

ARTÍCULOS ORIGINALES

Nivel de dificultad y poder de discriminación del tercer y quinto examen parcial de la cátedra de cito-histología 2007 de la carrera de medicina de la UMSA

Dr. Julio Hernán Pérez Tapia*, Dr. Nelson Acuña Aguilar**, Dr. Elvis Ruben Arratia Cuela***.

RESUMEN

Objetivo:

Determinar el nivel de dificultad y el índice de discriminación del tercer y quinto examen parcial de Cito-Histología.

Diseño de investigación:

Descriptivo de corte transversal.

Lugar:

Cátedra de Cito-Histología

Muestra:

388 exámenes de estudiantes que dieron el tercer parcial y 350 el quinto

Método:

Los datos se recolectaron a través de un formulario, el nivel de dificultad y poder de discriminación se obtuvieron por formulas establecidas.

Resultados:

La media de dificultad del tercer parcial fue de 0,34, siendo medianamente difícil, la media del quinto fue 0,54 siendo de dificultad media. La media del índice de discriminación del tercer parcial fue 0,57 y 0,45 del quinto, el coeficiente de discriminación fue 0,44 para el tercero y 0,54 para el quinto, así, ambos parciales tuvieron una calidad excelente.

Conclusiones:

Según el nivel de dificultad del tercer parcial fue más difícil que el quinto, éste último estuvo dentro de parámetros esperados, el tercer parcial no tuvo reactivos fáciles. Ambos parciales discriminaron positivamente, una pregunta del tercer parcial debería descartarse y dos del quinto revisarse a profundidad.

Palabras clave:

Evaluación, nivel de dificultad, poder de discriminación.

ABSTRACT

Objective:

To determine the level of difficulty and the power of discrimination of the third and fifth partial exam in Cytohistology

Research Design:

Descriptive, cross sectional

Location:

Chair of Cytohistology

Sample:

The students who took the third and fifth partial exams: 388 for the third and 350 for the fifth.

* Ortopedia-Traumatología. JEFE DE DEPARTAMENTO CIENCIAS MORFOLOGICAS FACULTAD DE MEDICINA-UMSA

** Medico Residente II – Imagenología. Hospital de Clínicas

*** Médico - Cirujano

Method:

The data were collected by means of a form, the difficulty level and power of discrimination were calculated by established formulas

Results:

The average difficulty of the third partial exam was 0.34, qualified as moderately difficult. The average difficulty of the fifth partial exam was 0.54 which qualifies for half difficult. The average index of discrimination of the third partial exam was 0.57 and of the fifth 0.45. The discrimination quotient was 0.44 for the third and 0.54 for the fifth exam, and both partials were of excellent quality.

Conclusions:

Considering the difficulty level, the third partial exam was more difficult than the fifth the latter being within expected parameters. Both partials discriminated positively, one question of the third partial should be removed, and two questions of the fifth test should be reviewed in depth.

Key Words :

Assessment, difficulty level, power of discrimination.

INTRODUCCIÓN

La evaluación educativa es parte fundamental del proceso enseñanza-aprendizaje ¹. Es un proceso que involucra la elaboración, aplicación y análisis de los instrumentos de medición. La función principal de un instrumento de medición educativo cuando se crea como medida para inferir las capacidades de las personas, es ofrecer información para la correcta toma de decisiones ².

El examen escrito es el tipo de evaluación más utilizado en la Facultad de Medicina; nuestra cátedra de citohistología no es la excepción, la prueba que se utiliza para evaluar el nivel de conocimientos de la materia de citohistología en los estudiantes son los tests objetivos, utilizando preguntas de selección múltiple. Este tipo de pruebas es preferido por la facilidad, la rapidez y la objetividad de su calificación ³⁻⁶.

Cuando se termina de elaborar un instrumento evaluativo, surge el problema de verificar de que manera la información obtenida refleja el nivel de competencia del estudiante, aparecen entonces conceptos para evaluar la calidad técnica de un examen, entre estos se encuentra el análisis de reactivos (preguntas), y son tres los indicadores que se utilizan para describir las características psicométricas de los reactivos de un test objetivo: su nivel de dificultad, su poder de discriminación y el funcionamiento de sus distractores ³, de estos los dos primeros son los más usados por los autores ². Tenemos entonces no sólo la posibilidad, sino también la obligación, de someter al juicio crítico de la psicometría la calidad de elaboración de nuestras pruebas.

