

EVALUACIÓN EN ASIGNATURAS DE MOTORES TÉRMICOS EMPLEANDO TESTS DE NIVEL DE CONFIANZA

PATINHO VILAS, David⁽¹⁾; GONZÁLEZ GIL Arturo⁽²⁾

PAZ PENÍN, Concepción⁽¹⁾; MÍGUEZ ÁLVAREZ, Carla⁽¹⁾

patinho@uvigo.es

⁽¹⁾Universidad de Vigo, Escuela de Ingeniería Industrial

⁽²⁾Centro Universitario de la Defensa en la Escuela Naval Militar de Marín

RESUMEN

La evaluación del conocimiento adquirido por los estudiantes es probablemente una de las tareas más complejas a las que se deben enfrentar los profesores. Dicha complejidad radica en la necesidad de proponer pruebas justas y adecuadas que permitan una evaluación objetiva de los resultados. Si además el grupo de trabajo es muy numeroso, el tiempo empleado en la preparación y corrección de dichas pruebas es elevado, lo que a menudo evita la posible realización de múltiples pruebas, reduciendo la valoración del alumnado a una única prueba cuya representatividad es, cuanto menos cuestionable. Por esta razón, al trabajar con grupos grandes se recurre a menudo a la realización de pruebas de evaluación tipo test, con respuesta múltiple de fácil y rápida corrección.

Uno de los inconvenientes de las pruebas tipo test de respuesta múltiple es establecer un sistema de puntuación adecuado. El método de evaluación propuesto en este trabajo para 3 materias de Máquinas y Motores Térmicos en los grados de Ingeniería Mecánica e Ingeniería de la Energía de la Universidad de Vigo se basa en los siguientes pilares. Por un lado la realización de diversas pruebas parciales a lo largo del curso con la idea ayudar al alumnado a que trabaje la materia poco a poco, a medida que avanza el curso. Para que esto no represente una carga de trabajo excesiva al docente, se propone el empleo exclusivo de pruebas tipo test. Tanto para la evaluación de los conocimientos teóricos como los prácticos (resolución de problemas y casos). Así mismo, para asegurarse que el alumno prepara toda la materia y para obtener información de cómo responde a aquellas preguntas de las que no está tan seguro, será obligatorio responder a todas las cuestiones. La variación con respecto a otros métodos radica en que, para evitar una exigencia excesiva y obtener información de cómo de bien o mal responden los alumnos ante cuestiones en las que no conocen con certeza la respuesta, se propone un sistema complementario de nivel de confianza semejante al presentado en [3]. El alumno responderá el test indicando además el nivel de confianza que tiene en su respuesta, con 3 niveles (30%, 70% y 100%) y obtendrá una puntuación positiva o negativa consecuenta con la certeza y confianza mostrada.

Palabras clave: Evaluación, Test, Confianza, Ingeniería, Motores Térmicos.

1. Introducción

Uno de los grandes inconvenientes de las pruebas tipo test de respuesta múltiple es establecer un sistema de puntuación adecuado [1-2]. Dependiendo de la elección de la puntuación el examen será más o menos complejo para el alumno, lo que necesariamente no informará mejor al docente a cerca de su conocimiento de la materia. Si sólo se puntúan adecuadamente los aciertos pero no se tienen en cuenta los fallos la posibilidad de éxito del alumnado no cualificado puede aumentar en exceso. La opción más adoptada suele ser la puntuación positiva de los aciertos y la negativa de los fallos, pudiendo ésta última ser minorada si se estima conveniente. En ambos casos, es habitual que las preguntas no contestadas no tengan ningún efecto, de modo que se favorece el hecho de que el alumno responda sólo aquellas cuestiones que ciertamente conoce. Por desgracia, ambos sistemas expuestos no conceden al docente ningún tipo de información acerca de cómo respondería el alumnado frente aquellas cuestiones en las que tiene dudas o ninguna idea. Esto resulta poco conveniente, especialmente en asignaturas técnicas de ingeniería. A lo largo de su vida profesional los ingenieros se encuentran muchas veces ante problemas o proyectos cuyo modo de resolución exacto desconocen a priori. En esas ocasiones deberán tomar decisiones y realizar cálculos con poca información y seguridad en base a su intuición y guiándose por sus conocimientos generales. Por esta razón puede resultar interesante evaluar la capacidad del alumnado ante situaciones similares con un test.

Con esa idea, hay un último tipo de puntuación que valora positivamente los aciertos, negativamente los fallos pero obliga al alumnado a contestar todas las cuestiones. Este es probablemente el test más completo que se puede realizar, pero a la vez el más exigente para los estudiantes.

