

## **DISEÑO E IMPLANTACIÓN DE UN SISTEMA DE CONTROL Y MEJORA DE CALIDAD DE LAS PRUEBAS OBJETIVAS TIPO TEST: UN CASO EN LA LICENCIATURA DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS<sup>1</sup>**

*Victor G. Aguilar Escobar\**, *Pedro Garrido Vega\** y  
*Carmen Rocío Fernández Camacho\*\**

\* Dpto. Economía Financiera y Dirección de Operaciones,  
Facultad de Ciencias Económicas y Empresariales

\*\* Dpto. de Métodos de Investigación y Diagnóstico en Educación,  
Facultad de Ciencias de la Educación

Universidad de Sevilla

### **RESUMEN**

La evaluación constituye uno de los pilares básicos del proceso educativo. Sin embargo, es un aspecto al que en ocasiones no se le presta la suficiente atención, lo que afecta a la calidad de esta actividad docente. Este trabajo es fruto de la preocupación de los autores por esta cuestión y describe un proyecto de innovación docente llevado a cabo durante el curso 2005/06 en la asignatura de Dirección y Gestión de la Producción/Operaciones (I) de la licenciatura de Administración y Dirección de Empresas. El objetivo básico del mismo es el diseño e implantación de un sistema de control y mejora continua de la calidad de las pruebas objetivas (preguntas de elección múltiple) que venimos usando para evaluar la parte teórica de nuestra asignatura. Para ello, se han llevado a cabo diversas actividades que incluyen la creación de una completa base de datos de preguntas, el establecimiento de un sistema de indicadores de calidad y la elaboración de test online para el autoaprendizaje de los alumnos.

### **SUMMARY**

The evaluation constitutes one of the basic pillars of the educational process. Nevertheless, in some occasions the sufficient attention is not lent to it and, consequently, its quality can be affected by this. This paper reflects the authors concern about this question and describes a project of educational innovation carried out during academic year 2005/06 in the Production / Operations Management (I) course of the degree on Business Administration. The basic objective of the project is the design and implementation of a control and improvement system of the quality of the multiple choice tests that we are using to evaluate the theoretical part of our course. For it, diverse activities have been carried out, including the creation of a complete data base of questions, the establishment of a system of quality indicators and the elaboration of tests online for the self-learning of the students.

---

<sup>1</sup> Aguilar Escobar, V. G., Garrido-Vega, P., & Fernández Camacho, C. del R. (2008). Diseño e implantación de un sistema de control y mejora de calidad de las pruebas objetivas tipo test: Un caso en la Licenciatura de administración y dirección de empresas, En *Experiencia de Innovación Universitaria: Curso 2005/2006*, Universidad de Sevilla, Vol. 2, Págs. 9-26.



## INTRODUCCIÓN

El proceso de enseñanza consta de tres fases fundamentales: la planificación y programación de la enseñanza, la realización de las actividades docentes y la evaluación del proceso de enseñanza-aprendizaje. La evaluación, por tanto, es una actividad sistemática integrada en el proceso educativo y que forma parte necesaria de cualquier proceso de enseñanza-aprendizaje.

La finalidad de la evaluación es servir de diagnóstico, no sólo del rendimiento del alumno, sino también del grado de cumplimiento de los objetivos y de la validez de la técnica y el método de enseñanza. Indirectamente, es un indicador del rendimiento del profesor. En consecuencia, la mejora de la calidad de la enseñanza pasa también por la mejora del sistema global de evaluación. De hecho, la evaluación está muy ligada a la calidad, hasta el punto de ser una de las cuatro actividades que conforman del círculo de calidad de Deming (planear, ejecutar, evaluar y ajustar). La evaluación cobra pleno sentido dentro de este círculo, no si se la considera de forma independiente y aislada del resto de la actividad educativa. En este último caso, la evaluación sólo repercute sobre el alumno y no sobre el propio proceso de enseñanza-aprendizaje. Lamentablemente esta es la situación en muchos casos dentro de la universidad. El objetivo no ha de ser evaluar por evaluar, para cumplir con el requisito de otorgar una calificación a los alumnos al final del curso, sino también para comprobar la eficacia de la acción docente y contribuir al perfeccionamiento de la misma.

