamente y, por tanto, no ha demostrado haber alcanzado los objetivos propuestos. En este caso, deberiamos reflexionar sobre las causas de este resultado:

1. Puede ser debido a cuestiones propias del instrumento de evaluation:

— ¿Esta bien formulada la cuestion en su enunciado (doble negation, ambigüedades...)?

— Los distractores utilizados en la respuesta son adecuados? o, por el contrario, inducen al error o confusion.

— Ese item, ¿se corresponde con un objetivo importante del programa?

— ¿Se han fijado previamente los objetivos de aprendizaje cuyo logro se quiere comprobar?

— Ese item, ¿es representativo de los contenidos basicos propuestos? o, por el contrario, corresponde a una materia secundaria o adicional a la que no se le dedico demasiado tiempo o importancia durante el curso.

— ¿Esta claramente especificado que debe realizar el alumno ante la pregunta?

2. Si tras pensar en ello, la pregunta se confecciona correctamente, cabria preguntarse ¿por qué no aprendieron los alumnos la materia correspondiente a ese item? En definitiva: ¿por qué no alcanzaron los objetivos docentes programados? Las respuestas a estas preguntas, quizas, puedan encontrarse en la revision de algunos de los siguientes aspectos relacionados con la planificacion docente:

— Estrategias didacticas no adecuadas para transmitir la information.

— Experiencias de aprendizaje desarrolladas que no hayan sido eficaces.

Llegados a este punto, el profesor deberia plantear la modification de algun aspecto de la planificacion docente para mejorar los logros o resultados. Por ejemplo, trabajar sobre un tema de forma más exhaustiva, ampliando el tiempo dedicado y estimulando el trabajo autnomo del alumno en la biblioteca o Internet; fomentar el trabajo en grupo, etc.

Del desarrollo anteriormente expuesto en los puntos a) y b), podriamos pensar que lo más adecuado para un buen examen test seria utilizar todos los items con dificultad media. ¿Seria esto conveniente?

En cuanto a la distribution de las preguntas, respecto a su indice de dificultad, el profesor *Perez Juste* [16] recomienda colocar un pequeño porcentaje de items faciles al inicio del examen, para que no se reduzca la motivation o no se produzca una frustration inicial en los sujetos que menos conocimientos tienen de los temas. A continuation, se aumenta progresivamente la dificultad de las preguntas que debe aproximarse a un indice de 0.50. Para finalizar la prueba, se puede completar la distribution introduciendo al-

gunos ítems difíciles y muy difíciles, según se describe (tabla 3).

Con este tipo de reflexiones se puede llegar a conocer las posibles causas o problemas que existan en la evaluación y en todo el proceso enseñanza-aprendizaje. Obviar estas posibles fuentes de error produciría una alteración en la validez y fiabilidad del propio instrumento de evaluación y, con ello, de sus resultados.

ÍNDICE DE DISCRIMINACIÓN de un ítem

Diferencia los exámenes de los que responden correctamente a esa pregunta y los que no. En las pruebas objetivas, un ítem será discriminador si todos o casi todos los integrantes del grupo superior la contestan correctamente, mientras los alumnos que menor puntuación han obtenido, no lo contestan o lo contestan de forma incorrecta.

$$De = \frac{N^\circ}{(AS-AI)}$$

N° = Numero de alumnos de un subgrupo

AS = Aciertos del grupo superior

AI = Aciertos del grupo inferior

La interpretación de estos resultados es relativa y cada docente puede valorar los resultados en función de diferentes categorías, aunque por lo general se acepta [16] que (tabla 4:

TABLA 3.