En 1985, la Asociación Americana de Psicología, en colaboración con la Asociación Americana de Investigación Educativa y el Consejo Nacional para la Medición en Educación, publicó el manual "Estándares

para las pruebas educativas y psicológicas", documento que enfatiza la importancia de tomar en cuenta el bienestar de las personas que hacen una prueba y evitar el mal uso de los instrumentos de evaluación ². Mientras que, en países desarrollados, es obligatorio que estos criterios de calidad se satisfagan, en Bolivia es inexistente esta normatividad.

El propósito de este estudio es dar a conocer los resultados de los indicadores psicométricos de los reactivos del tercer y quinto examen parcial de la Cátedra de Cito-Histología, los mismos que fueron tomados el primero como una evaluación diagnóstica y el segundo como una evaluación de proceso, los relacionados con su nivel de dificultad y poder de discriminación, que son los indicadores básicos que se utilizan para el análisis de los ítems en un test objetivo. Esto para llegar a mejorar los instrumentos evaluativos que se realizan en la cátedra.

MÉTODOS

Tipo de estudio:

Descriptivo de corte transversal.

Universo:

420 exámenes de los alumnos inscritos en la cátedra de histología

Muestra:

388 pruebas de los estudiantes que rindieron el tercer examen y 350 el quinto examen parcial.

Procedimiento:

Los pasos para la administración y calificación del examen, fueron los mismos que define la cátedra para todos los parciales: 1) cada docente que dio las clases magistrales elaboró las preguntas y respuestas correspondientes al tema; 2) se publicó la fecha, lugar y hora del examen; 3) el estudiante respondió el

examen y 4) los profesores de la cátedra calificaron los exámenes según un patrón de respuestas realizado anteriormente por el docente que impartió el tema. Luego del tiempo correspondiente para reclamos de los estudiantes por sus notas, se procedió a la recolección de la información en una planilla de recolección de datos.

Para lograr nuestro objetivo, en primer lugar, describiremos lo que en la literatura especializada se entiende técnicamente por dificultad y discriminación de un ítem, así como las formas de calcularlos.

ANÁLISIS DE REACTIVOS

Existen varias maneras equivalentes de calcular el nivel de dificultad y el índice de discriminación, así como su evaluación^{2,7-12}, en este trabajo utilizaremos los criterios los siguientes: (cuadro 1)

Nivel de Dificultad.- Llamado también índice de facilidad¹⁰, es la proporción de personas que responden correctamente un reactivo de una prueba. Este índice nos muestra qué tan fácil (valores cercanos a 1) o difícil (valores cercanos a 0) ha resultado la pregunta para el total de esos alumnos.

Usualmente, a esta proporción se le denota con una p . La obtendremos por la siguiente fórmula²:

$$p_i = \frac{A_i}{N_i} \quad (1)$$

Cuadro N° 1
Nivel de dificultad y evaluación del reactivo

p=	Evaluación del reactivo
> 0.86	Altamente fáciles
0.74 – 0.86	Medianamente fáciles
0.53 – 0.73	Dificultad media
0.33 – 0.52	Medianamente difíciles
< - 0.32	Altamente difíciles

Índice de Discriminación.- Llamado también índice de homogeneidad^{8,9}, es la capacidad de un ítem respecto del objetivo que esta siendo evaluado. Así, un buen ítem debe ser contestado en proporción mayor por

los estudiantes con las mejores calificaciones. Para calcular este índice utilizaremos la siguiente fórmula²:

$$D_i = \frac{GA_{aciertos} - GB_{aciertos}}{N_{grupomayor}} \quad (2)$$

Entre más alto es el índice de discriminación, el reactivo diferenciará mejor a las personas con altas y bajas calificaciones.

El cuadro N° 2, muestra los valores D y su correspondiente interpretación (regla de “dedo”). Asimismo, en la tabla se señalan las recomendaciones para cada uno de estos valores.