En nuestra asignatura de Dirección y Gestión de la Producción / Operaciones I, (DGPO-I), como en muchas otras de la Licenciatura de Administración y Dirección de Empresas, LADE, se vienen usando desde hace ya bastantes años las pruebas objetivas de respuesta múltiple (pruebas tipo test) para la evaluación de la parte teórica del contenido docente. No perdemos de vista que la evaluación es un proceso sistemático que implica tres fases (Zabalza, 2001): recogida de información, valoración de la misma y toma de decisiones. Y, por tanto, evaluar no es medir, sino que ésta forma parte de aquélla y constituye sólo un medio. En este sentido, las pruebas de elección múltiple (PEM) son un instrumento más de medición y obtención de datos. En nuestro caso, son utilizadas para obtener información sobre el nivel de conocimientos teóricos de los alumnos. Esta parte pondera el 40% sobre la calificación global, correspondiendo el 60% restante a la parte práctica que valoramos a través de varios supuestos (normalmente tres).

La evaluación debe cumplir cuatro condiciones básicas de acuerdo con Marcelo (2001): ser útil para los individuos implicados (profesores y alumnos), factible (que pueda llevarse a cabo sin dificultad), ética y exacta. Las pruebas objetivas PEM cumplen con creces estos criterios. Pueden ser útiles en el sentido de ayudar a identificar los puntos fuertes y débiles, tanto para los profesores en su actividad docente como para los alumnos en su aprendizaje. Resultan de las más factibles, especialmente en condiciones de masificación. Y, también resultan éticas y exactas, por la eliminación de la subjetividad y la igualdad de trato que aseguran en la calificación de los alumnos.

Las pruebas objetivas tipo test (preguntas de elección múltiple PEM) constituyen un instrumento de evaluación ampliamente utilizado a nivel universitario. Existen otros tipos de pruebas objetivas como las de emparejamiento, las de verdadero-falso, de elección múltiple con varias respuestas correctas, de ordenación, de clasificación, etc. No conviene sobreestimar ni

infravalorar este tipo de pruebas ya que cuentan con numerosas posibilidades y ventajas pero, al mismo tiempo, también importantes limitaciones y desventajas.

Como es conocido, los exámenes tipo test presentan como ventaja su carácter objetivo, puesto que la puntuación no depende del profesor que corrige, ni está sujeta a interpretaciones subjetivas. También permiten realizar pruebas más exhaustivas que otros tipos, ya que pueden introducirse preguntas que exploren todos o muchos de los temas y contenidos de una asignatura. Esto aumenta el grado de validez de la prueba frente a otros tipos. Al no exigir al alumno que se detenga en la escritura, sino en la lectura que es una actividad mucho más rápida, facilita que se centre en el desarrollo del pensamiento. Una ventaja más es la facilidad con la que se corrigen y, por tanto, la rapidez con la que pueden ofrecerse los resultados a los alumnos. También para los alumnos son más fáciles de realizar, ya que sólo requieren marcar la opción correcta y no han de estructurar y desarrollar ideas sobre un tema, para luego redactar la respuesta con sus propias palabras. Por último, como apunta Gómez de Terreros (1998), este tipo de pruebas permiten valorar el conocimiento, la comprensión, la aplicación, el análisis, la síntesis y otros niveles pedagógicos necesarios para estimar el grado de aprendizaje en ámbito cognitivo.

No obstante, son también muy considerables sus inconvenientes y limitaciones. Así, entre las limitaciones de este tipo de pruebas podemos señalar que no permiten la libertad de expresión, ni el pensamiento divergente por parte de los alumnos. De la misma forma, no permiten valorar aspectos como “la originalidad, la expresión verbal, la escritura y la capacidad de reacción ante problemas reales” (Gómez de Terreros, 1998, pág. 114). Por otro lado, no ofrecen información de la causa de la respuesta errónea, ni del proceso de pensamiento que lleva al sujeto a su elección. También habría que mencionar que hay determinados contenidos que se prestan mal a ser estructurados en forma de preguntas de este tipo. Además, este tipo de pruebas son más proclives a prácticas no éticas durante el examen por parte de los alumnos. También requieren mayor esfuerzo por parte de los profesores para su elaboración, ya que son relativamente complejas.