Ítems	Distribución
Muy fáciles	10%
Fáciles	20%
Dificultad Media	40%
Difíciles	20%
Muy difíciles	10%

(Perez Juste, 1989)

En la interpretación de estos resultados, se puede hablar de:

a. *Discriminación positiva o alta discriminación*: cuando la mayoría de los alumnos del grupo superior responden correctamente a la pregunta y la mayoría, o todos los del grupo inferior, responden de forma incorrecta.

b. *Discriminación nula*: si a un ítem concreto responden correctamente el mismo número de alumnos del grupo superior y del inferior. En este caso este ítem, no discrimina entre los alumnos que han adquirido conocimientos de los que no lo han hecho.

c. *Discriminación negativa o absurda*: se da cuando los alumnos del grupo inferior responden correctamente a la cuestión y los del grupo superior no lo hacen o la contestan de forma errónea. Este hecho, en la interpretación que se está realizando, no tendría sentido.

Como queda reflejado, los valores intermedios obtenidos en el cálculo del índice de dificultad favorecen valores máximos del índice de discriminación.

Con estos dos índices, dificultad y discriminación, se puede efectuar un análisis del banco de preguntas y, con ello, saber si es conveniente modificar ítems, suprimirlos o mantenerlos en las pruebas. Si una pregunta concreta diferencia a los alumnos que tienen

TABLA 4. Índice de discriminación de un ítem

Valor De	Interpretación
0.40- 1:	ítem excelente o muy bueno
0.30 -0.39:	Ítem bueno aunque mejorable
0.20-0.29:	Ítem regular. Debe mejorarse
0 - 0.19:	ítem malo. Rectificarlo o anularlo
-1 -0:	ítem absurdo. Anularlo

(Ebel y Frisbie 1986)

un dominio sobre esos conocimientos y a los que no lo poseen, posiblemente tengamos un ítem adecuado para comprobar el grado de logro de un determinado objetivo propuesto.

También se puede confeccionar una ficha específica para cada objetivo de aprendizaje, con sus ítems para valorarlo y con los datos obtenidos tras el examen. Así, se puede observar qué ítem es el más adecuado para valorar la consecución de ese objetivo.

Sampascual, [17] establece que indirectamente los resultados de este análisis sirven también para:

- a. Conocer cómo se desarrolla el proceso enseñanza-aprendizaje.
- b. Realizar una valoración del programa y de su metodología didáctica.
- c. Adquirir experiencia para no cometer los mismos errores.

CONCLUSIONES:

— En el proceso de recogida de información sobre los aprendizajes de los estudiantes siempre existe algún porcentaje de error. El profesor debe tratar de identificar las posibles causas e intentar corregirlas.

— Cuando en un examen test, se utilice el sistema de elección de una única respuesta para valorar el grado de adquisición de objetivos del ámbito cognoscitivo, debemos procurar que la prueba tenga la mayor calidad posible. Es decir, que mida realmente lo que pretende medir, validez, que no incida en contenidos u objetivos banales que ni siquiera hemos planteado, y que la calificación obtenida por el alumno sea proporcionada con el grado de estudio y la dificultad de la prueba, ausente de errores (fiabilidad).

— Existe un alto grado de aceptación de este tipo de pruebas por parte de profesores

y alumnos [18]. Entre otras causas, porque cuando las preguntas de una prueba están bien confeccionadas, ésta posee un elevado nivel de fiabilidad y validez, demostrada por numerosos autores (*Guilbert* 1994, *Thyne* 1978, *Hubbard* 1978, *Sachs* 1975)

— La validez y fiabilidad de una prueba objetiva, y por tanto gran parte de su calidad, pasa por tener bien estructurados los objetivos y los contenidos del programa docente.

— Los objetivos del programa sirven de criterio para evaluar, posteriormente, las experiencias de aprendizaje y el comportamiento del estudiante [6].

— El análisis de los ítems no consiste únicamente en obtener cifras numéricas sino que además es necesario dedicar tiempo a la reflexión y el análisis de esos resultados matemáticos, reuniendo cierta dosis de autocrítica.

— Sin lugar a duda, llevar a cabo el análisis de la calidad de una prueba conlleva tiempo y esfuerzo pero revierte directamente en la calidad del proceso enseñanza-aprendizaje. [18],

— La prueba debe incluir un número razonable de ítems para conseguir una muestra representativa de los resultados que se miden.

— Para obtener una discriminación máxima, los elementos del examen deben ser moderadamente difíciles [19].

