

EFFECTO SOBRE EL RENDIMIENTO ACADÉMICO DE LA EVALUACIÓN CONTINUA LLEVADA AL EXTREMO: PRUEBAS SEMANALES ACUMULATIVAS

Desamparados Blázquez
Universitat Politècnica de València
mdeblzso@ade.upv.es

Josep Domènech Universitat
Politécnica de València
jdomenech@upvnet.upv.es

Raúl Peña-Ortiz
Universitat Politècnica de València
rpenya@upvnet.upv.es

María Cinta Vincent Vela
Universitat Politècnica de València
mavinve@iqn.upv.es

Jaime Lloret
Universitat Politècnica de València
jlloret@dcom.upv.es

Elena de la Poza Universitat
Politécnica de València
elpopla@esp.upv.es

Resumen

La forma de evaluación de una asignatura es un arma poderosa que puede utilizarse para dirigir el proceso de aprendizaje. Un sistema de evaluación continua permite proporcionar una retroalimentación continua a los estudiantes, a la vez que aumenta su implicación en el aprendizaje y su motivación.

Este artículo propone un método de evaluación continua basado en pruebas semanales acumulativas, y analiza el efecto de su aplicación en el rendimiento académico de los estudiantes. Los resultados de esta experiencia demuestran que, comparado con la evaluación por examen final, los estudiantes están ahora más motivados, asisten más a clase (ya que es necesario para hacer las pruebas semanales), aprueban más, obtienen mejores notas y, por tanto, aprenden más.

Introducción

El método de evaluación utilizado en un curso puede impulsar la motivación de los estudiantes. La motivación es un elemento clave, la base para el aprendizaje, por lo que es de suma importancia considerarla en el diseño de métodos de evaluación. De hecho, citando a Biggs et al. (2007), "qué y cómo aprenden los estudiantes depende en gran medida, de cómo piensan que van a ser evaluados".

Un examen, como parte del procedimiento de evaluación, es más que simplemente tomar una muestra del aprendizaje de los estudiantes. Las pruebas evaluativas fomentan el aprendizaje, incluso en el caso en el que no se proporciona ninguna realimentación explícita, (Dempster 1991), aunque proporcionar retroalimentación rápida y adecuadamente aumenta la motivación y el aprendizaje en mayor medida (Race 1995, Dochy 2008, Kuo 2009).

Las pruebas de evaluación frecuentes no sólo ofrecen a los estudiantes información periódica sobre su rendimiento, sino que también proporcionan información valiosa sobre el desarrollo del curso a los evaluadores (Yorke 2003). A pesar de las ventajas potenciales de realizar exámenes frecuentes, la literatura no ofrece resultados concluyentes sobre los detalles de implementación de un sistema de evaluación con exámenes frecuentes (Casem 2006, Kling et al. 2005, Kuo 2009). Esto significa que se debe investigar aún más en este campo.

En este trabajo se propone un nuevo método de evaluación en el que el conocimiento de los estudiantes se evalúa semanalmente. Es decir, los contenidos evaluados son

acumulativos: los estudiantes se examinaron de todos los contenidos explicados desde el inicio del semestre en cada una de las pruebas de evaluación. La evaluación del mismo contenido una y otra vez tiene como objetivo promover el aprendizaje de larga duración. Este sistema de evaluación frecuente y acumulativa se aplicó a la asignatura de Microeconomía, una asignatura de primer año del Grado en Administración y Dirección de Empresas, en consonancia con el nuevo modelo de enseñanza y aprendizaje que requiere el Espacio Europeo de Educación Superior (EEES). Este nuevo paradigma educativo se caracteriza por el diseño y la implementación de sistemas de evaluación continua que fomentan una participación más activa en los estudiantes. Ambos objetivos están cubiertos por el método ideado de evaluación masiva.

Así, comparamos este esquema de evaluación con el método de examen final, que se aplicó en años anteriores. Con el fin de llevar a cabo la comparación, se presta atención a algunos de los indicadores de rendimiento académico: el examen y los resultados finales, la tasa de abandono y el índice de aprobados. Los resultados muestran que el nuevo método de evaluación mejora significativamente los indicadores analizados.

