



El efecto de incentivos negativos en un sistema online de valoración docente del profesorado

Ana I. Cid-CidUniversidad Rey Juan Carlos (España) ✉ **Santiago Leguey**Universidad Rey Juan Carlos (España) ✉ **Jaime Prieto**Universidad Rey Juan Carlos (España) ✉ **Rocío Guede-Cid**Universidad Rey Juan Carlos (España) ✉ <https://dx.doi.org/10.5209/rced.93797>

Recibido: Enero 2024 • Evaluado: Abril 2024 • Aceptado: Junio 2024

Resumen: Diversos estudios han señalado que la administración online de las encuestas de valoración docente arroja tasas de respuesta sensiblemente inferiores a las que se obtenían cuando su administración era en papel en el aula. Entre los métodos propuestos para su incremento destaca el uso de incentivos de distinto tipo que actúan como motivadores extrínsecos para el alumnado, ya sean de carácter positivo (p.ej.: publicación anticipada de las calificaciones) o negativo (p.ej.: permitir la realización de los exámenes). El objetivo del presente estudio fue analizar el efecto que tuvo sobre la tasa de respuesta de las encuestas de valoración docente de una universidad la incorporación de un incentivo negativo (obligatoriedad de cumplimentar la encuesta para la consulta de las calificaciones) al migrar de un sistema de administración en papel (sin incentivo de respuesta) a un sistema de administración online. Se analizaron las valoraciones docentes de 35.497 asignaturas pertenecientes a 18 semestres, en 9 de ellos se utilizaron cuestionarios en papel administrados presencialmente en el aula y en los otros 9 cuestionarios online. Los principales resultados del estudio mostraron que la introducción del incentivo contribuyó a incrementar la tasa de respuesta, pero tuvo efectos no deseables como el aumento de estudiantes que respondían con el mismo valor a todos los ítems de las encuestas. Los resultados obtenidos tras la introducción del incentivo negativo de respuesta podrían sugerir una posible falta de interés de los estudiantes a la hora de cumplimentarlas, siendo su único objetivo completarlas cuanto antes con un patrón de respuesta de mínimo esfuerzo respondiendo a todos los ítems con el mismo valor para poder acceder a la calificación de la asignatura. El trabajo discute las posibles implicaciones de ello y señala posibles líneas de actuación para los gestores de las instituciones universitarias a este respecto.

Palabras clave: Evaluación del profesorado; educación superior; encuestas de valoración docente; cuestionario online; alumnado.

ENG The effect of negative incentives in an student evaluation of teaching online system

Abstract: Several studies have shown that online administration of Students' Evaluation of Teaching (SET) surveys typically results in significantly lower response rates compared to paper-based administration conducted in the classroom. To address this issue, institutions have implemented various incentives as extrinsic motivators for students, either of positive (e.g., early release of grades) or negative nature (e.g., making survey completion a prerequisite for accessing exam results). This study aimed to analyze the impact of introducing a negative incentive (requiring students to complete the survey in order to access their grades) on the response rates of SET surveys during a university's transition from a paper-based system (without incentives) to an online one. The analysis covered teaching evaluations from 35,497 course subjects across 18 semesters, with nine semesters using paper surveys administered in class and the other nine using online surveys. The main findings of the study showed that the introduction of the incentive contributed to higher response rates but also led to undesirable effects, such as an increased number of students providing uniform responses across all survey items. These results suggest a possible lack of genuine engagement, with students completing surveys hastily and with minimal effort solely to access their grades. The study discusses the implications of these findings and offers recommendations for university administrators regarding the design and implementation of SET survey strategies.

Keywords: teacher evaluation; higher education; student evaluation of teaching; online survey; students.

Sumario: 1. Introducción. 2. Metodología. 2.1. Muestra y procedimiento de recogida de información. 2.2. Instrumento. 2.3. Análisis de los datos. 3. Resultados. 3.1. Tasa de respuesta. 3.2. Intravariabilidad. 3.3. Tamaño del efecto. 4. Discusión. 5. Conclusiones. 6. Referencias bibliográficas.

Cómo citar: Cid-Cid, A. I.; Leguey, S.; Prieto, J.; Guede-Cid, R. (2025). El efecto de incentivos negativos en un sistema online de valoración docente del profesorado. *Revista Complutense de Educación* 36(3), 313-322

1. Introducción

Las encuestas de valoración docente se han utilizado desde principios del siglo XX como herramienta para garantizar la calidad de la educación superior (Sullivan et al., 2024) y como instrumento en los procesos de promoción del profesorado y de evaluación de las titulaciones (Abellán-Roselló y Fernández-Rodicio, 2023). A través de estas encuestas los estudiantes evalúan a sus profesores en distintos aspectos de la enseñanza, normalmente con cuestionarios estandarizados de escala tipo Likert que pueden completarse con preguntas abiertas (Castro-Morera, et al., 2020; Spooren et al., 2007). Tradicionalmente, estas encuestas se administraban a los alumnos en papel en el aula al final del periodo académico, y los resultados podían servir a los profesores para mejorar su práctica docente en el futuro (Jiang y Xiong, 2021). Con el desarrollo de la tecnología, estos cuestionarios comenzaron a administrarse electrónicamente a finales del siglo pasado, hasta convertirse en la principal forma de administración a día de hoy (Nederhand et al., 2023). Se prevé que el protagonismo de la valoración docente *online* siga aumentando (Campos et al., 2022), observándose la necesidad de ampliar las investigaciones sobre este sistema (Iancu et al., 2024). Cabe destacar que, en general, los estudiantes prefieren el método de administración *online* a la hora de evaluar a sus profesores, puesto que perciben un mayor anonimato al no sentirse presionados por la presencia de los profesores en clase (Stowell et al., 2012). Por el contrario, algunos profesores son reacios a cambiar a un entorno *online* por diversas razones, como la obtención de tasas de respuesta más bajas y la creencia de que los resultados son menos precisos (Crews y Curtis, 2011).