— La adaptación de la docencia universitaria al Espacio Europeo de Educación Superior exige, entre otras muchas cosas, la aplicación del crédito europeo, que no solo constituye una transformación matemática sino que implica necesariamente un cambio en la metodología didáctica. Este cambio, inexorablemente generará modificaciones en la evaluación del aprendizaje. En relación a estos cambios, cabría plantearse el prota-

gonismo de las pruebas objetivas en estos nuevos planteamientos.

— «*Evaluar para conocer, y conocer para actuar*. Es fundamental establecer conexiones entre el proceso de enseñanza-aprendizaje, la evaluación y la toma de decisiones pues, de lo contrario, el análisis de los ítems no tendría sentido sin una actuación posterior de mejora [20],

BIBLIOGRAFÍA

1. Mateo, J.: La evaluación de los aprendizajes de los alumnos. En: La evaluación educativa, su práctica y otras metáforas. Barcelona: Editorial ICE-Horsori. p. 57-91. 2000.
2. Contreras E. Técnicas de evaluación en la enseñanza universitaria. Seminario para la formación inicial del profesorado universitario. Universidad Miguel Hernández. Elche, 27 y 28 de Marzo de 2003.
3. Navarro, N. Instrumentos de evaluación aplicados a los programas del área de la salud. En: Manual de evaluación para carreras de la salud. Temuco (Chile): Ediciones Universidad de la Frontera. p. 32-40. 2002.
4. Tenbrink, T.D.: Construction del test por el profesor. En: Evaluación: Guía práctica para profesores. Madrid. Narcea S.A.p. 313-359. 1981.
5. Mujika, JR: Análisis de ítems. En: Análisis de ítems y de test con ITEMAN. Bilbao. Servicio editorial del País Vasco, p. 101-127. 1998.
6. Muniz, J.: Teoría clásica de los test. Madrid. Pirámide. 1992.
7. Bordas, I; Cabrera, F.: La evaluación del alumnado en la Universidad. EDUCAR. 28: 61-82, 2001.
8. García, M.: Calificación. En: Planificación educativa en ciencias de la salud. Barcelona. Masson, p.213-229. 1995.
9. Tindal, G.A; Morston, D.B.: Classroombased assessment. Evaluating instructional outcomes. Ohio. Merrill Publishing Company. 1990.
10. Crocker, L; Algina, J.: Introduction to classical and modern test theory. New York. Rinehart and Winston. 1986.
11. Beuchert, A.K; Mendoza, J.L.: A Montecarlo comparison of ten item discrimination indices. Journal of Educational Measurement. 16: 109-118. 1979.
12. Englehart, M.D.: A comparison of several item discrimination indices. Journal of Educational Measurement. 2: 69-76. 1967.
13. Kelly, T.L. The selection of upper and lower groups for the validation of test items. Journal of Educational Psychological. 30: 17-24. 1939.
14. Guilbert, J.J.: Docimología. En: Guía pedagógica para el personal de salud. Valladolid. OMS-ICE Valladolid, p. 404-457. 1989.
15. Ebel, R.L; Frisbie, D.A.: Essentials of educational measurement. New Jersey. Prentice Hall. 1986.
16. Pérez, R.: Pedagogía experimental: la medida en educación. Madrid. UNED. 1989.
17. Sampascual, G.: Las pruebas objetivas, un procedimiento para evaluar el rendimiento escolar. Madrid. Anaya. 1978.
18. De Juan, J.: La evaluación. En: Introducción a la enseñanza universitaria. Didáctica para la formación del profesorado. Dykinson s.l. Cap 8. p. 101-108. 1996.
19. Pallares, M.: Instrumentos de medida, atributos y clases. En: Técnicas e instrumentos de evaluación. Barcelona. CEAC Ediciones. p.7-26. 1977.
20. Gómez de Terreros, L: Análisis evaluativo de la calidad de la prueba objetiva tipo test (pregunta de elección múltiple). Revista de Enseñanza Universitaria, 13:105-111, 1998.