



**Carlos Maté Jiménez**

Profesor propio de la ETSI (ICAI) de la Universidad Pontificia Comillas de Madrid. Doctor en Ciencias Matemáticas. Ha escrito varios libros sobre estadística y publicado diversos artículos sobre métodos estadísticos aplicados.

## Sistema automatizado para la mejora del rendimiento en pruebas objetivas. Implicaciones en la selección de recursos humanos



**José María Labernia Salvador**

Ingeniero Informático del ICAI. Actualmente trabaja en el área de Enterprise Risk Services (ERS) de Deloitte.

La evaluación del aprendizaje basada en pruebas objetivas o tipo test ha sido ampliamente estudiada a lo largo del siglo XX desde diferentes perspectivas.

La evolución de las TIC en los últimos 10 años proporciona en la actualidad un marco idóneo para el desarrollo de herramientas automatizadas, a través de redes informáticas, que permitan poner a disposición de los usuarios de estos exámenes (alumnos y profesores) nuevos sistemas de ayuda, con alto grado de interactividad. El objetivo de dichos sistemas es orientar a los usuarios sobre la evolución del aprendizaje en un área de conocimiento y, como consecuencia de ello, mejorar el rendimiento en dichos exámenes.

En este contexto se enmarca SAAPET (Sistema Avanzado de Apoyo a la Preparación de Exámenes Tipo Test). Sus características y posibilidades futuras en la selección de recursos humanos son presentadas en este trabajo.

Comentarios a: [comentarios@icai.es](mailto:comentarios@icai.es)

## Algunas cuestiones teóricas sobre las pruebas tipo test

Se considera que una prueba de conocimientos o habilidades se puede denominar “objetiva” o “tipo test” cuando está formada por preguntas de las llamadas de elección múltiple. Para responder a una pregunta se ha de seleccionar una o varias de las alternativas que se presentan. La mayor parte de las cuestiones relativas a las pruebas objetivas se pueden consultar en Morales (2000). Dichas pruebas son habituales en las universidades, las empresas o como parte del proceso de obtención del carné de conducir. Como ejemplo del interés creciente en todo lo que rodea a estas pruebas se encuentra el trabajo de Simkin y Kuechler (2005), que incluye una completa revisión de la literatura y analiza la conexión entre estas pruebas y la comprensión del estudiante.

En lo que sigue, consideraremos que los parámetros que controlan las personas que diseñan la prueba son los siguientes.

- **Número total de preguntas del test.** Se notará por  $N$ , siendo  $N > 1$ .
- **Número de alternativas en cada pregunta.** Se notará por  $r$ , siendo  $r > 1$ . Diferentes estudios coinciden en proponer preguntas con  $r = 3$ .
- **Incorporación de la corrección por adivinación.** Consiste en la posibilidad de incluir en la calificación de la prueba una corrección para reducir el efecto de acertar por azar (por ejemplo, con cuatro alternativas se puede responder lanzando dos monedas y eligiendo cada alternativa en función de los 4 resultados del experimento).

Un análisis detallado sobre los parámetros anteriores se puede encontrar en Morales (2000) y en todas las referencias allí citadas.

## Modelado de los parámetros de una prueba de elección múltiple desde la perspectiva del evaluado

En todo lo que sigue se supone que las pruebas consideradas están formadas por preguntas con una única respuesta. Para modelar estas pruebas objeto de estudio, desde la perspectiva del usuario que se evalúa, se utiliza el modelo de probabilidad de la distribución binomial, basado en el concepto de pruebas de Bernoulli. Se llama prueba de Bernoulli a todo experimento aleatorio (responder a una pregunta de

test) con dos resultados, llamados **éxito** (acertar la respuesta correcta en una pregunta) y **fracaso** (fallar o dejar en blanco una pregunta), que se notarán por  $A$  y  $\bar{A}$ , respectivamente, y que acontecen con probabilidades  $p$  y  $q$ , siendo  $p + q = 1$ .