Cuadro N° 2
Poder de discriminación de los reactivos según su valor D

D =	Calidad	Recomendaciones
> 0.39	Excelente	Conservar
0.30 – 0.39	Buena	Posibilidades de mejorar
0.20 – 0.29	Regular	Necesidad de revisar
0.00 – 0.20	Pobre	Descartar o revisar a profundidad
< - 0.01	Pésima	Descartar definitivamente

Otro indicador que se utiliza es el punto de correlación biserial (r_{pbis}), la ventaja es que toma en cuenta todas y cada una de las personas evaluadas, mientras que con índice de discriminación, sólo se toma 54% de ellas (27% más alto y 27% más bajo). La ecuación para obtener este indicador, es la siguiente²:

$$r_{pbis} = \frac{\bar{x}_1 - \bar{x}_0}{s_x} * \sqrt{\frac{n_1 n_0}{n(n-1)}} \quad (3)$$

El índice de dificultad se calculó con la ecuación (1), el índice de discriminación con la ecuación (2) y el coeficiente de discriminación con la ecuación (3). Este procedimiento se realizó manualmente utilizando una calculadora, el software de la hoja de cálculo Excel 2003 y el paquete estadístico SPSS 11.5

RESULTADOS

De los 420 alumnos inscritos en la Cátedra de Histología 388 rindieron el tercer examen parcial y 350 el quinto,

cada parcial tuvo 35 preguntas de opción múltiple, el cuadro N° 3 muestra los temas de cada parcial.

Cuadro N° 3
Temas y número de preguntas por parcial

TERCER EXAMEN PARCIAL		QUINTO EXAMEN PARCIAL	
Temas	N° de Preguntas	Temas	N° de Preguntas
Sangre y hematopoyesis	10	Sistema reproductor femenino	6
Sistema Respiratorio	10	Sentidos especiales	7
Sistema Circulatorio	9	Oído	7
Sistema linfoide	6	Ojo	9
		Sistema reproductor masculino	6
TOTAL	35		35

Fuente propia

La media del nivel de dificultad del tercer examen parcial fue de 0.34, la oscilación estuvo entre 0.9 y 0.67. La media de dificultad del quinto parcial fue 0.54 y su oscilación estuvo entre 0.21 y 0.90. La media de ambos parciales fue de 0.44. La media del índice de discriminación del tercer parcial fue 0.57, los valores oscilaron entre 0.00 y 0.84. La media del índice de discriminación del quinto parcial fue 0.45 y osciló entre 0.07 y 0.76, la media de ambos fue de 0.51. La media del coeficiente de discriminación (rpbis) del tercer parcial fue de 0,44, los valores oscilaron entre -0,01 y 0,71. La media de discriminación del quinto parcial fue 0,45 y los valores estuvieron entre 0,2 y 0,95. (Cuadro N°4)

Cuadro N° 4
Medias y desviaciones estándar del nivel de dificultad (p), índice de discriminación (D) y coeficiente de discriminación (rpbis), por temas y parciales.

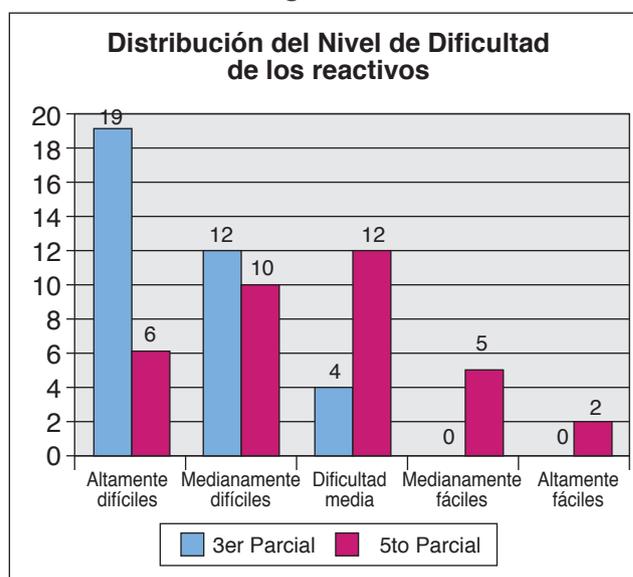
Tercer Parcial	Temas	P		D		rpbis		
		N°	Media	D.E.	Media	D.E.	Media	D.E.
	Sangre y Hematopoyesis	10	0,37	0,12	0,64	0,12	0,52	0,11

Sistema Respiratorio	10	0,34	0,21	0,52	0,19	0,39	0,19
Sistema Circulatorio	9	0,35	0,12	0,53	0,22	0,40	0,18
Sistema linfoide	6	0,27	0,08	0,60	0,08	0,45	0,06
Total	35	0,34	0,14	0,57	0,17	0,44	0,15
Quinto Parcial		P		D		rpbis	
Temas	N°	Media	D.E.	Media	D.E.	Media	D.E.
Sistema reproductor femenino	6	0,64	0,21	0,32	0,04	0,50	0,18
Sentidos especiales	7	0,45	0,15	0,49	0,08	0,51	0,13
Oído	7	0,61	0,17	0,52	0,16	0,69	0,19
Ojo	9	0,63	0,22	0,45	0,20	0,57	0,24
Sistema reproductor masculino	6	0,31	0,09	0,44	0,21	0,38	0,16
Total	35	0,54	0,21	0,45	0,16	0,54	0,20