La mayoría de los estudios que han comparado ambas formas de administración muestran las ventajas e inconvenientes de cada una de ellas, centrándose en dos cuestiones fundamentales. Por un lado, en analizar las posibles diferencias entre las valoraciones medias que obtiene el profesorado según sea su administración en el aula o de forma *online*. Y por otro, en examinar las posibles diferencias en las tasas de respuesta que se obtienen con cada forma de administración. En cuanto al impacto en las valoraciones medias, no existe un consenso claro en la literatura. Algunas investigaciones han encontrado puntuaciones medias más bajas en la administración *online*, sugiriendo que los estudiantes que responden sin presión en un entorno digital son más estrictos a la hora de evaluar a sus profesores (Baldo et al., 2020). No obstante, existen otros estudios que muestran puntuaciones medias más altas, lo que podría sugerir que la utilización de este sistema *online* predispone a los estudiantes a realizar evaluaciones más favorables, en comparación con el uso tradicional del papel (Avery et al., 2006). Asimismo, otros trabajos no muestran diferencias significativas entre las puntuaciones medias cuando se comparaban los entornos *online* y presenciales (Guder y Malliaris, 2010).

En lo que se refiere a la tasa de respuesta, la mayoría de los estudios concluyen que esta es menor cuando la administración de las encuestas de valoración docente es *online* (Guder y Malliaris, 2010; Stowell et al., 2012). Por ello, el estudio de los factores que pueden explicar las tasas de respuesta más bajas y cómo aumentar el porcentaje de respuestas ha sido objeto de investigación recurrente en los últimos años (Goodman et al., 2014; Sullivan et al., 2024). La literatura ha puesto de relieve diferentes formas de aumentar la tasa de respuesta en la valoración docente *online*. Así, motivar a los estudiantes a completar las encuestas cuando las reciban, explicándoles la importancia de estas como posible mecanismo de mejora de la práctica docente, y mostrándoles que el proceso garantiza la confidencialidad y el anonimato, conducen a un aumento de la tasa de respuesta (Crews y Curtis, 2011; Goodman et al., 2014). Asimismo, el hecho de que se familiaricen con el entorno de la encuesta *online* tiene un efecto positivo en la tasa de respuesta (Crews y Curtis, 2011; Nulty, 2008). Cuando varias de estas acciones se realizan de forma combinada, su eficacia aumenta (Ballantyne, 2003).

Por otra parte, algunas instituciones han optado por la incorporación de incentivos de respuesta como estrategia para aumentar las tasas de respuesta cuando las encuestas se administran *online* (Alvero et al., 2019; Stowell et al., 2012). Los incentivos pueden ser positivos o negativos, dependiendo de si los estudiantes están motivados para conseguir algo que quieren o para evitar una consecuencia negativa, respectivamente. Los incentivos positivos más eficaces incluyen la obtención de puntos extra en sus calificaciones o la publicación anticipada de las calificaciones (Berk, 2012). Por su parte, los incentivos negativos contemplan entre otros hacer obligatorias las encuestas para permitir la realización de exámenes o la consulta online de las calificaciones (Crews y Curtis, 2011; Anderson, Cain y Bird, 2005). En este sentido, varios estudios han destacado el uso de incentivos como la forma más eficaz de aumentar las tasas de respuesta en la administración *online* de las encuestas de valoración docente (Crews y Curtis, 2011; Goodman et al., 2014; Zipser y Mincieli, 2018).

Pero ¿podría el uso de incentivos de respuesta tener consecuencias, no sólo en el deseable aumento de la tasa de respuesta de las encuestas *online*, sino también en cómo los estudiantes responden a las encuestas? En concreto y en lo que a los incentivos negativos se refiere, ¿podría afectar al modo en que los estudiantes completan las encuestas de valoración docente el uso de ciertos incentivos negativos (p.ej.: la

necesidad de completar la encuesta para la consulta de calificaciones) que hagan que las encuestas *online* se conviertan en prácticamente obligatorias en comparación con la administración tradicional en papel y en clase en las que no había ningún tipo de incentivo? Este fenómeno puede ser examinado mediante el análisis de la distribución de los rangos de respuesta. Para su análisis contamos con un indicador específico como es la variabilidad intracuestionario (también conocido como intravariabilidad de respuestas), que refleja cómo de variables son las respuestas de un encuestado cuando responde a un cuestionario (Lam y Green, 2023). Este indicador reflejaría el grado de variabilidad de las respuestas que cada estudiante da a los diferentes ítems que les son presentados en las encuestas de valoración docente de sus profesores. Esta medida es empleada habitualmente como indicador del nivel de esfuerzo e interés (o cuidado) que los encuestados muestran a la hora de cumplimentar un cuestionario (Dunn, et al., 2018), permitiendo así discriminar entre respondientes conscientes (del inglés *conscientious responders*; aquellos que responden de buena fe e interés a la encuesta) y respondientes aleatorios (del inglés *random responders*; aquellos que responden a los ítems de los cuestionarios sin realmente pararse a analizarlos (Marjanovic et al., 2015; Osborne y Blanchard, 2011). De entre los posibles métodos para computar la intravariabilidad de respuestas en un cuestionario destaca el uso del rango de variación o recorrido estadístico, entendido como la diferencia entre la puntuación máxima y la puntuación mínima otorgada a los diferentes ítems de la encuesta por parte de cada encuestado. Se trata de una medida de dispersión absoluta cuya principal ventaja es que, de forma sencilla, permite conocer la amplitud de los datos de una distribución (Casado, 2007). En contraposición, y es preciso señalarlo, se trata de una medida que por sí sola proporciona poca información sobre la distribución interna de los datos (Casado, 2007). Otros autores con este mismo propósito han utilizado la varianza intracuestionario o intravarianza (Steedle et al., 2019).

Hasta donde nuestro conocimiento alcanza, no existen estudios previos que hayan examinado los interrogantes anteriormente planteados, por lo que responder a ellos podría contribuir a avanzar en el conocimiento sobre las encuestas de valoración docente como elemento principal usado por las universidades para la valoración del desempeño de sus docentes. Con este contexto como motivación, el propósito de esta investigación fue analizar cómo el hecho de cambiar de un sistema tradicional de administración de encuestas de valoración docente en clase en papel, sin incentivos de respuesta, a un sistema de administración *online* fuera de clase que incluye el uso de un incentivo de respuesta negativo, puede afectar a la tasa de respuesta y al modo en que los estudiantes responden a las encuestas.

2. Metodología

2.1. Tipo de estudio, muestra y procedimiento de recogida de información

Se planteó una Investigación cuantitativa de tipo exploratorio, en la que sin hipótesis previas se describen los comportamientos de los indicadores de interés. Aunque el carácter es longitudinal, ya que se presentan los resultados de 9 cursos académicos, el estudio se focaliza en observar las diferencias entre lo que ocurre antes y después del momento en el que se produce un cambio en el sistema de recolección de las encuestas. Respecto a la selección de individuos para el estudio, se trata de un diseño censal puesto que se pretende obtener la valoración de todos los estudiantes que han cursado cada asignatura a lo largo del periodo de análisis.