Para cuantificar estas pruebas y extenderlas a una serie de repeticiones del experimento (un grupo de preguntas de un test), es necesario formalizar la definición de la correspondiente variable aleatoria (v.a. de ahora en adelante). Como el interés al analizar una prueba de Bernoulli se centra en el suceso éxito, la v.a. asignará el valor 1 al éxito y el 0 al fracaso. Sea dicha v.a.  $X$ . Se tiene que la función de masa de probabilidad de esta v.a.  $X$  será  $P(X = x) = p^x(1 - p)^{1-x}$ ,  $x = 0, 1$ .

Como esta v.a. se refiere a pruebas de Bernoulli, se conoce con el nombre de **distribución de Bernoulli** de parámetro  $p$ . Por ejemplo, si una persona habitualmente deja sin contestar el 20% de las preguntas de un test, para modelar su comportamiento de no responder a una pregunta consideramos una distribución de Bernoulli de parámetro  $p = 0,2$ .

Una generalización inmediata y fundamental de la distribución de Bernoulli surge cuando un experimento aleatorio con dos posibles resultados se repite varias veces. Supongamos pruebas de Bernoulli como las anteriores donde la probabilidad de éxito en cada prueba es  $p$  y se desarrollan  $n$  pruebas independientes (se responde a  $n$  preguntas en un examen tipo test), de manera que el resultado de cada prueba (pregunta) no tiene influencia sobre el resultado de cualquier otra prueba (pregunta).

La v.a.  $X$  –número de éxitos que aparecen al desarrollar estas  $n$  pruebas– podría tomar cualquier valor desde 0 hasta  $n$ . Nuestro interés se centra en calcular la probabilidad de obtener exactamente  $X = x$  éxitos en  $n$  pruebas. Se puede demostrar que dicha probabilidad es

$$P(X = x) = \binom{n}{x} p^x(1 - p)^{n-x}, x = 0, 1, \dots, n$$

se dirá que  $X$  sigue una ley binomial de parámetros  $n$  y  $p$ , notándose como  $X \approx B(n, p)$ .

Por ejemplo, si una persona habitualmente deja sin contestar el 20% de las preguntas y se enfrenta a un examen con 50 preguntas, entonces la v.a. que modela el número de preguntas que dejará en blanco será una  $B(50, 0,2)$ . Si acierta el 30% de las preguntas,

la v.a. que modelaría el número de preguntas que acierta en dicho examen será una  $B(50, 0,3)$ ; mientras que una  $B(50, 0,5)$  nos daría la ley que gobierna al número de preguntas que dicha persona falla.

De una ley de probabilidad interesan diferentes medidas. Por ejemplo, para identificar su comportamiento central, su media y moda. En el caso de una  $B(n, p)$  se verifica que la media es  $\mu = n \cdot p$ , mientras que la moda viene dada por

$$Mo = \begin{cases} [(n+1)p] \text{ si } (n+1)p \text{ no es entero} \\ (n+1)p \text{ y } (n+1)p - 1 \text{ si } (n+1)p \text{ es entero} \end{cases}$$

siendo  $[x]$ =parte entera de  $x$ . Por ejemplo, en la situación anterior se espera que una persona deje sin contestar  $\mu = 50 \cdot 0,2 = 10$  preguntas; mientras que el número más probable de preguntas que acertará será.  $Mo = [51 \cdot 0,3] = 15$ .

Para adecuar la estructura binomial a las preguntas tipo test es necesario hacer algunas hipótesis que pasamos a enumerar. Vamos a ir comentando estas hipótesis para el caso de un temario general de ofimática sobre Microsoft donde se distinguen cuatro partes: Word, Excel, Powerpoint y Access.

- El nivel de conocimientos del examinado o la estructura de la prueba permiten configurar  $k$  bloques de conocimiento sobre el área considerada. En el ejemplo  $k = 4$ .
- En cada bloque de conocimiento de los  $k$  bloques, las preguntas se consideran independientes y existen las siguientes probabilidades:
  - $p_{aj}$  = Probabilidad de acierto de una pregunta en el bloque  $j$ .
  - $p_{fj}$  = Probabilidad de fallo de una pregunta en el bloque  $j$ .
  - $p_{ncj}$  = Probabilidad de no contestar una pregunta en el bloque  $j$ . verificando

$$p_{aj} + p_{fj} + p_{ncj} = 1, \text{ con } j = 1, \dots, k$$

En el ejemplo tendríamos un total de cuatro bloques (las 4 partes) con tres probabilidades en cada bloque que suman 1.

