

LA UTILIZACIÓN DE LAS PRUEBAS OBJETIVAS DE RESPUESTA MÚLTIPLE CON SOPORTE INFORMÁTICO, CÓMO SISTEMA DE EVALUACIÓN Y ENTRENAMIENTO-SEGUIMIENTO DEL ALUMNADO.

Ariza A, Sánchez-Aguayo I, Daza P, Moreno FJ, Torreblanca J.
Departamento de Biología Celular, Universidad de Sevilla.

ANTECEDENTES

En la actividad docente encontramos tres momentos bien definidos, la planificación o programación de la asignatura, la realización en sí (clases, seminarios, prácticas, etc.), y por último el proceso de evaluación o control de los resultados obtenidos (2, 13).

En el campo de la Biología existe bastante consenso en lo que hay que enseñar (programación). La reflexión sobre el proceso enseñanza-aprendizaje es más de tipo operativo. Es más ¿cómo enseñar? que ¿qué enseñar?

Existen muchas herramientas para exponer los conocimientos, tanto en el aula, transparencias, videos, presentaciones, etc.; como telemáticamente. Esta labor creemos que, generalmente, es gratificante.

Otra labor más ardua y tediosa, pero necesaria, es la corrección de exámenes para la evaluación. Es la fase final de un proceso en donde se comprueba si el diseño de enseñanza que se ha llevado a cabo es correcto o no, es decir, si los objetivos que se marcaron al comienzo se han logrado o no. Si el alumno ha conseguido los objetivos marcados, diremos que la programación ha sido planteada de forma correcta; si no es así tendremos que modificar y corregir lo que sea necesario (13). Por tanto la evaluación es una actividad sistemática integrada en el proceso educativo, cuya finalidad es servir de diagnóstico, no solo del rendimiento del alumno, sino de todo el proceso de enseñanza-aprendizaje (8), extendiéndose actualmente a otros objetivos como proyectos, programas, materiales de enseñanza o instituciones educativas (9).

Son muchos los procedimientos de evaluación que podemos utilizar, siendo válidos todos aquellos que sean capaces de poner de manifiesto si las actividades del profesor y del alumno llevan al logro de los objetivos propuestos (13).

Un elemento esencial en la evaluación es la medición, aunque en muchas ocasiones ambos términos se confunden, la medición es una fase previa que proporciona objetividad a la evaluación (7). Una de las técnicas de medición utilizadas con más frecuencia en el campo de la enseñanza son las pruebas objetivas. Éstas consisten en una serie variable de preguntas (ítems) formuladas de manera que la respuesta sea siempre breve y concreta lo que permite una calificación cuantitativa rápida.

El automatizar los procesos en las pruebas objetivas ayuda a estandarizarlas, haciendo más fácil y rápida su realización y permitiendo que las diferencias registradas en los resultados se deban a rendimientos diferentes de los alumnos en la dimensión medida y no en otra (5).

En el mercado existe diferente software para realización de pruebas objetivas (4). Nosotros venimos utilizando desde hace más de 10 años un sistema, propio, de generación y corrección de pruebas objetivas. La primera versión (1995) estaba realizada con el lenguaje de programación Clipper y funcionaba bajo el sistema operativo MS-DOS, posteriormente se reprogramó con el lenguaje Visual Dbase para que funcionara en el entorno Windows 95/98 (5).

La primera razón que nos llevó a utilizar las pruebas objetivas con soporte informático fue el elevado número de alumnos que teníamos en algunas asignaturas y el hecho de realizar pruebas objetivas desde hacía algún tiempo.

Curso	% de Alumnos no presentados
90-91	22.85
91-92	37.03
92-93	39.13
93-94	33.33
94-95	6.45

Ese conjunto de preguntas utilizado fue el germen de la 1ª base de datos del programa. Posteriormente, cuando el número de alumnos disminuyó, utilizamos la posibilidad que nos brindaba el programa de generar y corregir exámenes con facilidad, para ofrecer a los alumnos la libertad de examinarse el día que ellos quisieran. De esta forma disminuimos de manera considerable el número de alumnos no presentados en la asignatura (curso 94-95, tabla) (13).

ESTADO ACTUAL

La continua evolución del programa ha aumentado su versatilidad, y ahora además de generar y corregir exámenes nos permite, entre otras cosas, medir la propia bondad de las pruebas que utilizamos. El hecho de disponer de software que permita construir y mantener un banco de ítems y aumentar la eficiencia en la medición y evaluación, hace que el profesor no sólo pueda ahorrar tiempo, si no que también pueda ofrecer a sus alumnos posibilidades que sin Exawin serían muy difícil de llevar a cabo (escoger sistema de corrección, evaluación continua, etc.).

