

ANÁLISIS DE LA POSICIÓN DE LOS ÍTEM EN RELACIÓN CON LOS ÍNDICES DE DIFICULTAD DE LOS MISMOS.

Inmaculada Sánchez-Aguayo, Antonio Ariza*, Inmaculada Domínguez, Petra Amores, José Torreblanca.

Dpto. Biología Celular, Facultad de Biología; *Dpto. Didáctica de las Matemáticas, Facultad de Ciencias de la Educación.

Universidad de Sevilla.

En el curso de los últimos años hemos desarrollado un software de asistencia a la docencia (Exawin) para el que hemos creado y validado diferentes bases de preguntas, tanto teóricas como prácticas. Uno de los valores de Exawin es el de generar exámenes a través de la elección aleatoria de las preguntas. Incluso se pueden generar copias de un mismo examen cambiando aleatoriamente el orden de las preguntas y de las respuestas.

Diversos autores sostienen que la disposición al azar de las preguntas tiene más inconvenientes que ventajas, pues el orden de los ítems influye en la calificación final del examen. Otros autores opinan que la elección aleatoria de los ítems no solo no es mala sino que es recomendable.

Siendo esta cuestión esencial en la forma de generar los exámenes por Exawin hemos querido comprobar cuál de las dos visiones es correcta. Para ello hemos calculado y representado gráficamente los índices de dificultad (ID) de las preguntas de 5 exámenes iguales con preguntas colocadas aleatoriamente. No hemos encontrado ningún ítem cuya gráfica presente una pendiente positiva o negativa en todos sus tramos. Por el contrario, la pendiente de las distintas gráficas no ha seguido ninguna pauta fija. Nos encontramos con gráficas en su mayoría con pendientes cambiantes, gráficas en dientes de sierra, o pendiente cero que nos estará indicando que el ítem tiene la misma dificultad en distintas posiciones. De estos resultados podemos concluir que la posición del ítem en la prueba no influye en la dificultad del examen.

Parte del trabajo y su divulgación han sido sufragados con ayudas del I Plan Propio de Docencia de la Universidad de Sevilla.

ANTECEDENTES

A la hora del diseño de pruebas objetivas de respuesta múltiple, son muchos los autores que hacen recomendaciones sobre los ítems. Así se nos indica como deben de redactarse tanto los enunciados como las respuestas y como deben de ser los distractores. Además, se nos indica cual sería el número idóneo de ítems e incluso en que orden deben de preguntarse. En este sentido Margarita Oncins de Frutos, en su recomendación metodológica para la realización de encuestas, considera una serie de normas que pueden ayudar a situar mejor las preguntas dentro de un cuestionario, como son entre otras:

- Colocar antes las preguntas que pueden ser influenciadas por las respuestas a otras preguntas.
- Las preguntas generales deben ir delante de las específicas
- Evitar la monotonía, alterando el orden lógico de temas similares

De igual forma, F. Javier Murillo Torrecilla de la Universidad Autónoma de Madrid, considera que un elemento básico de un cuestionario es el orden de los ítems y que para establecer la ordenación y disposición de los mismos, al no existir criterios generales, podría tenerse en cuenta las siguientes recomendaciones:

- Incluir las cuestiones identificativas al principio o, excepcionalmente al final, y sólo aquellas que tengan relación directa con el tema estudiado y que, además, vayamos a analizar.
- Presentar las cuestiones agrupadas por temáticas. Saltar de un tema a otro supondría un gran esfuerzo al alumno.
- Preguntar de lo general a lo particular (es la llamada técnica embudo), de lo simple a lo complejo.
- Situar juntas las preguntas de un mismo tipo, y siempre con el mismo orden de respuesta.

Asimismo, en el Servicio de innovación educativa de la Universidad de Navarra, se recogen las propuestas de la Dra Charo Reparaz (2006) sobre “Como preparar exámenes de tipo test de forma adecuada”, la cuál considera que existen varios criterios de ordenación de los ítems según:

- Tipos de elementos (verdadero-falso, elección múltiple, etc.). Para presentar al principio de cada bloque instrucciones necesarias para resolverlo y facilitar su corrección ya que cada tipo de elemento puede tener reglas específicas.
- Orden de dificultad. Situar los elementos más fáciles al principio da ánimos para seguir adelante.
- Seguir el orden que la estructura de la materia exige; lo que puede facilitar las contestaciones.
- Al azar. Presenta más inconvenientes que ventajas

De estas orientaciones se puede deducir que el orden de los ítems influye en la calificación final del examen. Es decir, que la dificultad de un ítem, varía en función de su

colocación en la prueba. Los situados al principio les resultarán al alumno más fáciles, ya que se encuentra más fresco y puede dedicarles más tiempo, Los últimos resultarán más difíciles ya que el alumno dispondrá de menos tiempo y estará más agobiado por el peso de la prueba. Si los ítems se eligen al azar los alumnos pueden desorientarse y contestar peor a las preguntas.