- En cada bloque de conocimiento de los  $k$  bloques se considera que la prueba presenta  $n_j$  preguntas, con  $j = 1, \dots, k$ . Se definen las siguientes variables aleatorias dependientes:
  - $A_j$  = Número de aciertos en el bloque  $j$ .
  - $F_j$  = Número de fallos en el bloque  $j$ .

- $NC_j$  = Número de preguntas sin contestar en el bloque  $j$ .

A partir de lo anterior, la teoría sobre la distribución binomial permite garantizar que  $A_j$  sigue una ley  $B(n_j, p_{aj})$  y  $F_j$  seguirá una ley  $B(n_j, p_{fj})$ , con  $j = 1, \dots, k$ .

Para poder utilizar los resultados asociados a la ley binomial es necesario conocer los valores de  $p_{aj}$  y  $p_{fj}$ , para lo cual supondremos que el usuario al ejecutar el bloque  $j$  presenta los siguientes resultados:

- $n_{aj}$  número de preguntas "acertadas" en el bloque  $j$ .
- $n_{fj}$  número de preguntas "falladas" en el bloque  $j$ .
- $n_{ncj}$  número de preguntas "no contestadas" en el bloque  $j$ .

Es trivial que  $n_j = n_{aj} + n_{fj} + n_{ncj}$ , siendo  $j = 1, \dots, k$ .

A partir de estos datos podremos desglosar tres tipos de probabilidades  $p$  en el bloque  $j$ :

$$p_{aj} = \text{Probabilidad de acierto} = \frac{n_{aj}}{n_j}$$

$$p_{fj} = \text{Probabilidad de fallo} = \frac{n_{fj}}{n_j}$$

$$p_{ncj} = \text{Probabilidad de no contestar} = \frac{n_{ncj}}{n_j}$$

El sistema que describimos en el siguiente epígrafe necesita de estas probabilidades, o mejor dicho de sus estimaciones, para ofrecer información al usuario.

## Características del sistema SAAPET

Ya existen herramientas disponibles en el mercado que facilitan la creación de páginas web a un administrador para publicar preguntas de tipo test; es el caso de Hot Potatoes, que crea código fuente de una manera transparente y que ayuda a la ejecución y resolución de dichas preguntas de tipo test. SAAPET no sólo ofrece una gestión de las preguntas de tipo test de un tema en concreto, sino que proporciona una serie de datos estadísticos personales a sus usuarios, referenciándolos con los de su grupo de estudio, que se detallan a continuación.

Al comienzo de la ejecución de la aplicación aparece en pantalla un menú principal



Figura 1. Menú principal

(ver Figura 1) con dos áreas claramente diferenciadas: las opciones *offline* y las opciones *online*. Las primeras permiten al usuario conocer más a fondo los exámenes de tipo test y los parámetros que influyen en el resultado de los mismos. Las opciones *online*, por el contrario, proporcionan preguntas de tipo test para que el usuario las conteste y realizan diferentes estudios estadísticos para cada usuario y para todos los usuarios que están dados de alta en un temario o grupo.

### Opciones "offline"

Bajo dicha área, el usuario puede observar cuáles son los factores que influyen a la hora de obtener una nota en especial. Los parámetros del sistema son siempre externos y elegidos por el usuario, proporcionando así claridad y simplicidad al mismo. No hay ningún parámetro interno que el sistema utilice sin que el usuario lo sepa, y todos los cálculos y medidas estadísticas son conocidos. En concreto, se utiliza una v.a.  $X$  como "el número de aciertos en el examen tipo test". Esta v.a. es una binomial (número de preguntas,  $p$  de acertar):  $X \approx B(N^\circ \text{ preg's}, p)$ .