Fuente propia

En la figura N° 1 se puede observar la distribución del nivel de dificultad de ambos parciales, donde muestra que el tercer parcial presenta una mayor cantidad de reactivos altamente difíciles y el quinto parcial reactivos de dificultad media.

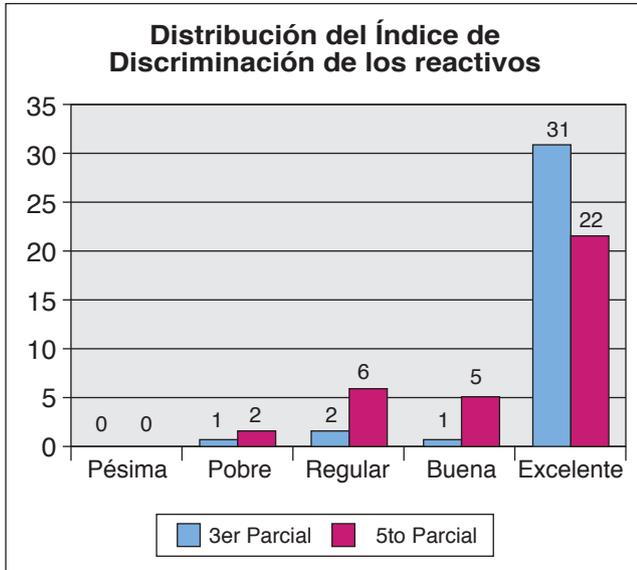
Figura N° 1



Fuente propia

La figura N° 2 muestra la distribución del índice de discriminación, donde se evidencia que la gran mayoría de los reactivos de ambos parciales discrimina excelentemente.

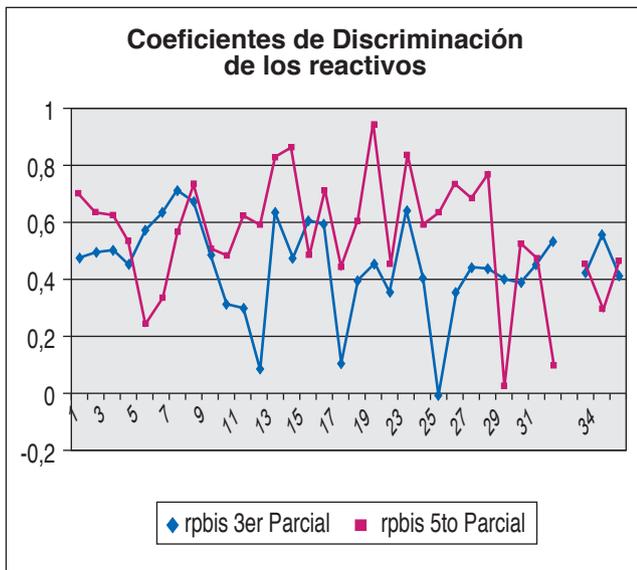
Figura N° 2



Fuente propia

La figura N° 3 se aprecia el comportamiento discriminativo de ambos parciales con el coeficiente de discriminación, se puede observar que la mayoría de los reactivos discriminan excelentemente.

Figura N° 3



Fuente propia

Según la información presentada en los cuadros y gráficos, podemos clasificar a los reactivos del tercer y quinto parcial de Histología de acuerdo a su:

Niveles de dificultad: El tercer parcial tuvo un 54.28% de reactivos altamente difíciles ($p < 0.32$); 34.28% medianamente difíciles ($p = 0.33$ a 0.52); y 11.42% de dificultad media ($p = 0.53$ a 0.73), no existiendo preguntas medianamente fáciles ($p = 0.74$ a 0.86) ni preguntas altamente fáciles ($p > 0.86$). Las preguntas del tema sistema linfóide fueron las más difíciles (media 0.27).

El quinto examen parcial tuvo un 17.14% de reactivos altamente difíciles, 28.57% de reactivos medianamente difíciles; 34.28% de dificultad media; 14.28% de medianamente fáciles y; 5.71% altamente fáciles. Los reactivos del tema del sistema reproductor masculino fueron los más difíciles (media 0.31).