En contraposición a estas orientaciones anteriores, Neira y colaboradores entre otros, consideran que los alumnos pueden predecir, en parte, las preguntas formaran parte de cuestionario de evaluación, su estilo, su forma, etc. y nos indica que la mejor solución para evitarlo es disponer de una amplia base de datos con las preguntas de la asignatura y elegir las de forma totalmente aleatoria. Es decir, que la elección aleatoria de los ítems no solo no es mala sino que es recomendable.

OBJETIVO

Puesto que diversos autores sostienen que la disposición al azar de las preguntas tiene más inconvenientes que ventajas, ya que el orden de los ítems influye en la calificación final del examen, y otros opinan que la elección aleatoria de los ítems no solo no es mala sino que es recomendable, hemos querido comprobar cuál de las dos visiones es más correcta. Para ello hemos valorado y analizado los datos extraídos del examen final de la asignatura “*Lenguaje informático para la introducción de nociones matemáticas básicas en educación infantil*”, optativa de tercer curso de la Diplomatura de Maestros de Educación Infantil durante los cursos académicos 07-08, 08-09 y 09-10 y de una asignatura obligatoria del Grado en Biología “*Principios, Instrumentación y Metodologías en Biología Celular y Microbiología*”(PIM) que se ha impartido por primera vez en el curso actual(09-10) como consecuencia de la implantación del Grado en Biología en la Universidad de Sevilla. Consecuentemente, los datos sobre esta asignatura solo se corresponden con un curso académico.

INTRODUCCIÓN Y METODOLOGÍA

En el curso de los últimos años hemos desarrollado un software de asistencia a la docencia (Exawin)(Ariza, 2006) para el que hemos creado y validado diferentes bases de preguntas, tanto teóricas como prácticas. Este software es una aplicación que nos permite entre otras acciones, almacenar ítems en base de datos ordenados por temas, objetivos o cualquier tipo de ordenación que quiera el profesor. Uno de los valores de Exawin es el de generar exámenes a través de la elección aleatoria de las preguntas (pruebas objetivas), incluso se pueden generar copias de un mismo examen cambiando aleatoriamente el orden de las preguntas y de las respuestas.

El programa Exawin es una herramienta de medición y evaluación realizada en lenguaje Visual Basic que funciona bajo los sistemas operativos Windows 95 / 98 / 2000 / Xp y en su versión actual sirve para:

- Crear y mantener bases de datos de alumnos.
- Crear y mantener bases de datos de preguntas.
- Generar pruebas objetivas.

- Corregir exámenes, tanto colectivos como individuales, utilizando diferentes criterios.
- Analizar y comparar los ítems.

Las posibilidades antes enumeradas van a permitir comprobar la validez de las preguntas y en caso de ser necesario, modificarlas. El análisis de los ítems, por su parte, permite estudiar la eficacia de las preguntas utilizadas en función de los objetivos propuestos. Este análisis se basa en el estudio de las propiedades estadísticas de los ítems siendo los índices más relevantes:

- El índice de dificultad (ID): proporción de alumnos que fallan un ítem entre aquellos que han intentado resolverlo (6).
- El índice de facilidad (IF): proporción de alumnos que aciertan entre aquellos que han intentado resolverlo. Ambos son complementarios, si no hay preguntas en blanco, y nos indicarían el grado de dificultad/facilidad de un ítem o examen (6).
- El índice de discriminación, es la correlación entre las puntuaciones de los alumnos en el ítem y sus puntuaciones en la prueba (12). Mide la eficacia del ítem para diferenciar entre alumnos “buenos” o eficaces y “malos” o ineficaces en relación a la resolución de la prueba realizada (5).

Una vez creado el examen de opción múltiple, el programa Exawin, nos puede generar hasta cinco copias del mismo examen con los ítems, la respuesta verdadera y los distractores cambiados de sitio.