La muestra comprendió 18 semestres consecutivos de administración de encuestas de valoración docente del profesorado en una universidad pública española, para un total de 9 cursos académicos, del curso 2010-11 al curso 2018-19. Durante los 18 semestres la encuesta de valoración docente fue la misma, cambiando únicamente su forma de administración. En los primeros 9 semestres, su administración fue en papel en clase, a través de unas hojas de lectura óptica que eran completadas a mano por los estudiantes con un bolígrafo. Las encuestas eran administradas durante una sesión de clase entre dos y cuatro semanas antes del periodo de exámenes. Los estudiantes no tenían ningún tipo de incentivo por completar las encuestas, más allá de emplear ese tiempo en que se detenía la clase para valorar cómo había sido el desempeño de su profesor durante la asignatura que iba a concluir. En los siguientes 9 semestres (a partir del segundo semestre del curso 2014-15), la administración de la encuesta migró a un sistema *online* (los estudiantes cumplimentaban por su cuenta las encuestas web fuera de clase) e incorporó un incentivo negativo, por el cual su cumplimentación era condición previa imprescindible para acceder a la consulta *online* de las calificaciones. El enlace para acceder a las encuestas era enviado por correo electrónico a los estudiantes, estando disponibles para ser respondidas entre dos y cuatro semanas antes del final del semestre. Cuando el estudiante accedía a la plataforma para consultar la calificación en una asignatura el sistema comprobaba si había completado la encuesta de valoración del profesor antes de mostrar la nota, rediriéndole al formulario de respuesta en caso de no haber sido así. Aunque todas las preguntas contaban con la opción de respuesta “No deseo contestar”, la gran mayoría decidía responder (un 93,4% contestó a todas las preguntas y un 99,2% contestó al menos a la mitad de ellas).

Se recogieron un total de 1.603.593 respuestas correspondientes a 35.497 asignaturas.

2.2. Instrumento

El cuestionario administrado (recordemos que se trató del mismo cuestionario durante todos los cursos, cambiando únicamente su forma de administración) constaba de una selección de preguntas extraídas del instrumento de 31 ítems desarrollado y validado inicialmente por Spooren et al. (2007), uno de los instrumentos de valoración docente más extendidos y empleados internacionalmente en el ámbito de la

educación superior. En concreto, el cuestionario de valoración docente empleado constaba de 10 ítems relativos a los aspectos que la universidad consideraba esenciales a la hora de valorar la calidad docente. La escala de respuesta era tipo Likert de 5 puntos y oscilaba entre 1: 'Totalmente en desacuerdo' y 5: 'Totalmente de acuerdo'. Los alumnos podían evaluar a sus profesores en todas las asignaturas en las que estuvieron matriculados cada curso.

2.3. Análisis de los datos

Para cada curso académico se obtuvieron tres indicadores de calidad de la encuesta de valoración docente: la tasa de respuesta, la distribución de los rangos de respuesta de las encuestas y la medida del tamaño del efecto de la intravariabilidad en las respuestas.

En cuanto a la tasa de respuesta, ésta representaba la proporción de encuestas realizadas. Es decir, el número de estudiantes que respondían a la encuesta dividido entre el número total de matriculados en la asignatura. La tasa de respuesta es considerada como uno de los indicadores preferentes a la hora de estudiar la precisión de cualquier encuesta y es un elemento de análisis principal en el campo de estudio de las encuestas de valoración docente (Biemer y Lyberg, 2003).

En cuanto a la distribución de los rangos de respuesta de las encuestas, se examinó la intravariabilidad de las respuestas dadas por cada estudiante en cada encuesta de valoración docente (variabilidad intracuestionario). Para ello se computó el rango de variación o recorrido como la diferencia entre la puntuación máxima y la puntuación mínima otorgada a los diferentes ítems de la encuesta por parte de cada estudiante. Como se trataba de encuestas de escala tipo Likert de 1 a 5 puntos únicamente son posibles 5 rangos diferentes de respuesta (de rango 0 a rango 4; p.ej., el rango=0 representaría a un estudiante que valora todos los ítems de la encuesta con la misma puntuación; el rango=4 representaría una encuesta en la que al menos una respuesta tiene puntuación 1 y al menos otra puntuación 5). Esto permitió, por un lado, examinar la evolución de la variabilidad intracuestionario a lo largo de los años y en función del sistema de administración empleado (papel u *online*). Y, por otro, clasificar los cuestionarios en dos grupos en función de si todos los ítems recibían la misma puntuación o no (al menos dos ítems obtenían puntuaciones distintas). La intravarianza y el rango de los cuestionarios del primer grupo sería cero (intravariabilidad nula), siendo para el segundo grupo mayores que cero (intravariabilidad positiva). De entre estas dos medidas presentadas anteriormente en el marco teórico introductorio (intravarianza y rango) se decidió emplear el rango por su simplicidad a la hora de presentar los resultados. Además, el uso del rango de variación en las respuestas permitió clasificar todos los cuestionarios en 5 categorías ligadas a los 5 posibles rangos de respuesta, facilitando la interpretación y presentación de los resultados.

Una vez establecidos los dos anteriores grupos (cuestionarios con intravariabilidad nula vs. cuestionarios con intravariabilidad positiva), el objetivo pasaba por comparar ambos grupos antes y después de la aplicación del cuestionario *online* con el incentivo negativo. Para ello, en primer lugar, se realizó un contraste de diferencias de medias entre los dos grupos mediante la prueba *t* de Student. El resultado de la prueba fue significativo y próximo a cero en todos los cursos analizados ($p=0.000$), por lo que no facilita apreciar la evolución de las posibles diferencias entre los dos grupos a lo largo de los años. En este contexto, se decidió emplear el tamaño del efecto como medida para analizar la magnitud de la diferencia entre las puntuaciones de los dos grupos, constituyendo una medida complementaria al contraste de hipótesis (Kelley y Preacher, 2012). En particular, se empleó el coeficiente eta cuadrado (η^2) para su estimación. Es preciso destacar que el coeficiente eta cuadrado (η^2) es considerado como una medida habitual y apropiada para calcular el tamaño del efecto en la investigación educativa (Richardson, 2011).