El resto del trabajo se organiza de la siguiente manera. La sección 2 presenta una revisión de la literatura relacionada. La sección 3 presenta el contexto en el que se aplicó el sistema de evaluación frecuente. En la Sección 4 se explica el método de investigación, detallando las variables específicas consideradas en el estudio, así como el análisis realizado. Sección 5 discute los resultados del método de evaluación nuevo en contraste con los resultados del régimen de examen final aplicado el año anterior. Por último, se ofrecen algunas conclusiones en la Sección 6.

Antecedentes

En estudios previos, Race (1995) y Brown (2005) resaltan la importancia de la evaluación del alumnado como la tarea más importante del profesor en el proceso de aprendizaje de sus alumnos. Por lo tanto, es de importancia capital idear un método de evaluación eficaz que permita a los estudiantes alcanzar un aprendizaje de calidad y elevar su rendimiento y participación en el curso.

Tradicionalmente, la evaluación se ha clasificado en "sumativa" o en "formativa". Según Trotter (2006), los métodos de evaluación sumativa son aquéllos que están centrados en medir los logros de los estudiantes hasta un punto determinado. Por el contrario, la función principal de la evaluación formativa es proporcionar información a los estudiantes sobre su desempeño, como se describe en Yorke (2003).

Sin embargo, evaluación sumativa y formativa no son excluyentes, ya que un mismo método de evaluación puede presentar ambas perspectivas en mayor o menor grado (Brown et al. 1997). La evaluación proporciona información sobre el rendimiento de los estudiantes tanto a ellos mismos como al profesorado, por medio de una doble retroalimentación (Brown et al. 1997, Yorke 2003 y Marriott 2009). Los comentarios procedentes de la evaluación juegan un papel importante en la educación, ya que ayuda a los estudiantes a conocer su nivel de logro personal y también al profesorado a reflexionar sobre los resultados de sus técnicas de enseñanza, permitiendo modificar su aprendizaje y/o métodos de enseñanza respectivamente (Yorke 2003, Nicol et al. 2006). Dochy (2008) establece que la evaluación puede, de hecho, ser utilizada como un medio para consolidar y dirigir el aprendizaje.

El propósito esencial de la retroalimentación es ayudar a los estudiantes a aprender y mejorar su rendimiento en una tarea (Chickering et al. 1987, Brown et al. 1997). Estudios como los de Crooks (1998) y Nicol et al. (2006) muestran evidencia sobre la efectividad de la retroalimentación para promover el aprendizaje. Así, para que la retroalimentación sea efectiva,

constructiva y formativa, ha de ser adecuada, precisa, oportuna, clara, y se debe centrar en el progreso personal del alumno, Crooks (1988), Brown et al. (1997), Gibbs et al. (2005).

Si se lleva a cabo una conversación con el alumno sobre la información proporcionada a éste tras la evaluación, se incrementa la calidad de la retroalimentación, Race (1995), y las garantías que los estudiantes presten atención a esos comentarios porque, según explica Gibbs et al. (2005), los alumnos participan en un diálogo formativo. En consecuencia, esta práctica favorece su proceso de aprendizaje. Ahora bien, la retroalimentación ha de ser continua, para que sea efectiva, de forma que los estudiantes reflexionen sobre su progresión, y cuenten con tiempo suficiente para identificar y corregir cualquier deficiencia (Chickering et al. 1987, Crooks 1988, Gibbs et al. 2005, Marriott 2009).

La evaluación de los estudiantes mediante evaluación continua es una poderosa práctica que sustenta a la enseñanza y aprendizaje de calidad, permite medir el progreso del estudiante y ofrecer un apoyo adecuado a través de la retroalimentación oportuna y eficaz, impulsando la reflexión personal y la participación en el proceso de aprendizaje (Marriott 2009).

Por esta razón, la aplicación de un sistema de evaluación frecuente permite a los estudiantes recibir información periódica y puntual. Es así que en un contexto de metodología de evaluación frecuente, Kling et al. (2005) describe que la retroalimentación ayuda a los estudiantes a mejorar su aprendizaje, lo que lleva a mejorar su rendimiento académico y aumentar su compromiso. Un requisito para que esto sea así es que la retroalimentación se centre en comentarios positivos sobre el rendimiento (Nicol et al. 2006).

