

# HOW TO INTERPRET THE STUDENTS' MULTIPLE-CHOICE TEST RESULTS

Peiró, R.<sup>1</sup>; Guijarro, E.<sup>1</sup>; Verdeguer, M.<sup>1</sup> and Reig, C.<sup>1\*</sup>

<sup>1</sup>Universitat Politècnica de València (SPAIN)

ropeibar@btc.upv.es, esguitar@doe.upv.es, merversa@eaf.upv.es, mareiva@prv.upv.es

## Resumen

El Espacio Europeo de Educación Superior (EEES) es un área de la organización educativa iniciada con la Declaración de Bolonia, fomentando la evaluación del proceso de aprendizaje de los estudiantes. Esta evaluación ha de ser continua, formativa, variada, justa, conocida y exitosa. La prueba final, tanto oral como escrita, ha sido utilizada generalmente no sólo en la metodología tradicional de la enseñanza universitaria, sino también en el EEES. La prueba final se considera un buen complemento a la evaluación continua que permite evaluar el conocimiento de la asignatura. La prueba objetiva es una de las pruebas utilizadas para evaluar a los estudiantes. A pesar de que la corrección es completamente objetiva, la formulación de las preguntas y la nota mínima necesaria para superar la prueba son decisiones subjetivas de los profesores.

Fruticultura es una asignatura optativa, de 6 ECTS, ubicada en el cuarto año de Ingeniería agrónoma de la Escuela Técnica Superior de Ingenieros Agrónomos de la Universitat Politècnica de València. En este trabajo se analizan las pruebas objetivas realizadas en dicha asignatura. Concretamente, se analizan un total de 40 preguntas de opción múltiple a partir de las respuestas de 50 alumnos correspondientes al curso académico 2011/2012. Todas las preguntas de opción múltiple tienen cuatro opciones de respuesta siendo sólo una la correcta. Todas las preguntas tienen la misma ponderación, y la puntuación bruta se calcula para cada sección en función del número de preguntas que el estudiante contesta correctamente (1) o incorrectamente (0).

Este trabajo realiza un análisis descriptivo de cada una de las preguntas de opción múltiple, así como los coeficientes de dificultad, la discriminación y la fiabilidad, para evaluar la calidad de las preguntas formuladas en las preguntas objetivas. Los diferentes resultados se obtuvieron en función de la sección. Las secciones teóricas obtienen un mayor porcentaje de respuestas correctas que las secciones prácticas / aplicativas. Los coeficientes de prueba estimados estuvieron dentro del rango de los coeficientes aceptables para evaluar el rendimiento académico.

El objetivo de este estudio es evaluar la calidad de las preguntas objetivas tipo test utilizados como parte de la evaluación con el fin de mejorar su calidad, si es necesario, y para obtener información acerca de los aciertos y errores de los alumnos.

Palabras Clave: Multiple-choice questions, evaluation, improve quality questions.

## 1 INTRODUCCIÓN

Uno de los principales cambios que han tenido que asumir tanto profesores como alumnos en el sistema actual de créditos ECTS que se ha implantado como consecuencia del Espacio Europeo de Educación Superior (EEES) es la reducción de la presencialidad del alumno en el aula, así como un cambio en el proceso de enseñanza-aprendizaje. En este nuevo marco del proceso de enseñanza-aprendizaje se incentiva realizar una evaluación del proceso de aprendizaje del alumnado, de manera que la evaluación del alumno debe ser continua, formativa, variada, justa, conocida y adaptada al contexto (Morales 2007; Morales, 2009). Resultados previos indican que si la evaluación se realiza a lo largo del periodo de enseñanza-aprendizaje (docente) es mucho más útil, en vez de sólo al finalizar éste, ya que permite corregir los errores detectados y reforzar los conceptos, aptitudes y actitudes que se observan, así como ajustar o modificar el ritmo del proceso de enseñanza-aprendizaje (Secretaría de Estado de Universidades, 2010). Sin embargo, un examen o prueba final, oral o escrita, al finalizar el periodo se considera un buen complemento a la evaluación continua ya que permite valorar de forma conjunta los conocimientos de la materia, permitiendo obtener una visión global (Santos, 1999). Entre las pruebas que se recomiendan para facilitar el aprendizaje en la evaluación se recomiendan (Secretaría de Estado de Universidades, 2010) las pruebas objetivas (verdadero/falso o de elección múltiple), las pruebas de respuesta breve, las pruebas de respuesta

larga o de desarrollo, las pruebas orales (individuales, en grupo, presentaciones...), los trabajos y proyectos, los informes o memorias de prácticas, las pruebas de ejecución de tareas reales o simuladas, la autoevaluación (oral, escrita, individual, en grupo), las escalas de actitudes y las técnicas de observación.