El método de evaluación propuesto en este trabajo se basa en los siguientes pilares. Por un lado la realización de diversas pruebas parciales durante el transcurso del curso con la idea de ayudar al alumnado a que trabaje la materia poco a poco. Para que esto no represente una carga de trabajo excesiva al docente, se propone el empleo exclusivo de pruebas tipo test para la evaluación de los conocimientos teóricos y prácticos (resolución de problemas y casos). Así mismo, para asegurarse de que el alumnado prepara toda la materia y para obtener información de cómo responde a aquellas preguntas de las que no está tan seguro, será obligatorio responder a todas las cuestiones. La variación con respecto a otros métodos radica en que para evitar una exigencia excesiva y obtener información de cómo de bien o mal responden ante cuestiones en las que no conocen con certeza la respuesta, se propone un sistema complementario de nivel de confianza semejante al presentado por Pablos Pons et al. [3]. Se responderá el test indicando además el nivel de confianza que tiene en su respuesta, con 3 niveles (30%, 70% y 100%) y obtendrá una puntuación positiva o negativa consecuenta con la certeza y confianza mostrada.

2. Grupos de trabajo y metodología

2.1. Grupo de estudio pequeño (G1)

La primera materia en la que se ha probado este método es Motores y Turbomáquinas Térmicas (curso 2012/13) del grado de Ingeniería de la Energía de la Universidad de Vigo. Se trata de una asignatura obligatoria del tercer curso cuyo tamaño medio de grupo para el estudio fue de 30 individuos. Al tratarse de unos estudios ya adaptados al EEES, el alumnado está acostumbrados a una docencia más guiada y con evaluación continua, de modo que es la situación ideal para la que se ha planificado este método.

A lo largo del cuatrimestre se realizaron tres pruebas parciales tipo test. Cada una correspondiente a un bloque de la asignatura. Al tratarse del primer grupo de estudio en el que se intenta poner a punto la metodología, se pidió al alumnado que contestara a los cuestionarios de dos modos diferentes. El primero más tradicional en el que las respuestas correctas valían 1 punto y las falladas -1 punto. En este test no era obligatorio contestar todas las cuestiones, y las que dejaban en blanco simplemente no puntuaban. En el segundo test, con las mismas cuestiones debían contestar obligatoriamente a todas las preguntas e indicar a su vez la confianza que tenían en la respuesta con tres niveles de

aproximación (100%, 70%, 30%). La puntuación asignada a priori era la misma para los casos negativos que positivos (+1, +0.7, +0.3, -1, -0.7, -0.3). Uno de los primeros objetivos de este grupo estudio era ajustar las puntuaciones del modo más adecuado.

Para que la evaluación no se viera perjudicada por la realización de estas pruebas, se informó al alumnado que la nota final que obtendría sería la mejor de los 2 test realizados. De este modo, al tratarse de una situación beneficiosa para ellos se consiguió la participación de todo el grupo sin ningún tipo de problemas. Por supuesto, las respuestas deberían ser las mismas en ambos cuestionarios, excepto cuando en el primero se dejaba en blanco. Esto es necesario para evitar que aumenten estadísticamente sus probabilidades de éxito al disponer de 2 intentos en la misma prueba.

Como reseña final, cabe destacar que el tipo de test empleado facilitaba 4 respuestas para la elección, de las cuales sólo 1 era correcta. Las preguntas tenían indistintamente contenido teórico y problemas, cuya resolución debe ser entregada conjuntamente con el examen. Así mismo, para superar la materia mediante la evaluación continua era necesario superar los 3 parciales (nota mínima de 5 sobre 10). El resumen de las características del grupo puede verse en la Tabla 1.

2.2. Grupo de estudio grande (G2)

Una vez evaluado el método de trabajo en el grupo 1 se probó en una asignatura con un mayor alumnado y cuya evaluación se lleva a cabo exclusivamente con un examen final. La materia Ingeniería Térmica I (2012/13) es troncal del título LRU de Ingeniero Industrial de la Universidad de Vigo. Se trataba de una asignatura de 4º curso con cerca de 379 estudiantes matriculados de los cuales se presentaron al primer examen 284. Al igual que en el caso anterior el alumnado debía cubrir 2 test diferentes (tradicional y con confianza) de los cuales se les atribuiría la mejor nota. En esta ocasión la puntuación del test de confianza ya había sido ajustada previamente a los valores (+1, +0.7, +0.3, -1, -0.5, -0.2). También se trabaja con 4 respuestas válidas de las cuales sólo una es correcta.

La gran diferencia con el caso anterior es que en esta ocasión el test únicamente evaluaba la parte teórica de la asignatura, aportando un 35% de la nota final. Así, las 2 pruebas realizadas (parcial 1 y 2) se corresponden a las convocatorias oficiales de la asignatura y no realmente a pruebas parciales. Es decir, en el segundo parcial sólo participaban aquel alumnado que no había superado la materia en la primera convocatoria. Los detalles resumidos pueden verse en la Tabla 1.

