

# LA EVALUACION COMO SISTEMA DE INCENTIVACION DOCENTE PROGRAMA “EVADORA”

María Antonia MURCIA-TOMAS<sup>1</sup>, Magdalena MARTÍNEZ-TOMÉ<sup>2</sup>, Antonia María JIMÉNEZ MONREAL<sup>3</sup>, Ana María RIVAS-VELASCO<sup>4</sup> y Luis GARCIA-DIZ<sup>5</sup>

Dpto. de Tecnología de Alimentos, Nutrición y Bromatología<sup>1,2,3</sup>, Dpto. de Nutrición y Bromatología<sup>4</sup>, Dpto. de Nutrición I<sup>5</sup>  
Universidad de Murcia<sup>1,2,3</sup>, Universidad de Granada<sup>4</sup>, Universidad Complutense de Madrid<sup>5</sup>

## Resumen

*La evaluación puede convertirse en una poderosa herramienta de incentivación del estudio si el estudiante cree en su justicia y eficacia. ¿Para que esforzarse si el sistema empleado en medirlo no es fiable? Desear suerte a una persona que acude a un examen es un hábito que demuestra hasta qué punto examinandos, y personas cercanas a ellos, saben que el azar es un elemento determinante en la calificación final. Lo más importante que ha de medir un docente es el nivel de preparación de sus alumnos y ¿está seguro de la validez del sistema de medida que emplea? Los alumnos no.*

*Si el estudiante participara en todas las fases del sistema de evaluación confiaría en él y la relación esfuerzo/calificación se convertiría en un elemento que uniría el esfuerzo que los docentes queremos inculcar, con la calificación que el estudiante quiere obtener.*

*La aplicación informática desarrollada, “EVADORA”, escoge las preguntas aleatoriamente en la primera prueba, para hacerlo en las siguientes según la evolución de cada alumno, presenta los resultados en tiempo real, y permite al examinando estudiar sobre los errores cometidos.*

*El análisis del comportamiento durante las pruebas otorga al docente información con la que revisar y mejorar sus clases.*

**Palabras clave:** *Incentivación, evaluación, exámenes personalizados, test, informática, TIC, autoaprendizaje, aprendizaje guiado por errores.*

(\*) Donde pone alumno en el texto entiéndase como genérico refiriéndose a uno u otro sexo en todas las ocasiones.

## 1. Introducción

El sistema de evaluación es uno de los motores fundamentales, sino el único, que impulsa a los alumnos de una asignatura a estudiar sus contenidos. Los profesores queremos que APRendan, ellos, en su gran mayoría, lo único que quieren es APRobar ; dos objetivos distintos que parecen el mismo, solo al comienzo. Hay que mostrar al alumno la rentabilidad del esfuerzo de estudiar (aprender) para aprobar, demostrando la justicia y objetividad del sistema de evaluación, para desterrar la costumbre de desear suerte cuándo alguien acude a un examen.

Algunos de los efectos indeseables de las pruebas destinadas a medir conocimientos es que sus resultados se afectan demasiado por elementos ajenos a lo que quieren medir: la ansiedad con la que se acude a la prueba, la falta de confianza en la rentabilidad del esfuerzo, el acúmulo de esfuerzo en los días en los que se concentran los exámenes, el estado de ánimo del sujeto, etc. En resumen:

1. La evaluación depende de una muestra escasa de conocimientos de cuerpo global de la asignatura.

2. La selección de las preguntas de un examen dependen del criterio del profesor.
3. Su calificación depende, en gran medida, del profesor que la realiza.
4. La fecha del examen puede condicionar la calificación obtenida.
5. Los alumnos suspensos tienen múltiples oportunidades de mejorar su calificación sin que influya en la final las anteriores.

Frente a estos problemas, el sistema que hemos desarrollado permite la participación protagonista del alumno en la evaluación de sus conocimientos teóricos, dando solución a los anteriores problemas de la siguiente manera:

1. Los exámenes objetivos (pruebas test con 3 contestaciones alternativas) tienen 60 preguntas y si repiten las pruebas cada alumno contestaría, 120, 180, ...
2. Son los propios alumnos los que proponen las preguntas que consideran que deberían conocer y un sistema aleatorio es el que las escoge para la prueba de evaluación.
3. La calificación es objetiva e independiente del evaluador que la realiza.
4. La fecha de la prueba la pueden elegir los alumnos a su gusto, evitando solapamientos o cercanías con otros exámenes, situaciones personales difíciles, etc.
5. El número de repeticiones las elige el alumno en función de sus objetivos marcados.