En consecuencia, generalizando todos los casos posibles, para acertar un número de preguntas mayor o igual a "t" tendremos que calcular la probabilidad de que  $X$  sea mayor o igual a  $t$ . Como  $X$  se aproxima a una binomial sería:

$$p(x \geq t) \equiv \sum_{j=t}^{N^\circ \text{ preg's}} \binom{N^\circ \text{ preg's}}{j} p^j \cdot (1-p)^{N^\circ \text{ preg's}-j}$$

Una de las opciones *offline* es la de poder visualizar gráficamente la importancia que tienen los parámetros: número de preguntas del test y la probabilidad de acertar cada pregunta. Para un número de preguntas fijas, se podrá calcular la probabilidad de obtener

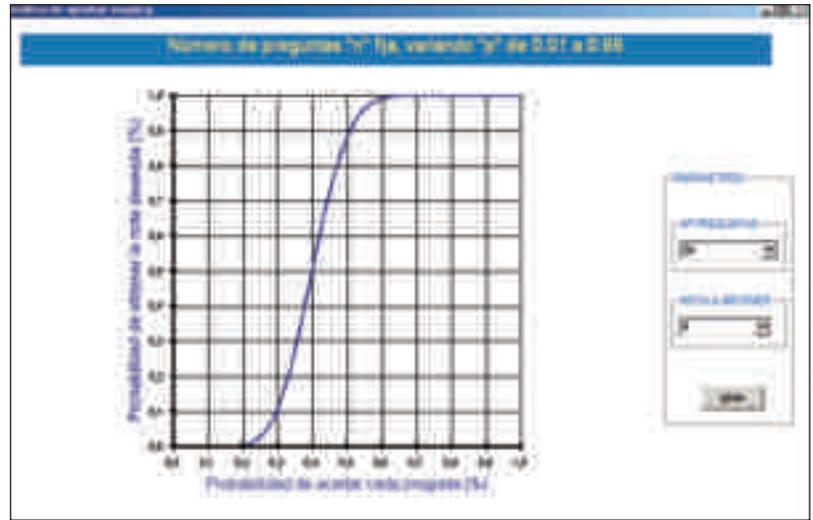


Figura 2. Probabilidad de obtener una nota determinada en función del n° de preguntas, la nota a obtener y la probabilidad de acertar cada pregunta

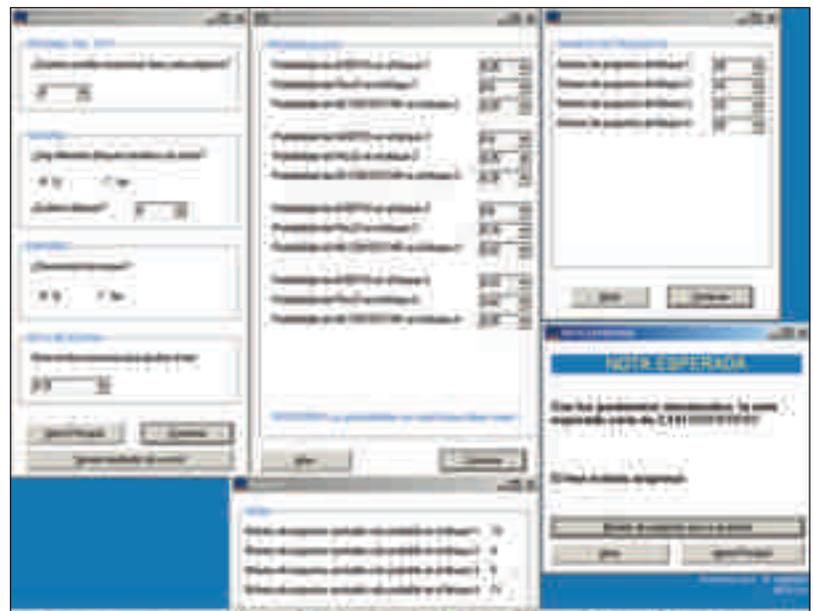


Figura 3. Parámetros para hallar la nota esperada

una nota determinada en una prueba según la probabilidad ( $p$ ) de acertar cada pregunta (ver Figura 2). El usuario interactuará con las gráficas para sacar conclusiones de qué es lo que más le conviene dependiendo del número de preguntas que vaya a tener un test, de la nota que desea obtener o de la probabilidad de acierto ( $p$ ) que espera obtener en dicho test.

