

ISSN: 3101-0148

DOI: <https://doi.org/10.33732/ridau.41>

Evaluación con retroalimentación relativa. Un experimento aleatorio en la Universidad Rey Juan Carlos

Evaluation with Relative Feedback. A Randomised Experiment at Rey Juan Carlos University

Cristian Macías Domínguez

Universidad Rey Juan Carlos. España.

<https://orcid.org/0000-0003-2385-5880>

Rosa Santero Sánchez

Universidad Rey Juan Carlos. España.

<https://orcid.org/0000-0002-1071-4280>

Fecha de recepción: 11/04/2024

Fecha de aceptación: 07/06/2024

Cómo citar: Macías Domínguez, C.; Santero Sánchez, R. (2025). Evaluación con retroalimentación relativa. Un experimento aleatorio en la Universidad Rey Juan Carlos. *Revista de Innovación Docente en el Aula Universitaria (RIDAU)*, 1, 39-61. <https://doi.org/10.33732/ridau.41>

RESUMEN

La retroalimentación sobre las calificaciones de las actividades y exámenes permite a los estudiantes tener información sobre cómo su trabajo se traduce en una puntuación, que puede afectar a su esfuerzo posterior y, por tanto, a su rendimiento. Esta investigación tiene como objetivo determinar si un mayor conocimiento en la evaluación continua a lo largo del semestre aumenta el rendimiento al final del curso. Para ello, se lleva a cabo un experimento de control aleatorio en la asignatura de Análisis Económico-Financiero II del grado de Marketing en la Universidad Rey Juan Carlos. El grupo de tratamiento recibió retroalimentación relativa respecto al percentil en

la distribución de calificaciones de su grupo-clase, mientras que el grupo de control solo obtenía su calificación numérica individual.

Los resultados determinan que las calificaciones medias entre ambos grupos son muy similares, lo que implica que la retroalimentación relativa no conlleva una mejora en los resultados del grupo tratado. Es necesario avanzar en el análisis de factores moderadores y ampliar la muestra para poder contrastar si los efectos sobre los valores medios están ocultando efectos contrapuestos entre los cuartiles de mayor y menor rendimiento.

PALABRAS CLAVE: Retroalimentación relativa; Experimento aleatorio; Rendimiento académico; Evaluación; Educación superior.

ABSTRACT

Feedback on grades from activities and exams allows students to have information about how their effort translates into a score, which can affect their endeavour in the future and, therefore, their performance. This research aims to determine whether increased knowledge in ongoing assessment throughout the semester enhances performance at the end of the course. To this end, a randomized control experiment is conducted in the subject of Financial Economic Analysis II in the Marketing degree at Rey Juan Carlos University. The treatment group received relative feedback regarding their percentile rank within the grade distribution of their class group, while the control group only received their individual numerical grade.

The results indicate that the average grades between both groups are very similar, suggesting that relative feedback does not lead to an improvement in the results of the treated group. Further analysis of moderating factors and expanding the sample size is necessary to determine if the effects on mean values are obscuring opposing effects between higher and lower performing quartiles.

KEYWORDS: Relative performance feedback; Randomized field experiment; Academic performance; Evaluation; Higher education.

1. INTRODUCCIÓN

El siglo XXI ha sido el periodo de mayor transformación en el sistema universitario español de las últimas décadas. El proceso de convergencia hacia el Espacio Europeo de Educación Superior ha venido a replantear la forma de organizar la enseñanza universitaria. El nuevo diseño de titulaciones oficiales ha ido de la mano de un cambio de paradigma, donde el estudiantado debe ser protagonista de su propio

aprendizaje y donde el profesorado se convierte en un facilitador y conductor de su proceso formativo.

Este cambio afecta a todos los elementos curriculares (López-Pastor, 2012). La normativa universitaria (Real Decreto 1393/2007, ahora derogado, y el actual Real Decreto 822/2021) fijan los requisitos formales para el diseño de una titulación y los principios rectores para los planes de estudio, destacando la necesidad de buscar una coherencia y alineación entre los objetivos formativos, las competencias y resultados de aprendizaje de la titulación, así como con los sistemas de evaluación asociados (art. 4.c Real Decreto 822/2021). Por tanto, las materias y asignaturas diseñadas en los nuevos títulos universitarios deben contar con un objetivo claro que establezca los resultados de aprendizaje asociados al perfil de egreso, y para ello, se incluyen diferentes actividades académicas con sus correspondientes sistemas de evaluación y calificación.

El enfoque metodológico y los procesos de evaluación son los dos elementos que más ajustes están experimentando. En el caso de los sistemas de evaluación, el proceso de convergencia da más importancia a realizar una evaluación orientada a la mejora del aprendizaje del estudiantado y no solo como control del resultado, lo que conlleva a centrarse en evaluaciones continuas y formativas, en lugar de evaluaciones finales y sumativas (López-Pastor, 2012).

La normativa universitaria actual recoge la obligación de contar con un sistema de evaluación que finalice con una calificación numérica (art. 9.2 Real Decreto 822/2021). Este tipo de evaluación tiene un objetivo de control de los resultados obtenidos por el alumnado, para determinar la superación de las materias o asignaturas. Esta forma de entender la evaluación parece dejar de lado el carácter formativo de la misma (Gill-Flores, 2012). Sin embargo, la experiencia docente defiende que una actividad de evaluación puede contribuir al aprendizaje si proporciona información que docentes y estudiantes puedan utilizar como retroalimentación para evaluarse a sí mismos y a los demás y para ajustar el proceso de enseñanza-aprendizaje. Así, se convierte en “evaluación formativa” cuando las pruebas se utilizan para adaptar la labor docente a las necesidades de aprendizaje del estudiantado (Black et al., 2004).

La retroalimentación en la evaluación puede considerarse como parte del proceso de una evaluación formativa (Kingston & Nash, 2011). Gran parte de la literatura académica sobre evaluación se ha centrado en distinguir entre evaluación sumativa y formativa, y sobre las fortalezas que tiene la evaluación formativa y sus efectos positivos sobre los rendimientos del alumnado (Black & Wiliam, 1998; Dunn & Mulvenon, 2009; Hattie & Timperley, 2007). Centrado en el nivel universitario, Nyquist (2003) revisa trabajos centrados en evaluación formativa diferenciando por el grado de retroalimentación, teniendo en cuenta una amplia gama de posibilidades, desde una retroalimentación débil (solo reciben calificación individual) hasta una evaluación formativa fuerte (reciben además de los resultados, explicaciones sobre los mismos y actividades específicas para mejorar sus resultados). La medición de cada

tipo de retroalimentación sobre los resultados académicos muestra que la evaluación formativa fuerte es la que mayores efectos provoca.