La calificación se obtiene, tras la introducción de las respuestas en el programa, de manera rápida y objetiva. Exawin corrige exámenes, individuales y colectivos, utilizando diferentes criterios y eliminando el acierto ciego. Se obtiene una puntuación bruta de cada prueba que es multiplicada por un factor en función de la nota máxima asignada por el profesor para dicho examen.

Criterios de corrección disponibles:

- De Lord o clásico (acierto-errores). La fórmula sería $\text{Puntuación Bruta} = \text{Aciertos} - \text{Errores/Opciones} - 1$ (Ariza 2001, Lord 1980)
- Sólo aciertos. La fórmula sería $\text{Puntuación Bruta} = \text{Aciertos} - N_{\text{ítems/Opciones}}$ (Ariza 2001)
- Por niveles de dificultad. La puntuación bruta se obtendría del sumatorio del valor de las respuestas acertadas y de las respuestas falladas. El valor de cada pregunta variará según un nivel de dificultad que asignará el programa en función de la dificultad de cada una ellas. Esta dificultad se basará en las respuestas obtenidas en la prueba. Es decir, la puntuación dependerá de las características del grupo (Ariza, 2001)

De esta manera, obtuvimos las notas de cada alumno, con sus fallos, aciertos y respuestas en blanco, la nota media de cada grupo, el cálculo de los índices de dificultad y discriminación de cada ítem y la media del grupo.

Por otra parte, el programa Exawin, nos ofrece la posibilidad de realizar un análisis comparativo, ítem a ítem, de los índices de dificultad y discriminación entre las cinco copias del examen indicándonos el lugar que ocupa cada ítem en los distintos exámenes. Esto nos permitió hacer gráficas comparativas de cada ítem y comprobar como variaban tanto la dificultad como la discriminación. Según sea la variación en estos índices podremos afirmar que tipo de influencia tiene el orden de colocación de los ítems en la dificultad de los mismos.

Si aceptamos la hipótesis de que el orden influye en la dificultad de los ítems y que esta aumenta cuanto más atrás esté el ítem, al realizar la gráfica por cada ítem pero colocándolos no según el examen al que pertenezca sino en función del orden de aparición de menor a mayor indistintamente del examen, se obtendría una gráfica siempre con una pendiente mayor que cero, ya sea en una sola línea o en tramos. Si la pendiente fuera negativa, en un porcentaje estadísticamente significativo, la hipótesis sería falsa lo cual implicaría que el orden no influiría en la dificultad de los ítems.

RESULTADOS

El trabajo se ha realizado analizando los datos obtenidos durante tres cursos académico, 07-08, 08-09 y 09-10 y corresponden al examen final de la asignatura “*Lenguaje informático para la introducción de nociones matemáticas básicas en educación infantil*” optativa de tercer curso de la diplomatura de Maestros de Educación Infantil y de “*Principios, Instrumentación y Metodología en Biología Celular y Microbiología*” del Grado de Biología.

En la primera asignatura, el número de alumnos que realizaron el examen estaba comprendido entre 100 y 150 por curso. En los dos primeros, se hicieron 4 grupos de 25-35 alumnos y en el tercero cinco grupos de aproximadamente 30 alumnos. Cada grupo, del mismo curso, realizó una copia idéntica del mismo examen con los ítems en diferentes posiciones.

Los datos recogidos en las tablas 1, 2 y 3 se corresponden con la nota media sobre 5 de cada examen, el índice de facilidad/dificultad medio y el índice de discriminación medio de cada examen. Los datos están agrupados por cursos.

TABLA 1. Examen primera convocatoria curso 07_08			
Lenguaje Informático para la intr. de nociones matem. básicas en Educación Infantil			
	Nota Media	IF/ Idf	Ids (10%)
Examen_1	2,53	0,59	0,50
Examen_2	2,57	0,59	0,48
Examen_3	2,68	0,64	0,50
Examen_4	2,53	0,58	0,72
Medias	2,57	0,60	0,55

Examen primera convocatoria curso 08-09			
Lenguaje Informático para la intr. de nociones matem. básicas en Educación Infantil			
	Nota Media	IF/ Idf	Ids (10%)
Examen 1	3,79	0,81	0,52
Examen 2	3,55	0,80	0,42
Examen 3	3,95	0,83	0,33
Examen 4	3,84	0,81	0,50
Medias	3,70	0,81	0,44