Una de las pruebas más utilizadas en la universidad son las pruebas objetivas ya que permiten cubrir la mayoría de contenidos propuestos en una asignatura si se dedica tiempo a la elaboración del examen; se pueden evaluar conceptos y procedimientos, incluyendo operaciones cognitivas de mayor dificultad. Para que las pruebas objetivas sean un procedimiento adecuado de evaluación deben recoger toda la información que se considere relevante y significativa para determinar el grado de conocimiento adquirido por los estudiantes. Es una prueba que implica mucho tiempo de elaboración; sin embargo, implica menor tiempo en su aplicación y corrección. Normalmente se considera que las preguntas tipo test comprueban mayoritariamente los conocimientos de memoria, aunque esto depende del tipo de preguntas que se formulen (Secretaría de Estado de Universidades, 2010). También es habitual asumir que este tipo de pruebas evitan el sesgo en las evaluaciones debidas a la subjetividad de los evaluadores; hay que tener en cuenta que la corrección y puntuación de las preguntas es objetiva (por ejemplo, dos preguntas correctas restan una mal contestada) pero tanto la formulación de la pregunta (que y como se pregunta) como dónde está el mínimo para el aprobado son decisiones subjetivas del profesor. Por tanto, la calidad de las preguntas (objetivas o de otro tipo) es un tema relevante si se tiene en consideración que el qué y cómo estudia un alumno depende del tipo de prueba y de pregunta esperado. Se ha propuesto el análisis de las preguntas objetivas para mejorar la calidad de éstas, de manera que faciliten la autoevaluación del profesor e informen más específicamente del proceso de aprendizaje de los alumnos (Morales, 2012).

## 2 INNOVACIÓN DOCENTE

Fruticultura es una asignatura obligatoria para todas las especialidades de 6,0 créditos que se imparte en el segundo ciclo de la Titulación de Ingeniero Agrónomo de la Universitat Politècnica de Valencia. A pesar de ser una asignatura de una titulación del plan antiguo, desde hace varios años se está intentado adaptar a las nuevas necesidades del EEES debido a que será una asignatura optativa de 4,5 ECTS en el Grado en Ingeniería Agroalimentaria y del Medio Rural, a excepción de los alumnos que elijan la intensificación de "Producción Vegetal" para los que será obligatoria de la intensificación. El presente curso, 2012-2013, es el primero en el que la asignatura se imparte en el Grado, ya que es una asignatura de tercer curso. Aunque año tras año se ha variado el sistema de evaluación de la asignatura para hacerlo más adecuado a una enseñanza formativa a lo largo del curso, los aspectos a evaluar en los últimos años han sido la presentación de actividades y problemas, la asistencia y participación en clases teóricas y prácticas, diversas pruebas de conocimientos (tipo test) a lo largo del curso y un examen final de conocimientos teóricos y prácticos.

El examen final de conocimientos de la asignatura se realiza con una intención integradora de todos los aspectos del curso y está compuesta de distintos apartados con preguntas tipo test de respuestas múltiple, preguntas de desarrollo corto, tareas realizadas durante el curso y preguntas de identificación del material vegetal (especies y/o variedades). Las preguntas cortas de teoría suponen el 45% de la nota final y el reconocimiento del material vegetal el 30%. Para poder aprobar la asignatura se exige la obtención de un 5 sobre 10 para poder mediar con el resto de apartados de la evaluación. Un 20% de la nota corresponde al tipo test de respuestas múltiple y el 5% restante a las tareas y actividades que se han desarrollado durante el curso.