## **JUSTIFICACIÓN Y OBJETIVOS**

En nuestra asignatura, hemos percibido como los resultados han ido mejorando con la experiencia, pero aún persiste una cierta dificultad en la confección de los exámenes tipo test y en las puntuaciones obtenidas. Adicionalmente, en la medida que las preguntas de respuesta múltiple pasan a formar parte de un banco de preguntas que son utilizadas en sucesivas convocatorias, los errores que estas contengan o la falta de capacidad discriminatoria y evaluativa de las mismas, exigen el establecimiento de un sistema que controle la calidad de cada ítem y permita mejorar de forma continua la calidad del banco de preguntas y, por extensión, de los exámenes teóricos.

Por estas razones, nos planteamos establecer un sistema global de mejora y control de calidad para este tipo de pruebas que permitiera cumplir más fielmente los objetivos de los exámenes. Con ello, esperábamos conseguir pruebas más fiables y discriminatorias, además de favorecer el aprendizaje de los alumnos.

A la vista de la situación anterior, hemos realizado durante el curso académico 2005-2006 un proyecto de innovación educativa financiado por el Instituto de Ciencias de la Educación (ICE) de la Universidad de Sevilla consistente en el diseño e implantación de un sistema de control de

calidad para las pruebas objetivas tipo test de la asignatura DGPO-I de 4º curso de LADE. Los objetivos planteados han sido los siguientes:

1. Mejorar la evaluación de los alumnos en lo que se refiere a la parte teórica de la asignatura.
2. Dar seguridad y fiabilidad a la prueba, tanto desde el punto de vista de los alumnos como del profesor.
3. Mejorar la capacidad de los alumnos para afrontar las pruebas tipo test de la asignatura.
4. Transmitir nuestra experiencia a otros profesores interesados en la mejora de sus pruebas objetivas.

## METODOLOGÍA

Para registrar, ordenar, relacionar, analizar y utilizar la información relativa a las preguntas y los exámenes de Dirección de Operaciones, se ha diseñado y desarrollado una base de datos relacional mediante Microsoft Access. El diseño se ha basado en el análisis del sistema a través del modelo entidad-relación. El diseño ha usado toda la potencialidad de Microsoft Access incluyendo la utilización del lenguaje de interrogación SQL, diseño de macros y lenguaje de programación Visual Basic. En la Figura 1 se observa el formulario principal para la introducción de preguntas en la base de datos

Examen	Idpregunta	Orden tipo A:	Orden tipo B:	Bien:	Mal:	No coate	Total	Índice de dificultad	Índice de discriminación 27%	A:	B:	C:	D:	Índice de resalto
Obligatoria Facultad Diciembre 1999/2000	80	5	0											
Obligatoria Facultad Diciembre 2005/2006	9	11	16	5	3	24	61,5%	37,5%	5	16	0	0	25,3%	
Obligatoria Facultad Primer Parcial 2005/2006	16	0	19	120	19	24	163	70,7%	-5,8%	14	120	2	3	15,0%

Figura 1. Formulario “Preguntas” de la base de datos de preguntas de DGPO-I

Las respuestas obtenidas en los exámenes han sido tratadas con diferentes programas (SPSS, METRIX y EXCEL, fundamentalmente) para valorar estadísticamente la calidad de los ítems, la fiabilidad de las pruebas y obtener al final un banco de preguntas convenientemente testadas.

Para la realización de los ejemplos de auto-evaluación se ha utilizado un software específico, “Hot-Potatoes”, y se han publicado en la página web de la asignatura a través del programa “Front-Page”.

## ACTIVIDADES REALIZADAS

Hemos dividido nuestro proyecto en 4 grupos de actividades:

### 1. DEFINICIÓN DE UNA PROPUESTA TEÓRICA SOBRE LOS EXÁMENES TIPO TEST DE NUESTRA ASIGNATURA.

Este grupo de actividades comprende:

- a) Revisión bibliográfica (Teoría Clásica de los Test, Psicometría, Didáctica universitaria, etc.)
- b) Definición de una propuesta teórica sobre:
  - las características que deben tener las preguntas de elección múltiple,
  - los errores que se deben evitar,
  - los indicadores de calidad adecuados
  - las características del examen en su conjunto (número de preguntas necesario para otorgar fiabilidad al test, nº de opciones por pregunta, fórmula de valoración de la prueba, etc..).