TABLA 3. Examen primera convocatoria curso 09_10			
Lenguaje Informático para la intr. de nociones matem. básicas en Educación Infantil			
	Nota Media	IF/ Idf	Ids (10%)
Examen 1	3,36	0,72	0,38
Examen 2	3,25	0,72	0,48
Examen 3	3,42	0,74	0,38
Examen 4	3,45	0,75	0,43
Examen 5	3,57	0,77	0,33
Medias	3,41	0,74	0,40

En las tablas 1, 2 y 3 se aprecia que las notas medias no muestran diferencias significativas respecto a la media, manteniéndose en el mismo rango y nunca son superiores al 5% por lo que estas diferencias son estadísticamente despreciables. Además, la nota no varía sustancialmente de un examen a otro. Por tanto la diferente posición de los ítems en los exámenes no afecta al resultado final de los mismos.

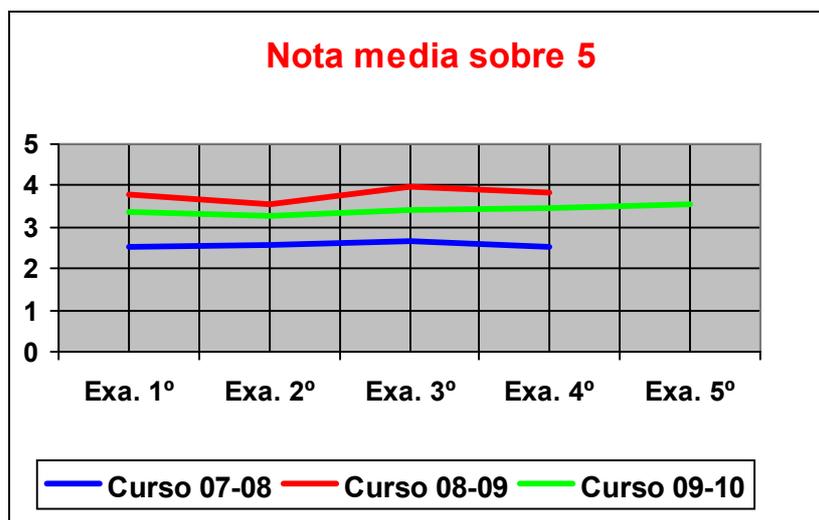
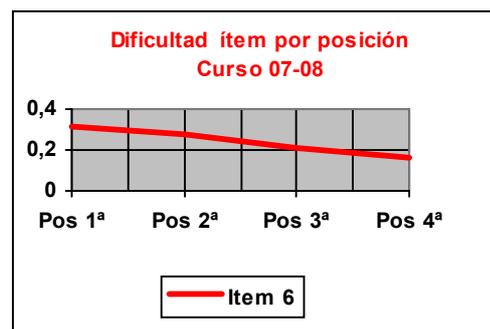
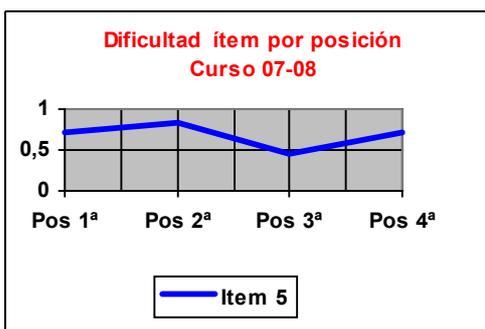
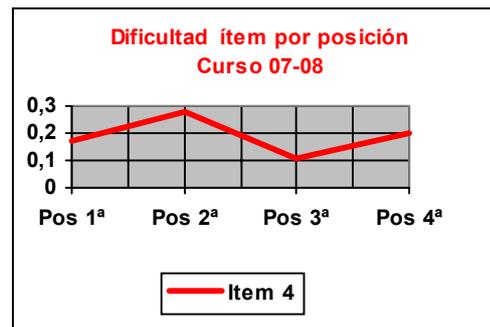
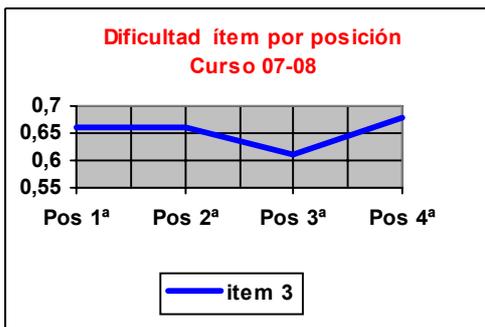
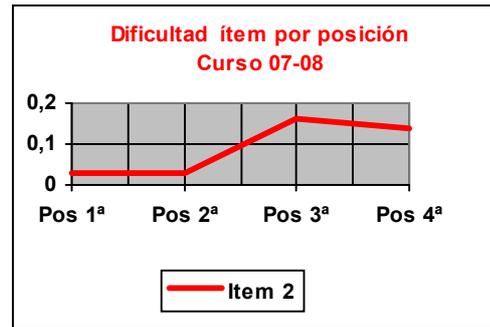
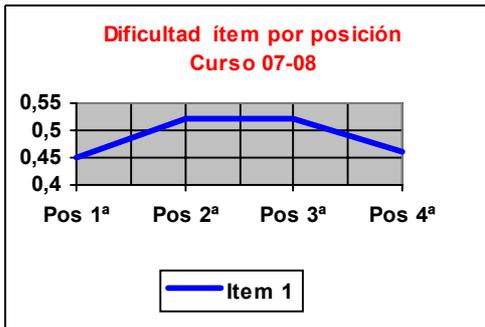
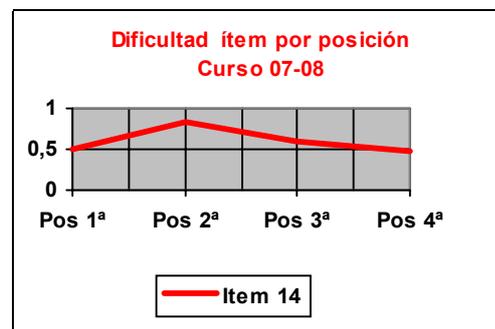
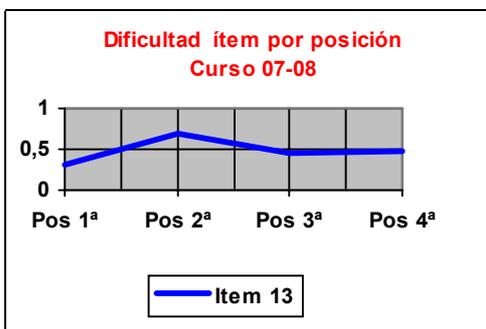
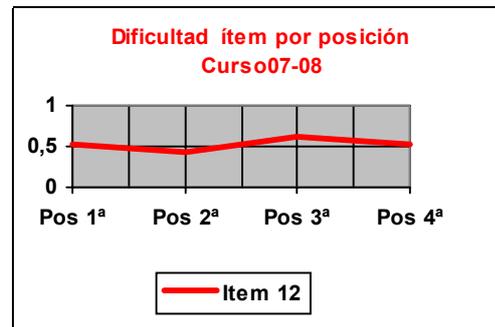
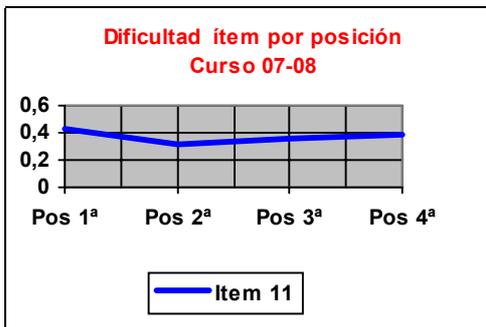
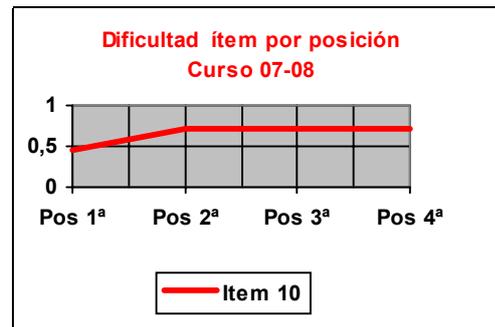
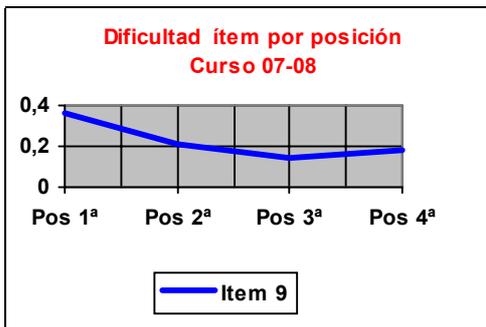
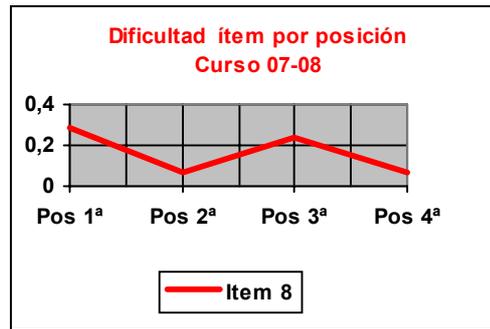
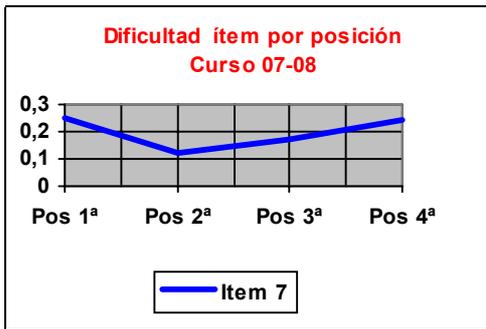