El objetivo del presente trabajo es evaluar la calidad de las preguntas objetivas de respuesta múltiple que se utilizan para la evaluación de la asignatura con la finalidad de mejorar su calidad (si es necesario) y para tener una información más específica sobre los aciertos y errores de los alumnos, de manera que ayude a mejorar la calidad de la enseñanza del profesor y del aprendizaje de los alumnos. Para ello se han analizado las repuestas de 30 preguntas de la parte del examen final tipo test de respuestas múltiples de un total de más de 70 alumnos.

En cada una de las preguntas se especifica que sólo hay una respuesta correcta. La clave de corrección para las preguntas ha sido la siguiente: las respuestas acertadas suman 1 punto, las respuestas erróneas o no contestadas ni suman ni restan. Se han realizado los análisis descriptivos de cada una de las preguntas del tipo test. Para realizar los diferentes análisis de fiabilidad se han tabulado las respuestas obtenidas por todos los alumnos y se han ordenado los alumnos en función del número de respuestas correctas (no por la puntuación total de la prueba) y se seleccionan aproximadamente el 25% con puntuación más alta (grupo superior formado por 14 alumnos) y el 25%

con puntuación más baja (grupo inferior formado por 14 alumnos). El índice de dificultad se define como la suma de alumnos del grupo superior (AS) y del grupo inferior (AI) que aciertan entre la suma del número de alumnos que forman cada grupo (N);  $AS+AI/2N$ . El índice de discriminación 1 es la diferencia entre dos proporciones, los alumnos que aciertan del grupo superior menos los del grupo inferior:  $(AS/N)-(AI/N)$ ; como los denominadores son iguales (idéntico número de alumnos en cada grupo) la fórmula queda simplificada. El índice de discriminación 2, menos utilizado, indica la proporción de aciertos en el grupo superior con respecto al número total de acertantes  $(AS/AS+AI)$ . El índice de Cronbach se realizó utilizando el paquete estadístico SPSS (2007).

### 3 RESULTADOS Y DISCUSIÓN

En la Tabla 1 se observa el porcentaje de alumnos que han contestado cada una de las opciones de cada pregunta. Los alumnos han contestado aproximadamente el 82% de las preguntas correctamente, siendo el porcentaje de preguntas contestadas erróneamente del 18%, ya que ninguno de los alumnos ha dejado en blanco ninguna pregunta. La pregunta con menor porcentaje de respuestas correctas ha sido la pregunta 5; el 53% de los alumnos la ha contestado correctamente. Sin embargo los alumnos que no han contestado correctamente a esta pregunta solo han escogido dos de las tres opciones erróneas, descartando totalmente la opción D. Además, en esta pregunta, los alumnos que han contestado erróneamente han elegido mayoritariamente una de las opciones como correcta (43% de la A frente al 4% de la B). Las preguntas 1, 7, 11, 16, 17 y 24 han sido contestadas correctamente por menos del 75% de los alumnos. De estas preguntas, a excepción de la 16, los alumnos que han contestado erróneamente han elegido en diferentes proporciones todas las opciones posibles. Dado que las preguntas formuladas en estos temas implican un alto grado de comprensión y análisis de conceptos aplicados se esperaba, en general, un menor porcentaje de acierto en las respuestas. El resto de preguntas, 23, han sido contestadas correctamente por aproximadamente el 86% de los alumnos. En seis de estas preguntas solo una de las opciones posibles ha sido elegida por los alumnos que han contestado erróneamente, siendo también seis el número de preguntas en las que todas las opciones han sido elegidas por alguno de los alumnos. Doce de las veintitrés preguntas restantes están relacionados con aspectos teóricos básicos; concretamente siete de ellas con la definición y relación de conceptos y las otras cuatro con la comprensión y/o razonamiento de aspectos teóricos. Por último, destacar que ninguna de las preguntas ha sido contestada correctamente por todos los alumnos.

Tabla 1. Porcentaje de alumnos que han respondido a cada una de las opciones múltiples (A, B, C y D) de las 30 preguntas (P) de las que consta el test.