La relación entre la retroalimentación y la evaluación del desempeño ha sido analizada en diferentes ámbitos de conocimiento, incluyendo el educativo, aunque el desarrollo mayor se ha producido en el entorno laboral (Gill et al., 2018; Kosfeld & Neckermann, 2011; Tran & Zeckhauser, 2012) donde la posición relativa en una clasificación puede dar acceso a un premio, un ascenso o una bonificación. Estos logros pueden ser públicos y conocidos por todos, y conllevar un reconocimiento social, o ser particulares y formar parte de la autoestima de la persona que lo recibe, aunque ambos tipos de logros no son fácilmente separables.

Centrando el interés en el ámbito educativo, y en concreto, en el universitario, la presente investigación examina si la utilización de retroalimentación relativa en estudiantes universitarios tiene efectos positivos sobre su rendimiento académico. Para ello, se diseña un experimento aleatorio en una asignatura de grado en la Universidad Rey Juan Carlos, con un grupo de tratamiento y un grupo de control, para incluir en la evaluación de varias actividades a lo largo del curso 2021-2022, una retroalimentación individual, con su calificación, y relativa, a través del percentil en el que se encuentra respecto al grupo-clase. Esta retroalimentación relativa se ofreció al grupo de tratamiento mientras que el grupo de control solo obtiene la calificación individual en valor absoluto. El sistema de evaluación se complementa con un cuestionario pre y postest que recoge información personal y motivacional, además de ítems relativos a las expectativas que el estudiantado tiene sobre la asignatura y sus propios resultados en la misma.

El artículo se desarrolla por secciones. A continuación, se presenta una breve revisión de literatura sobre la relación entre la evaluación formativa y la retroalimentación y los resultados de aprendizaje. En la tercera sección, se detalla el objetivo de la investigación, y en la cuarta, la metodología del experimento realizado. En la quinta sección se detalla la muestra y en la sexta se presentan los principales resultados, finalizando la investigación con la discusión y las conclusiones.

2. REVISIÓN DE LA LITERATURA

En la literatura empírica parece haber consenso en cuanto a la defensa de la evaluación formativa para mejorar el rendimiento de los estudiantes, especialmente aquellos que tienen peores resultados, y para mejorar las prácticas pedagógicas de los docentes, sin embargo, parece que faltan investigaciones de calidad que puedan ser generalizables (Dunn & Mulvenon, 2009; Kingston & Nash, 2011). En la bibliografía sobre evaluación formativa, el artículo seminal de Black y William (1998) se cita con frecuencia como prueba de que la evaluación formativa mejora el rendimiento del alumnado, sin embargo, otros autores no respaldan de forma concluyente este resultado, y como reconocen Dunn y Mulvenon (2009) deberían haber recomendado

que se siguiera investigando, puesto que los estudios analizados tenían algunas características muy específicas que podrían no ser generalizables. En concreto, se trata del metaanálisis de Fuch y Fuch (1986), centrado en educación especial, con un 83% de personas con discapacidad entre el estudiantado analizado.

Posteriormente, Hattie y Timperley (2007) analizaron una docena de metaanálisis que incluían 196 estudios y casi 7.000 efectos relacionados con la provisión de retroalimentación. Concluyeron que la eficacia de esta dependía del tipo de feedback (positivo o negativo), del contexto en el que se proporciona (por ejemplo, el momento de la retroalimentación) y de los diferentes patrones del enfoque de motivación que tienen los estudiantes respecto a su aprendizaje (orientados o elusivos al éxito).

Centrando la revisión de literatura empírica en iniciativas en la etapa universitaria, y en concreto, en investigaciones experimentales, encontramos tres iniciativas de interés. Una universidad pública del Reino Unido evaluó un experimento aleatorio en el que se permitía que cada departamento fuese libre para retroalimentar a sus estudiantes (Bandiera et al., 2015). Su hipótesis de investigación es que una retroalimentación continua de las calificaciones en términos absolutos mejora el desempeño futuro del estudiante. El resultado de la investigación determina que dar a conocer a los alumnos las calificaciones de forma continua tiene un efecto positivo en las calificaciones de las pruebas posteriores: el impacto medio corresponde a una mejora del 13% de una desviación estándar en las calificaciones de las pruebas. El estudio determina que el impacto de la retroalimentación es más fuerte para los estudiantes más brillantes y para los estudiantes que tienen menos información inicial sobre el entorno académico. Además, sugiere que la retroalimentación no tiene efectos negativos en las calificaciones de las pruebas de ningún estudiante, esto implica que la política óptima es siempre brindar retroalimentación, asumiendo que el objetivo de un departamento es maximizar el rendimiento académico de los estudiantes.

Otro estudio que refuerza la teoría de la retroalimentación en la educación superior fue realizado en Vietnam sobre estudiantes universitarios en la asignatura de inglés (Tran & Zeckhauser, 2012). Se dividió la clase en tres grupos y se les realizaban pruebas quincenalmente: un grupo de control compuesto por alumnos a los que se les dio la calificación absoluta en cada prueba realizada; un grupo de tratamiento que además de la calificación se les dio de forma privada la nota respecto a la clase; y otro grupo de tratamiento al cual se les informó de la calificación y la nota respecto a la clase de forma pública, al finalizar el curso de inglés se les realizó el examen TOEIC. Se demostró que el primer grupo de tratamiento superó en 10 puntos porcentuales al grupo de control, demostrando que esta mejora apoya firmemente la hipótesis de que el conocimiento del rango respecto al resto de compañeros motiva a las personas incluso cuando no aporta beneficios tangibles. Además, el segundo grupo de tratamiento que recibió el rango de forma pública superó a los que solo recibieron el rango de forma privada. Los coeficientes sugieren que las calificaciones fueron

un 45% mayor sobre el grupo de control lo que implica una mayor motivación que el primer grupo de tratamiento.

En España, varios estudios realizados en la Universidad Carlos III de Madrid (Azmat et al., 2019; Cabrales et al., 2019), determinan que los rendimientos académicos de los estudiantes de los grupos de tratamiento -recibiendo retroalimentación- y de control en el primer año de carrera, son similares. Una vez que los estudiantes tratados son proporcionados con su información de rango -retroalimentación relativa-, se observa una disminución significativa en su rendimiento en relación con los del grupo de control. Este resultado refuerza otros estudios realizados que han corroborado que el exceso de confianza provoca una reducción sistemática en las calificaciones (Moore & Cain, 2007). En particular, durante su segundo año los estudiantes tratados aprueban, en promedio 0,4 puntos menos que los estudiantes del grupo de control. Su puntuación media para ese año se reduce alrededor de 0,20 desviaciones estándar. Esta reducción viene fundamentada por el exceso de confianza de cara a los exámenes finales según estos autores. Estos autores completan su investigación con un análisis del nivel de satisfacción, y concluyen que esta aumenta para el grupo de tratamiento.