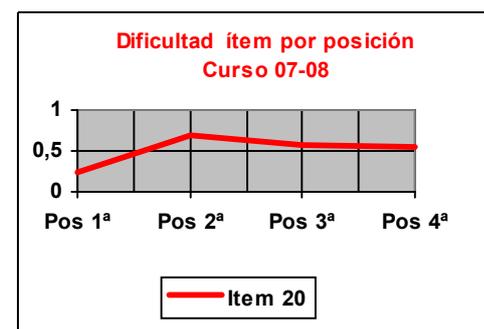
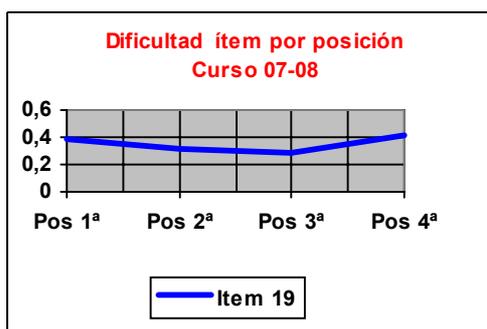
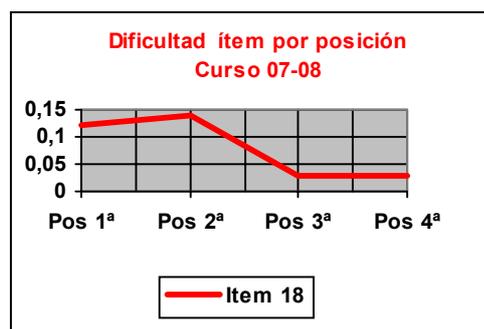
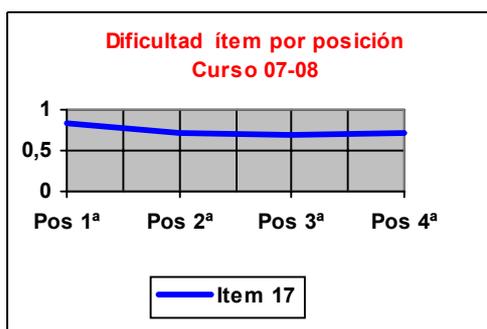
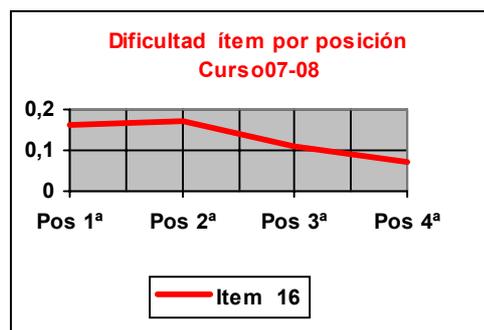
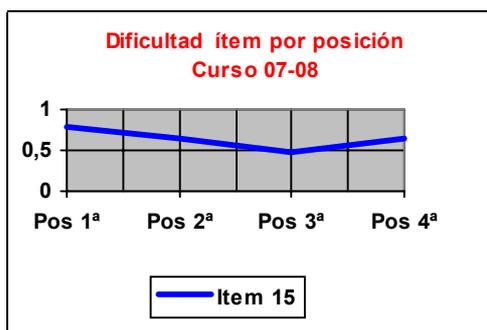
Tabla 4.- Análisis comparativo entre las notas medias de los tres cursos académicos, no apareciendo diferencias significativas entre ellos.