3. Resultados

La Tabla 1 muestra los datos descriptivos de las encuestas de valoración docente analizadas. El 34,4% de las encuestas correspondieron a los nueve primeros semestres, recogidas en el aula y en papel, frente al 65,6% de encuestas correspondientes a los 9 semestres restantes, completadas *online* fuera del aula.

Tabla 1. Distribución de las encuestas de valoración docente recogidas.

| Curso académico | Número de encuestas recogidas | Número de clases | Número medio de encuestas por clase |
|-----------------|-------------------------------|------------------|-------------------------------------|
| 2010-11 | 92.341 | 3.192 | 28,9 |
| 2011-12 | 117.881 | 3.533 | 33,4 |
| 2012-13 | 130.305 | 3.693 | 35,3 |
| 2013-14 | 136.670 | 3.724 | 36,7 |
| 2014-15 | 169.086 | 3.691 | 45,8 |
| 2015-16 | 215.889 | 3.527 | 61,2 |
| 2016-17 | 224.630 | 3.788 | 59,3 |
| 2017-18 | 242.249 | 4.978 | 48,7 |
| 2018-19 | 274.542 | 5.371 | 51,1 |

3.1. Tasa de respuesta

En la Figura 1 puede observarse cómo la tasa de respuesta se incrementó considerablemente con la implantación del sistema *online*. En el curso 2010-11 la tasa de respuesta fue del 39,6%, porcentaje que fue aumentando gradualmente hasta el primer semestre de 2014-2015, cuando alcanzó el 55,7%. En el segundo semestre de 2014-15, el primero en el que se utilizó el sistema *online*, la tasa de respuesta se disparó hasta el 80,6%. La tasa de respuesta siguió aumentando en los dos cursos siguientes, alcanzando valores máximos del 87,1% y el 86,8%, respectivamente. En los dos últimos cursos analizados se produjo un ligero descenso en la tasa de respuesta, pero manteniéndose en valores cercanos al 80%.

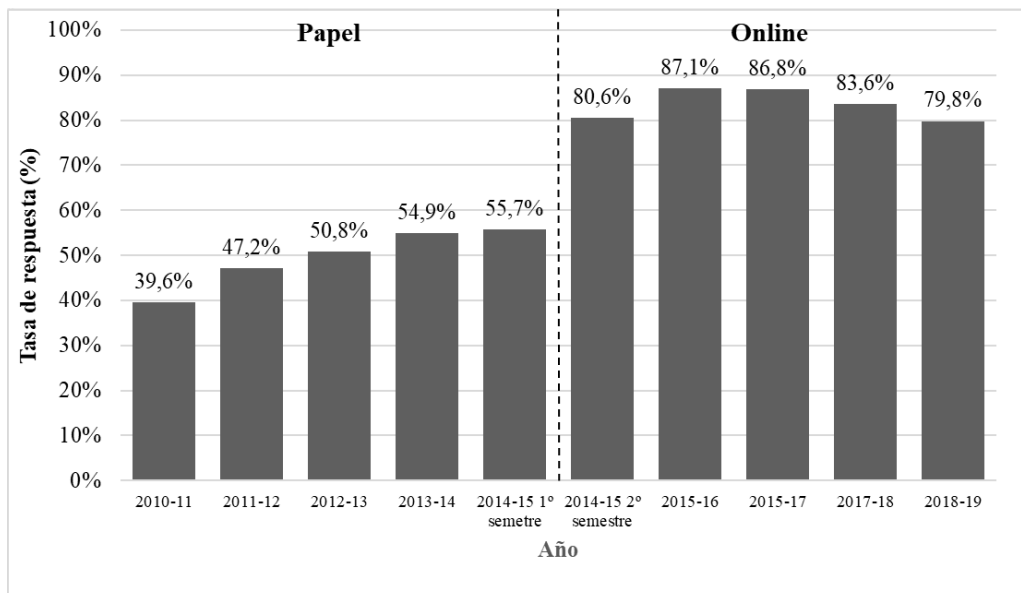


Figura 1. Evolución de la tasa de respuesta.

3.2. Intravariabilidad

La Tabla 2 muestra la evolución de los rangos de respuesta a lo largo de los años, evidenciando una clara tendencia a la reducción de los mismos. Se observa, en este sentido, como el porcentaje de respuestas de rango 0 aumentó con el tiempo, alcanzando cifras superiores al 50% en los dos últimos cursos analizados. El aumento fue especialmente pronunciado entre los dos semestres en los que se llevó a cabo la transición del sistema de recogida de información en el aula al sistema *online*, pasando de un 15% de respuestas de rango 0 en el primer semestre a un 33% en el segundo semestre. En los cuatro años en los que las encuestas de valoración docente fueron presenciales y se recogieron en papel, no se encontraron grandes variaciones en la distribución de los rangos de respuesta (ligero aumento en el rango 0 y ligero descenso en el rango 2). Con la administración de la encuesta *online* fuera de clase, el aumento del porcentaje de respuestas del rango 0 fue tan pronunciado que provocó una disminución de los porcentajes del resto de los rangos.

Tabla 2. Evolución de los rangos de respuesta.

| Curso académico | Rango=0 | Rango=1 | Rango=2 | Rango=3 | Rango=4 |
|-----------------|---------|---------|---------|---------|---------|
| 2010-11 | 13% | 30% | 36% | 16% | 6% |
| 2011-12 | 15% | 31% | 35% | 15% | 5% |
| 2012-13 | 16% | 30% | 34% | 15% | 5% |
| 2013-14 | 16% | 30% | 33% | 15% | 6% |
| 2014-15 | 24% | 29% | 30% | 13% | 5% |
| 2015-16 | 42% | 26% | 20% | 10% | 3% |
| 2016-17 | 47% | 23% | 18% | 9% | 3% |
| 2017-18 | 51% | 22% | 16% | 9% | 2% |
| 2018-19 | 53% | 21% | 16% | 9% | 2% |

3.3. Tamaño del efecto

La Tabla 3 muestra la evolución del tamaño del efecto a lo largo de los cursos analizados. El coeficiente experimentó un fuerte descenso con el sistema de recogida de información *online*. En el curso en que se produjo la migración del sistema, la proporción pasó del 32% en el primer semestre (en clase y en papel) al 22% en el segundo semestre (fuera de clase y *online*). En los cursos siguientes, en los que la administración de las encuestas de valoración docente fue totalmente *online*, se situó en torno al 18%.