La diversidad de estrategias utilizadas dentro de la evaluación formativa es amplia, y permite adecuarse a los estilos de aprendizaje del alumnado (Castejón et al., 2009). La puesta en marcha de estas estrategias forma parte de las políticas educativas, y como tal, debe existir un procedimiento de evaluación de las políticas, que tengan en cuenta tanto el diseño, como su implementación y ejecución, como sus resultados e impactos. En el ámbito de la Economía de la Educación, los experimentos aleatorios se consideran un buen instrumento de evaluación de políticas y se han concretado en una herramienta crucial para evaluar la eficacia de las intervenciones (Styles y Torgerson, 2018; Williams y Kane, 2009). En estudios como el de Connelly et al. (2018), se destaca la necesidad de que las revisiones sistemáticas de los experimentos aleatorios no se centren únicamente en los resultados, sino también en el proceso de implementación de las intervenciones. Esta perspectiva resalta la importancia de comprender no solo si una intervención funciona, sino también cómo y por qué lo hace. Incluso en casos donde los resultados son nulos, se evidencia que la implementación de esas medidas no ha tenido el efecto deseado (Abadie, 2020).

3. OBJETIVOS

La retroalimentación, como parte del proceso de la evaluación formativa, es muy común en entornos educativos, especialmente en niveles previos a la universidad. A menudo se proporciona una retroalimentación sobre el desempeño absoluto o en relación con algún grupo de referencia relevante (Cabrales et al., 2019).

Desde una perspectiva teórica, el impacto de la retroalimentación relativa del desempeño sobre el esfuerzo es ambiguo. La reacción de las personas depende de sus

creencias previas sobre sus propias capacidades y las de los demás, de la información inferida de la retroalimentación, así como de otras motivaciones (Azmat et al., 2019; Azmat e Iriberry, 2010). Por ejemplo, si la capacidad es complementaria al propio esfuerzo con el fin de lograr un resultado particular, las noticias positivas (negativas) sobre la propia capacidad harán que los individuos trabajen más (menos). De forma adicional, las personas podrían preocuparse por su posición relativa, mostrando una motivación “competitiva” en sus preferencias, tal vez porque los responsables académicos, o los reclutadores corporativos cuando se incorporen al mercado laboral, valoren el desempeño relativo por encima del absoluto. Si ese es el caso, saber que la capacidad de los demás es menor (mayor) de lo que inicialmente se pensaba podría hacer que las personas ejerzan un nivel menor (mayor) de esfuerzo. Desde un punto de vista teórico, los incentivos asociados al conocimiento del percentil que ocupa el/la estudiante respecto al grupo-clase, puede influir en un exceso de confianza ocasionando una reducción en el rendimiento a lo largo del curso.

El objetivo principal de esta investigación se centra en evaluar el efecto que proporciona una retroalimentación relativa sobre los resultados académicos en un grupo de estudiantes de la Universidad Rey Juan Carlos (URJC). Para alcanzar este objetivo, se diseña un experimento aleatorio con grupo de tratamiento y control, y se busca contrastar si existen diferencias estadísticamente significativas entre los rendimientos medios de la asignatura en la que se aplica una retroalimentación relativa (a través del percentil de posición) en el grupo de tratamiento frente a otro de control, seleccionados estos de forma aleatoria.

La revisión de la literatura empírica no muestra un consenso en cuanto a que una información más precisa del estudiante respecto a sus calificaciones tenga un efecto positivo sobre sus rendimientos académicos. Elementos como la motivación, las expectativas, la capacidad, el esfuerzo, los resultados previos o los estilos de aprendizaje, deben tenerse en cuenta por ser factores moderadores del desempeño.

Por ello, la hipótesis de investigación será que la retroalimentación relativa es una información adicional que va a ejercer de empuje motivacional durante el semestre para mejorar las calificaciones de cara a las pruebas finales. Esto implica que los alumnos que conocen su percentil a lo largo del semestre son capaces de sacar mejores notas al final del mismo.

Esta investigación pretende mostrar evidencias sobre dos cuestiones, para comprender el efecto de proporcionar retroalimentación a los alumnos. La primera se centra en si los estudiantes que conocen su rango porcentual respecto a la clase persiguen maximizar sus propios resultados, o si también se preocupan por su posición relativa de manera competitiva. Esta puede deberse a que el desempeño relativo se materializa en diferentes resultados o los individuos se preocupan inherentemente por la posición relativa en el desempeño. La segunda, considera que las creencias previas de los individuos con respecto a sus habilidades y las de los demás compañeros

ayudará a entender, si, en ausencia de información, es posible sobrestimar su posición relativa. En un contexto donde los individuos tienen preferencias competitivas, aprendiendo que la habilidad de uno, con respecto a la de los demás, es más alta de lo que inicialmente se creía, conduce a un menor esfuerzo y, a la inversa, aprender que la capacidad relativa de uno es menor de lo que inicialmente se creía puede llegar a sobrestimar un mayor esfuerzo (Cabrales et al., 2019).

4. METODOLOGÍA

4.1 Descripción del diseño del estudio

La investigación que se ha llevado a cabo entra dentro del tipo de investigación experimental. Investigaciones recientes como las realizadas por Williams y Kane (2009) o Styles y Torgerson (2018) exploran la relevancia de los experimentos aleatorios en el ámbito de la Economía de la Educación en el siglo XXI. El Randomized Controlled Trial (RCT) o experimento aleatorio emerge como una forma innovadora de investigación que mitiga el sesgo en el análisis de las causas, asegurando una evaluación más precisa de los resultados. En una época en la que se persiguen respuestas sobre la eficacia y la efectividad de políticas y prácticas educativas, los RCT desempeñan un papel esencial al posibilitar estimaciones imparciales sobre los rendimientos promedio de los estudiantes, siempre y cuando la muestra sea lo suficientemente amplia.

En nuestro caso, se trata de un experimento aleatorio, donde un grupo-clase se divide en dos grupos: un grupo de tratamiento y un grupo de control. La forma de determinar la inclusión de cada estudiante a uno u otro grupo ha sido realizada de forma aleatoria. Todas las personas matriculadas en la asignatura donde se realiza el experimento han firmado un formulario que garantiza la voluntariedad de pertenecer al experimento aleatorio y en el cual serán informados de manera concisa sobre el objetivo del mismo, garantizando la privacidad en todo momento de sus datos personales. Para poder realizar el experimento aleatorio con alumnos de la URJC fue preciso pedir permiso al Comité de Ética Investigador de la Universidad.

Al grupo de tratamiento se les proporciona la calificación de cada actividad, como habitualmente, además de añadirles el percentil en el que se encuentran sus calificaciones con respecto a la clase durante las pruebas realizadas a lo largo del semestre. Por otro lado, al grupo de control solo se les proporciona la calificación individual en valor absoluto.