Asimismo, realizamos la representación gráfica de cada uno de los ítem que formaban parte de los exámenes valorados durante los tres cursos académicos con objeto de definir la pendiente de la misma que nos indicaría la dificultad del ítem según la posición que ocupara. Si la pendiente fuera negativa, en un porcentaje estadísticamente significativo, implicaría que el orden no influiría en la dificultad de los ítems.

Durante el curso 2007-2008, las gráficas de los 20 ítems que formaban el test fueron las siguientes:

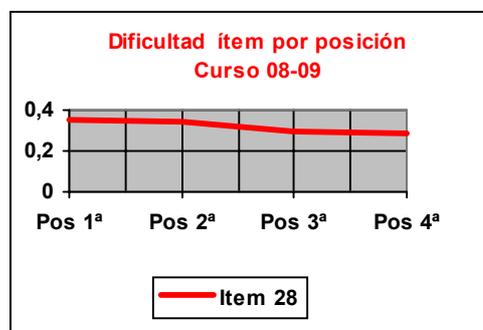
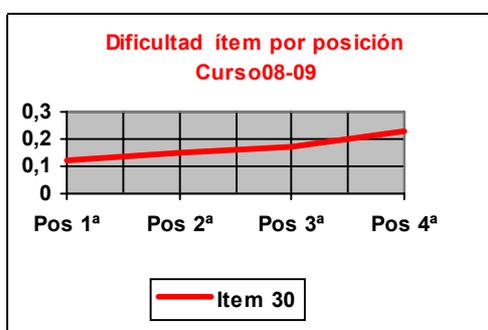




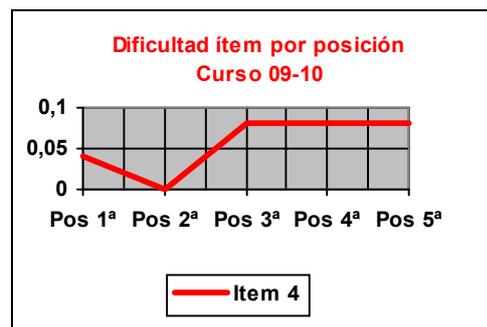
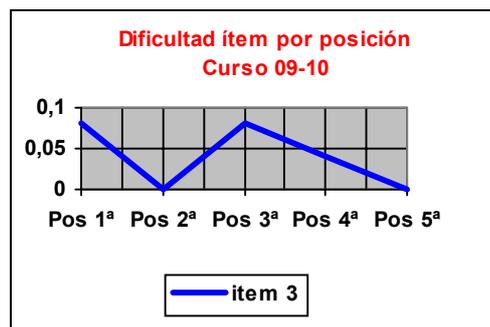
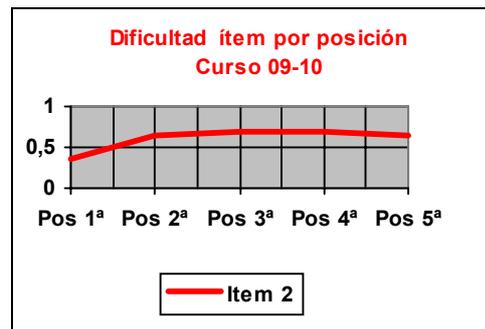
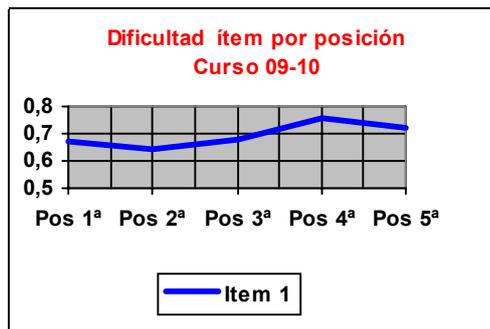