Si el usuario quiere tener una mayor precisión y saber la nota esperada que obtendrá en un examen, se le pedirán los siguientes parámetros (ver Figura 3): el número de opciones de respuesta que tiene el test, el número de bloques o temarios diferenciados que hay en un test (de 1 a 10), la posibilidad

de que en la corrección se descuenten o no los errores y una nota mínima de corte para pasar el test.

Una vez introducidos dichos parámetros, un nuevo formulario pedirá al usuario que introduzca la probabilidad de acertar, fallar o no contestar las preguntas de tipo test en cada bloque (estas probabilidades podrán ser halladas con el sistema gracias a uno de los módulos de las opciones *online*). Una vez introducidas las probabilidades, el programa le pedirá al usuario que introduzca el número de preguntas para cada bloque temático. Tras este último pará-

metro, el sistema generará la nota esperada en un examen de tipo test con dichas características, y el número de preguntas más probables que acertará en cada bloque.

#### NOTA ESPERADA

Para determinar dicha nota esperada en la escala de 0 a 10, partimos de la hipótesis de la independencia de preguntas para los  $k$  bloques que pudieran existir y de los desarrollos de la sección 2. La nota esperada con dichos parámetros depende de si los errores descuentan o no.

(a) Descontando errores en el bloque  $j$ , con  $j = 1, \dots, p$ ; sería:

$$NE_{jde} = \frac{n_j}{n_{\text{totales Test}}} \times \left( p_{aj} \cdot \frac{p_{aj}}{\text{número\_opciones\_respuesta\_bloque\_j-1}} \right) \times 10$$

(b) Sin descontar errores en el bloque  $j$ , con  $j = 1, \dots, p$ ; sería:

$$NE_{jde} = \frac{n_j}{n_{\text{totales Test}}} \times (p_{aj}) \times 10$$

La nota esperada en una prueba es la suma de las notas esperadas en los bloques de dicha prueba.

#### Opciones "online"

En este módulo del programa, siempre que haya una conexión de red con el servidor central, el usuario podrá preparar un examen de tipo test y ver su situación actual con respecto a la de su grupo. Cuando el usuario se da de alta en el sistema, indicará en el registro cuáles son los temarios a los que quiere pertenecer (actualmente se ha implantado el sistema en ICAI con temarios de la asignatura de estadística, pero el sistema es muy flexible y permite fácilmente incorporar otros temarios como, por ejemplo, inglés, o aplicaciones ofimáticas como Word, Excel, Access...). Al pertenecer a un temario, podrá conectarse al mismo y responder preguntas de test averiguando su probabilidad de acierto, fallo o de no contestar a las preguntas.

Esta flexibilidad permitirá introducir nuevas preguntas en el sistema con total transparencia para el usuario, que simplemente se limitará a seguir contestando preguntas en orden secuencial. Cuando llegue a la última pregunta, volverá de nuevo a la primera para iniciar un nuevo ciclo de preguntas (ver Figura 4). El usuario podrá