Las decisiones tomadas sobre estos aspectos han sido las siguientes. Por un lado, hemos elevado el número de preguntas de nuestros test a 30 (en los últimos años usábamos sólo 20) aumentando con ello la fiabilidad y validez de las pruebas. Por otro lado, la distribución de preguntas por temas, e incluso por apartados dentro de los temas, se hace de acuerdo con la importancia asignada a los mismos en las clases. Los ítems se ordenan de acuerdo al orden seguido en la impartición de los temas. Cuidamos la apariencia de los test para que resulten claros y cómodos de leer. En la redacción de las preguntas hemos tratado de seguir una serie de recomendaciones generales que hemos identificado a partir de la revisión bibliográfica:

- Deben ser claras, comprensibles y adaptadas a los alumnos a los que va dirigido el test.
- Deben referirse a contenidos relevantes y estar ligadas a lo enseñado en las clases.
- Cada pregunta debe centrarse en un sólo aspecto temático (concepto, técnica, teoría, etc.). Se deben evitar las preguntas demasiado generales que mezclan varios aspectos temáticos distintos.

- Deben evitarse preguntas concatenadas, cada una debe ser independiente y no debe dar información que permita responder correctamente otra distinta.
- Debe usarse un lenguaje claro, huyendo de la ambigüedad y la complejidad.
- Deben evitarse los juicios de valor.
- Se han de evitar los errores gramaticales y ortográficos y las abreviaturas.
- Conviene minimizar el tiempo de lectura, buscando la mayor concisión posible y eliminando información irrelevante. Esto implica también no utilizar más palabras de las necesarias y evitar adjetivos y adverbios superfluos.
- No se deben incluir preguntas que puedan entenderse como “preguntas con trampa”, es decir, aquéllas con intención de confundir a quien lo responde, con contenidos irrelevantes, con opciones respuestas muy parecidas, que tengan varias respuestas correctas, etc.

Además, existen otros criterios a tener en cuenta más específicos en relación con los enunciados, las opciones de respuesta, la respuesta correcta y las alternativas incorrectas o distractores:

#### Enunciado

- Debe ser claro y preciso, conteniendo la idea central por la que se pregunta y toda la información necesaria para responder.
- Se debe escribir en formato de frase completa mejor que incompleta pero si se hace de esta última forma, no se deben dejar espacios en blanco ni al comienzo ni en medio de la frase.
- Se debe formular el enunciado preferiblemente en forma afirmativa. Si se usan formas negativas, las partículas de negación deben ir en negrita o mayúsculas.
- No ha de darse ninguna pista semántica o gramatical que permita identificar la opción correcta.

#### Respuestas

- El número de respuestas alternativas debe ser idealmente tres. En todo caso, hay que tener en cuenta que este número tiene consecuencias sobre la dificultad, el tiempo requerido y la fiabilidad de la prueba.
- Intentar que todas las opciones tengan cierta homogeneidad y una longitud parecida.
- No incluir unas opciones en otras.
- Evitar el uso de expresiones del tipo “Ninguna de las opciones es correcta” o “Todas las otras opciones son correctas”.
- No dar pistas sobre la opción correcta como, por ejemplo, incluir expresiones del enunciado en la opción correcta o usar opciones incorrectas que resulten absurdas.

- Evitar el uso de calificativos absolutos como siempre, nunca, totalmente, etc., ya que las opciones que las contienen suelen ser incorrectas y su presencia ayuda a descartarlas.
- Cuando las opciones son números, deben aparecer ordenadas de mayor a menor o viceversa. De igual forma si son fechas, deben ser ordenadas cronológicamente. Si son intervalos numéricos, es conveniente que no se solapen entre sí. Siempre debe indicarse la unidad de medida.

#### Respuesta correcta

- Sólo una alternativa correcta.
- La posición de la respuesta correcta debe ir alternando a lo largo del test.
- Su longitud debe ser parecida a la de las otras. A veces, por redactarla con mayor precisión que las demás, se ofrece una pista indeseada.