2.3. Grupo de consolidación (G3)

Finalmente, una vez ajustado y evaluado el método, se aplica como medida única de evaluación a una materia completa. El tercer grupo de estudio es la materia Motores y Máquinas Térmicos (2013/14) del grado de Ingeniería Mecánica de la Universidad de Vigo. Las características de este grupo son muy semejantes a la del G1. El contenido y descriptores de las guías docentes son semejantes y el número de asistentes (35 en este caso) también. La diferencia radica en que en este caso ya sólo se evaluará al alumnado con el test de confianza.

Se divide la materia en 3 bloques y cada uno de ellos se evalúa en un parcial a lo largo del curso que contiene preguntas teóricas y problemas. Se presentarán 4 opciones posibles de respuestas de las cuales sólo una es correcta. El modo de puntuación simplemente se hereda de cómo había sido fijada en los grupos anteriores. Básicamente este grupo permitirá el análisis e impresiones del alumnado una vez sometida su evaluación única y exclusivamente a esta nueva metodología. Al igual que en los casos anteriores puede verse el resumen y detalles del grupo en la Tabla 1.

Tabla 1. Características de los grupos de estudio

	G1	G2	G3
Título	Ingeniería de la Energía	Ing. Industrial	Ingeniería Mecánica

Curso	3° Grado			4° LRU		4° Grado		
Trad.	Si			Si		No		
Total	Si			No (35%)		Si		
Prueba	G 1.1	G 1.2	G 1.3	G 2.1	G 2.2	G 3.1	G 3.2	G 3.3
N	3 0	3 0	3 0	2 84	1 03	3 5	3 3	3 4
n	6 00	6 00	6 90	3 408	1 030	8 75	8 25	6 80
Trad.=Posibilidad de realizar también el test tradicional; Total=La puntuación total de la materia depende del test (%); N=Número de alumnos presentados; n=Número de preguntas respondidas por test								

2.4. Metodología

La metodología experimental empleada con cada grupo ya ha sido descrita en los epígrafes anteriores. Simplemente se aclara en esta sección que para el tratamiento, análisis y procesado de datos se ha empleado un código propio desarrollado en la plataforma comercial Matlab. El código desarrollado permite automatizar la corrección de los exámenes una vez introducidas las matrices de solución y respuesta, así como el posterior análisis y consulta de las estadísticas más relevantes.

3. Resultados y discusión

Las Tablas 2 y 3 muestran las estadísticas de los 2 primeros grupos de estudio G1 y G2 que sirven de puesta a punto del método. En la Tabla 2 se aprecia que los resultados parecen mejorar ligeramente a medida que el alumnado conoce mejor el tipo de examen, pero el efecto es pequeño y también podría ser atribuido a la propia naturaleza o dificultad del test.

Tabla 2. Resultados globales de los grupos G1 y G2

	G1 .1	G1 .2	G1 .3	G2 .1	G2 .2
Nº de Tests	30	30	30	28 4	10 3
R2 (%)	16. 7	43. 3	40. 0	36. 6	55. 3
Aprob. R1 (%)	70. 0	80. 0	83. 3	25. 0	64. 1
Aprob. R2 (%)	63. 3	76. 6	76. 6	18. 3	58. 3
R2 (%)= Porcentaje de test con mejor puntuación con el nuevo método					

Tabla 3. Estadísticas de los grupos G1 y G2

	G 1.1	% G 1.2	% G 1.3	% G 2.1	% G 2.2
Nº de Tests	3	3	3	2	1

	0		0		0		84		03	
N° Preguntas	600		600		690		3408		1030	
Seguras 100%	380	63	442	74	484	70	1760	52	641	62
No seguras	220	37	158	26	206	30	1648	48	389	38
Seguras fall.	69	18	58	13	43	9	677	38	86	13
No seg. acert.	138	63	91	58	128	62	696	42	230	59
No seg. 70%	123	56	91	58	112	54	963	58	228	59
No seg. 30%	97	44	67	42	94	46	685	42	161	41
No entendió	1	3	1	3	0	0	10	4		0

En términos generales, la primera materia tiene un elevado índice de aprobados (75%) y aunque los resultados son peores con el nuevo test, la disminución en aprobados es menor del 10%. Esto es asumible ya que se entiende que al contestar todas las preguntas, la exigencia de la prueba aumenta. Para el otro grupo (G2) los resultados son semejantes, pero con un nivel de aprobados bastante inferior.