Contexto

El estudio se ha desarrollado en la Universitat Politècnica de València y se refiere a los métodos de evaluación que se emplearon durante el curso académico 2010-11 en Microeconomía II. Esta asignatura se imparte en el segundo semestre del primer año del Grado en Administración y Dirección de Empresas. Este programa consta de 240 créditos, repartidos en cuatro cursos académicos. Microeconomía II es la continuación natural a la Microeconomía I. Microeconomía II se centra en el análisis de la toma de decisiones de los agentes, y en la revisión de modelos de mercados de competencia imperfecta y en la comprensión del modelo de equilibrio general, así como los fallos del mercado y el papel del Estado. El curso tiene un peso de seis créditos, lo que corresponde a 60 horas de clase y hasta 120 horas de trabajo autónomo de los estudiantes. Así, el curso se divide en tres partes: clases magistrales, seminarios y prácticas de laboratorio, aunque en este trabajo nos centramos sólo en la evaluación de los contenidos impartidos en las clases magistrales.

La asignatura constaba de tres horas de clase a la semana, en la que los estudiantes estaban repartidos en dos grupos: un grupo de 52 estudiantes y otro de 23. Además de las clases, los alumnos asistían a dos sesiones de laboratorio y dos seminarios, donde el grupo de 52 estudiantes se dividía en dos. Así, las clases ponderan el 85% de la puntuación total del curso, mientras que las prácticas y seminarios conjuntamente ponderan el 15% restante. Para aprobar el curso, los estudiantes están obligados a obtener el 50% de la puntuación, así como lograr al menos el 40% de la puntuación de cada parte y asistir a un mínimo del 50% de las actividades presenciales programadas. Las clases fueron evaluadas mediante pruebas semanales escritas a mano durante casa sesión. De esta forma, se realizaron doce pruebas en total, y la puntuación de esta parte del curso se calculó como el promedio del rendimiento de las diez mejores pruebas de cada alumno. Estas pruebas se caracterizan por tres elementos: en primer lugar, su duración era de 30 minutos, una condición que promueve el desarrollo de la capacidad de resolver problemas con rapidez y evita que el cansancio. En segundo lugar, el carácter acumulativo del contenido ayudó a los estudiantes a consolidar los conocimientos adquiridos, al tener que estudiar semana tras semana lo evaluado en pruebas anteriores. Por

último, se programaron dos tipos de pruebas: pruebas tipo test y pruebas de problemas, que fueron alternadas semanalmente. Cada prueba de test contenía entre diez y doce preguntas de opción múltiple, y las pruebas de problemas consistían en hacer un problema práctico similar a los explicados en clase.

Ambos grupos de estudiantes realizaron el mismo examen el mismo día. Las respuestas correctas fueron publicadas inmediatamente después de las pruebas. Las preguntas del examen que la mayoría de los estudiantes hicieron mal en los exámenes, así como las dudas relacionadas fueron comentadas el día siguiente de clase. Las notas de cada prueba se publicaron entre el mismo día de la prueba y dos días después. La rápida publicación de las respuestas y los comentarios tras cada prueba trataron de proporcionar información sobre el desempeño del estudiante, lo que les permitió identificar y corregir sus debilidades a tiempo.

Metodología

Variables

Para analizar los resultados del sistema de evaluación descrito anteriormente, se compararon los resultados académicos de los estudiantes con los resultados de los alumnos del año anterior, que fueron evaluados utilizando un método de examen final. A continuación, se describen las variables consideradas en el análisis:

- EXAM_SCORE: Es una variable continua que representa qué proporción de la calificación máxima obtuvo cada estudiante en un examen o una prueba determinada. Por lo tanto, toma valores entre 0 y 1. Tal como está definido, los estudiantes evaluados por el método de examen final sólo tienen una calificación (la correspondiente al examen), mientras que los estudiantes evaluados por el sistema de evaluación frecuente pueden tener hasta 12 resultados, dependiendo de cuántos de ellos hicieron.

- FINAL_SCORE: Esta variable continua refleja qué proporción de la calificación máxima obtuvo cada estudiante en la asignatura, aplicando el método de evaluación correspondiente. Es decir, esta variable toma el mismo valor que EXAM_SCORE para los estudiantes que fueron evaluados por el método de examen final, pero para los estudiantes del sistema de evaluación frecuente adopta el promedio de las calificaciones de sus pruebas (después de descartar los dos peores resultados obtenidos).