#### Respuestas incorrectas o distractores

- Usar distractores coherentes y verosímiles, esto implica, que deben parecer correctos para los que saben menos, pero han de ser descartados por los que tienen más conocimientos.
- Para crear un buen distractor podemos usar los errores más comunes de los estudiantes.
- También puede ser útil usar expresiones técnicas propias de la materia que se evalúa y que puedan parecer correctas a aquellos alumnos que no han estudiado suficiente.
- Otra opción para elaborar buenos distractores es emplear afirmaciones que sean verdaderas, pero que no respondan a lo que se les pide.
- Evitar alternativas graciosas puesto que se descartan automáticamente.

## 2. CREACIÓN UNA BASE DE DATOS DE PREGUNTAS TEST PARA LA ASIGNATURA lo que a su vez ha comprendido:

- a) Diseño de la base de datos.
- b) Introducción de ítems (preguntas) en la base de datos a partir de las pruebas de años anteriores. Disponemos a Julio de 2006 de 393 ítems lo que parece un número suficiente para poder seleccionar las 30 preguntas necesarias para cada examen.

## 3. ESTABLECIMIENTO DE UN SISTEMA DE CONTROL DE CALIDAD EN LA EVALUACIÓN DE LAS PRUEBAS TEÓRICAS TIPO TEST A LAS QUE SOMETEMOS A NUESTROS ALUMNOS.

Este sistema ha consistido en:

- a) Selección de ítems por parte de uno o varios profesores utilizando el banco de preguntas de convocatorias anteriores o confeccionando preguntas nuevas. La selección se hace de acuerdo a unos criterios objetivos que se explican más adelante.

Como se ha comentado, se ha buscado una distribución equilibrada de las preguntas por temas del programa. Asimismo, las preguntas de cada tema se han referido a distintos apartados del mismo.

- b) Revisión colegiada por parte del conjunto de profesores para evitar los errores de formulación más frecuentes (doble negación, opciones más correctas, etc.), refinar los enunciados de las preguntas y de sus respuestas alternativas, evitación de duplicidades, etc.
- c) Realización de la prueba.
- d) Corrección de la prueba mediante lectora óptica de marcas facilitada por el ICE. Esto ha implicado el diseño y adquisición de las hojas de respuesta y la necesidad de adiestramiento a los alumnos para su correcta confección. Curiosamente, esta última actividad ha implicado ciertas dificultades pero, con la colaboración de los alumnos, éstas han ido disminuyendo.
- e) Obtención de resultados de control de calidad. En ellos se pueden distinguir entre *indicadores individuales* de cada ítem (índice de dificultad, índice de discriminación, índice de homogeneidad e índice de reparto) e *indicadores globales* para el examen (validez, fiabilidad, dificultad promedio, discriminación promedio, y, como medidas de la forma de la distribución, la asimetría y la curtosis). De esta manera, tratamos de resumir el comportamiento y la calidad psicométrica de los ítems en cada una de las pruebas analizadas. Los principales conceptos utilizados para las pruebas en su conjunto han sido los siguientes:

- *Validez*: hace referencia a la coherencia entre lo que se consigue con la prueba y lo que se desea obtener. Refleja, por tanto, el grado en que la prueba cumple con lo que se pretende y mide lo que se quiere medir. Se suele distinguir entre la validez interna y externa. Las técnicas que suelen utilizarse para medirla son la validez aparente, la validez de contenido, la validez concurrente y la validez predictiva. En nuestro caso, el elemento fundamental que otorga validez a la prueba es la confección cuidadosa de las preguntas y la revisión colegiada de los exámenes.
- *Fiabilidad*: se refiere al grado de consistencia o precisión de la medición. Una prueba fiable sería aquella que ofrece el mismo valor de la medición cuando se dan las mismas condiciones. La fiabilidad, por tanto, está relacionada con la estabilidad, repetibilidad y precisión de una prueba. Para calcular la fiabilidad pueden emplearse técnicas estadísticas como el test-retest o el coeficiente de Cronbach, entre otras. En nuestro caso hemos elegido esta última.

Entre los indicadores individuales, los más utilizados son el índice de dificultad y el índice de discriminación. También es posible utilizar otros, como el índice de homogeneidad o el de reparto.