## **2. Desarrollo de la aplicación informática “Evaluación Incentivadora-EVADORA”**

La aplicación informática, desarrollada en Visual Basic 6.0 para entornos Windows, permite que los alumnos se examinen cuando quieran, tantas veces como decidan y personaliza el examen de cada uno de ellos, adaptándolos a su histórico, si existe. Esto la convierte en una buena herramienta de trabajo con la que vencer gran parte de los elementos negativos que tienen las pruebas de evaluación, haciendo rentable a los ojos del estudiante el binomio estudio/calificación, y retroalimenta al profesor con una información muy valiosa de cómo han comprendido sus alumnos los diferentes conceptos de la asignatura, para que mejore su programa docente en base a una realidad observada y no a la supuesta.

Aprender y aprobar son dos palabras que deberían ir siempre de la mano, y que entre docentes y discentes hemos convertido en elementos casi contrapuestos por culpa en gran parte de los sistemas de evaluación. El objetivo de todos los miembros de una institución docente debería ser la primera de ella, y disponer de un buen sistema de medida de ella para certificar quien obtiene la segunda. Cuando las circunstancias hacen que se pueda obtener o no el aprobado contaminando el aprendizaje realizado con la suerte, oportunidad de la fechas del examen, proximidad con otras evaluaciones, situación afectiva personal, grado de ansiedad, escaso número y poca representatividad de la preguntas de las prueba, etc... , algo falla.

Los sistemas de evaluación de los conocimientos teóricos de una asignatura habituales en nuestras universidades suelen ser instrumentos de medir poco adecuados para realizar el trabajo que tienen encomendado. Este hecho, así percibido por la inmensa mayoría de los alumnos y alumnas, no es el mejor acicate para estimular el esfuerzo en aprender, puesto que en la mayoría de los casos no va a ser medido correctamente. ¿Quién no ha deseado suerte a una persona que va a examinarse? Somos conscientes, dando por hecho como algo inevitable, que el azar solo tendrá un peso importante en la calificación obtenida en cualquier prueba de evaluación de conocimientos.

Todo un conjunto de estrategias subyacentes plasmadas en el diseño y desarrollo de una herramienta específica denominada “Sistema de evaluación incentivadora” tienen como objetivo que los estudiantes participen en la gestación y funcionamiento del sistema de evaluación, de forma que creyendo en su capacidad para medir el esfuerzo realizado, se convierta en si mismo en algo que estimule el estudio de la asignatura. Desterramos frases como “¡Que mala suerte tuve en el examen!”.

El programa informático se ha desarrollado para corregir algunos de los problemas que se achacan siempre a los exámenes tradicionales de evaluación de conocimientos teóricos como:

1. La evaluación depende de una muestra escasa de conocimientos de cuerpo global de la asignatura.
2. La selección de las preguntas de un examen dependen del criterio del profesor.
3. Su calificación depende del profesor que la realiza.
4. La fecha del examen puede condicionar la calificación obtenida.
5. Los alumnos suspensos tienen múltiples oportunidades de mejorar su calificación sin que influya en la final las anteriores.

Estos hechos, difíciles de solucionar con los sistemas tradicionales de pruebas escritas, pueden abordarse con la ayuda prestada con el auxilio de las nuevas tecnologías de la información (TIC) de forma que un programa informático se encargue de la gestión de la evaluación de los conocimientos teóricos de principio a fin.