- WITHDRAW: Es una variable binaria que indica si un estudiante terminó el curso. En el método de examen final, un estudiante se considera no presentado cuando éste no realiza el examen final. Sin embargo, en el año en que se aplicó el sistema de evaluación frecuente, se considera el abandono del estudiante cuando no realiza el 50% de las pruebas de clase.

- PASS: Esta variable binaria representa si un estudiante aprueba la asignatura. Siguiendo el sistema de puntuación español, un estudiante aprueba el curso si obtiene al menos el 50% de la puntuación de la asignatura.

- YEAR: Es una variable binaria que toma el valor 0 en el año con el método de examen final (2009-2010) y el valor 1 en el año en que se adoptó el sistema de evaluación frecuente (2010-2011).

Análisis

Para analizar cómo el cambio del método de evaluación afecta el rendimiento académico de los estudiantes se realiza, en primer lugar, un análisis descriptivo de las variables. A continuación, para comprobar si las diferencias entre ambos sistemas de evaluación son estadísticamente significativas, se desarrolla el análisis de la varianza (ANOVA)

para cada variable considerada. Todos los datos fueron analizados mediante el paquete estadístico *Statistical Package for the Social Sciences* (SPSS) versión 16.0 para Windows.

Resultados

La Tabla 1 muestra las medias y desviaciones típicas de las variables estudiadas para cada sistema de la evaluación. Los estadísticos descriptivos muestran mejores resultados para el sistema de evaluación frecuente que para el sistema de examen final en las cuatro variables. A continuación, se analizan en detalle estas diferencias.

Al examinar las diferencias en la variable EXAM_SCORE, la media para los estudiantes que siguen el sistema de evaluación continua supera en 13 puntos porcentuales a la de los alumnos que siguieron el método de examen final.

Se observa que el número de pruebas es mayor en el sistema de evaluación continua, debido a su naturaleza. Es decir, en el año de evaluación final, cada estudiante sólo tiene una puntuación, mientras que los estudiantes que siguen el sistema de evaluación continua podrían tener un máximo de 12 notas. El análisis de varianza mostró que las diferencias de EXAM_SCORE entre los sistemas de examen final y de evaluación frecuente son estadísticamente significativas. Estos resultados muestran que los cambios en el sistema de evaluación son eficaces para mejorar el rendimiento de los estudiantes.

Las diferencias en la variable FINAL_SCORE son también similares. Aquí, el nuevo método de evaluación proporciona un aumento de cerca de 17 puntos porcentuales en comparación con el método de examen final. Puesto que el resultado final en el sistema de evaluación continua se calcula como el promedio de las mejores diez pruebas, las diferencias en la variable de FINAL_SCORE son mayores que en el EXAM_SCORE. Como era de esperar, el análisis de varianza mostró que las diferencias entre los dos métodos son estadísticamente significativas.

Los valores medios de la variable WITHDRAW muestran que la aplicación del sistema de evaluación continua también contribuyó a reducir la tasa de abandono del 9% al 5%. Sin embargo, no es posible extraer conclusiones acerca de la efectividad del método de evaluación en su reducción ya que la diferencia entre los sistemas de evaluación no es estadísticamente significativa.

Entre las variables estudiadas, el porcentaje de aprobados es la que muestra mayor mejoría. Mientras que sólo el 43% de los alumnos aprobaron la asignatura utilizando el método de examen final, el promedio de la variable PASS (es decir, el porcentaje de aprobados) alcanzó el 93% con el sistema de evaluación continua. Como era de esperar, el análisis ANOVA reveló que esta diferencia es estadísticamente significativa.