Índice de discriminación: El tercer parcial tuvo 2.85% de reactivos con pobre poder de discriminación ($D = 0.00-0.19$); 5.71% regular poder de discriminación ($D = 0.20$ a 0.29); 2.85% buen poder de discriminación y 88.57% excelente poder de discriminación ($D > 0.40$). El tema de tejido sanguíneo fue el que mejor discriminó (media de 0.64).

El quinto examen parcial tuvo un 5.71% de reactivos con pobre poder de discriminación; 17.14% reactivos con regular poder de discriminación; 14.28% buen poder de discriminación; 62.8% un excelente poder de discriminación, no existiendo reactivos de pésimo poder de discriminación. Los reactivos de oído y sentidos especiales discriminaron mejor (media 0.52 y 0.49 respectivamente).

Coefficiente de discriminación: El tercer parcial contó con 2.85% reactivos con pésimo poder de discriminación ($D < -0.01$); 5.71% pobre poder de discriminación ($D = 0.00-0.19$); 17.14% buen poder de discriminación ($D = 0.30 - 0.39$) y 74.28% excelente poder de discriminación ($D > 0.40$), no existiendo valores de regular nivel de discriminación ($D = 0.20$ a 0.29). También el tema de tejido sanguíneo discriminó mejor (0.52).

El quinto parcial tuvo 5.71% reactivos de pobre poder de discriminación; 5.71% regular poder de discriminación; 8.57% buen poder de discriminación y; 80% excelente poder de discriminación, no existiendo ningún reactivo con pésimo poder de discriminativo. Los temas de oído y ojo fueron los que mejor discriminaron (0.69 y 0.57).

DISCUSIÓN Y CONCLUSIONES

Una de las formas fundamentales de conocer la calidad, si las preguntas están adecuadamente elaboradas,

de los reactivos (preguntas) de una prueba o test de opción múltiple es poniéndolos a prueba y analizando el comportamiento de éste en forma empírica. Dos indicadores fundamentales para realizar el análisis de los ítems de un examen son: el nivel de dificultad y el poder discriminativo ².

El nivel de dificultad del tercer parcial (0.34) es más difícil que el nivel de dificultad esperado en la literatura (0.5-0.6) ², sin embargo el quinto parcial tuvo una media de 0.54, un nivel adecuado. La distribución de la dificultad de los reactivos del tercer examen sólo tiene preguntas altamente difíciles, medianamente difíciles y de dificultad media, no existiendo ningún porcentaje de preguntas de dificultad media, medianamente fáciles ni altamente fáciles, siendo recomendable que exista este tipo de preguntas en un cierto porcentaje ^{2,15}, el cuál si lo tuvo el quinto parcial, cubriendo prácticamente toda la gama de valores de *p*, además, los valores se aproximaron en parte a los valores de las referencias (5% de reactivos fáciles, 20% medianamente fáciles, 50% con una dificultad media, 20% medianamente difíciles y 5% difíciles) ², estos valores no fueron planificadas por la cátedra, cada docente que impartió la clase magistral realizó las preguntas y respuestas de ese determinado tema.

Pasando al índice de discriminación del examen, podemos decir que el 91.42% presentan niveles buenos a excelentes, valor superior a otros estudios ², sólo un 2.85% discriminó negativamente que se asemeja a la literatura ². El quinto parcial discriminó favorablemente en un 77.08%. En ningún caso existió la calidad de pésima discriminación.

Los coeficientes de discriminación también mostraron resultados similares a los índices, tuvieron un 91.42% de nivel bueno a excelente, 8.56% discriminó negativamente y uno de ellos tuvo una calidad pésima. Los reactivos del quinto parcial tuvieron un 88.57% de discriminación buena a excelente sólo 5.71% discriminó negativamente. Estos resultados son similares a los encontrados en la literatura ².

Por todo lo mencionado podemos señalar que el nivel de dificultad del tercer examen parcial fue más difícil que el quinto, el cual si estuvo dentro de los parámetros esperados en la literatura, ya que más de la mitad de las preguntas (54.28%) fueron altamente difíciles. Tomando el primero como evaluación diagnóstica y el segundo como evaluación de proceso se puede decir que un parcial fue más difícil de responder para el estudiante, será por la complejidad de los temas?, por qué el alumno no se dedicó lo suficiente a ese parcial?, son preguntas que surgen y que se intentaran responder en las siguientes gestiones al realizar un estudio similar.