4.2 Selección de los participantes

El experimento se ha realizado en el segundo semestre del curso académico 2021-2022 en el Doble Grado de Marketing y Publicidad y Relaciones Públicas

del Campus de Fuenlabrada de la URJC. Dicho experimento se lleva a cabo en la asignatura de Análisis Económico Financiero en Marketing II siendo esta asignatura del segundo año y semestre para los alumnos que realizan la asignatura únicamente de Marketing mientras que es del tercer curso para los alumnos de la doble titulación.

Todos los alumnos fueron incluidos en la base de datos del experimento aleatorio, aunque solo recibieron la retroalimentación relativa los que dieron su consentimiento informado solicitado por el Comité de Ética Investigador de la Universidad Rey Juan Carlos.

4.3 Procedimiento

En una fase inicial del experimento, en el mes de enero, al comienzo de la asignatura, se realiza un cuestionario (ver Figura 1) para conocer aspectos sociodemográficos de las personas matriculadas en la asignatura que va a conformar la muestra y, sobre todo, las motivaciones, creencias y expectativas previas del estudiantado, lo que permite valorar los resultados del experimento.

Figura 1. Cuestionario realizado a los alumnos de la Universidad Rey Juan Carlos

Marque con una X donde corresponda y complete los espacios acimi que se solicitan.		
Fecha:	Año de inicio de estudios universitarios:	
Edad:	Género: <input type="checkbox"/> Masculino <input type="checkbox"/> Otra <input type="checkbox"/> Femenino	
Mes de 13 nacimiento:	País de origen:	
Número de hermanos: <input type="checkbox"/> Ninguno <input type="checkbox"/> Dos hermanos <input type="checkbox"/> Un hermano <input type="checkbox"/> Más de dos hermanos	Modalidad de entrada a la universidad: <input type="checkbox"/> Selectividad/EVAU <input type="checkbox"/> Otra <input type="checkbox"/> Formación profesional	
Número de materias matriculadas en el presente curso: <input type="checkbox"/> Menos de 6 asignaturas <input type="checkbox"/> Más de 9 asignaturas <input type="checkbox"/> Entre 6 y 9 asignaturas	Zona geográfica de su domicilio habitual: <input type="checkbox"/> Madrid-Capital <input type="checkbox"/> Fuera de la C. Madrid <input type="checkbox"/> Municipio de la C. Madrid	
¿Ha recibido usted una beca para realizar los estudios en la Universidad? <input type="checkbox"/> Sí <input type="checkbox"/> No	¿Dispone de carné de conducir? <input type="checkbox"/> Sí <input type="checkbox"/> No	¿Dispone de coche? <input type="checkbox"/> Sí <input type="checkbox"/> No

(Continuada)

Figura 1. Cuestionario realizado a los alumnos de la Universidad Rey Juan Carlos (Continuada)

Marque con una X donde corresponda y complete los espacios acimi que se solicitan.	
Fecha:	Año de inicio de estudios universitarios:
Dispone de ordenador particular: <input type="checkbox"/> Sí <input type="checkbox"/> No	Cuenta con conexión a internet de forma habitual: <input type="checkbox"/> Sí <input type="checkbox"/> No
Trabaja actualmente: <input type="checkbox"/> Sí Número de horas a la semana: <input type="checkbox"/> No	Asiste a clases extracurriculares: <input type="checkbox"/> Sí <input type="checkbox"/> No
Nivel de estudios del padre: <input type="checkbox"/> Estudios superiores (Universidad o FP Superior finalizados) <input type="checkbox"/> Estudios medios (FP Media o Bachillerato finalizados) <input type="checkbox"/> Estudios básicos (Primaria, EGB o ESO finalizados) <input type="checkbox"/> Sin estudios	Nivel de estudios de la madre: <input type="checkbox"/> Estudios superiores (Universidad o FP Superior finalizados) <input type="checkbox"/> Estudios medios (FP Media o Bachillerato finalizados) <input type="checkbox"/> Estudios básicos (Primaria, EGB o ESO finalizados) <input type="checkbox"/> Sin estudios
Utilizando una escala de 0 a 1, siendo cero el rendimiento académico más bajo y 1 el más alto, ¿en el momento de su acceso a la universidad, en qué categoría/grupo de los siguientes cree que estaría? <input type="checkbox"/> Entre 0,00 y 0,25 <input type="checkbox"/> Entre 0,25 y 0,50 <input type="checkbox"/> Entre 0,50 y 0,75 <input type="checkbox"/> Entre 0,75 y 1,00	

Fuente: Elaboración propia.

Durante el semestre, se realizaron dos pruebas de evaluación, en base a estas pruebas se les notificó a los alumnos del grupo de tratamiento el percentil en el que se encontraban con respecto a la clase. La aleatorización se llevó a cabo mediante el programa estadístico STATA14, mientras que el análisis descriptivo y los contrastes de hipótesis se realizaron con SPSS27.

La notificación se realizó por el correo interno de la URJC de manera individual a cada uno de los estudiantes del grupo de tratamiento, junto con una breve explicación de lo que suponía encontrarse en dicho percentil. Finalmente, en el mes de mayo, se realizó el examen final de la asignatura en la denominada convocatoria ordinaria.

Además, una vez realizado los exámenes finales, los estudiantes respondieron una encuesta sobre la satisfacción del curso y su posible mejora académica al conocer su posición relativa respecto a la clase (ver Figura 2).

Figura 2. Encuesta sobre la satisfacción al conocer el percentil durante el curso

Experimento aleatorio sobre el rendimiento relativo retroalimentado	
Experimento aleatorio sobre el rendimiento relativo retroalimentado en la Educación Superior en la URJC	
* Indica que la pregunta es obligatoria	
Correo *	
¿Has recibido la información del percentil sobre tu nota en la asignatura en relación al conjunto de la clase? *	
<input type="radio"/> Sí <input type="radio"/> No	
¿Te ha parecido una información interesante? *	
¿Entiendes el significado de tu percentil? *	
¿Es una información sencilla de entender? *	
<input type="radio"/> Sí <input type="radio"/> No	
¿Esperabas quedar en una posición mejor, igual o peor de la que has quedado? *	
En concreto ¿en qué posición pensabas que habías quedado (en percentiles)? *	
¿Consideras que has mejorado tu percentil a lo largo del curso con respecto al percentil que ya conoces? *	
¿Crees que conocer el percentil a lo largo del curso puede suponer una motivación para mejorar en la calificación final? *	
<input type="radio"/> Sí <input type="radio"/> No	

Fuente: Elaboración propia.

4.4 Medidas de resultados

a) Variables dependientes

El principal indicador de resultados (output) es el rendimiento académico de los estudiantes de la asignatura Análisis Económico Financiero en Marketing II, a través del cual se va a aproximar el impacto generado por la retroalimentación relativa. La asignatura tiene un rango de puntuación que oscila entre el cero y el diez (variable cuantitativa), siendo este último valor la nota máxima que puede alcanzar un alumno y la calificación de cinco, la nota mínima para superar la asignatura.