Tabla 3. Evolución del tamaño del efecto.

| Academic year | η^2 |
|---------------|----------|
| 2010-11 | 34% |
| 2011-12 | 36% |
| 2012-13 | 33% |
| 2013-14 | 33% |
| 2014-15 | 27% |
| 2015-16 | 18% |
| 2016-17 | 17% |
| 2017-18 | 17% |
| 2018-19 | 19% |

4. Discusión

El propósito del presente trabajo fue analizar cómo afectaba a la calidad de los resultados de las encuestas valoración docente el cambio del sistema tradicional de administración de encuestas de la valoración docente en papel de forma presencial y sin incentivos de respuesta, a un sistema *online* no presencial y que incluía el uso de un incentivo de respuesta negativo: obligación de completar la encuesta para poder consultar las calificaciones de la asignatura. Para acometer dicho análisis se obtuvieron tres indicadores de calidad de las encuestas de valoración docente estudiadas, a saber: la tasa de respuesta, la distribución de los rangos de respuesta de las encuestas y la medida del tamaño del efecto de la intravariabilidad en las respuestas.

En cuanto a las tasas de respuesta, tratar de conseguir un aumento en ellas ha sido una de las principales preocupaciones de académicos, investigadores e instituciones cuando se realizan encuestas de valoración docente *online*, pues tal y como sucede en el análisis de los resultados de cualquier encuesta, la preferencia será contar con el mayor número posible de encuestas recogidas para obtener estimaciones más precisas y menos sesgadas (Wu et al., 2022). Sin embargo, en el caso particular de las encuestas de valoración docente, la tasa de respuesta que se considera aceptable dependerá de múltiples factores, desde el propio tamaño del grupo de clase, hasta el porcentaje de asistencia del alumnado, lo que dificulta establecer una tasa concreta (Crews y Curtis, 2011). Así, algunos autores afirman que serían aceptables tasas de respuesta del 20% para clases de 100 alumnos (Nulty, 2008), mientras que otros consideran deseables tasas de respuesta en torno al 60-70% (Stowell, et al., 2012).

Los resultados de nuestro estudio mostraron un incremento claro tras la migración del sistema en papel al sistema *online*, con promedios superiores a las tasas de respuesta deseables para los criterios más exigentes. El resultado difiere del de otros estudios que reflejaron tasas de respuesta menores cuando la administración de las encuestas de valoración docente migraba a un sistema online (p.ej.: Dommeyer et al., 2004; Guder y Malliaris, 2010). Sin embargo, en estos estudios no había un incentivo de respuesta, como sí sucede en la investigación realizada. Es, de hecho, a partir de investigaciones como las anteriormente citadas cuando se plantea la necesidad de incorporar incentivos de respuesta en forma de motivadores extrínsecos como estrategia para aumentar las tasas de respuesta en la administración *online* de las encuestas de valoración docente (Alvero et al., 2019; Goodman et al., 2014; Stowell et al., 2012). Los resultados de nuestro estudio ratifican la importancia del uso de incentivos en línea con los hallazgos de Zipser y Mincieli (2018) que encontraron un incremento en la tasa de respuesta superior al 20% al incorporar un incentivo que permitía la consulta adelantada de la calificación de la asignatura si se completaba la encuesta de valoración docente de la misma.

La necesidad de completar las encuestas para poder consultar las calificaciones fue ya sugerida como posible incentivo para aumentar la tasa de respuesta en una revisión de la literatura sobre encuestas de valoración docente online publicada en 2005, cuando todavía el protagonismo era para las encuestas en papel (Cain y Bird, 2005). El incentivo implementado y cuyas consecuencias analizamos en esta investigación, tiene un antecedente en el trabajo de Crews y Curtis (2011) en el que proponían como posible medida para incrementar la tasa de respuesta no proporcionar la calificación de la asignatura hasta que el 80% hubiera completado la encuesta de valoración docente de la asignatura. Señalan los autores que se trataba de una medida que ya era implementada por uno de los docentes encuestados, si bien dicho profesor no había cuantificado su posible eficacia. Sin embargo, estos autores, no habían evaluado el efecto de la incorporación de un incentivo negativo en otros indicadores de calidad, como se discute seguidamente.

Cuando se comparan los indicadores de calidad entre el sistema presencial sin incentivo y el online con incentivo, se observó cómo la tasa de respuesta aumenta sensiblemente con la introducción del sistema online, mientras que disminuye la variabilidad intracuestionario. Estos dos indicadores son habitualmente empleados a la hora de analizar la calidad de las respuestas de una encuesta (Dunn et al., 2018; Kelley y Preacher, 2012; Osborne y Blanchard, 2011). La disminución de la variabilidad intracuestionario se manifestó en el incremento del porcentaje de respuestas con rango de variación cero. Se han encontrado dos explicaciones complementarias de este fenómeno, por una parte, el desinterés por parte de los estudiantes a la hora de completar las encuestas de valoración docente, que no tendrían la motivación suficiente para prestar atención a las respuestas que dan en cada ítem. Nichols et al. (1989) catalogan esta situación como

falta de respuesta al contenido y la definen como el hecho de responder a la encuesta sin tener en cuenta el contenido del ítem en sí. Es decir, el estudiante responde a la encuesta solo, sin instrucciones por parte de una persona que administre los cuestionarios y con distracciones propias del entorno *online*, que lleva a los jóvenes a realizar varias tareas a la vez con la consecuente pérdida atencional (Huang et al., 2012; Van Der Schuur et al., 2015). Por otra parte, las respuestas de rango cero podrían deberse a un patrón de respuesta rápida, que sería aquella en la que un estudiante valora al profesor mediante una única puntuación que repite en todos los ítems de la encuesta de valoración docente y que se produce cuando los estudiantes quieren responder en el menor tiempo posible a la encuesta por tratarse de un requerimiento obligatorio para consultar la nota (Curran, 2016).