En este examen no encontramos ningún ítem que cumpla la hipótesis de partida, la pendiente nunca es mayor que cero en todos los tramos. El ítem que más se acercaría a la hipótesis de partida sería el ítem 10 de la primera copia del examen 07-08 cuya pendiente es mayor o igual que cero en todos los tramos, lo que implica que la dificultad del ítem se mantiene o va aumentando conforme el ítem se retrasa en la prueba. Además, nos encontramos con que el ítem 6 de la primera copia del mismo examen, mantiene, en todos los tramos, una gráfica con pendiente menor que cero lo cual nos llevaría a afirmar que la dificultad del ítem disminuye cuando aumenta su posición en el examen que es justo, lo contrario de la hipótesis de partida. El resto de los ítems presentan gráficas con pendiente cambiante de positiva a negativa o viceversa pasando por cero en más de una ocasión. Esto último, estaría indicando que la dificultad es la misma. La mayoría de las gráficas presentan una pendiente próxima a cero lo cual implica que su dificultades similar en todos los exámenes, lo cual no impide la existencia de picos que marcaría una dificultad muy diferente en alguno de los exámenes.

En el curso 08-09, el ítem 30 de la primera copia del examen dio una gráfica en la que todos los tramos la pendiente es mayor que cero lo que implica que la dificultad del ítem va aumentando conforme el ítem se retrasa en la prueba, pero, es el único ítem del examen 08-09 que cumple la hipótesis de partida. El ítem 11, presenta una pendiente igual a cero en el primer tramo y mayor que cero en los restantes, lo que supondría una aproximación a la hipótesis inicial. Además, nos encontramos con que el ítem 17 y el ítem 28 de la primera copia del mismo examen, mantienen una gráfica con pendiente menor que cero en todos los tramos lo cual, nos llevaría a afirmar, que la dificultad del ítem disminuye cuando aumenta su posición en el examen que es justo lo contrario que pasa con los ítems 11 y 30. El resto de los ítems presentan gráficas con pendiente cambiante de positiva a negativa o viceversa pasando por cero en más de una ocasión. Esto último, estaría indicando que la dificultad es la misma. Las gráficas adjuntas están dispuestas en el mismo orden en el que se nombran en el párrafo anterior.



En el examen del curso 09-10, no hemos encontrado ningún ítem cuya gráfica presente una pendiente positiva en todos sus tramos. Tampoco hemos encontrado ningún ítem que presente pendiente negativa en todos sus tramos. Por el contrario, la pendiente de las distintas gráficas no ha seguido ninguna pauta fija y, nos encontramos con gráficas en su mayoría con pendientes cambiantes de positivas a negativas o viceversa, gráficas en dientes de sierra, pasando, en más de una ocasión, por la pendiente cero que nos estará indicando que el ítem tiene la misma dificultad en distintas posiciones. Es decir, que en este examen ningún ítem satisfaría la hipótesis de partida. Como ejemplo se muestran las gráficas de los 4 primeros ítems.



En la asignatura *Principios, Instrumentación y Metodología en Biología Celular y Microbiología*, valoramos diferentes copias de un mismo examen con los mismos 20 ítems en posiciones diferentes que se distribuyeron entre 370. La diferencia entre las notas medias no fue estadísticamente significativa y la diferente posición de los ítems no afectó a la valoración final del examen. Por ello consideramos que los datos analizados en este único curso son consecuentes con los obtenidos en el análisis comparativo de tres cursos realizado en la asignatura de Lenguaje Informático.

CONCLUSIÓN: Según nuestra experiencia y como muestran los datos presentados, la posición del ítem en la prueba no influye en la dificultad del examen.

BIBLIOGRAFÍA

- Ariza, A. y colaboradores (2001). Evaluación y Medición: comparación de cuatro maneras de medición asistidas por ordenador un estudio estadístico. *Píxel-bit. Revista de Medios y Educación*. 16, 121-130.
- Ariza, A. (2006). *Exawin 5.1 Pruebas Objetivas de respuesta múltiple con soporte informático*. Ice. Universidad de Sevilla.
- Lord, F.M. (1980) *Applications of item response theory to practical testing problems*. Hisdale, Nueva York, academic Press.
- Neira T.R. y col. (1993) *Instrumentos de evaluación del aprendizaje*. Aula Abierta. (ICE Universidad de Oviedo).

Oncins de Frutos, Margarita: http://www.jmcprl.net/ntp/@datos/ntp_283.htm.
http://www.insht.es/InshtWeb/Contenidos/Documentacion/FichasTécnicas/NTP/Ficheros/201a300/ntp_283.pdf.

Raparaz, R. (2006): <http://www.unav.es/quimicayedafologia/Documento%20test.pdf>