La Tabla 3 muestra también cómo el índice de respuestas marcadas como seguras pero que el alumno falla no es demasiado elevado. Normalmente oscila entre 0-18%. Este punto es importante y habla del buen diseño de las pruebas. Las cuestiones no deben ser demasiado engañosas como para generar demasiadas respuestas 100% erróneas. Del mismo modo, existe un porcentaje elevado de acierto (en torno a 65%) en preguntas no seguras, lo que indica que el alumnado tenía una buena idea de por dónde se movía. Se recuerda que en el supuesto de que el alumno no tuviera la más mínima idea y hubiera contestado aleatoriamente el porcentaje de aciertos se acercaría al 25% por razones estadísticas (1 respuesta correcta entre 4 posibles).

Ajustado el método, se aplicó al grupo consolidado (G3) como método único de evaluación. Las estadísticas pueden verse en la Tabla 4. El índice de aprobados en esta materia también es bastante elevado, cercano al 70%. Los datos obtenidos son semejantes a los ya mostrados para los grupos anteriores. Un bajo índice de fallos a preguntas seguras (15%), en torno a un 50% de acierto en las preguntas no seguras, siendo conscientes de ello en la mayoría de los casos, ya que escogían más la opción del 70% que la de 30% cuando no lo tenían claro. Como avance en esta materia, también cabe destacar que todos los alumnos fueron capaces de cubrir adecuadamente y entender el proceso de evaluación, cosa que en las anteriores ocasiones siempre generó algún problema.

Tabla 4. Estadísticas del grupo G3

	G	%	G	%	G	%
	3.1		3.2		3.3	

N° de Tests	3		3		3	
	5		3		4	
N° Preguntas	8		8		6	
	75		25		80	
Seguras 100%	6	7	6	7	4	7
	22	1	42	8	98	3
No seguras	2	2	1	2	1	2
	53	9	83	2	82	7
Seguras fall.	7	1	7	1	7	1
	1	1	4	2	1	4
No seg. acert.	1	5	1	5	8	4
	38	5	01	5	4	6
No seg. 70%	1	5	1	5	1	6
	37	4	01	5	17	4
No seg. 30%	1	4	8	4	6	3
	16	6	2	5	5	6
No entendió	0	0	0	3	0	0

4. Conclusiones

Las conclusiones del trabajo son bastante positivas ya que a lo largo de toda la experiencia el método permitió a los docentes realizar diversas pruebas parciales durante el cuatrimestre (3 en muchos casos), sin tener que dedicar excesivo tiempo a la corrección y elaboración de las pruebas. La base de datos de preguntas se va actualizando y mejorando, de modo que de cara al futuro cada vez será más sencillo aplicar el método. Por tanto, se permite la evaluación múltiple con recursos mínimos y revisiones de exámenes sencillas.

Los resultados obtenidos no distan demasiado a los que produciría una puntuación clásica en pruebas tipo test, pero las nuevas pueden proporcionar más información acerca de los conocimientos reales del alumno. Está obligado a preparar toda la materia y a dar respuestas aunque no tenga clara su certeza, desarrollando su razonamiento crítico y entrenándole en la toma de decisión. Además las puntuaciones obtenidas con el nuevo sistema generan mayor nivel de detalle en la nota, permitiendo clasificar mejor el resultado del alumnado.

Como puntos débiles, el método puede ser muy dependiente de la complejidad de las preguntas. Cuestiones muy intrincadas puedan dar la falsa sensación de seguridad y luego ser erróneas, lo que tiene un importante impacto sobre el resultado final. No obstante, la puntuación clásica no está exenta de este factor.

En lo que refiere a los alumnos, existe cierta reticencia al nuevo método debido a su novedad. No están acostumbrados a este tipo de evaluación ni tampoco a verse obligados a contestar cuando no tienen clara la respuesta. Esto puede provocar en ellos la sensación de que cuando fracasan (mal resultado) la culpa no es suya, sino del método. Esa reacción es peligrosa ya que cuesta hacer entender al alumno que debe preparar mejor la materia, ya que el problema no es del método, como se puede demostrar con los resultados de sus compañeros.

Una consecuencia del razonamiento anterior es que el alumnado menos cualificado a veces se preocupa más de elegir sus respuestas en función de lo que considera más seguro matemáticamente y pierde más el tiempo preocupado en buscar el aprobado que en responder adecuadamente a las preguntas. Aunque este comportamiento también tiene lugar en los test con puntuación normal, aquí puede verse acrecentado dado el incremento en el número de variables que entran en juego.

5. Referencias

- [1] ALONSO, C. y Otros. (1992). *Principios comunes para la evaluación de los resultados cognitivos de la formación*. Comisión de las Comunidades Europeas. Programa EUROTECNET. Barcelona: Universidad de Barcelona
- [2] LECLERCQ, D. (1983). *Evaluation in education: an international review series*. Oxford, England: Pergamon Press.
- [3] DE PABLOS, J. y OTROS. (1993). *La evaluación del alumno de la Universidad: el proyecto CERT*. Revista de enseñanza universitaria, número 6, pp 49-71.