El análisis del esfuerzo o interés que los participantes en cualquier encuesta ponen a la hora de responder a las preguntas que en ella se plantean y cómo esto puede influir en los datos y conclusiones que se extraen ha sido objeto de múltiples estudios, pues se trata de uno de los principales instrumentos para la recogida de datos en investigación. En este sentido, diferentes autores han subrayado que las respuestas dadas por participantes desmotivados, que responden sin esfuerzo o interés, aun en porcentajes bajos, pueden tener un impacto negativo en la utilidad de los datos recogidos (Huang et al., 2012; Maniaci y Rogge, 2014), siendo necesario disponer de métodos para su detección (Curran, 2016; Huang et al., 2015). Para Marjanovic et al. (2015), las diferencias que se producen entre aquellos que son concienzudos en sus respuestas y el resto son fácilmente detectables y requieren de un análisis específico, tal y como se hace en esta investigación.

Los resultados muestran que con el cambio de sistema aumenta la proporción de respuestas de rango cero y se produce una disminución de las diferencias de puntuación entre el grupo de cuestionarios con intravarianza nula y el de intravarianza positiva, tal y como ocurre en otras investigaciones que han estudiado las amenazas que las respuestas descuidadas o aleatorias pueden tener en la validez y tamaño del efecto en las encuestas (Credé, 2010; Osborne y Blanchard, 2011). El coeficiente empleado para medir el tamaño del efecto disminuyó cuando se migró al sistema online. Es decir, las distancias entre el promedio de puntuaciones de aquellos que contestaron a todas las preguntas con el mismo valor (intravariabilidad nula) y las de aquellos que no lo hicieron así (intravariabilidad positiva) se redujeron. Para interpretar el significado del hecho se debe considerar que los cuestionarios de intravariabilidad nula comprenden tanto a aquellos estudiantes que desean otorgar la máxima o mínima puntuación a su docente (están muy satisfechos o insatisfechos con su labor) y por tanto marcan el máximo o mínimo valor en todos los ítems de la encuesta, como a aquellos estudiantes que emplean la misma puntuación en toda la encuesta para otorgar una puntuación global a su profesor o profesora (p.ej.: un estudiante que considera de forma global la labor del docente como de “2 sobre 5” y decide marcar un 2 en todos los ítems de la encuesta sin prestar especial atención a lo preguntado en cada una de ellas). Es decir, se trataría de estudiantes que usan un patrón de respuesta de menor esfuerzo, equivalente a una única pregunta que resumiera de forma global su opinión sobre el desempeño docente, en línea con algunos estudios que afirman que la evaluación de la enseñanza por parte de los alumnos podría sustituirse por una única medida sin apenas pérdida de predictibilidad (Clayson, 2022). Las puntuaciones de los estudiantes que desean otorgar la máxima o mínima puntuación estarían más alejadas de las del resto de estudiantes que las de aquellos que emplean el patrón de respuesta de menor esfuerzo. Así, la disminución observada del tamaño del efecto podría explicarse si el incremento de las respuestas de intravariabilidad nula se debe mayoritariamente al incremento de la proporción de estudiantes cuyo patrón de respuesta es el de menor esfuerzo.

El estudio presenta algunas limitaciones. La primera de ellas se refiere al hecho de basarse en una muestra de una única universidad con un incentivo de participación tan específico como la obligatoriedad de respuesta, lo que ha de conducir a interpretar los resultados con cautela, no debiendo extrapolar los resultados a otras instituciones que quizás apliquen otros instrumentos y/o incentivos de respuesta en sus encuestas de valoración docente. Otra de las limitaciones pasa por el hecho de contar con una muestra de datos que se detuvo, desde la fecha de publicación del presente trabajo, hace cinco cursos académicos (curso 2018-19). Sería interesante, de cara a un futuro próximo, poder complementar el estudio con los datos relativos a los cursos 2019-20 y siguientes, probablemente considerando ya únicamente la administración online del cuestionario desde el segundo semestre del curso 2014-15. Esto permitiría examinar, entre otras cuestiones, si el descenso en la tasa de respuesta observado desde el curso 2016-17 es puntual o muestra una tendencia clara sobre la que, por ejemplo, los organismos encargados de la valoración del desempeño de los docentes debieran actuar. Otra posible línea de investigación para el futuro pasaría por la incorporación de otros incentivos de participación (p.ej.: obtención de unas décimas adicionales en la calificación final, reconocimiento de créditos extra, cupones descuento para utilizar en el campus) y posterior análisis del efecto que puedan tener en la tasa de respuesta y en el modo en que los estudiantes responden a las encuestas de valoración docente. Por otra parte, el hecho de haber basado el estudio en encuestas de valoración docente con escala de respuesta de tipo Likert puede representar una limitación. En este sentido, existen investigaciones que han subrayado que emplear escalas de medida de tipo Likert en las encuestas de valoración docente puede ocasionar la aparición de cansancio en los estudiantes por lo repetitivo de la escala, más aún cuando han de completar encuestas de varias asignaturas (Cañadas y Cuétara, 2018; Moreno Olivios, 2018). Este hecho puede influir en el patrón de respuesta de los estudiantes, pudiéndoles llevar a cumplimentar las encuestas con rapidez y sin prestar la debida atención. Un efecto no deseado que puede verse aumentado cuando se incorporan incentivos de respuesta que hagan que los estudiantes quieran completar la encuesta cuanto antes para obtener la recompensa deseada (Matosas-López et al., 2019b). A este respecto, podría resultar interesante analizar el comportamiento de los estudiantes ante encuestas de valoración docente que empleen otras escalas de medida, como podrían ser aquellas que empleen escalas de conducta BARS

(Behaviorally Anchored Rating Scale), en las que se muestran ejemplos de comportamiento concretos de desempeño docente (no descriptores genéricos) que son valorados por los estudiantes como buenos o deficientes (Martin-Raugh et al., 2016; Matosas-López et al., 2019a). Asimismo, en línea con estudios previos publicados que han analizado las puntuaciones de carácter extremo y su comportamiento en las encuestas de valoración docente (Prieto et al., 2023; Leguey et al., 2023), sería deseable investigar si la incorporación de incentivos puede afectar a las puntuaciones que un docente, que imparte la misma asignatura en cursos consecutivos, obtiene a lo largo de los años.