Para definir el experimento y generar la retroalimentación relativa, se han clasificado los estudiantes por percentiles:

- Percentil 25: Aquellos alumnos que han tenido notas por debajo del 25% de los estudiantes de la clase.
- Percentil 50: Aquellos alumnos que han tenido notas por debajo del 50% de los estudiantes de la clase.
- Percentil 75: Aquellos alumnos que han tenido notas por debajo del 75% de los estudiantes de la clase.
- Percentil 100: Aquellos alumnos que han tenido notas por debajo del 100% de los estudiantes de la clase.

b) Variables independientes

El cuestionario realizado al principio de curso permite obtener datos socio-demográficos de los estudiantes con el fin de conocer la muestra empleada en la investigación, siendo las principales variables:

- Edad: Variable cuantitativa que indica el número de años del estudiante.
- Género: variable cualitativa, con dos posibles valores: 1-Masculino y 2-Femenino.
- Número de hermanos: variable cualitativa con los valores siguientes: 0-Ningún hermano, 1-si tiene un hermano, 2-en caso de tener dos hermanos y 3-para más de dos hermanos.
- Modalidad de entrada a la universidad: variable cualitativa que toma valor 1-si la entrada fue por selectividad/EVAU, 2-entrada por Formación Profesional y 3-otras opciones.
- Número de materias matriculadas: variable cualitativa que toma valor 1-si el alumno está matriculado en menos de 6 asignaturas en el curso actual, 2-entre 6 y 9 asignaturas y 3-para más de 9 asignaturas matriculadas.
- Zona geográfica del domicilio: variable cualitativa que toma valor 1-si el alumno reside en Madrid Capital, 2-si el domicilio está en algún municipio de la Comunidad de Madrid y 3-si vive fuera de la Comunidad.

4.5 Análisis de datos y estadística

Para dar respuesta al objetivo principal del experimento, es necesario en primer lugar aleatorizar el grupo-clase para obtener un grupo de tratamiento y de control.

Para proporcionar la retroalimentación relativa, se ha utilizado la calificación de las actividades y se han calculado los percentiles.

Se hace un análisis descriptivo con estadísticos centrales y de dispersión de las calificaciones pre y post experimento, además de realizar un contraste de medias, para comparar si la diferencia de resultados académicos es estadísticamente significativa entre el grupo de tratamiento y el de control (Macías & Santero, 2022).

5. RESULTADOS

El estudio se realiza en la Universidad Rey Juan Carlos de Madrid, concretamente en la asignatura Análisis Económico Financiero en Marketing II del segundo año del grado de Marketing y del tercer año del doble grado en Marketing y Publicidad y Relaciones Públicas en el campus de Fuenlabrada. La muestra participante en el experimento aleatorio aparece reflejada en la Tabla 1.

Tabla 1. Resumen de la muestra

	Marketing	Marketing y Publicidad	Totales
Alumnas	32	16	48
Alumnos	20	1	21
Totales (N)	52	17	69
Repetidores	5	0	5

Fuente: Elaboración propia a partir de las listas de alumnos.

Los alumnos objeto de estudio completaron un cuestionario a principios del semestre para conocer con detalle la muestra seleccionada. Los resultados del cuestionario muestran que la edad media de los alumnos es 19,8 años, tienen un número medio de hermanos de 1,1 y la distribución por género es de 49 alumnas (71%) y 20 alumnos (29%). Más del 80% de los estudiantes acceden a la universidad mediante la prueba de selectividad siendo un 18,2% los estudiantes que residen en Madrid capital, un 61,4% tienen su domicilio en municipios de Madrid y un 20,5% lo hacen en localidades fuera de la Comunidad de Madrid. Todos los alumnos cuentan con internet y un ordenador personal por lo que pudieron seguir sin problemas las clases el curso anterior de forma telemática.

En dicho cuestionario, se les preguntó también sobre la estimación que creían a principios del semestre sobre su nota en relación con los cuartiles. Los datos muestran que un 9,1% consideraba que su nota se encontraría en el segundo cuartil, un 54,5% aseguró que su nota se encontraría en el tercer cuartil y un 36,4% que sus notas se encontrarían en el cuarto cuartil. Ningún alumno indicó estar en el primer cuartil siendo estas las notas más bajas.

El grupo de tratamiento fue seleccionado de manera aleatoria entre todos los estudiantes matriculados, y se agrupó a un total de 35 personas.

Se realizaron dos pruebas durante el semestre, todos los alumnos conocieron las calificaciones numéricas. Los alumnos del grupo de tratamiento junto con la calificación recibieron un mensaje por correo interno de la universidad con el percentil en el que se encontraba su nota con respecto a la clase y una breve explicación de lo que eso significaba. En el mes de mayo se realizó el examen de la convocatoria oficial y posteriormente al grupo de tratamiento se les envió un cuestionario para determinar si el haber conocido su percentil a lo largo del semestre supuso una motivación extra para aumentar el esfuerzo de cara al examen oficial.

Para iniciar el experimento es necesario conocer en qué posición relativa se encuentran los estudiantes del grupo de tratamiento en términos medios con respecto a las notas del grupo-clase. La Tabla 2 indica los percentiles más importantes en la distribución junto con las calificaciones correspondientes a dicho percentil de todos los alumnos, antes de realizar el experimento, en las pruebas realizadas durante el semestre y el porcentaje acumulado.

Tabla 2. Relación percentil-calificación antes del experimento

Percentil	Calificación	Porcentaje acumulado
10	1,1	9,3%
20	2,0	18,5%
30	2,8	31,5%
40	3,7	40,7%
50	4,5	53,8%
60	5,0	61,1%
70	5,7	70,4%
80	7,0	83,3%
90	9,2	90,7%
100	10,0	100%

Fuente: Elaboración propia a partir de las notas obtenidas.

Los resultados muestran que antes de realizar el experimento aleatorio en torno al 61,1% de los alumnos se encontraban por debajo del cinco, la nota mínima para aprobar la asignatura. La Tabla 3 indica las medias, las desviaciones típicas y las medianas de ambos grupos antes de realizar el experimento aleatorio. Los resultados muestran que el valor promedio del grupo-clase es de 4,65 puntos sobre 10, con un valor medio de 4,47 en el grupo de tratamiento y de 4,81 en el grupo de control. Existe una diferencia entre ambos grupos en sus calificaciones medias en torno a 0,34 puntos, aunque las diferencias no son estadísticamente significativas.