Para ello, diseñamos y realizamos una aplicación informática, en Visual Basic 6.0 para entornos Windows, que permite que los alumnos se examinen cuando quieran, tantas veces como decidan, personalizando el examen de cada uno de ellos, adaptándolos a su histórico, si existe. Esto la convierte en una buena herramienta de trabajo con la que vencer gran parte de los elementos negativos que tienen las pruebas de evaluación, haciendo rentable a los ojos del estudiante el binomio estudio/calificación, y retroalimenta al profesor con una información muy valiosa de cómo han comprendido sus alumnos los diferentes conceptos de la asignatura, para que mejore su programa docente en base a una realidad observada y no a la supuesta. Tratando de dar solución a los anteriores problemas, la aplicación los aborda:

1. Los exámenes objetivos (pruebas test con 3 contestaciones alternativas) tienen 60 preguntas y si repiten las pruebas cada alumno contestaría, 120, 180, ...
2. Son los propios alumnos los que proponen las preguntas que consideran que deberían conocer y un sistema aleatorio informatizado es el que las escoge para la prueba de evaluación.
3. La calificación es objetiva e independiente de quien la realiza.
4. La fecha de la prueba la pueden elegir los alumnos a su gusto, evitando solapamientos o cercanías con otros exámenes, situaciones personales difíciles, etc.
5. El número de repeticiones las elige el alumno en función de sus objetivos personales.

Todos los alumnos han de presentar, como mínimo, 10 preguntas con sus contestaciones correctas sobre el cuerpo de doctrina de la asignatura, escritas en un lenguaje académico pero inteligible, sobre la parte de la asignatura que ellos mismos decidan, con la obligatoriedad de acompañarlas de la fotocopia de la fuente bibliográfica de donde ha sido obtenida, lo que da como subproducto el que redacten y manejen libros en unos momentos en los que apuntes e internet lo cubren todo. ¿Cómo desconfiar de algo que ha creado uno mismo?

Si el profesor indica que él cubrirá con sus preguntas las zonas de la asignatura de las que no reciba propuestas de los alumnos, ellos se reunirán para coordinar su trabajo y evitar dejar huecos que haya de cubrir el profesor. Trabajar “contra el profesor” suele ser un incentivo, frente a seguir los “consejos paternalistas” que este sugiera, y hay que aprender a utilizar esta motivación. Al final dispondrán de todo el banco de preguntas del examen, que se intercambiarán “a escondidas” y encuadernarán para estudiar. Creerán que de esa forma solo tienen que estudiar las preguntas que caen en el examen; la cuestión es que 10 preguntas propuestas por cada alumno, en un grupo de 70 personas suman 700 “ítem” de la asignatura. Las 700 cosas escritas más importantes de una asignatura, organizadas por temas y encuadernadas, se llama libro de texto, y eso es lo que estudian, sin darse cuenta y sin la carga negativa que tiene esa palabra entre muchos de ellos. La realidad convierte a estas 700 preguntas esperadas en más de 1000 porque muchas personas presentan más de las 10 solicitadas como mínimos, por las razones que sean.

El profesor se reserva la redacción de los cebos. Esto obliga a los alumnos a estudiar con mayor seguridad los conceptos positivos de la asignatura, evitando el esfuerzo de memorizar falsedades. Con este proceder, el profesor puede evaluar la claridad de los conceptos incluyendo cebos con los errores más comunes o las posibles fuentes de confusión de algunos de ellos. Se prueba la solidez de los conceptos evaluados frente a los elementos de confusión más habituales. Su análisis posterior puede orientar sobre el tiempo que se ha de dedicar en clase a evitar estas confusiones.