6. Conclusiones

La presente investigación arroja conclusiones interesantes en un campo que no había sido suficientemente estudiado como es el efecto que pueden tener los incentivos negativos en la tasa de respuesta de las encuestas de valoración docente del profesorado universitario. En este sentido, y pese a que los resultados del estudio muestran un deseable aumento en la tasa de respuesta tras la migración a un sistema de encuesta *online* con incentivo de respuesta que requería al estudiante valorar al docente para acceder a las calificaciones finales de las asignaturas, se encontraron otras consecuencias no deseables, como son la reducción de la variabilidad intracuestionario y del tamaño del efecto. Este fenómeno, que ha sido estudiado en otros ámbitos (p.ej.: encuestas en el área de recursos humanos de las empresas; Dunn et al., 2018), debería invitar a la reflexión por parte de las instituciones, pues podría manifestar una falta de interés de los estudiantes a la hora de cumplimentar las encuestas, siendo su único objetivo completarlas cuanto antes para poder alcanzar el incentivo acceder así a la calificación de la asignatura. Esto podría sugerir que los estudiantes no tienen la motivación suficiente para dedicar atención a las respuestas que dan en cada ítem. Fruto del análisis realizado, así como del marco teórico elaborado, es posible plantear una serie de recomendaciones al respecto, que pasarían por la incorporación de nuevos mecanismos complementarios a la hora de que los estudiantes valoren al profesorado. Se podrían así, por ejemplo, realizar entrevistas individualizadas a un determinado número de estudiantes seleccionados de forma aleatoria en cada clase, lo que permitiría obtener una visión de carácter cualitativo que complementase la valoración cuantitativa aportada por las encuestas. Asimismo, podría ser interesante encuestar al alumnado en varios momentos durante el desarrollo de la asignatura (no únicamente al final), para contar con una visión más global y poder apreciar posibles variaciones en la valoración de los estudiantes (¿podrá un estudiante acabar valorando peor o mejor a su profesor por haber obtenido una peor o mejor calificación en una actividad de evaluación realizada a mitad de curso?). En definitiva, consideramos conveniente reflexionar a este respecto y encaminarse hacia un sistema de valoración del desempeño del profesorado por parte de los alumnos más amplio y de carácter holístico.

7. Referencias bibliográficas

- Abellán-Roselló, L., y Fernández-Rodicio, C. I. (2023). Análisis de la satisfacción de la actividad docente medida a través del alumnado en contextos universitarios siguiendo el Modelo de Calidad de la Situación Educativa. *Revista Complutense de Educación*, 34(4), 821-831. <https://dx.doi.org/10.5209/rced.80155>
- Alvero, A. M., Mangiapanello, K., y Valad, J. (2019). The effects of incentives, instructor motivation and feedback strategies on faculty evaluation response rates in large and small class sizes. *Assessment y Evaluation in Higher Education*, 44(4), 501-515. <https://doi.org/10.1080/02602938.2018.1521913>.
- Anderson, H. M., Cain, J., y Bird, E. (2005). Online student course evaluations: review of literature and a pilot study. *American Journal of Pharmaceutical Education*, 69(1), 34-43. <https://doi.org/10.5688/aj690105>.
- Avery, R. J., Bryan, W. K., Mathios, A., Kang, H., y Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations?, *Journal of Economic Education*, 37(1), pp. 21-37. <https://doi.org/10.3200/JECE.37.1.21-37>.
- Ballantyne, C. (2003). Online evaluations of teaching: An examination of current practice and considerations for the future. *New Directions for Teaching y Learning*, 96, 103-12. <https://doi.org/10.1002/tl.127>.
- Baldo, C. M., Snyder, J., y Holguin, A. (2020). Revisiting the Online versus Face-to-Face teaching evaluations. *International Journal of Education and Development using Information and Communication Technology*, 16(2), 144-157.
- Berk, R. A. (2012). The Top 20 Strategies to Increase the Online Response Rates of Student Rating Scales. *International Journal of Technology in Teaching and Learning*, 8(2), 98-107.
- Biemer, P. P., y Lyberg, L. E. (2003). *Introduction to Survey Quality*, New York: John Wiley y Sons.
- Campos, E., Núñez, S. D., Enriquez-de-la-O, J. F., Castaño, R., Escamilla, J., y Hosseini, S. (2022). Educational model transition: Student evaluation of teaching amid the COVID-19 pandemic. *Frontiers in Education*, 7, 1-11. <https://doi.org/10.3389/educ.2022.991654>.
- Cañadas, I., y Cuétara, I. De. (2018). Estudio psicométrico y validación de un cuestionario para la evaluación del profesorado universitario de enseñanza a distancia. *Revista de Estudios de Investigación en Psicología y Educación*, 5(2), 102-112. <https://doi.org/10.17979/reipe.2018.5.2.3701>.
- Casado, J. (2007). *Manual básico de Estadística*. Colección de Libros de autor: Instituto Nacional de Estadística.
- Castro-Morera, M., Navarro-Asencio, E., y Blanco-Blanco, Á. (2020). The quality of teaching as perceived by students and university teachers: analysis of the dimensionality of a teacher evaluation questionnaire. *Educación XX1*, 23(2), 41-65. <https://doi.org/10.5944/educXX1.25711>
- Clayson, D. (2022) The student evaluation of teaching and likability: what the evaluations actually measure. *Assessment y Evaluation in Higher Education*, 47(2), 313-326. [10.1080/02602938.2021.1909702](https://doi.org/10.1080/02602938.2021.1909702)

- Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement*, 70, 596-612. <https://doi.org/10.1177/0013164410366686>.
- Crews, T. B., y Curtis, D. F. (2011). Online course evaluations: Faculty perspective and strategies for improved response rates. *Assessment y Evaluation in Higher Education*, 36(7), 865-878. <https://doi.org/10.1080/02602938.2010.493970>.
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4-19. <https://doi.org/10.1016/j.jesp.2015.07.006>.
- Dommeyer, C. J., Baum, P., Hanna, R. W., y Chapman, K. S. (2004). Gathering Faculty Teaching Evaluations by In class and Online Surveys: Their Effects on Response Rates and Evaluations. *Assessment y Evaluation in Higher Education*, 29(5), 611-623. <https://doi.org/10.1080/02602930410001689171>.
- Dunn, A. M., Heggstad, E. D., Shanock, L. R., y Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology*, 33, 105-121. <https://doi.org/10.1007/s10869-016-9479-0>.
- Goodman, J., Anson, R., y Belcheir, M. (2014). The Effect of Incentives and Other Instructor-Driven Strategies to Increase Online Student Evaluation Response Rates. *Assessment y Evaluation in Higher Education*, 40(7), 958-970. <https://doi.org/10.1080/02602938.2014.960364>.
- Guder, F., y Malliaris, M. (2010). Online and Paper Course Evaluations. *American Journal of Business Education*, 3(2), 131-138. <https://doi.org/10.19030/ajbe.v3i2.392>.
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., y DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27, 99-114. <https://doi.org/10.1007/s10869-011-9231-8>.
- Huang, J.L., Liu, M., y Bowling, N.A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, 100, 828-845. <https://doi.org/10.1037/a0038510>.
- Iancu, D. E., Maricuțoiu, L. P., y Ilie, M. D. (2024). Student Evaluation of Teaching: The analysis of measurement invariance across online and paper-based administration procedures of the Romanian version of Marsh's Student Evaluations of Educational Quality scale. *Studies in Educational Evaluation*, 81, 101340. <https://doi.org/10.1016/j.stueduc.2024.101340>.
- Jiang, H. C., y Xiong, Y. (2021). The indicators, characteristics and enlightenment of students' evaluation of teaching in Japanese national universities: taking education as an example. *Higher Education Research*, 42, 103-109.
- Kelley, K., y Preacher, K. J. (2012). On effect size. *Psychological Methods*, 17(2), 137-152. <https://doi.org/10.1037/a0028086>.
- Lam, T. C. M., y Green, K. E. (2023). *Survey Development: A Theory-Driven Mixed-Method Approach*.
- Leguey, S., Cid-Cid, A. I., Guede-Cid, R. y Prieto, J. (2023). An Exploratory Analysis of Major Dropdowns in Student Evaluation of Teaching Ratings in Higher Education. *Multidisciplinary Journal of Educational Research*, 13 (1), 91-113. <https://doi.org/10.17583/remie.10419>.
- Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., y Greenglass, E. (2015). The interitem standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences*, 84, 79-83. <https://doi.org/10.1016/j.paid.2014.08.021>.
- Martin-Raugh, M., Tannenbaum, R. J., Tocci, C. M., y Reese, C. (2016). Behaviorally anchored rating scales: An application for evaluating teaching practice. *Teaching and Teacher Education*, 59, 414-419. <https://doi.org/10.1016/j.tate.2016.07.026>.
- Maniaci, M.R., y Rogge, R.D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61-83. <https://doi.org/10.1016/j.jrp.2013.09.008>.
- Matosas-López, L., Aguado-Franco, J. C., y Gómez-Galán, J. (2019a). Construcción de un instrumento con escalas de comportamiento para la evaluación la calidad docente en modalidades blended learning. *Journal of New Approaches in Educational Research*, 8(2), 148-172. <https://doi.org/10.7821/naer.2019.7410>.
- Matosas-López, L., Romero-Ania, A., y Cuevas-Molano, E. (2019b). ¿Leen los universitarios las encuestas de evaluación del profesorado cuando se aplican incentivos por participación? Una aproximación empírica. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*. 17(3), 99-124. <https://doi.org/10.15366/reice2019.17.3.006>.
- Moreno Olivos, T. (2018). La evaluación docente en la universidad: Visiones de los alumnos. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 3(16), 87-102. <https://doi.org/10.15366/reice2018.16.3.005>.
- Nederhand, M., Auer, J., Giesbers, B., Scheepers, A., y Van der Gaag, E. (2023). Improving student participation in SET: effects of increased transparency on the use of student feedback in practice. *Assessment & Evaluation in Higher Education*, 48(1), 107-120. <https://doi.org/10.1080/02602938.2022.2052800>.
- Nichols, D. S., Greene, R. L., y Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology*, 45(2), 239-250. [https://doi.org/10.1002/1097-4679\(198903\)45:2<239::aid-jclp2270450210>3.0.co;2-1](https://doi.org/10.1002/1097-4679(198903)45:2<239::aid-jclp2270450210>3.0.co;2-1).
- Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment y Evaluation in Higher Education*, 33, 301-314. <https://doi.org/10.1080/02602930701293231>.
- Osborne, J. W., y Blanchard, M. R. (2011). Random responding from participants is a threat to the validity of social science results. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2010.00220>.

- Prieto, J., Guede-Cid, R., Cid-Cid, A. I. y Leguey, S. (2023). Major increases in teachers' performance evaluations: Evidence from student evaluation of teaching surveys. *Tuning Journal for Higher Education*, 10 (2), 105-125. <https://doi.org/10.18543/tjhe.2299>
- Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2), 135-147. <https://doi.org/10.1016/j.edurev.2010.12.001>.
- Spooren, P., Mortelmans, D., y Denekens, J. (2007). Student evaluation of teaching quality in higher education. Development of an instrument based on 10 Likert scales. *Assessment and Evaluation in Higher Education*, 32, 667-679. <https://doi.org/10.1080/02602930601117191>.
- Steedle, J. T., Hong, M., y Cheng, Y. (2019). The effects of inattentive responding on construct validity evidence when measuring social-emotional learning competencies. *Educational Measurement: Issues and Practice*, 38(2), 101-111. <https://doi.org/10.1111/emip.12256>.
- Stowell, J. R., Addison, W. E., y Smith, J. L. (2012). Comparison of online and classroom-based student evaluations of instruction. *Assessment y Evaluation in Higher Education*, 37(4), 465-473. <https://doi.org/10.1080/02602938.2010.545869>.
- Sullivan, D., Lakeman, R., Massey, D., Nasrawi, D., Tower, M., y Lee, M. (2024). Student motivations, perceptions and opinions of participating in student evaluation of teaching surveys: a scoping review. *Assessment & Evaluation in Higher Education*, 49(2), 178-189. <https://doi.org/10.1080/02602938.2023.2199486>.
- Van Der Schuur, W. A., Baumgartner, S. E., Sumter, S. R., y Valkenburg, P. M. (2015). The consequences of media multitasking for youth: A review. *Computers in Human Behavior*, 53, 204-215. <https://doi.org/10.1016/j.chb.2015.06.035>.
- Wu, M.-J., Zhao, K., y Fils-Aime, F. (2022). Response rates of online surveys in published research: A meta-analysis. *Computers in Human Behavior Reports*, 7, 100206. <https://doi.org/10.1016/j.chbr.2022.100206>
- Zipser, N., y Mincieli, L. (2018). Administrative and structural changes in student evaluations of teaching and their effects on overall instructor scores. *Assessment y Evaluation in Higher Education*, 43(6), 995-1008. <https://doi.org/10.1080/02602938.2018.1425368>