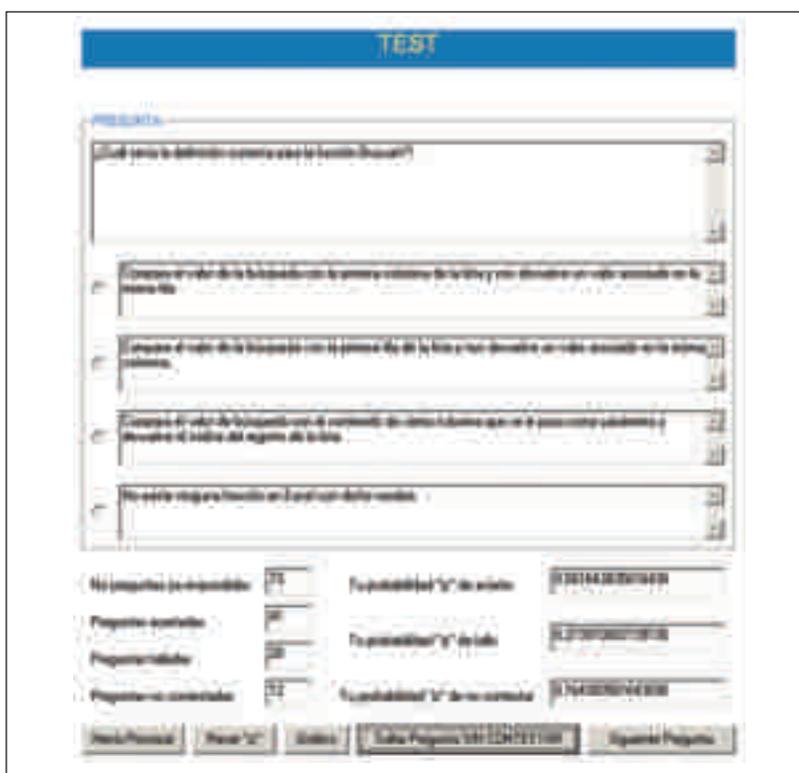


Figura 4. Menú de respuestas en cada pregunta

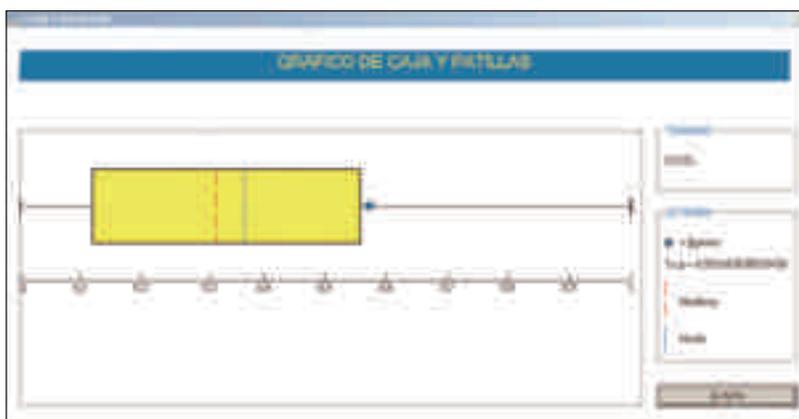


Figura 5. Gráfico de cajas del grupo y posición del alumno

optar por responder a la pregunta, comunicándole el sistema la respuesta correcta si la falla y registrándola en otra tabla para repetirla más adelante, o no contestar la pregunta, si considera que dicha pregunta no pertenece a su temario lectivo o por cualquier otra razón. En cualquier momento podrá ver cómo reacciona ante las preguntas, con un intuitivo gráfico de sectores donde se recogen las acertadas, falladas y no contestadas.

Una vez que el usuario ha respondido a diferentes preguntas del sistema, y se ha involucrado en el mismo, podrá obtener diferentes estudios para cada uno de los temarios a los que pertenece. Si el usuario quiere ver las notas más representativas de la clase, se le mostrará un formulario con información como número de alumnos dados de alta, probabilidad "p" de acertar una pregunta máxima en el grupo, probabilidad "p" mínima en el grupo, probabilidad "p" media del grupo, etc.

Pero no sólo se puede obtener un resumen con el número de alumnos que tienen una probabilidad mayor o menor a un valor determinado, sino que puede solicitar al programa diez gráficos en 2 y 3 dimensiones de la distribución de probabilidad de todos los alumnos en dicho grupo.

Si lo que se necesita es tener un conocimiento de los centiles o percentiles de un grupo o temario específico, se podrá solicitar dicho valor al servidor y conocer de esa manera el ranking de un grupo y la posición actual del usuario con respecto a los demás. Si aun así el usuario no tiene una idea clara de cómo se encuentra con respecto al grupo, el sistema es capaz de generar un gráfico de caja y patillas (ver Figura 5) donde los cuartiles, la media de las probabilidades y la posición del usuario siempre quedan representados, de manera que el usuario puede valorar en el 25% donde se encuentra, de los cuatro 25% en los que se puede dividir el rango de valores.

Existen dos versiones del sistema. La "versión alumno" descrita anteriormente y la "versión profesor", que permite introducir nuevas preguntas en cualquiera de los bloques sin que los usuarios sufran ningún trastorno y obtener otros estudios diferentes sobre cuáles son las preguntas más falladas dentro de cada temario, el resultado de las encuestas y sugerencias que aportan los usuarios al sistema, o información sobre

cuáles son los temarios más fallados y conflictivos (ver Figura 6).