- *Índice de dificultad*: pretende determinar en qué medida una pregunta es fácil o difícil. Se mide a partir del número de aciertos de los evaluados. A mayor índice más fácil es la pregunta. Como criterio de selección exigiremos a las preguntas a seleccionar el que su índice de dificultad

(realmente es de facilidad) sea mayor que 0,1 y menor que 0,9 para evitar las preguntas muy difíciles o muy fáciles.

- *Índice de discriminación*: permite asegurarnos de que la pregunta es acertada más por aquellos individuos que tienen mayores niveles de conocimientos. A mayor índice más discriminatoria es la pregunta. Consideramos validas, de acuerdo a este siguiendo un criterio, los ítems con índices de discriminación mayores de 0,15. Para este índice consideramos apropiada la escala ofrecida en Gómez de Terreros (1998):
  - 0,35 o más = Pregunta excelente.
  - 0,25 a 0,34 = Pregunta buena.
  - 0,15 a 0,24 = Pregunta límite (para revisar)
  - Menos de 0,15 = Pregunta mala. Debe ser eliminada o reexaminada.
- *Índice de homogeneidad*: establece la coherencia de cada ítem con el conjunto de la prueba.
- *Índice de reparto*: Mide el grado en que las respuestas de los evaluados se distribuyen entre las opciones de un ítem. Este índice toma valor 1 si todas las opciones han sido elegidas por el mismo número de alumnos.

f) Introducción de los datos anteriores en la base para la selección futura de preguntas.

4. ELABORACIÓN DE TEST TIPO PARA EL AUTOAPRENDIZAJE POR PARTE DE LOS ALUMNOS. Consideramos que ésta es una de las actividades que más pueden ayudar a los alumnos a mejorar su rendimiento en la parte teórica. Para ello, al final de cada tema o bloque temático de la asignatura, hemos dedicado algún tiempo en clase a realizar ejercicios test. Los alumnos reflexionan sobre las distintas preguntas durante unos minutos y a continuación discutimos sobre cada una de ellas, señalando las razones por las que cada opción es cierta o falsa. Ello les sirve a los alumnos para repasar la teoría y para familiarizarse con el tipo de preguntas que se encontrarán más tarde en los exámenes de la asignatura. Para conseguir mejores resultados de esta actividad, es conveniente hacer una buena selección de las preguntas que se trabajan en clase. A partir de este proyecto, nuestra intención es que esta selección se haga también teniendo en cuenta los datos de dificultad y discriminación que obtengamos de nuestro estudio de las preguntas de examen. Como complemento a esta actividad, en el marco de este proyecto de innovación docente, hemos procedido a elaborar preguntas test para el autoaprendizaje del alumno. Dos de los profesores participantes en este proyecto siguieron un curso de Hot Potatoes, con el objeto de conocer en profundidad esta aplicación informática que permite elaborar distintas modalidades de pruebas para el autoaprendizaje. Como resultado de ello, este curso hemos añadido a la página web de nuestra asignatura ([www.us.es/adgpo1](http://www.us.es/adgpo1)) conjuntos de ejercicios de autoevaluación para dos temas concretos del programa, que incluyen pruebas de tipo test, de rellenar huecos, de emparejamientos, de ordenación de frases o palabras y crucigramas. En el futuro, pretendemos extender esto al resto de los temas del programa.

5. DIFUSIÓN DE NUESTRA EXPERIENCIA a los miembros de nuestro departamento y otros profesores interesados. En este último caso, se han mantenido reuniones con miembros de otros grupos con actividades de innovación en la que se ha expuesto nuestro proyecto.

### ALGUNOS RESULTADOS DEL CONTROL DE CALIDAD DE LOS EXAMENES DEL CURSO 2005-2006

Presentaremos a continuación algunos indicadores de la calidad de los exámenes en su conjunto obtenidos en el curso académico citado (hasta junio) y para los aspectos comentados anteriormente, dejando para aportaciones futuros un análisis más pormenorizado por ítems particulares.