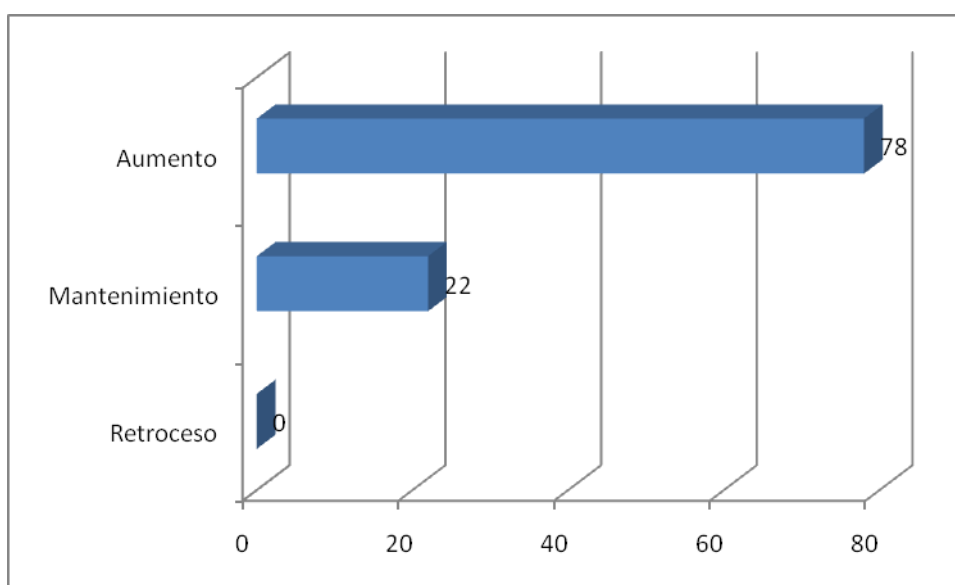
Un aspecto de vital importancia es la calificación de los resultados de las pruebas de tipo test. Este hecho ha de ser perfectamente explicado para que los estudiantes sean conscientes de la equidad de la calificación cuando se ha de eliminar el componente de azar de una calificación (la buena o mala suerte que se pretende erradicar). En las pruebas con un determinado número de ítem con posibilidades de contestación verdadero o falso, todo el mundo intuye que el 50% de las contestaciones correctas pueden acertarse solo por azar y que por tanto la calificación de conocimientos se realizará a partir de esta cifra, es decir, sobre la mitad ítem restantes. Visto de otra forma, contestar la mitad de las preguntas sucederá cuando el alumno conozca la mitad de las respuestas de la base de datos, contestando el resto al azar. Ambos efectos sumados, un 50% debido al conocimiento y el 50% del resto (un 25%) debido al azar, conforma el 75 % que se alcanza cuando se es merecedor de un 5 sobre 10. En pruebas de  $n$  preguntas con 3 contestaciones de las que siempre 1 y solo 1 es verdadera, el azar permite acertar la tercera parte del total, y cuando se conocen la mitad de las preguntas de la base de datos, se acertaría un 79.2 % de las mismas. Que obtener un 5 suponga contestar acertadamente 48 de 60 preguntas han de saberlo y creerlo, y aunque parezcan muchas, en test verdadero/falso, hubieran sido 45 de las 60. El modelo elegido cuenta tan solo los aciertos, considerando fallos las contestaciones erróneas y la falta de contestaciones. Una demostración estadística de este hecho sobre la pizarra y el uso de una demostración práctica con simulación informática en clase, son imprescindibles. En la confianza absoluta en la equidad del sistema se basa su funcionamiento.

### **3. Uso y resultados de la aplicación informática EVADORA**

Cada estudiante accede a un ordenador que le confecciona en tiempo real un test de 60 preguntas con 3 posibles respuestas de las que una y solo una es correcta (de entre las que ellos mismos propusieron), y dos cebos o contestaciones incorrectas (puestas por el profesor y desconocidas por ellos mismos). Tanto el orden en el que se presentan las preguntas como las tres respuestas alternativas, es aleatorio, de forma que aunque alguna pregunta pueda volver a presentarse en un examen posterior, la posición que ocupe la respuesta correcta no tiene porque coincidir. Memorizar que a tal pregunta la contestación correcta es la segunda no tiene sentido.

Cada pregunta se presenta ante el alumno durante 1 minuto como máximo, teniendo tiempo suficiente para leer su enunciado y tres contestaciones alternativas, elegir unas de ellas y confirmar su elección, pasando a la siguiente. El tiempo sobrante en cada una no es acumulable en las restantes, y no puede retroceder para releer las preguntas ya contestadas, De esa forma, el tiempo máximo posible de examen de una hora, 60 preguntas x 1 minuto para cada una de ellas, se reduce notablemente al tiempo medio de contestación de cada una de ellas, siendo en la práctica de unos 20 minutos (tiempo medio de respuesta por pregunta, 20 segundos). Con esta cadencia, no hay tiempo material de "copiar" durante el transcurso del examen, por lo que la necesidad de "vigilancia" del mismo es casi inexistente, permitiendo compatibilizar su realización con el desempeño de otras funciones al profesor, elemento indispensable para permitir que el alumno se examine cuando quiera y tantas veces como desee sin que ello impida al profesor el desempeño de sus otras funciones. Además, cada estudiante puede repasar sus exámenes anteriores aprendiendo de sus fallos, para reorientar su estudio hacia las partes de la asignatura que menos conoce. El programa EVADORA genera las repeticiones de exámenes retirando de la base a sortear las preguntas que ya ha contestado correctamente en ocasiones anteriores, de modo que además de nuevos ítem pueden volver a parecer los que no sabía.