El programa Exawin 5.1 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:

1. Crear y mantener bases de datos de alumnos
2. Crear y mantener bases de datos de preguntas
3. Generar pruebas objetivas
4. Corregir exámenes, tanto colectivos como individuales, utilizando diferentes criterios.
5. Analizar los ítems

Las posibilidades antes enumeradas se traducen en las siguientes ventajas:

1. Rapidez en la confección de pruebas por el profesor y en su realización por parte del alumno
2. Comprobar la validez de las preguntas y en su caso modificarlas. El análisis de los ítems permite estudiar la eficacia de las preguntas utilizadas, para conseguir los objetivos propuestos. Este análisis se realiza por el estudio de las propiedades estadísticas de los ítems. Los índices más relevantes pueden ser:
 - a. Índice de dificultad (ID), proporción de alumnos que fallan un ítem entre aquellos que han intentado resolverlo (6).

- b. 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).
 - c. 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).
3. Calificar objetiva y rápidamente
- a. 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. (para más detalles sobre calificación consultar 3 y 5)
 - i. Los criterios que utiliza son:
 - 1. De Lord o clásico (acierto-errores). La fórmula sería $Puntuación_{Bruta} = \text{Aciertos} - \text{Errores/Opciones} - 1$ (11, 3)
 - 2. Sólo aciertos. La fórmula sería $Puntuación_{Bruta} = \text{Aciertos} - N_{\text{items/Opciones}}$ (3)
 - 3. 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 (ver tabla) (3).

	Niveles de Dificultad				
	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
Respuesta Correcta	1	0.9	0.8	0.6	0.4
Respuesta Errónea	-0.4	-0.6	-0.8	-0.9	-1

- 4. Por niveles de confianza (CERT). Parecido al anterior, la corrección de los efectos del azar se deja al alumno, estableciendo 6 niveles de confianza que varían el valor del acierto y el fallo (ver tabla) (1, 2, 9, 10).

	Niveles de Confianza					
	0	1	2	3	4	5
Respuesta Correcta	13	16	17	18	19	20
Respuesta Errónea	4	3	2	0	-6	-20

- 5. Por niveles de confianza CERT modificados. Igual al anterior pero con la modificación del número de niveles de confianza y del valor asignado a ellos (ver tabla) (5).

	Niveles de Confianza		
	0	1	2
Respuesta Correcta	8	10	12
Respuesta Errónea	-1	-2	-5

6. En todos los criterios anteriores se pueden escoger dos niveles de rigurosidad, 0 y 1, dependiendo de que se considere o no, las respuestas en blanco como una opción más de cada pregunta.
4. Confeccionar las listas de los distintos exámenes
5. Aplicar esta utilidad a un número elevado de alumnos
6. Realizar un número elevado de exámenes
7. Implicar más al alumno gracias a un seguimiento más exhaustivo
8. Barajar exámenes, tanto enunciados como respuestas, con lo que se dificulta la copia a pesar de utilizar las mismas preguntas
9. Almacenar ítems para la elaboración posterior de distintas pruebas

PERSPECTIVAS

Tras varios años utilizando el programa para realizar los exámenes de la parte teórica de distintas asignaturas, la incorporación de mejoras al mismo nos abre las siguientes perspectivas:

1. **Aplicación a exámenes de prácticas.** La inclusión de imágenes en las preguntas, hasta ahora, hacía que la base de datos se volviera muy “pesada” y el programa se hacía lento y poco operativo. En la actualidad el lenguaje de programación hace que Exawin pueda contener, en sus bases de datos, cientos o miles de imágenes sin dificultar su funcionamiento, pues sólo la correspondiente a la pregunta que se está visualizando es la que realmente está manejando el programa. Esta posibilidad, en combinación con la proyección de la pantalla a través de un cañón de imagen, nos va a permitir probarlo también en la elaboración y realización de los exámenes prácticos (de visu) en el presente curso. De esta forma se podría abordar el componente visual tan necesario en el aprendizaje de nuestras asignaturas, y que hasta hace poco el programa no podía realizar.
2. **Sistema de seguimiento-entrenamiento del alumno.** El desarrollo de experiencias piloto para la adaptación al crédito ECTS en la licenciatura de Biología nos ha llevado a la aplicación de usos docentes que intentan fomentar el trabajo autónomo y el desarrollo de capacidades por parte del alumno. Éste se debe convertir en el principal autor de su formación. En nuestra asignatura propusimos a los alumnos un trabajo tutelado que sería evaluado junto al examen y tendría un valor del 10% de la nota. Tras experimentar durante dos cursos con este modelo, hemos podido comprobar que, en muchas ocasiones, el alumno decide renunciar al 10% de la nota que supone el trabajo tutelado antes que emplear el tiempo necesario para poder preparar el tema propuesto. De hecho las notas conseguidas en ese apartado son inferiores a las del resto del examen. Ante estos resultados estamos estudiando la posibilidad de hacer un seguimiento más cercano del alumno a partir de una evaluación continua de los temas explicados. Para este fin utilizaríamos las posibilidades que nos ofrece el programa Exawin a la hora de generar y corregir pruebas, en particular podría ser interesante la utilización de los niveles de confianza a la hora de medir el aprendizaje, pues nos daría, tanto al profesor como al alumno, una estimación de lo aprendido y de la confianza en ese aprendizaje. Para ello estamos estudiando la

posibilidad de utilizar un Índice de Realismo, ya expuesto en el Proyecto CERT, que nos indica en qué medida la confianza que el alumno tiene en sí mismo se corresponde o no a la realidad. Es decir, si la probabilidad de acierto predicha por él mismo se acerca o no a los aciertos obtenidos en la realidad. Además, para facilitar esta función de entrenamiento estamos pensando en incorporar un campo “memo” a las bases de datos de las asignaturas y, por tanto a los exámenes. En dicho campo se recogería la justificación de la respuesta correcta, lo que acentuaría la función de adiestramiento que queremos darle al programa ya que, no solo le daría una nota si no que además le justificaría las respuestas, correctas o incorrectas.

3. Establecimiento de pruebas con un nivel de dificultad predeterminado.

El cálculo del nivel de dificultad, de las distintas preguntas de un examen que realiza el programa, abre la posibilidad de utilizar el almacenamiento de este dato para filtrar las preguntas que puedan ser utilizadas para generar una prueba posterior. Esta posibilidad, fácil de realizar y que ha estado siempre accesible a las distintas versiones del programa podría ser una forma de ir estableciendo los diferentes niveles de aprendizaje de los alumnos.

Como conclusión podemos decir que la utilización de nuevas tecnologías, en este caso de un programa que gestiona bases de preguntas y genera exámenes, no realiza nada diferente a lo que un profesor pueda hacer. No obstante, el coste en tiempo para brindar lo mismo sin ayuda del programa sería inviable. De esta forma nos abre posibilidades implícitas en nuestra manera de trabajar que no se podrían ofrecer sin estar auxiliado por este software.

Bibliografía:

1. Alonso, C. Y otros. Principios comunes para la evaluación de los resultados cognitivos de la formación. Comisión de las Comunidades Europeas. Programa Eurotecnet. 1992
2. Ariza, A. y otros. CERT: un modelo matemático y tecnológico de la evaluación. Píxel-Bit. Revista de Educación. 11, 51-56, 1998
3. Ariza, A. y otros. 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, 2001
4. Ariza, A. El software en la realización de pruebas objetivas. Revista Espacio y Tiempo. 19, 131-142, 2005
5. Ariza, A. Pruebas objetivas de respuesta múltiple con soporte informático. ISBN 84-86849-46-2. ICE Universidad de Sevilla, 2006
6. Cabrera F.A. y Espín J.V. Medición y evaluación educativa. Barcelona PPU 1986
7. De Juan, J. Introducción a la enseñanza universitaria. Didáctica para la formación del profesorado. Madrid, Dykinson. 1996
8. De Ketele J.M. Observar para educar. Observación y evaluación en la práctica educativa. Madrid, Visor Libros. 1984
9. De Pablos, J. y otros. La evaluación del alumno en la universidad: el proyecto CERT. Revista de Enseñanza Universitaria. 6, 49-71, 1993

10. Leclercq, D. Evaluation in education: An international review series. Oxford, England, Pergamon Press. 1983
11. Lord. F.M. Applications of item response theory to practical testing problems. Hisdale, Nueva York, academic Press. 1980
12. Muñiz J. Teoría clásica de los tests, Madrid, Ed. Pirámide S.A., 2003
13. Torreblanca J. y otros. La utilización de las bases de datos como herramientas de evaluación. *Píxel-bit. Revista de Medios y Educación.* 7, 73-82, 1996