Examen:	Nº de alumnos	Índice de dificultad medio	Índice de discriminación medio	% ítems con índice dificultad adecuado	% ítems con suficiente índice discriminación	% de preguntas a revisar	% de preguntas excelentes
<b>3ª Conv. (Diciembre)</b>	24	48,2%	0,278	90%	73,3%	26,7%	36,7%
<b>1er Parcial</b>	367	66,8%	0,347	96%	96%	8%	52%
<b>2º Parcial</b>	293	55,9%	0,292	100%	83,3%	16,7%	33,3%
<b>1ª Conv. (Junio)</b>	169	56,6%	0,255	86,7%	73,3%	26,7%	26,7%

**Tabla 1 Resultados de las pruebas realizadas durante el curso académico 2005-2006.**

En esta tabla, el índice de dificultad se ha calculado dividiendo los aciertos entre el total de sujetos que han realizado la prueba y multiplicándolos por cien. El índice de discriminación para cada pregunta se ha obtenido por diferencia entre la puntuación obtenida por el grupo fuerte (27% de alumnos con mejores puntuaciones) y la puntuación obtenida por del grupo débil (27% de alumnos con peores puntuaciones). La fórmula correspondiente es:

$$\text{Índice de Discriminación} = (F-D)/N$$

En nuestro caso F = (Aciertos Grupo Fuerte – Fallos Grupo Fuerte por 0,25)/Número de alumnos de cada grupo, D = (Aciertos Grupo Débil– Fallos Grupo Débil por 0,25)/Número de alumnos de cada grupo.

Como se desprende de los resultados expuestos en la tabla anterior, podemos decir que los valores de dificultad están relacionados con el tipo de examen siendo mejores los resultados obtenidos en los parciales que en las convocatorias de junio y diciembre. Esto está claramente relacionado con la diferente tipología de los alumnos de unas y otras convocatorias ya que los alumnos con mejores resultados aprueban la asignatura por parciales, acudiendo a Junio, Septiembre o Diciembre los alumnos que no han podido aprobar por curso. Por este motivo, los

exámenes parciales son más discriminatorios que las convocatorias finales.

En la tabla hemos añadido dos columnas que indican, respectivamente, el porcentaje de preguntas con adecuado índice de dificultad (entre 10% y el 90%) y el porcentaje de preguntas con suficiente índice de discriminación (mayor que 0,15). Nuevamente nos encontramos valores mayores en los parciales que, como hemos aclarado, son los exámenes más representativos. En nuestro sistema de control de calidad las preguntas que no cumplan cualquiera de los dos criterios señalados anteriormente son objeto de revisión para su rechazo o reformulación. En el examen de diciembre, por ejemplo, hay 8 preguntas con insuficiente índice de discriminación y 2 preguntas que tienen inadecuado índice de dificultad. Estas últimas tienen también insuficiente índice de dificultad por lo que en total tendríamos 8 preguntas a revisar, lo que supone un 26,7% del examen. En el primer parcial, donde el análisis se limita a 25 preguntas comunes a todos los grupos y no a las 30 habituales, nos encontramos con una pregunta con inadecuado índice de dificultad y otra con inadecuado índice de discriminación lo que nos lleva a 2 preguntas a rechazar o reformular radicalmente. En cualquier caso, el valor de estos índices y la decisión correspondiente, dependerán del tipo de examen y del número de alumnos presentados.

Por último las preguntas excelentes son aquellas con índice de discriminación superior a 0,35 siempre que se cumpla la condición de que el índice de dificultad superior al 10% e inferior al 90%. Nuestro objetivo, en el futuro, sería que todas las preguntas fuesen de esta naturaleza.