P	A (%)	B (%)	C (%)	D (%)	P	A (%)	B (%)	C (%)	D (%)
1	11	21	<b>66</b>	2	16	0	34	4	<b>62</b>
2	0	<b>82</b>	9	9	17	13	<b>72</b>	9	6
3	4	<b>96</b>	0	0	18	15	2	<b>83</b>	0
4	<b>85</b>	15	0	0	19	2	2	0	<b>96</b>
5	43	4	<b>53</b>	0	20	<b>98</b>	0	2	0
6	6	15	<b>79</b>	0	21	<b>85</b>	9	6	0
7	8	14	<b>70</b>	8	22	8	<b>92</b>	0	0
8	8	13	<b>77</b>	2	23	4	<b>77</b>	19	0
9	<b>94</b>	2	2	2	24	<b>74</b>	9	2	15
10	6	<b>89</b>	2	4	25	11	<b>85</b>	4	0
11	<b>68</b>	19	11	2	26	8	<b>75</b>	8	9
12	7	9	9	<b>75</b>	27	<b>77</b>	8	11	4
13	<b>91</b>	0	7	2	28	<b>87</b>	2	11	0
14	2	<b>98</b>	0	0	29	0	6	<b>94</b>	0
15	17	0	<b>83</b>	0	30	0	<b>94</b>	2	4

Nota: La respuesta correcta está señalada en negrita.

Al realizar estos mismos análisis descriptivos en cada uno de los grupos, se observa que los alumnos del grupo superior, 14 alumnos, han respondido correctamente el 96% de las preguntas, mientras que el grupo inferior, formado también por 14 alumnos, han respondido correctamente al 62% de las preguntas (Tabla 2 y 3, respectivamente). En el grupo superior todas las opciones correctas de cada pregunta han sido elegidas por la mayoría de los alumnos (Tabla 2). Dos tercios de las preguntas han sido contestadas correctamente por todos los alumnos del grupo superior (20/30), a excepción de la pregunta, 5, en la que al menos el 79% de estos alumnos han contestado correctamente a las

cuestiones. Solo en el caso de la pregunta 6, los alumnos que han respondido erróneamente lo han hecho eligiendo dos de las posibles opciones erróneas, ya que en el resto de preguntas sólo lo hicieron eligiendo una.

Tabla 2. Porcentaje de alumnos del grupo superior que han respondido a cada una de las opciones múltiples (A, B, C y D) de las 30 preguntas (P) de las que consta el test.

P	A (%)	B (%)	C (%)	D (%)	P	A (%)	B (%)	C (%)	D (%)
1	0	7	<b>93</b>	0	16	0	7	0	<b>93</b>
2	0	<b>86</b>	14	0	17	0	<b>100</b>	0	0
3	0	<b>100</b>	0	0	18	0	0	<b>100</b>	0
4	<b>93</b>	7	0	0	19	0	0	0	<b>100</b>
5	43	0	<b>57</b>	0	20	<b>100</b>	0	0	0
6	7	14	<b>79</b>	0	21	<b>100</b>	0	0	0
7	0	0	<b>100</b>	0	22	0	<b>100</b>	0	0
8	0	0	<b>100</b>	0	23	0	<b>100</b>	0	0
9	<b>100</b>	0	0	0	24	<b>93</b>	0	0	7
10	7	<b>93</b>	0	0	25	7	<b>93</b>	0	0
11	<b>100</b>	0	0	0	26	0	<b>100</b>	0	0
12	0	0	0	<b>100</b>	27	<b>100</b>	0	0	0
13	<b>100</b>	0	0	0	28	<b>93</b>	0	7	0
14	0	<b>100</b>	0	0	29	0	0	<b>100</b>	0
15	0	0	<b>100</b>	0	30	0	<b>100</b>	0	0

Nota: La respuesta correcta está señalada en negrita.

Tabla 3. Porcentaje de alumnos del grupo inferior que han respondido a cada una de las opciones múltiples (A, B, C y D) cada una de las 30 preguntas (P) de las que consta el test.