Cada alumno puede examinarse cuantas veces quiera (y lo hacen unas 4 veces, como cualquiera que suspenda repetidamente en el sistema tradicional, dos cursos académicos con dos convocatorias por curso, solo que esta vez son los que aprueban los que usan esa posibilidad: los buenos estudiantes se encuentran con las mismas posibilidades que los no tan buenos), dejando transcurrir al menos 2 días entre prueba y prueba, para evitar usar el sistema como si fuera un juego de azar. Queríamos desterrar la percepción de la suerte como elemento influyente en la obtención de calificación. Subproducto de este sistema de repetición y mantenimiento de la calificación como media de las 2 mejores obtenidas, es el reestudio de la asignatura para mejorar la nota final, sin la sensación de riesgo de perder lo ya obtenido. Estudiar lo ya estudiado. Consolidar los conocimientos ya adquiridos. Solo esta consecuencia del sistema de evaluación lo justificaría en sí mismo. En vez de olvidar lo ya aprobado, el alumno lo reestudia y consolida. Las calificaciones obtenidas en las repeticiones han sido siempre ascendentes, con lo que el propio sujeto percibe la relación esfuerzo/calificación como un binomio incentivador.

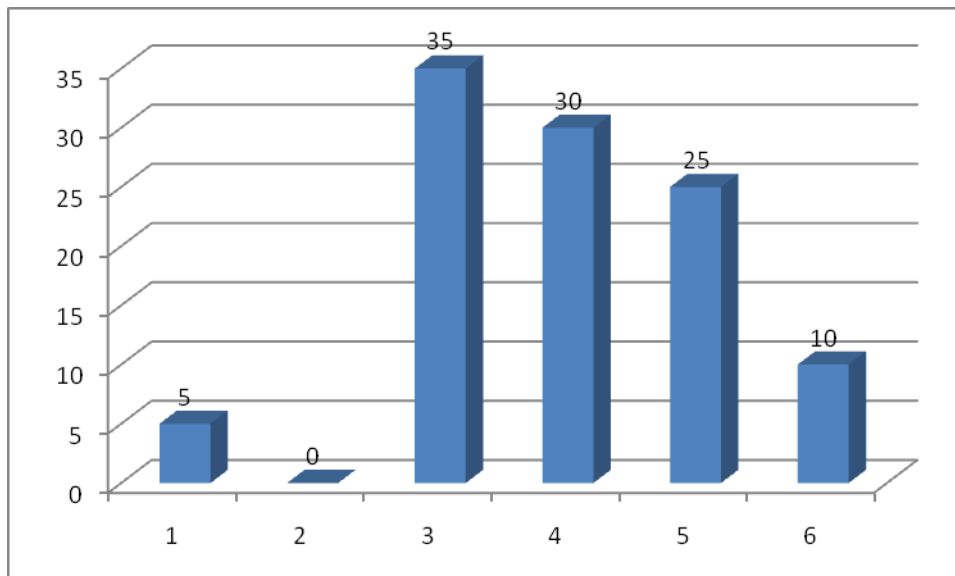


*Consecuencia de la segunda repetición sobre la calificación final obtenida.*

A priori podría pensarse que las repeticiones indefinidas manteniendo la calificación media de las 2 mejores daría pie a un uso perverso del sistema, llevando a algún alumno a repetir alocadamente nuevas pruebas porque ya no perdería nota a ver si la mejoraba por suerte. El estudio de los datos demuestra que esto no sucedió ya que mayoritariamente elevaron la calificación obtenida anteriormente en un 78% de las ocasiones y cuando no se mejoró es porque se mantuvo, el 22% restante, no reduciéndose en ningún caso. Una de las posibles explicaciones de este hecho observado sería que los estudiantes volvían a estudiar los contenidos de la asignatura para ver si mejoraban su nivel de conocimientos, y a base de estudiar lo ya sabido, consolidan sus conocimientos y cada vez dominan con mayor soltura la disciplina.