Tabla 1 – Análisis descriptivo y resultados ANOVA

	Examen final (Año = 0)			Sistema evaluación frecuente (Año = 1)			F value (df = 1)
	N	Media	d.e	N	Media	d.e.	
EXAM_SCORE	133	0.470	0.189	751	0.594	0.243	31.257*
FINAL_SCORE	133	0.470	0.189	71	0.634	0.116	44.203*
WITHDRAW	146	0.090	0.286	75	0.050	0.226	0.885
PASS	133	0.430	0.497	71	0.930	0.258	63.059*

* $P < 0.01$

Conclusiones

Este artículo ha evaluado un sistema evaluación frecuente, el cual ha reflejado unos mejores resultados con respecto al método de examen final. A la luz de los resultados, se puede afirmar que el sistema de evaluación frecuente conlleva varios efectos positivos, que resumimos a continuación.

Por un lado, promovió un estudio profundo del tema, ya que obligó a los estudiantes a revisar los contenidos semana tras semana. Asimismo, se incrementó la asistencia a clase, ya que los estudiantes deben asistir a clases para realizar las pruebas.

Por otra parte, la tasa de abandono también disminuyó con el nuevo sistema de evaluación frecuente, aunque el análisis estadístico no permite afirmar esto contundentemente. Hay que tener en cuenta, además, que éste es un punto difícil de mejorar, ya que la automotivación, la actitud de los estudiantes y las dificultades especiales que puedan tener también influyen en la decisión de abandono.

En promedio, el sistema evaluación frecuente mostró mejores resultados: puntuación media más alta y mayor tasa de aprobados. Tres razones principales pueden explicar este resultado: i) un elevado número de pruebas elimina el efecto del azar asociado a un único examen final, ii) el aumento de la motivación de los alumnos asociado al mayor número de oportunidades para demostrar su conocimiento y, iii) mejor conocimiento de los contenidos debido al proceso de revisión continua. Así, llegamos a la conclusión de que la aplicación de un sistema evaluación frecuente es un medio eficaz para dirigir el aprendizaje de los alumnos, mejorar su rendimiento y mantenerles motivados.

Sin embargo, su aplicación y eficacia depende de otros factores, como la dificultad de la materia, su naturaleza (ciencias, ciencias sociales...) y el número de estudiantes en la asignatura. Todos estos factores deben tenerse en cuenta en el diseño de un método de evaluación continua.

Bibliografía

- Biggs, J. and C. Tang (2007). Teaching for quality learning at university (Third Ed). Open University Press.
- Brown, G. A., J. Bull, and M. Pendlebury (1997). Assessing Student Learning in Higher Education (1st. Ed.). Routledge.
- Brown, S. (2005). Assessment for learning. Learning and Teaching in Higher Education 1, 81-89.
- Casem, M. L. (2006). Active learning is not enough. Journal of College Science Teaching 35.
- Chickering, A. W. and Z. F. Gamson (1987). Seven principles for good practice in undergraduate education. American Association for Higher Education Bulletin 39, 3-7.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. Review of Educational Research 58, 438-481.
- Dempster, F. N. (1991). Synthesis of research on reviews and tests. Educational Leadership 48, 71-76.
- Dochy, F. (2008). The Edumetric Quality of New Modes of Assessment: Some Issues and Prospects. Assessment, Learning and Judgement in Higher Education. Dordrecht: Springer Netherlands.
- Gibbs, G. and C. Simpson (2005). Conditions under which assessment supports students' learning. Learning and Teaching in Higher Education 1 (August 5, 2011), 3-31.
- Kling, N., D. McCorkle, C. Miller, and J. Reardon (2005). The impact of testing frequency on student performance in a marketing course. Journal of Education for Business 81, 67-72.
- Kuo, T. and A. Simon (2009). How many tests do we really need? College Teaching 57, 156 -160.
- Marriott, P. (2009). Students' evaluation of the use of online summative assessment on an undergraduate financial accounting module. British Journal of Educational Technology 40, 237- 254.
- Nicol, D. J. and D. Macfarlane-Dick (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education 31, 199-218.
- Race, P. (1995). The art of assessing 1. New Academic 4.
- Trotter, E. (2006). Student perceptions of continuous summative assessment. Assessment & Evaluation in Higher Education 31, 505-521.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education* 45, 477-501.

Cuestiones y/o consideraciones para el debate

Adaptación a los acuerdos de Bolonia y la evaluación continua en el aula universitaria en tiempos de crisis. La evaluación continua mejora el rendimiento, pero ¿cómo implementar sistemas de evaluación continua cuando el número de alumnos por aula aumenta y se reducen los recursos?