P	A (%)	B (%)	C (%)	D (%)	P	A (%)	B (%)	C (%)	D (%)
1	22	14	<b>64</b>	0	16	0	71	0	<b>29</b>
2	0	<b>72</b>	14	14	17	14	<b>43</b>	36	7
3	14	<b>86</b>	0	0	18	50	7	<b>43</b>	0
4	<b>71</b>	29	0	0	19	7	7	0	<b>86</b>
5	57	0	<b>43</b>	0	20	<b>93</b>	0	7	0
6	7	14	<b>79</b>	0	21	<b>50</b>	36	14	0
7	14	29	<b>50</b>	7	22	29	<b>71</b>	0	0
8	14	29	<b>50</b>	7	23	14	<b>22</b>	64	0
9	<b>86</b>	7	0	7	24	<b>29</b>	29	7	35
10	7	<b>79</b>	7	7	25	36	<b>64</b>	0	0
11	<b>36</b>	43	14	7	26	21	<b>43</b>	7	29
12	0	36	14	<b>50</b>	27	<b>43</b>	21	29	7
13	<b>93</b>	0	0	7	28	<b>64</b>	7	29	0
14	0	<b>100</b>	0	0	29	0	21	<b>79</b>	0
15	36	0	<b>64</b>	0	30	0	<b>86</b>	7	7

Nota: La respuesta correcta está señalada en negrita.

En el grupo inferior, por el contrario, solo una de las preguntas, la número 14, ha sido contestada correctamente por todos los alumnos. Además, sólo en veinticinco de las treinta preguntas la opción correcta ha sido elegida por la mayoría. De las cinco preguntas restantes en las que la opción mayoritaria no ha sido la correcta, la pregunta con menor porcentaje de alumnos que han contestado correctamente es la pregunta 23, con tan sólo un 22%. De estas cinco preguntas, en dos de ellas los alumnos han elegido únicamente una opción como la correcta, preguntas 5 y 16, mientras que en las otras dos o tres preguntas las respuestas han sido dadas por algún alumno del grupo inferior. A pesar de ello, aproximadamente un tercio de las preguntas han sido contestadas correctamente por más del 75% de los alumnos, siendo las opciones elegidas las correctas.

Tabla 4. Índices de dificultad (Df) y discriminación (Dc<sub>1</sub> y Dc<sub>2</sub>) de cada una de las 30 preguntas y del test.

P	Df (%)	Dc <sub>1</sub> (%)	Dc <sub>2</sub> (%)	P	Df (%)	Dc <sub>1</sub> (%)	Dc <sub>2</sub> (%)
1	79	29	59	16	61	64	76
2	79	14	55	17	71	57	70
3	93	14	54	18	71	57	70
4	82	21	57	19	93	14	54
5	50	14	50	20	96	7	52
6	79	0	50	21	68	64	74
7	75	50	67	22	86	29	58
8	75	50	67	23	61	79	82
9	89	21	56	24	61	64	76
10	86	14	54	25	79	29	59
11	71	57	70	26	71	57	70
12	75	50	67	27	71	57	70
13	96	7	52	28	79	29	59
14	100	0	50	29	89	21	56
15	82	36	61	30	93	14	54
				Test	79	34	62

Los índices de dificultad y discriminación se calculan utilizando únicamente los grupos superior e inferior. El índice de dificultad del test indica la proporción de aciertos en la muestra de alumnos; siendo en este caso de 79%. Todas las preguntas presentan un índice de dificultad superior al 50% y sólo una de ellas (la número 14) presenta un valor del 100%. Esta pregunta podría ser eliminada del test ya que todos los alumnos de ambos grupos la responden correctamente a excepción de uno de ellos. Habitualmente se utilizan dos índices de discriminación; Dc<sub>1</sub>, que es la diferencia entre los dos grupos evaluados y Dc<sub>2</sub>, que es la proporción de aciertos en el grupo superior con respecto al número total de aciertos. A mayor Dc<sub>1</sub> mayor diferencia en el número de preguntas acertadas entre los grupos superior e inferior, por lo que la pregunta es más discriminante. Además, este índice equivale a la estimación de la correlación ítem-total siendo un valor que varía entre -100 y 100. Si una pregunta es muy fácil (todos aciertan) o muy difícil (todos se equivocan) estas preguntas no discriminan ya que el valor obtenido será 0. Ninguna de las preguntas discrimina erróneamente, ya que ninguno de los valores obtenidos es negativo. Dos de las preguntas del test no tienen poder discriminante; la pregunta 14, que es respondida correctamente por todos los alumnos como se ha comentado anteriormente, y la pregunta 5, que es respondida correctamente por el mismo porcentaje de alumnos en ambos grupos. Hay que tener en cuenta que este resultado no implica que las preguntas sean malas, sino que no discriminan. Ninguna de las preguntas es capaz de discriminar correctamente a todos los alumnos ya que ninguna de las preguntas presenta un Dc<sub>1</sub>=100%. Una limitación de este índice es que este valor máximo solo se obtiene si todos los alumnos del grupo superior aciertan y todos los del inferior fallan. Por este motivo también se analiza el índice Dc<sub>2</sub>. Se suele considerar satisfactorio un Dc<sub>2</sub> superior al 50%, es decir, que más de la mitad de los alumnos que aciertan la pregunta pertenezcan al grupo que más sabe. Este índice es independiente del grado de dificultad de la pregunta, por lo que puede llegar al valor 100% si todos los acertantes, aunque sean pocos, pertenecen al grupo superior. En este test todas las preguntas obtienen un valor satisfactorio, al menos del 50%, siendo el promedio del índice de discriminación 2 del 62%. Ninguna de las preguntas presenta el máximo valor.