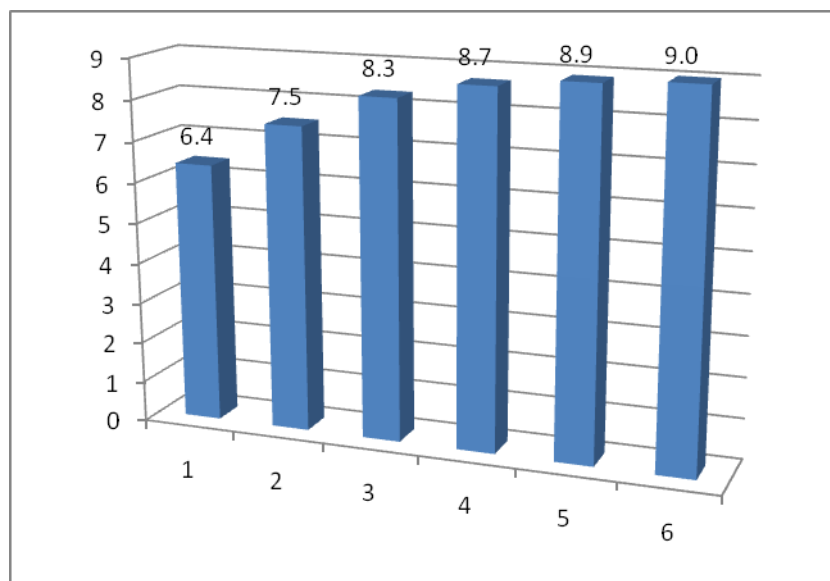
El histórico de cada sujeto permite que el sistema informático plante cada examen a partir de las preguntas que aun no ha contestado o que no sabía con anterioridad, incrementando el "screening" de conocimientos sobre la asignatura. Las preguntas que ya ha demostrado conocer, dejan de ser objeto de evaluación, debiendo demostrar sus conocimientos sobre otras partes de la asignatura en general o sobre las ya sabidas, pero evaluada en este nuevo caso a partir de otras preguntas. Cada nuevo examen explora una nueva parcela de conocimientos del alumno. Tras realizar 4 o 5 exámenes, promedio alcanzado por la mayoría de las personas que han participado en esta experiencia piloto, cada estudiante ha contestado a 300 preguntas con 3 contestaciones alternativas, es decir, han sido evaluados a partir de la contestación de 900 "ítem" de una misma

asignatura, con lo que se consigue un barrido amplísimo sobre los conocimientos teóricos y algunos aspectos de interpretación de la misma.



*Número de repeticiones realizadas por cada alumno.*

La posibilidad de repetición del examen fue usada mayoritariamente por el grupo de estudiantes, convirtiéndose en excepción los que tan solo realizaron un vez la prueba, conformándose con la calificación obtenida. Los que repitieron no se quedaron nunca en 2 pruebas, ya que a partir de ese momento tan solo se podría mejorar la calificación. La desaparición de la sensación de riesgo en la reducción de la nota final tuvo efectos claramente beneficiosos, quizás porque reducía notablemente la tasa de ansiedad con los que se acude a un examen o porque la repetición del estudio había consolidados bien los conocimientos básicos de la asignatura, o más probablemente por una combinación de ambos.



*Calificaciones medias obtenidas tras n repeticiones de las pruebas.*

Como consecuencia de este sistema implantado la calificación media del grupo obtenida tras la realización del primer examen fue de un 6.4, lo que hubiera supuesto un grupo con un APROBADO de media, mientras que al finalizar el curso y tras estudiar y examinarse repetidamente de la

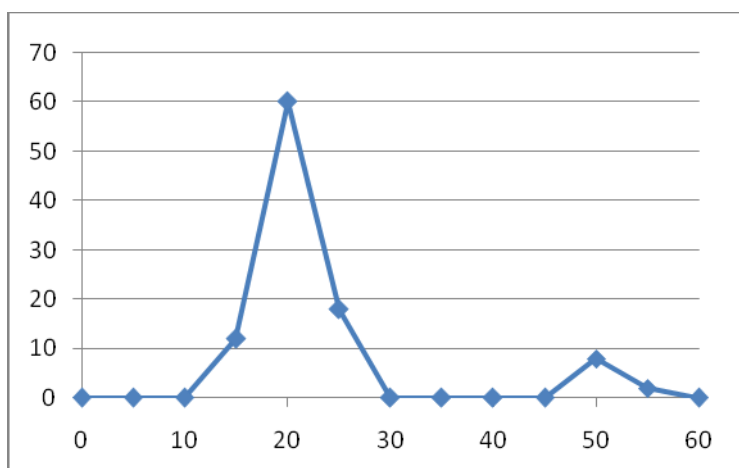
asignatura, la calificación media alcanzada por el grupo fue de SOBRESALIENTE. Si los suspensos de las primeras convocatorias ordinarias suben por lo menos a aprobados en las posteriores, estos alumnos de aprobado subieron a sobresaliente.