Tabla 3. Datos de los grupos antes de realizar el experimento aleatorio

Grupo	Observaciones	Media	Desv. Típica	Percentil 50
Grupo de control	28	4,81	2,73	4,23
Grupo de tratamiento	26	4,47	2,65	4,50
(N) Alumnos	54	4,65	2,67	4,46

Fuente: Elaboración propia a partir de las calificaciones obtenidas en estudiantes que han participado.

Sin embargo, el valor de la mediana del grupo de tratamiento es mayor, esto puede deberse a que la media recoge todos los valores extremos de la distribución y al ser una muestra tan reducida, el valor promedio puede verse altamente influenciado por la existencia de valores extremos.

La Tabla 4 muestra el número de alumnos y el porcentaje que representan sobre el total los alumnos del grupo de tratamiento con relación a los cuatro percentiles más importantes antes de realizar la intervención.

Tabla 4. Percentiles del grupo de tratamiento antes de la intervención

	Alumnos	Porcentaje acumulado (alumnos presentados)
Por debajo o igual percentil 25	7	20%
Por debajo o igual percentil 50	14	40%
Por debajo o igual percentil 75	22	62,85%
Por debajo o igual percentil 100	26	74,29%

Fuente: Elaboración propia a partir de los datos recogidos. Nota: El 25,71% restante no se presentaron al examen.

Los datos reflejan que un 40% de los alumnos del grupo de tratamiento se encontraban por debajo de la mediana de las calificaciones respecto a la clase, siendo este valor inferior a la mediana del inicio del ensayo que se ubicaba en 53,80%. Por otro lado, se aprecia que solo un 11,44% de los alumnos se encontraban entre las notas más altas del grupo.

En base a los datos obtenido en el cuestionario realizado a principios del semestre, se puede apreciar que un 20% de alumnos se encuentra en el segundo cuartil mientras que la estimación por parte de los alumnos fue un 9,10%. Un 22,85% se encuentra en el tercer cuartil mientras que la estimación realizada por los alumnos fue del 54,54%. Por último, la presencia en el cuarto cuartil es del 11,44% y la estimación realizada a principios del semestre por los alumnos fue del 36,36%. Estos resultados muestran las diferencias existentes entre la realidad y la estimación por parte de los alumnos según sus creencias y motivaciones al inicio del semestre.

Una vez realizada la retroalimentación relativa, esto es, después de la intervención del experimento, se vuelven a calcular los porcentajes acumulados obtenidos por los estudiantes del grupo de tratamiento en la prueba final de la asignatura, que se corresponde con el examen de la convocatoria ordinaria del mes de mayo (Tabla 5).

Tabla 5. Percentiles del grupo de tratamiento después de la intervención

	Alumnos	Porcentaje acumulado (alumnos presentados)
Por debajo o igual percentil 25	7	20,00%
Por debajo o igual percentil 50	12	34,29%
Por debajo o igual percentil 75	23	65,71%
Por debajo o igual percentil 100	29	82,86%

Fuente: Elaboración propia a partir de los datos recogidos. El 17,14% restante no se presentaron al examen.

Un primer resultado es que el número de alumnos presentados a este examen es superior a los presentados en las pruebas previas. Un segundo resultado es que aumenta el porcentaje acumulado en el percentil de calificaciones más altas. El número de alumnos por debajo de la mediana se reduce al 34,29% y se aprecia un ligero aumento del porcentaje de alumnos que se encuentra en el grupo con las mejores notas (17,15%).

Se puede destacar que la media del grupo de control es 0,391 superior a la media del grupo de tratamiento y con una desviación típica inferior al grupo tratado. En este caso la mediana está acorde con la media, ya que en el grupo de control es mayor que en el grupo de tratamiento. Los datos aparecen reflejados en la Tabla 6.

Tabla 6. Datos de los grupos después de realizar el experimento aleatorio

Grupo	Observaciones	Media	Desv. Típica	Percentil 50
Grupo de control	30	5,59	2,44	5,92
Grupo de tratamiento	29	5,20	2,49	4,78
(N) Alumnos	59	5,39	2,45	5,65

Fuente: Elaboración propia a partir de las calificaciones obtenidas en estudiantes que han participado.

Con el objetivo de analizar los datos, se realiza un estudio inferencial basado en el contraste de hipótesis de igualdad de medias. El contraste establece que las medias poblacionales de ambos grupos son iguales y utiliza las muestras para determinar si la evidencia observada es coherente con estos supuestos, es decir consiste

en averiguar si los datos observados en las muestras respaldan la hipótesis sobre las poblaciones. Junto al estadístico se indica el p-valor, el cual se emplea para indicar cuánto contradice la muestra actual la hipótesis alternativa.

Los resultados mostrados en la Tabla 7, determinan que, inicialmente (antes del experimento) no existen diferencias estadísticamente significativas entre ambos grupos, ya que como indican los resultados de significación mostrados en la tabla 7 son mayores a 0,05, lo que permite aceptar la hipótesis de igualdad de medias con una probabilidad superior al 95%. Este resultado previo muestra que ambos grupos tienen una distribución similar, requisito necesario para poder contrastar la hipótesis del efecto de la retroalimentación eliminando posibles sesgos de partida.

Tabla 7. Diferencias entre los grupos antes y después del experimento

Antes	Control	Tratamiento	Total	t	Significación
Media	4,814286	4,472115	4,649537	0,4669	0,6425
Std. Err.	0,515112	0,520126	0,363432		
Std. Dev.	2,725721	2,652134	2,670672		
(N) Alumnos	28	26	54		
Después					
Media	5,587	5,195632	5,394633	0,6091	0,5449
Std. Err.	0,445809	0,462985	0,319456		
Std. Dev.	2,441799	2,493252	2,453795		
(N) Alumnos	30	29	59		

Fuente: Elaboración propia a partir de los datos recogidos. Significaciones: <0,05 (*).

Los valores medios de las calificaciones después de realizar el experimento señalan mejores resultados en el grupo de control (5,6) que en el de tratamiento (5,2), aunque las diferencias entre ambos grupos no son estadísticamente significativas. En este ensayo inicial, se observa que ofrecer retroalimentación sobre el rendimiento relativo a los estudiantes no afecta a sus rendimientos.

La Tabla 8 refleja la matriz con los porcentajes de alumnos que cambian de un percentil a otro, cuando se compara los resultados de las pruebas realizadas a mediados del semestre con respecto al examen final del grupo de control. Se observa que los datos ubicados por encima de la diagonal principal son los porcentajes de alumnos que mejoran de percentil, mientras que los que están por debajo son aquellos que empeoran, siendo los que se encuentran en la misma diagonal aquellos que se mantienen constantes a lo largo del curso. Se puede destacar el porcentaje (8,70%) de alumnos del grupo de control que pasan del percentil 50 al percentil 100 siendo idéntico porcentaje de aquellos que empeoran sus calificaciones pasando del percentil 100 al percentil 75.