### El sistema SAAPET como instrumento de apoyo en la selección de recursos humanos

Siguiendo a Armstrong (2003), la gestión de recursos humanos es una aproximación estratégica a la estructura, motivación, desarrollo y gestión de los empleados de una organización, basada en los cuatro principios fundamentales siguientes.

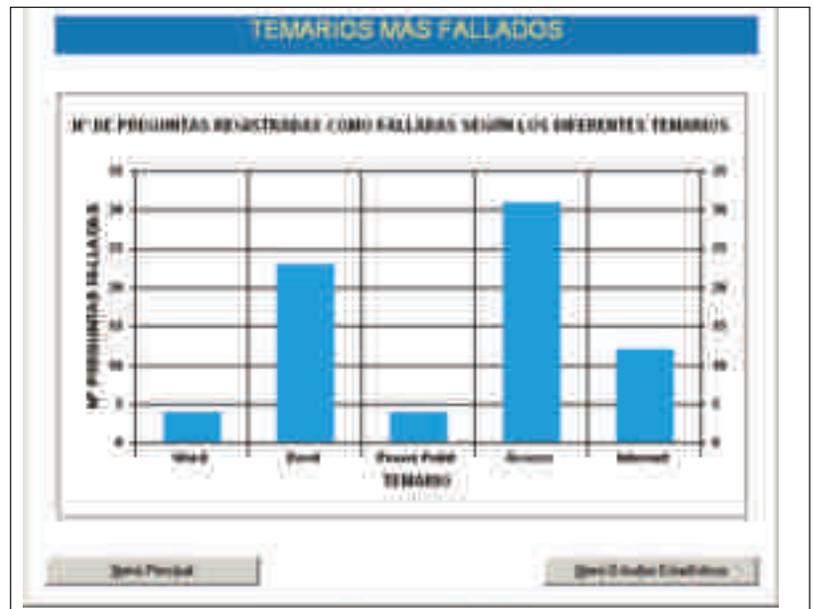


Figura 6. Gráfico que muestra los temarios más fallados

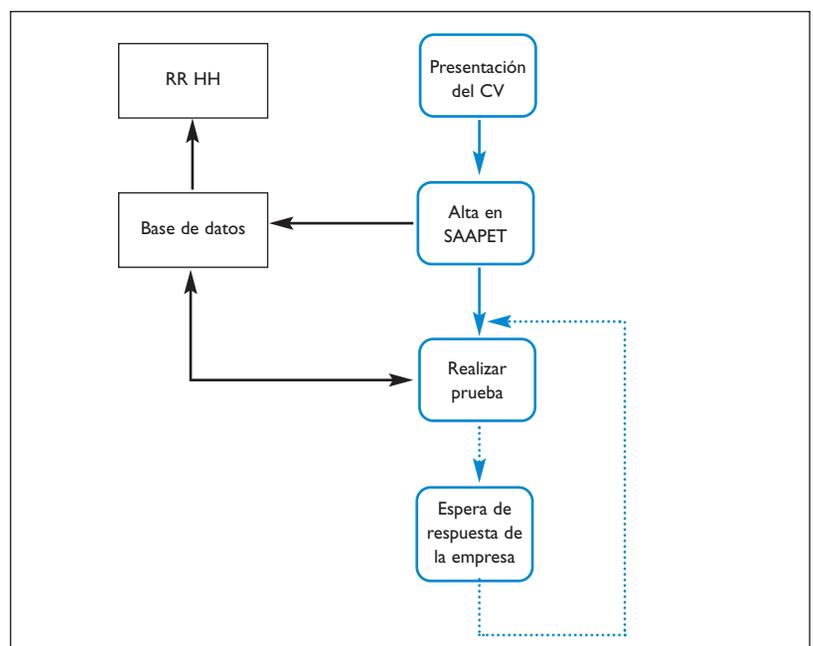


Figura 7. Organigrama de uso de SAAPET en RR HH

- La gestión efectiva de los RR HH es la clave para el éxito de una organización.
- El éxito se logra con mayor facilidad si las políticas de personal están estrechamente vinculadas con los objetivos corporativos de la organización.
- La cultura corporativa, los valores, el clima laboral y el comportamiento directivo que emanan de las acciones de la organización generan una influencia extraordinariamente determinante en el logro de la excelencia.
- La integración total de la mano de obra es esencial para el éxito.