#### 4. ¿La verdad, toda la verdad y nada mas que la verdad?

Los tiempos de respuesta, un elemento imposible de controlar en exámenes no informatizados son una importantísima fuente de información para el docente. El alumno, desconocedor de que se está controlando este parámetro, para que no intente manipularlo conscientemente, contesta rápidamente lo que cree saber, y espera casi hasta el final del minuto asignado en aquellas preguntas que duda.

La aplicación informática, EVADORA, señala el tiempo transcurrido sobre la pantalla del ordenador donde se muestran las preguntas, para que el alumno no tenga que distraerse mirando su reloj. Un contador descendente marca los segundos que le restan y una barra de color azul va creciendo en la parte inferior de la pantalla para informar sin distraer. A los 45 segundos, una señal acústica y el cambio de color de la barra creciente que marca el tiempo transcurrido, alertan al alumno de que entra en los últimos 15 segundos (tiempo medio empleado en leer y contestar cualquier pregunta). Este suceso “despierta” al alumno que espera la inspiración sobre lo que no sabe y le hace contestar sin mas dilación.

Aunque la bondad del sistema propuesto es la transparencia del mismo de cara al estudiante, algunos procesos han se ser estudiados con métodos ciegos o doble ciegos para asegurar la validez de los resultados. Ya que el tiempo empleado en responder cada pregunta no tiene efecto sobre la calificación obtenida, el que no sepa de la existencia del citado control, no entorpece la realización de las pruebas y permite un inestimable herramienta de retroalimentación para el docente.



*Tiempo de respuesta de las preguntas*

El análisis de las preguntas con tiempo de latencia superior a 50 segundos señala un 33% de aciertos frente a un 67% de fallos, justo lo esperado cuando se responde al azar. Las preguntas de tiempo de latencia alto, contestadas por fuera de esta proporción señalan preguntas de redacción muy larga o compleja que deben ser reescritas para adaptarlas a las condiciones del examen. Las preguntas contestadas rápidamente, tiempos de latencia cortos, calificadas peor que el conjunto global, indican conocimientos incorrectos bien implantados. Esto es lo peor que puede suceder. Es más difícil corregir un error de concepto, algo mal aprendido, que rellenar un hueco en el conocimiento de una persona. Un esfuerzo extraordinario ha de dedicarse a tratar de evitar la implantación de conceptos erróneos en la mente de los estudiantes. El análisis de los tiempos de

respuesta permiten mejorar la propia herramienta de evaluación y el contenido de las clases teórico/prácticas que se imparten.

## **5. Resumen de resultados y conclusión general**

Los resultados indican un excelente grado de aceptación por los alumnos de este sistema de evaluación que ha mostrado como inconveniente el que crea adicción, pudiendo ser contraindicado para el rendimiento en otras asignaturas a las que quitan tiempo de estudio porque es mucho más rentable el esfuerzo en aquellas en las que el sistema de evaluación es fiable; inconveniente que desaparecerá en el momento en que todas las asignaturas compartan sistemas equivalentes de valoración del esfuerzo realizado.

Después de realizado el trabajo surge un nuevo interrogante. ¿Se debe usar la penalización en las preguntas mal contestadas en las pruebas tipo test como un sistema de aprendizaje sobre las consecuencias negativas que tiene para la salud de los futuros pacientes los errores cometidos por el profesional? ¿Son equivalentes la falta de un conocimiento, contestación en blanco, que un conocimiento mal adquirido, contestación errónea?

Enseñar se apoya en aprender de los errores y rectificar. Cada vez que algo parece terminado, los resultados dan pie a nuevos planteamientos, a nuevas reformas.