Tabla 8. Porcentaje de alumnos que cambian de percentil en grupo control

Pruebas	Examen	P25	P50	P75	P100
P25		13,04%	4,35%	4,35%	0,00%
P50		0,00%	17,39%	4,35%	8,70%
P75		13,04%	4,35%	4,35%	0,00%
P100		0,00%	0,00%	8,70%	17,39%

Fuente: Elaboración propia a partir de los datos recogidos.

Analizando la matriz para el grupo de tratamiento (Tabla 9) se observa que los porcentajes de alumnos que cambian de percentil no es muy significativo a excepción del 8% correspondiente de aquellos alumnos que antes de la intervención se encontraban en el percentil 100 y han empeorado sus calificaciones bajando al percentil 75.

Tabla 9. Porcentaje de alumnos que cambian de percentil en grupo tratado

Pruebas	Examen	P25	P50	P75	P100
P25		16%	4%	4%	4%
P50		8%	8%	0%	4%
P75		4%	4%	16%	4%
P100		0%	4%	8%	12%

Fuente: Elaboración propia a partir de los datos recogidos.

Como se puede comprobar en ambas matrices (Tablas 8 y 9), la mayor parte de alumnos se mantienen en los mismos percentiles, reflejando que la intervención no obtuvo el efecto deseado en el grupo tratado.

Por otra parte, el análisis de las respuestas sobre el cuestionario realizado al final del curso por parte de los estudiantes que recibieron el percentil a lo largo del curso se aprecia que un 100% de los mismos manifiesta haber recibido el percentil, siendo comprensible para el 94,90% de los mismos. El cuestionario refleja que un 56,4% de los alumnos esperaba obtener una nota similar a la obtenida en el primer examen realizado mientras que un 23,10% esperaba obtener una nota superior, no obstante, el 51,30% de los alumnos del grupo de tratamiento considera que ha mejorado el percentil a lo largo del curso y un 74,42% consideran que tener constancia de este dato mejora la motivación para enfrentarse al examen final de la asignatura.

Cerca del 60% de los estudiantes del grupo de tratamiento (57,8%) manifiesta que el conocimiento de esta información es útil durante el curso de cara a afrontar el examen final.

6. CONCLUSIONES Y DISCUSIÓN

La eficacia de los servicios públicos, incluyendo el educativo, depende no solo del esfuerzo de quién lo presta sino también del que lo recibe (Bandiera et al., 2015). La función de producción de la educación considera el esfuerzo que realizan tanto docentes como estudiantes como un input, y la investigación educativa se preocupa por encontrar formas de motivación que mejore los resultados académicos y el desempeño del alumnado en todas las etapas educativas, elementos que se consideran el output de la producción educativa.

La retroalimentación sobre las calificaciones de las diferentes actividades académicas desarrolladas en una asignatura brinda a los estudiantes información sobre cómo su esfuerzo se traduce en calificaciones, y puede actuar como una señal para las puntuaciones futuras.

El análisis de la retroalimentación del desempeño, tanto en el entorno educativo como en el laboral, es un tema de interés tratado en la literatura científica. La provisión de retroalimentación sobre el propio desempeño en términos absolutos, como en términos relativos, es un elemento de interés que se estudiado desde diferentes ámbitos (psicológico, económico, educativo, sociológico, etc.).

En el ámbito educativo existen experiencias previas, tanto en etapas preuniversitarias como en universidades, aunque en España, solo se tiene conocimiento del experimento realizado en la Universidad Carlos III de Madrid. Utilizando esta como referencia, el experimento que aquí se presenta, realizado en la Universidad Rey Juan Carlos, trata de sumar otra experiencia universitaria al caso anterior.

La investigación experimental que aquí se presenta determina la importancia de la retroalimentación en los rendimientos relativos de los alumnos de la URJC, donde el azar es quien asigna a los estudiantes a pertenecer al grupo de tratamiento o de control. El experimento se desarrolla informando a los alumnos pertenecientes al grupo de tratamiento de su percentil en comparación con el grupo-clase, y en el caso de los estudiantes del grupo de control, no se le dio retroalimentación relativa, por lo que se mantuvieron ajenos a su posición relativa respecto a la clase.

Este experimento de control aleatorio sobre la retroalimentación relativa en alumnos de la asignatura de Análisis Económico Financiero en el grado de Marketing no ha encontrado diferencias estadísticamente significativas en las calificaciones medias finales de los alumnos entre el grupo de tratamiento, grupo que recibió retroalimentación relativa de sus calificaciones durante el segundo semestre del año, y las del grupo de control, por lo que se rechaza la hipótesis de partida. Estos resultados están en línea con la literatura existente que indican que proporcionar a los estudiantes información sobre su posición relativa en la cohorte tiene un efecto perjudicial inmediato en su rendimiento académico (Azmat et al., 2019; Cabrales et al., 2019). Estos autores, que hacen un seguimiento a lo largo de varios cursos, también señalan que los estudiantes cuyo desempeño se deterioró en respuesta a la información de

retroalimentación recuperan la mayor parte de la diferencia en las calificaciones con respecto al grupo de control relativamente rápido. Por tanto, de cara a seguir profundizando en el conocimiento de los mecanismos que impulsan el efecto de la información de retroalimentación sobre el desempeño, es muy relevante el poder diferenciar entre efectos inmediatos o a corto plazo y efectos a medio y largo plazo.

Los resultados de nuestra investigación se corresponden con un único semestre y es la prueba piloto de un estudio más amplio que está en proceso, por lo que la muestra inicial es reducida y se espera terminar con una mayor muestra, próxima a mil estudiantes, que permitirá analizar los resultados diferenciando por cuartiles o percentiles dentro de la distribución de calificaciones, y contrastar si los resultados que se obtienen en las calificaciones medias, difieren de si la comparativa se realiza entre los cuartiles de mejor y peor calificación, de forma que puedan encontrarse patrones diferentes.

Un valor añadido de este estudio fue la realización de un cuestionario al principio y al final del curso. El primero recoge información personal sobre el conocimiento y las expectativas de los estudiantes sobre sus calificaciones en términos relativos. Esto permitió observar que la estimación de los estudiantes difiere de la realidad al finalizar el semestre. En este contexto universitario hay que “orientar la evaluación hacia un proceso formativo, en contrapartida a una evaluación sumativa y finalista, con un alumnado activo y responsable de sus aprendizajes” (Fraile et al., 2013). Por otra parte, las respuestas por parte de los alumnos al cuestionario realizado al final del curso reflejan que la retroalimentación durante el semestre no implicaba una mayor motivación a la hora de estudiar de cara a los exámenes finales. Este resultado debe ser tomado con precaución, puesto que el número de respuestas del cuestionario no alcanzó un número significativo estadísticamente.