## **CONSIDERACIONES FINALES**

El proyecto de innovación desarrollado ha supuesto poner en marcha un sistema de mejora continua de la calidad de las pruebas teóricas de nuestra asignatura. Este sistema contribuye por tanto a la calidad del proceso global docente universitario pues afecta a un área significativa del mismo como es la evaluación. Aun cuando las pruebas teóricas ponderan sólo el 40% de la nota de nuestros exámenes, nos parece importante mejorar la calidad de las mismas, más si cabe en una asignatura en la que explicamos a nuestros alumnos como gestionar la calidad de las organizaciones. Ya en este primer curso de desarrollo e implantación, hemos apreciado ventajas significativas. Así, por un lado, hemos observado una reducción de la incertidumbre de los alumnos sobre este tipo de examen lo que ha redundado, entre otras causas, en el aumento significativo del porcentaje de alumnos presentados (7,2 puntos porcentuales respecto al curso anterior y 9,8 puntos respecto al 2003-2004). También se han constatado resultados más positivos entre los alumnos presentados habiendo superado la asignatura más del 88% de ellos. Desde el punto de vista del profesor, hemos reducido significativamente el tiempo de confección de los exámenes gracias a la base de datos desarrollada y también se ha reducido el tiempo de corrección lo que, junto a una mayor información, ha permitido centrarse en lo que debería ser el trabajo del docente: la mejora de la planificación del examen, la elaboración y selección de ítems y el control de calidad del examen y de cada uno de los ítems utilizados. Con ello, hemos elevado la calidad de nuestros exámenes. Como docentes, nos tenemos que convencer que, de la misma forma que una empresa intenta evitar que un producto defectuoso llegue al cliente, no deberíamos dejar que los ítems defectuosos lleguen a los alumnos y, mucho menos, que este error se repitiera una y otra vez. La "Teoría de los Test" unida a las herramientas informáticas nos permite hoy en día obtener una información que retroalimenta el sistema de control de calidad. Sin embargo, la validez de las pruebas precisa de un sistema global de mejora continua en la que los sistemas de revisión

cualitativa adquieren una gran importancia. Por último, hay que ser conscientes de que, como todo proceso de mejora continua, los resultados del mismo van más allá de un curso académico, siendo su utilidad real dependiente de la constancia y el interés de los sucesivos profesores en mantener y mejorar este sistema.

## **BIBLIOGRAFÍA**

- ARIZA GARCÍA, A.; DAZA NAVARRO, P. y TORREBLANCA LÓPEZ, J. (2001). Evaluación y medición: comparación de cuatro maneras de medición asistidas por ordenador. Un estudio estadístico. **Pixel-Bit, Revista de medios y educación**, 16. En [www.sav.us.es/pixelbit/](http://www.sav.us.es/pixelbit/), consultada el 09/09/2005.
- ARCE FERNÁNDEZ, C. (1994). **Técnicas de construcción de escalas psicológicas**. Madrid, Síntesis.
- GARCÍA CUETO, E. (1991). **Introducción a la Psicometría**. Madrid. Siglo XXI.
- GÓMEZ DE TERREROS SÁNCHEZ, I. (1998). Análisis evaluativo de calidad de la prueba objetiva tipo test (preguntas de elección múltiple). **Revista de enseñanza universitaria**, 13, 105-115.
- LORD, F.M. y NOVICK, M.R. (1968). **Statistical theories of mental test scores**. Reading, MA. Addison Wesley.
- MARTINEZ ARIAS, M.R. (1995). **Psicometría: Teoría de los tests psicológicos y educativos**. Madrid. Síntesis.
- MUÑIZ FERNÁNDEZ, J. (coord.) (1996). **Psicometría**. Madrid. Universitas.
- MUÑIZ FERNÁNDEZ, J. (1998). **Teoría Clásica de los Tests**. Madrid. Pirámide.
- NUNNALLY, J. C. y BERNSTEIN, J. (1995). **Teoría psicométrica**. México D.F. McGraw-Hill.
- RENOM PINSACH, J. (1992). **Diseño de tests**. Barcelona. Engine.
- RENOM PINSACH, J. (coord.) (1997). **Tratamiento informatizado de datos**. Barcelona. Masson.
- SANTISTEBAN REQUENA, C. (1990). **Psicometría: Teoría y Práctica en la construcción de los tests**. Madrid. Norma.
- TORREBLANCA LÓPEZ, J.; SANCHO MARTÍNEZ, M. y ARIZA GARCÍA, A. (1996). La utilización de bases de datos como herramienta de evaluación. **Pixel-Bit, Revista de medios y educación**, 7. En [www.sav.us.es/pixelbit/](http://www.sav.us.es/pixelbit/), consultada el 09/09/2005.
- YELA GRANIZO, M. (1980). **Introducción a la Teoría de los Tests**. Madrid. Facultad de Psicología, Universidad Complutense.