Tabla 5. Índices de Cronbach de cada una de las 30 preguntas y del test.

P	Cronbach	P	Cronbach	P	Cronbach	P	Cronbach	P	Cronbach	
1	0,801	7	0,776	13	0,796	19	0,783	25	0,781	
2	0,794	8	0,776	14	0,792	20	0,786	26	0,774	
3	0,784	9	0,783	15	0,779	21	0,777	27	0,770	
4	0,793	10	0,784	16	0,780	22	0,780	28	0,780	
5	0,804	11	0,774	17	0,778	23	0,764	29	0,786	
6	0,803	12	0,781	18	0,768	24	0,769	30	0,787	
									Test	0,789

Por último, utilizando las respuestas obtenidas por todos los alumnos en cada pregunta se ha calculado la fiabilidad del test, calculando el coeficiente alfa de Cronbach. Debido a que este análisis no posee un estadístico concreto dependiendo de la finalidad del análisis, se consideran diferentes

valores como aceptables; Tuckman (1999) sugiere que valores de Cronbach de 0.5 o superiores son aceptables para la evaluación de actitudes, mientras que 0.75 es el valor aceptable para evaluar el logro académico. Al realizar el análisis únicamente se ha eliminado una pregunta, porque la varianza era cero. En este caso el coeficiente de Cronbach para todo el test es de 0.789, por lo que es un valor superior al aceptable para la evaluación del logro académico. El alfa de Cronbach no mejoraba de manera importante si se eliminaba alguna de las preguntas (Tabla 5).

## 4 CONCLUSIONES

Existen diversos indicadores que permiten evaluar la calidad de las preguntas que se realizan en la evaluación mediante exámenes tipo test de respuesta múltiple. Con los resultados obtenidos en el presente trabajo se deberían revisar al menos cuatro de las treinta preguntas que se utilizan en la evaluación de los alumnos (pregunta 5, 6 14 y 20). La finalidad de la evaluación de las preguntas es triple; por una parte sirve como forma de autoevaluación del profesor (para poder ir mejorando en sus tareas), como feedback para los alumnos y como otro criterio que pueda influir en los criterios de calificación finales.

## References

- Morales, P (2007), “La fiabilidad de los tests y escalas”, <http://www.upcomillas.es/personal/peter/estadisticabasica/Fiabilidad.pdf>
- Morales, P (2009), “Análisis de ítems en las pruebas objetivas”, <http://cmapspublic2.ihmc.us/rid=1JD9TL205-115LBNW-X6P/AnalisisItemsPruebasObjetivas.pdf>
- Santos, MA (1999), “20 paradojas de la evaluación del alumnado en la universidad española”, *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 2, 1, 369-390.
- Secretaría de Estado de Universidades (2010), “¿Qué es Bolonia?”, Ministerio de Educación, [www.queesbolonia.es](http://www.queesbolonia.es).
- Tuckman, BW (1999), *Conducting educational research*, Wadsworth Group, Belmont.
- SPSS INC (2007), “SPSS”, Chicago, USA, <http://www.spss.com>