Este experimento piloto pretende aportar una serie de contribuciones a la literatura existente con relación a la retroalimentación de datos relativos por parte de los docentes a los estudiantes en la educación universitaria. La primera aportación del estudio refleja el impacto nulo en la mejoría de los rendimientos por parte del grupo tratado frente al grupo de control, en consonancia con otros estudios finalizados (Barankay, 2012; Blader et al. 2015; Cabrales et al., 2019; Macías 2023). Esto reforzaría la creencia de que el conocimiento del percentil antes del examen final pueda provocar una confianza adicional al estudiante y esforzarse menos ante este último examen de la asignatura. No obstante, merece la pena destacar que, en esta muestra, las notas del grupo de tratamiento siempre han estado por debajo de las notas del grupo de control, aunque las diferencias no eran estadísticamente significativas y el tamaño muestral es reducido.

Una segunda contribución importante por destacar es la alta percepción que tienen los alumnos sobre la motivación causada por el conocimiento del percentil de cara a mejorar las calificaciones en el futuro y la fácil interpretación de este dato por

parte de los alumnos. Esta información recogida por los cuestionarios realizados una vez calificada la asignatura arroja evidencia sobre la falta de información que reciben los alumnos durante el curso y cómo información adicional anima a los estudiantes y lo consideran como una herramienta útil para su actividad de aprendizaje. Este resultado también coincide con otros experimentos (Azmat et al., 2019; Cabrales et al., 2019; Macías, 2023), que señalan que la información mejora la satisfacción autoinformada de los estudiantes tratados.

Durante todo el curso se ha seguido la asistencia a clase para conocer si los alumnos retomaban la motivación por volver a las aulas tras la pandemia del COVID-19. Por ello, como última aportación del estudio se resalta la reducida asistencia media diaria, siendo esta de 38 alumnos, una cifra muy baja teniendo en cuenta que la matriculación de alumnos en la asignatura es de 69, lo que implica una ratio de asistencia del 55,60%, es decir, poco más de la mitad de los alumnos asisten a clase presencialmente habiendo un mínimo absoluto de asistencia en todo el semestre del 36,23%. Esta tendencia de la asistencia presencial se está estudiando en los semestres de cursos siguientes, para ver si la falta de asistencia puede ser uno de los motivos de los bajos resultados iniciales.

Una vez finalizado el estudio al completo (en mayo de 2024), se pretende aportar a la literatura existente, elementos para tener en cuenta con respecto a proporcionar a los alumnos una retroalimentación útil y eficiente, así como líneas de trabajo que puedan servir a los docentes para su planificación en relación con instrumentos y formas de evaluación. Una línea de trabajo futura se centrará en mejorar la implementación del cuestionario final, debido a que los alumnos del grupo tratado reciben el enlace para acceder al cuestionario una vez conocida su nota final y muchos de ellos, no tienen el interés de rellenar el mismo, por lo que esta valiosa información a posteriori es difícil de recuperar. Además, este tipo de experimentos sirve para cuestionar las diferentes concepciones de la evaluación de los aprendizajes, que como señala Hidalgo (2021) están limitadas y presionadas por el predominio de una cultura de resultados ligada a la evaluación sumativa, en lugar de a la formativa.

7. REGISTRO DEL EXPERIMENTO EN AMERICAN ECONOMIC ASSOCIATION

Este estudio fue registrado con la AEA Trial Registry (RCT ID: AEARCTR-0010116). Todas las expresiones son de los autores y no de ninguna de las instituciones a los que están afiliados.

8. BIBLIOGRAFÍA

Abadie, A. (2020). Statistical non significance in empirical economics. *American Economic Review*, 110(2), 193-208.

- Azmat, G., & Iriberry, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7-8), 435-452.
- Azmat, G., Bagues, M., Cabrales, A. & Iriberry, N. (2019). What You Don't Know... Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education. *Management Science*, 65(8) 3714-3736.
- Bandiera, O., Larcinese, V., & Rasul, I. (2015). Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics*, 34, 13-25.
- Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment. *The Wharton School, mimeo*.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Blader, S., Gartenberg, C., & Prat, A. (2015). The contingent effect of management practices. *Columbia Business School Research*, 15-48.
- Cabrales, A., Azmat, G., Bagues, M., & Iriberry, N. (2019). What you don't know... Can't hurt you? A natural field experiment on relative performance feedback in higher education. *Management Science*, 65(8), 3449-3947.
- Castejón, J., Capllonch, M., González, N. y López V. (2009). Técnicas e instrumentos de evaluación formativa y compartida para la docencia universitaria, Evaluación formativa y compartida en educación superior. *Narcea*, pp. 65-91.
- Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980-2016. *Educational Research*, 60(3), 276-291.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence on the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, 14(1), 1-11.
- Fraile, A., López-Pastor, V., Castejón, J., & Romero, R. (2013). La evaluación formativa en docencia universitaria y el rendimiento académico del alumnado. *Aula Abierta*, 41(2), 23-34.
- Fuch, L. S., & Fuch, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199-208.
- Gill, D., Kissová, Z., Lee, J., & Prowse, V. (2018). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*, 65(2), 494-507.
- Gil-Flores, J. (2012). La evaluación del aprendizaje en la universidad según la experiencia de los estudiantes. *Estudios Sobre Educación*, 22, 133-153.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 88-112.

- Hidalgo, M. E. (2021). Reflexiones acerca de la evaluación formativa en el contexto universitario. *Revista Internacional de Pedagogía e Innovación Educativa*, 1(1), 189-210.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37.
- Kosfeld, M., & Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3(3), 86-99.
- López-Pastor, V. (2012). Evaluación formativa y compartida en la universidad: clarificación de conceptos y propuestas de intervención desde la Red Interuniversitaria de Evaluación Formativa. *Psychology, Society, & Education*, 4(1), 117-130.
- Macías, C. (2023). Random experiment on relative performance feedback in higher education at URJC. In J. Sainz & I. Sanz (Eds.), *Addressing inequities in modern educational assessment* (pp. 45-802). Springer: Cham.
- Macías, C., & Santero, R. (2022). El sistema educativo en el ámbito rural español. Un análisis comparativo con la educación en la ciudad en España. *Revista Universitaria Europea*, 37, 215-248.
- Moore, D. A., & Cain, D. M. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes*, 103, 197-213.
- Nyquist, J. B. (2003). The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis. (Unpublished master's thesis). Vanderbilt University.
- Styles, B., & Torgerson, C. (2018). Randomised controlled trials (RCTs) in education research—methodological debates, questions, challenges. *Educational Research*, 60(3), 255-264.
- Tran, A., & Zeckhauser, R. (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics*, 96(9), 645-650.
- Williams, J., & Kane, D. (2009). Assessment and feedback: Institutional experiences of student feedback, 1996 to 2007. *Higher Education Quarterly*, 63, 264-286.