Está completamente asumido que dentro del primer principio se encuadran, entre otras, las tareas de selección, formación y evaluación de los RR HH. Veamos de qué forma puede ayudar SAAPET a la primera de esas tareas. Por limitaciones de espacio no se pueden abordar en este artículo las otras tareas, pero su utilidad en la formación de RR HH está fuera de toda duda.

El reclutamiento de candidatos y la selección de los RR HH siempre han preocupado a la dirección de las organizaciones (prueba actual de ello es el portal [www.infojobs.net](http://www.infojobs.net)). No obstante, sólo recientemente se ha puesto un mayor énfasis en lo que se ha dado en llamar “preempleo”, llegándose a afirmar que el coste de no contratar a la persona adecuada es el coste de la mediocridad y del fallo —véase Fister (2003)—. No es el objeto de este artículo entrar en una descripción de las funciones inherentes a la selección, para ello se puede consultar Puchol (2003).

Entendemos que un sistema como SAAPET puede ayudar a las tareas de reclutar y seleccionar candidatos. Por ejemplo, una posible opción sería que cuando un candidato presentara su CV, se le obligara a darse de alta en el sistema y a comprobar sus conocimientos a través de pruebas objetivas. Dicho candidato podría más tarde desde su casa mejorar su conocimiento en un área determinada realizando tantas pruebas como quisiera, y comprobar su evolución a medida que interactúa con el sistema. Un organigrama de este proceso se muestra en la Figura 7.

Si la candidata o candidato no resultan finalmente elegidos no habrán perdido el tiempo como en otros procesos de selección, puesto que ellos mismos podrán comprobar cómo de rápido ha crecido su conocimiento

en una determinada materia. Además, la empresa sería capaz de ver el interés y espíritu de superación que los candidatos presentan.

## Conclusiones e implicaciones

El sistema SAAPET se está probando en la Universidad Pontificia Comillas de Madrid, en ICAI, en la asignatura Métodos Estadísticos de la Ingeniería de tercer curso del título de Ingeniería Industrial. El prototipo está funcionando con éxito, describiendo todos los usuarios a SAAPET como un sistema de gran ayuda para ellos. Las encuestas registradas destacan la utilidad que el sistema tiene para el alumno, y a todos los encuestados les gustaría seguir utilizando el sistema en un futuro en otras asignaturas. Los alumnos valoran la herramienta como “muy buena” y “buena”, y la nota más repetida dada por los alumnos para la calidad del programa es de un 8 sobre 10, siendo la nota media 8,35.

Por otra parte, se puede afirmar que SAAPET es un instrumento de innovación pedagógica que está en sintonía con las líneas de actuación que marca la Declaración de Bolonia de 1999, las conferencias de ministros de Educación europeos de Praga en 2001, Berlín en 2003 y Bergen los pasados 19 y 20 de mayo de 2005, dentro de las reformas que se han de acometer en el denominado Espacio Europeo de Educación Superior (EEES).

Al cierre del artículo no se dispone de datos que permitan cuantificar la mejora que SAAPET aporta, pero sin duda éste es un terreno que abre enormes posibilidades de investigación a todas las personas involucradas tanto en la docencia de cualquier ámbito como en la gestión de RR HH. ■

## Bibliografía

Armstrong, M. (2003): *A Handbook of Human Resource Management Practice*, 9<sup>th</sup> Edition, Kogan Page.

Fister Gale, Sarah (2003): *Putting Job Candidates to Test*. “Workforce.com”, April 2003, 64-68.

Morales, P. (2000): *Las pruebas objetivas, modalidades y problemas*. “Praxis Educativa”, n° 2, 1-61, Universidad Rafael Landívar, Guatemala.

Puchol, L. (2003): *Dirección y Gestión de Recursos Humanos*. 5<sup>a</sup> Edición. Díaz de Santos. Madrid.

Simkin, M. G. y Kuechler, W. L. (2005): *Multiple-Choice Tests and Student Understanding: What Is the Connection?* “Decision Sciences Journal of Innovative Education”, Vol. 3, 1, 73-97.