El poder de discriminación de ambos exámenes fue en su mayoría excelente, solamente 1 pregunta del tercer examen debería descartarse definitivamente y dos descartarse o ser revisados a profundidad. Dos preguntas del quinto parcial deberían descartarse o ser revisados a profundidad.

Es necesario señalar que estos resultados sirven para la mejora en la calidad técnica, en el sentido de mejorar la elaboración de los reactivos, de los exámenes, también mejoramos la calidad docente-estudiantil, con lo que contribuimos a la visión facultativa "... prestar servicios idóneos en docencia...".

Los datos proporcionados beneficiará el proceso enseñanza aprendizaje en la cátedra, a los docentes, ya que estos conocerán el grado de dificultad de sus reactivos y si discriminan adecuadamente, para que mejoren los mismos, se revisará los reactivos señalados. Los estudiantes sabrán que los reactivos que ellos responden tienen una adecuada elaboración y que se sustentan bajo criterios establecidos.

Además fortalecen una de las principales labores, evaluar los conocimientos que los estudiantes adquirieron durante las horas de clases y de estudio en la cátedra, ya que con ello se podrá discriminar y por lo tanto seleccionar a los estudiantes que aprobaran la materia para seguir avanzando en su carrera universitaria.

REFERENCIAS

1. Venturelli J. Educación Médica Nuevos enfoques, metas y métodos. 2da ed. Washington: Organización Panamericana de la Salud; 2003.
2. Backhoff E, Larrazolo N, Rosas M. Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). Revista Electrónica de Investigación Educativa 2000; 2 (1)
3. Salas Perea R. S., La evaluación en la educación superior contemporánea. Bolivia. Editor Buddy Lazo de la Vega. 1998.
4. Mentruyt O. Acabar con la tragedia de la corrección Pruebas objetivas de rendimiento Tests de multiples choice. [monografía en línea]. Disponible en: <http://www.monografias.com/trabajos11/test/test.shtml>. Accesible desde el 27 de noviembre de 2007.
5. M. Eva Trigo, José López, Rafael J. Martínez y Rafael Moreno. Propiedades psicométricas de un conjunto de pruebas en función del contenido, tipo y número de opciones de respuesta. [artículo en línea]. Iber Psicología 2005 Disponible en: http://www.fedap.es/IberPsicologia/iberpsi10/congreso_lisboa/trigo/trigo.htm. Accesible desde el 16 de junio de 2007.
6. Norma Juárez Días González, Fausto García Carrasco. Taller Examen Escrito. [en línea]. Disponible en: <http://www.certificacionespecialistas.org.mx/pdfs/12.pdf>. Accesible desde el 17 de junio de 2007.
7. Informe del cuestionario. La enciclopedia libre. [en línea]. Disponible en: http://docs.moodle.org/es/Informe_del_cuestionario. Accesible desde el 30 de octubre de 2007.
8. Errores más frecuentes. [en línea]. disponible en: <http://www.uned.es/psico-3-psicometria/errores.htm#DOS>. Accesible desde el 5 de noviembre de 2007.
9. Glosario de términos (glosario en línea). Disponible en: <http://www.uam.es/docencia/ace/4-20Glosario.htm#Consistencia%20interna>. Accesible desde el 15 de octubre de 2007.
10. Raymundo Carrasco Soto. Condiciones socioeconómicas y demográficas del alumno como determinantes en el aprendizaje. [trabajo en línea]. Durango, México 2007. Disponible en: <http://www.psicologiacientifica.com/bv/imprimir-234-condiciones-socioeconomicas-y-demograficas-del-alumno-como-determinantes-en-el-aprendizaje.html>. Accesible desde el 15 de octubre de 2007.
11. Baldomero Carrera Santacruz, Miguel Romo Cedano. Examen de Admisión a la Universidad Iberoamericana. Un Programa de Evaluación para los Aspirantes de Primer Ingreso. [en línea]. Disponible en: http://www.anuies.mx/servicios/p_anuies/publicaciones/revsup/res047/txt6.htm#top. Accesible desde el 17 de junio de 2007.
12. Análisis de ítems ¿cómo saber si mi prueba de evaluación ha funcionado adecuadamente?. [en línea]. Disponible en: <http://www.uam.es/docencia/ace/24b-analisis.htm#preparación>. Accesible desde el 10 de octubre de 2007.